Basic Measure Theory

Tom Sawada

August 7, 2018

Why do we care about measure theory? -Sunday, 6.17.2018

Measure theory, in some sense is *modern calculus*. The notion of measure and Lebesgue integration replaces our old understanding of fundamental notions such as continuity, differentiability, and integrability. We can in fact, reformulate calculus[1] using this machinery, and therefore, we must reprove many of the theorems from classical calculus, including but not limited to: Fundamental Theorem of Calculus and Fubini's theorem. There will be new theorems that result from this as well, most notable being the convergence theorems.

So, how powerful is this reformulation? If it's strength is comparable to the old calculus, then surely it is not worth making a new theory. Here, we will see how the notions in measure theory resolves issues from classical problems.

1.) **Completeness.** Fourier theory gives an element in $\ell^2(\mathbb{Z})$ for every element in $\mathcal{R}$, i.e. Riemann integrable functions. However, there are plenty of elements in $\ell^2(\mathbb{Z})$ that do not correspond to elements in $\mathcal{R}$. How can we complete $\mathcal{R}$ so that these sequences have corresponding functions? Additionally, how do we integrate such functions?

2.) **Limit Theorems.** We can consider the limit of continuous functions $f_n : [0,1] \to [0,1]$. If they converge uniformly, then the limit as $n \to \infty$ is also continuous. But if we do not have this assumption, the limiting function could be very badly behaved. What method of integration makes the limit of the integral the integral of the limit?

3.) **Bounded Variations, Hausdorff Dimension.** We recall from multivariable calculus that a curve is *rectifiable* if the length of the polygonal approximations of the curve is bounded above. The necessary condition for this, as we shall see, is rectifiability. We also have the Pythagorean formula for the length of the curve. This does not hold in general, but holds for appropriate parametrizations of the curve. We also know from point-set topology that there are (nonrectifiable) curves that fill a square, and is therefore "2-dimensional." Rectifiable curves on the other hand are 1-dimensional. As we will see, we can appropriately define what is meant by a fraction dimensional curve.

4.) **Lebesgue Differentiation Theorem, Covering Theorems, Absolute Continuity.** The first fundamental theorem of calculus states that for a differentiable function $F : [a,b] \to \mathbb{R}$,

$$F(b) - F(a) = \int_a^b F'(x)dx \tag{1}$$

We would like to know what is the larger class of functions for which the above is applicable? There are pathological functions such as the Weierstrass nowhere differentiable function or the function

$$f(x) := \begin{cases} x^2 \sin\left(\frac{1}{x^2}\right) & x \neq 0 \\ 0 & x = 0 \end{cases} \tag{2}$$

[1] We reformulate everything after fundamental notions from point set topology such as open/closed sets, sequences, continuity, and limits, as well as the algorithms we use to compute limits, derivatives, and integrals. Even some of this topological notions are generalized to fit the new theory–we have the notions of measurable sets which replace the notion of closed rectangles, and measurable functions replace the notion of continuous functions.

The derivative exists on $[-1, 1]$ but is unbounded, hence f' is not Riemann integrable.[2] For the second fundamental theorem, i.e.

$$\frac{d}{dx}\int_0^x f(y)dy = f(x) \tag{3}$$

we would like to formulate the notion of differentiating an integral more carefully, and state this for the general notion of integrable functions considered above.

[2] Also see Volterra function.

Chapter 1. Measure Theory

Preliminary Results: Point-Set Topology. -Tuesday, 6.19.2018

We establish some basic results from point-set topology which are essential to measure theory. In measure theory, we often approximate point sets by simple points sets such as rectangles (in order to get the Borel and hence the Lebesgue measure), and so, it makes sense to establish some basic results on rectangles. A volume of a rectangle is very easy to define because we just take the product of the length of all sides. Let's establish some basic terminologies.

Come back to the opening remark of ch.1 when we get to measurable sets.

Very, Very Basic Point-Set Topology of $\mathbb{R}^n$.

A **point** is simply an element of $\mathbb{R}^n$, i.e. an n-tuple of real numbers. Addition and multiplication by real scalars are done by exploiting the real vector space structure of $\mathbb{R}^n$. The **Euclidean norm of a point** is

$$|x| := \sqrt{\sum_{i=1}^{n} x_i^2} \tag{4}$$

The **distance between the points** x **and** y is $d(x,y) := |x-y|$.

The **complement** of a set is as we know it from basic set theory. The **complement of** F **in** E is $E - F := E \cap F^c$. The **distance between the sets** E **and** F is given by

$$d(E,F) := \inf_{\substack{x\in E\\ y\in F}} |x-y| \tag{5}$$

The **open ball in** $\mathbb{R}^n$ **centered at** x **of radius** r is the set

$$B_r(x) := \{y \in \mathbb{R}^n : |x-y| < r\} \tag{6}$$

A subset $E \subset \mathbb{R}^n$ is **open** if for every $x \in E$, there is some $r > 0$ such that $B_r(x) \subset E$. A set is **closed** if its complement is open.

An easy theorem is that (arbitrary) union of open sets U_α is open (since for any point x, there is some U_α containing x and some $r > 0$ such that $B_r(x) \subset U_\alpha \subset \bigcup_\alpha U_\alpha$). Another easy theorem is that finite union of open sets is open (since for $x \in \bigcap_\alpha U_\alpha$, we can take $r := \min_\alpha r_\alpha > 0$ (where $B_{r_\alpha}(x) \subset U_\alpha$) to get $B_r(x) \subset \bigcap_\alpha U_\alpha$). Notice we cannot take infinite union because for instance, we can take $U_n := B_{1/n}(0)$ and we get $\bigcap_{n=1}^{\infty} U_n = \{0\}$ which is not an open set.

If we take complements of the above and used De Morgan's laws, we get that arbitrary intersections of closed sets are closed and finite unions of closed sets are closed.

A set E is **bounded** if there is some $r > 0$ and $x \in \mathbb{R}^n$ such that $E \subset B_r(x)$. A set is **compact** if it is closed and bounded. By Heine-Borel theorem, compactness is equivalent to having a finite cover for any 'open' cover.

A point x is a **limit point of a set** E if for every $r > 0$, $B_r(x) \cap E \neq \emptyset$, i.e. there are points in E which are arbitrarily close to x. It is easy to show that a set E is closed iff it contains all its limit points. If it is closed, then for any limit point x, there exists some $r > 0$ such that $B_r(x) \setminus \{x\} \subset E \neq \emptyset$. But E^c is characterized as the set of points for which such r does not exist, so E does indeed contain all limit points. If E is not closed, then the complement E^c is not open, so there is a point $y \in E^c$ for which for all $r > 0$, $B_r(y) \not\subset E^c$, hence y is a limit point of E which is not contained in E.

An **isolated point of a set** E is a point $x \in E$ which is not a limit point, i.e. there is some $r > 0$ such that $B_r(x) \cap E = \{x\}$.

A point $x \in E$ is an **interior point** of E if for some $r > 0$, $B_r(x) \subset E$. The **interior of the set** E is the set int(E) of all interior points of E. The **closure** $\overline{E}$ is the union of E with all the limit points of E. From the limit point characterization of closed sets, the closure of a set is a closed set. On the other hand, a closed set contains all its limit points, so the closure of a closed set is itself. The **boundary of a set** E is the set $\partial E := \overline{E} \setminus$ int(E). Notice that for any $E \subset \mathbb{R}^n$, a point $x \in \mathbb{R}^n$ is either an interior point of E, a boundary point of E, or an interior point of the complement E^c. Note that by definition of boundary, $\overline{E} = E \cup \partial E$, and so if x is not in $\overline{E}$, then it is in the open set $\overline{E}^c$, and in particular, in the interior of E^c. If $x \in \overline{E}$, then by definition of the boundary again, it is either in the interior of E or in the boundary. We see from here that the interior of a set is open since interior points are not in the closure of the complement. On the other hand, interior of an open set is itself, by definition of an open set. Therefore, the interior of a set is open iff the set is an open set.

A set E is **perfect** if it is closed and has no isolated points, or in other words, $x \in E$ iff x is a limit point of E.

Volumes of Rectangles and Cubes.

Definition 1. A **(closed) rectangle** $R \subset \mathbb{R}^n$ is the set

$$R := [a_1, b_1] \times ... \times [a_n b_n] \tag{7}$$

for $-\infty < a_i \leq b_i < \infty,\ i = 1, ..., n$. We call a rectangle a **cube** if $b_1 - a_1 = ... = b_n - a_n$.

When we say "cube" or "rectangle," we will assume that it is a *closed* cube or a rectangle.

Remark 2. Note that in our definition of a rectangle, R is

- closed, and
- the sides are parallel to the coordinate axis

In $\mathbb{R}$, the rectangles are just the closed, bounded intervals which are identical to the open balls in $\mathbb{R}$. (This is often what distinguishes $\mathbb{R}$ from $\mathbb{R}^2$.) For $\mathbb{R}^2$, they are four-sided rectangles.

Definition 3. The **volume of the rectangle** R is the number

$$|R| := \prod_{i=1}^{n} (b_i - a_i) \tag{8}$$

Definition 4. An **open rectangle** is the set

$$(a_1, b_1) \times ... \times (a_n, b_n) \tag{9}$$

Definition 5. A union of rectangles is **almost disjoint** if the interiors of the rectangles are disjoint.

Here is an easy lemma that can occasionally be helpful.

Proposition 6. Finite intersection of rectangles are rectangles.

PROOF 7. It suffices to show that the intersection of two rectangles are rectangles. (We can then iterate this to get the full result.) Let

$$R_1 := \prod_{d=1}^{n} [a_{1,d}, b_{1,d}],\ R_2 := \prod_{d=1}^{n} [a_{2,d}, b_{2,d}] \tag{10}$$

and assume that they intersect (i.e. $[a_{1,d}, b_{1,d}] \cap [a_{2,d}, b_{2,d}] \neq \emptyset$ for all d.) Consider the partition of the x_d-axis given by $a_{1,d}, b_{1,d}, a_{2,d}, b_{,d}$. We can order this to get $s_{1,d} \leq s_{2,d} \leq s_{3,d} \leq s_{4,d}$.[3] The key here is that $[s_{2,d}, s_{3,d}] = [a_{1,d}, b_{1,d}] \cap [a_{2,d}, b_{2,d}]$. WLOG, take $a_{1,d} \leq b_{2,d}$. Since $R_1 \cap R_2 \neq \emptyset$

$$a_{1,d} \leq a_{2,d} \leq b_{1,d} \leq b_{2,d} \tag{11}$$

for $d = 1, ..., n$. Thus, $[s_{2,d}, s_{3,d}] = [a_{1,d}, b_{1,d}] \cap [a_{2,d}, b_{2,d}]$.

Then we claim that

$$R_1 \cap R_2 = \prod_{d=1}^{n} [s_{2,d}, s_{3,d}] =: R \tag{12}$$

The RHS is a rectangle, so if we prove this equality, we are done. This falls right out of our definition of $s_{i,d}$. One inclusion is obvious; we must have $R_1 \cap R_2 \supseteq R$ since for $p := (p_1, ..., p_n) \in R$, $p_d \in [s_{2,d}, s_{3,d}] = [a_{1,d}, b_{1,d}] \cap [a_{2,d}, b_{2,d}]$ by how we defined $s_{i,d}$.

On the other hand, if $p \in R_1 \cap R_2$, then $p_i \in [a_{1,d}, b_{1,d}] \cap [a_{2,d}, b_{2,d}] = [s_{2,d}, s_{3,d}]$, so we are done.

$\square$

Now let's build some machinery.

Proposition 8 (Finite additivity of volume of almost disjoint rectangles.). If a rectangle R is the almost disjoint union of finitely many other rectangles, i.e. $R = \bigcup_{k=1}^{N} R_k$, then

$$|R| = \sum_{k=1}^{N} |R_k| \tag{13}$$

PROOF 9. Note that R_k gives a partition of R. The main idea is to consider refinement of this partition. Since rectangles are defined to have sides parallel to the coordinate axis, the following is perfectly well-defined. Consider the refinement $\tilde{R}_1, ..., \tilde{R}_M$ of R given by extending the sides of $R_1, ..., R_N$, and let $J_1, ..., J_N$ be a partition of M so that

$$R = \bigcup_{j=1}^{M} \tilde{R}_j \qquad \text{and} \qquad R_k = \bigcup_{j \in J_k} \tilde{R}_j \tag{14}$$

for $k = 1, ..., N$ are almost disjoint. In other words, J_k are the indices for which $\tilde{R}_j$ is a partition of R_k.

Then

$$|R| = \sum_{j=1}^{M} \left|\tilde{R}_j\right| = \sum_{k=1}^{N} \sum_{j \in J_k} \left|\tilde{R}_j\right| = \sum_{k=1}^{N} |R_k| \tag{15}$$

The first equality holds simply by looking at the picture. The others hold by how we defined $\tilde{R}_j$.

$\square$

Proposition 10 (Finite subadditivity of volume of rectangles.). If $R, R_1, ..., R_N$ are rectangles such that $R \subset \bigcup_{k=1}^{N} R_k$, then

$$|R| \leq \sum_{k=1}^{n} |R_k| \tag{16}$$

[3] This procedure amounts to what Stein-Shakarchi calls "extend the sides of the rectangles and consider the grid formed by them."

PROOF 11. The main idea is the same as in the previous proof. Once again, we take the refinement of the partition of $\bigcup_{k=1}^N R_k$. We get

$$|R| \leq \sum_{j=1}^{M} \left|\tilde{R}_j\right| \leq \sum_{k=1}^{N} \sum_{j \in J_k} \left|\tilde{R}_j\right| = \sum_{k=1}^{N} |R_k| \tag{17}$$

The first inequality follows from the inclusion in the hypothesis of the theorem. The second follows from the fact that $J_1, ..., J_N$ need not be disjoint anymore since the rectangles $R_1, ..., R_N$ need not be disjoint.

□

Here is a standard theorem from point set topology of $\mathbb{R}$. (Note that this does NOT hold in $\mathbb{R}^2$ as we will see later.)

Proposition 12. Every open set $\mathcal{O} \subset \mathbb{R}$ is a countable union of disjoint open intervals.

PROOF 13. The countability comes from the separability of $\mathbb{R}$ (i.e. the rationals $\mathbb{Q}$). The work is in showing the disjoint union.

For each $x \in \mathcal{O}$, let I_x denote the largest open interval containing x and contained in $\mathcal{O}$. More precisely, x is contained in some small interval contained in $\mathcal{O}$, and so, if

$$a_x := \inf\left\{a < x : (a, x) \subset \mathcal{O}\right\}, \qquad b_x := \sup\left\{b > x : (x, b) \subset \mathcal{O}\right\} \tag{18}$$

we must have $a_x < x < b_x$ (allowing for a_x, b_x to be $\pm\infty$). If we let $I_x := (a_x, b_x)$, then by construction $x \in I_x$ and $I_x \subset \mathcal{O}$. Thus, $\mathcal{O}$ is the union of I_x for all $x \in \mathcal{O}$.

To show that for distinct I_x, I_y are disjoint, we suppose that $I_x \cap I_y \neq \emptyset$. Then $I_x \cup I_y$ is an open interval in $\mathcal{O}$ containing x, and by maximality of I_x, it is contained in I_x, and likewise, in I_y. This implies that $I_x = I_y$ which gives us the claim.

For countability of $\{I_x\}$, notice that the open set I_x must intersect $\mathbb{Q}$, by denseness. But since the intervals are disjoint, each I_x contains a *distinct* rational, so there are countably many intervals.

□

It is natural to let the "length" of $\mathcal{O} = \bigcup_{k=1}^{\infty} I_k$ to be $\sum_{k=1}^{\infty} |I_k|$. The above thus gives the Borel measure of an arbitrary open set $\mathcal{O} \subset \mathbb{R}$ via σ-additivity (which we define later). However, we still do not know how to extend the notion of length to other sets on $\mathbb{R}$. Also, as remarked earlier, the above theorem does not hold for higher dimensions.

Proposition 14. [4] Proposition 12 is false for $\mathbb{R}^d,\ d \geq 2$:

1. An open disc in $\mathbb{R}^2$ is not the disjoint union of open rectangles.
2. An open connected set Ω is the disjoint union of open rectangles iff Ω is itself an open rectangle.[5]

PROOF 15 (Proof of 1.). The idea is that there will be a gap between the open rectangles.

WLOG, consider the unit disc $D := B_1(0)$, and assume for contradiction that it is an arbitrary union of open rectangles G_α. Since $\partial G_\alpha \neq \partial D$[6], there is $x \in \partial G_\alpha \cap D$ (i.e. a point on the boundary of the rectangle that does not lie on the boundary of the disc). But since x is a point in D, there is some G_β such that $x \in G_\beta$. Then x is a boundary point of G_α, so any open set containing x intersects G_α. So, $G_\alpha \cap G_\beta \neq \emptyset$ which is a contradiction.

[4] This proposition is the content of Exercise 1.12.

[5] Notice that the openness of Ω is redundant since a union of open rectangles must be open.

[6] One way to show this technical point is by observing that there are less than two points on ∂D sharing an x coordinate (or equivalently, a vertical line intersects a circle at most two points) whereas for a rectangle there can be uncountably many points (if the vertical line overlaps with one of the sides of the rectangle).

□

PROOF 16 (Proof of 2.). There is nothing to prove for the backward direction. For the forward direction, we want to show that if $\Omega = \bigcup_{\alpha \in A} G_\alpha$ is open connected, and G_α are disjoint, then in fact there is only one rectangle: $G_\alpha = \Omega$.

Suppose that $|A| \geq 2$. By connectedness, $\overline{G_{\alpha'}} \cap \bigcup_{\alpha \neq \alpha'} G_\alpha \neq \emptyset$. But then we can repeat the same argument as in the previous proof by taking $x \in \partial G_{\alpha'}$: there is some G_α for which $x \in G_\alpha$, but then $G_\alpha \cap G_{\alpha'} \neq \emptyset$ by definition of boundary, and so, this violates disjointness. Thus, $|A| = 1$ which gives the conclusion.

□

Here is a weaker result than proposition 12 for higher dimensions.

Proposition 17. Every open set $\mathcal{O}$ of $\mathbb{R}^n$ for $n \geq 1$ can be written as a countable union of *almost* disjoint *closed cubes*.

PROOF 18. We use the typical strategy to tackle theorems of this form: propose a construction for such a collection Q of cubes, and prove that the construction endows this collection with the desired properties.

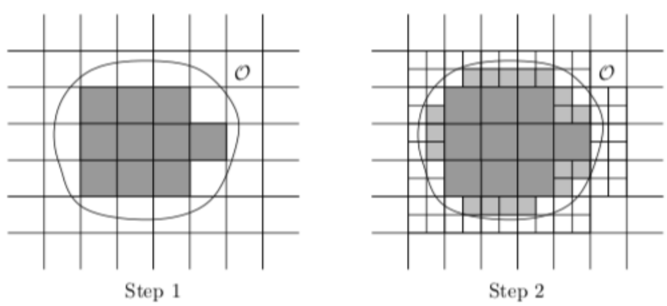
Let Q_1 be the set of all closed cubes of side length 1 contained in $\mathcal{O}$, and let Q_1' be the closed cubes that are not contained in $\mathcal{O}^c$ and not in Q_1.

Next, consider the partitions of the closed cubes in Q_1' into 2^2 cubes of side length $\frac{1}{2}$. Let Q_2 be the closed cubes which are contained in $\mathcal{O}$, and let Q_2' be those that are not contained in $\mathcal{O}^c$ and not in Q_2.

We let Q_N be the closed cubes given by partitioning each cube in Q_{N-1}' into 2^N cubes of side length 2^{-N+1} which are contained $\mathcal{O}$. We let Q_N' be the cubes in the partitions that are not contained in $\mathcal{O}^c$ and not in Q_N.

We then let $Q := \bigcup_{N=1}^\infty Q_N$.

Figure 1: First two steps for $\mathbb{R}^2$.



Q is obviously countable since it is a countable union of finite sets. The cubes are almost disjoint since for each N, the cubes in $Q_N \cup Q_N'$ are pairwise almost disjoint by inspection. (This follows from how we partitioned each cube.)

Finally, to see that the union of the the cubes in Q is $\mathcal{O}$, we note that if $x \in Q$, then we can partition all of $\mathbb{R}^n$ into cubes with side length 2^{-N} (so that it is a refinement of the original partition of side length 1 at the beginning of the proof), and so, in particular, x is contained in one of them. But (by openness of $\mathcal{O}$) there is some N such that this cube is contained inside $\mathcal{O}$. Then this cube is either contained in Q or is contained by some cube of side length 2^{-m} for $m < N$ by construction of Q.[7] Therefore, x is in some cube in Q, and so, indeed Q covers $\mathcal{O}$

[7] Notice that this part of the argument is really what drove the construction of Q.

□

Notice that using the above construction, we can assign the volume $\sum_{j=1}^{\infty} |R_j|$ to $\mathcal{O}$ (since it is plausible that the boundary contribute 0 volume). However, we did not show that this decomposition of $\mathcal{O}$ is *unique*. (This is actually easy to resolve by considering the refinement, once we have the formalism for this notion of volume.)

Here are some additional (albeit, geometrically obvious) lemmas.

Add some of the trivial geometric statements used in the exterior measure subsection that needs formal justifying. Make sure to remark on obviousness via geometric point of view.

Proposition 19. [8] Let $\epsilon > 0$ and Q be a cube in $\mathbb{R}^n$. Then there is some cube $\tilde{Q}_j$ contained in Q_j such that

$$|Q_j| \leq \left|\tilde{Q}_j\right| + \epsilon \tag{19}$$

This just means that there is a cube inside Q_j close enough to it in volume that it exceeds it in volume if given an ϵ-error.

PROOF 20. Wlog (via scaling), assume that Q_j is the cube with side lengths 1. Take N large enough so that $\frac{\epsilon}{2} > \frac{n}{N}$. Then take $\tilde{Q}_j$ be the cube with side lengths $1 - \frac{1}{N}$. Then

$$\begin{aligned}
|Q_j| - \left|\tilde{Q}_j\right| &= 1 - \left(1 - \frac{1}{N}\right)^n \\
&= 1 - \left(1 - \frac{n}{N} + O\left(\frac{1}{N^2}\right)\right) \\
&= \frac{n}{N} + O\left(\frac{1}{N^2}\right) \\
&\leq \epsilon
\end{aligned}$$

for a large enough choice of N. Notice we used the binomial theorem in the second line.

□

The Cantor Set.

The Cantor set is a very important example from point-set topology that is useful for constructing interesting examples in analysis (in addition to being an interesting example in itself).

Let's recall its construction. Let $C_0 := [0, 1]$. Take

$$\begin{aligned}
C_1 &:= [0, 1/3] \cup [2/3, 1] \\
C_2 &:= [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1] \\
&\vdots \\
&\vdots
\end{aligned}$$

and in general,

$$C_n := \bigcup_{k=0}^{3^{n-1}-1} \left(\left[\frac{3k}{3^n}, \frac{3k+1}{3^n}\right] \cup \left[\frac{3k+2}{3^n}, \frac{3k+3}{3^n}\right] \right)$$

[8] Note that this is different from the ϵ-definition of exterior measure since that would give the existence of Q_j given $\tilde{Q}_j$.

The **Cantor set** is then defined to be

$$C := \bigcup_{n=1}^{\infty} C_n \tag{20}$$

Notice that by construction, C_n is a decreasing sequence of compact sets, and so, their intersection is nonempty. A more transparent way to see this is to note that the endpoints of the intervals in C_n are in C.

Here are some properties of the Cantor set.

Proposition 21 (Properties of the Cantor set.). [9] Let C be the Cantor set. Then C is

1. compact
2. totally disconnected
3. perfect
4. uncountable
5. Lebesgue null set

PROOF 22 (Proof for 1 through 3.). The compactness is trivial since C is closed (since it is an arbitrary intersection of closed sets) and is bounded (since it is contained in $[0, 1]$).

Totally disconnectedness is also obvious since if $x, y \in C$ are distinct points, then for some k, $|x - y| > \frac{1}{3^k}$. Then x, y cannot belong in the same interval in C_k, so they belong in distinct connected components of C.

C is closed as noted before. On the other hand, if $x \in C$ then for any $k \geq 1$, there is some interval I_k in C_k so that $x \in I_k$. Taking one of its endpoints y_k, we then have $|x - y_k| \leq \frac{1}{3^k}$. Therefore, $y_k \to x$, and so, x is a limit point of C.

□

The fifth assertion is trivial if we observe that the intervals taken out at each step forms a geometric series, and thus, the total measure of these intervals is 1. (But of course, we have not defined what a measure is yet.)

For the fourth assertion, there are a few ways to prove it:

Proof 1. Cantor-Lebesgue function. The Cantor-Lebesgue function (see chapter 1 exercise 2) is a surjective map from the Cantor set onto $[0, 1]$, so the cardinality of the Cantor set must at least be the cardinality of $[0, 1]$. Therefore, the Cantor set is uncountable.

Proof 2. Cantor-like set. We can also show that the Cantor set is perfect, which is not too hard to show. (Indeed, this is how we show that the Cantor-like set is perfect.) It is closed, by construction, and we can take a point at each step in the construction so that the sequence converges to a point on the Cantor set.

We can see this proof as a special case of the property of the Cantor-like set, since (as we show in chapter 1 exercise 4 c) the Cantor-like set is perfect. In terms of generality, this is the most general proof out of the three (in that it extends to the Cantor-like set).

Proof 3. Ternary expansion. Alternatively, we can use the following propositions, and note that the set of ternary expansions in the proposition is uncountable (by the same reason that binary expansions of of numbers in $[0, 1]$ is uncountable).

[9] Exercise 1.1, 1.2.

Proposition 23 (Ternary Expansion Description of the Cantor Set.). x is a point in the Cantor set iff x has a ternary expansion

$$x = \sum_{k=1}^{\infty} a_k 3^k \tag{21}$$

where $a_k = 0, 2$.

PROOF 24. We assume that every number in $[0, 1]$ has a ternary expansion.[10]

Observe that

$$\begin{aligned}
x \in C &\iff \forall k \geq 0 \qquad x \in C_k \\
&\iff \forall k \geq 0 \qquad x \in \left(\left[\frac{3m}{3^k}, \frac{3m+1}{3^k} \right] \cup \left[\frac{3m+2}{3^k}, \frac{3m+3}{3^k} \right] \right) \text{ for } m = 0, ..., k-1 \\
&\iff \forall k \geq 0 \qquad x \notin \left[\frac{3m+1}{3^k}, \frac{3m+2}{3^k} \right] \text{ for } m = 0, ..., k-1 \\
&\iff \forall k \geq 0 \qquad a_k \neq 1
\end{aligned}$$

In the last step, we observed that

$$\begin{aligned}
x \in \left[\frac{3m+1}{3^k}, \frac{3m+2}{3^k} \right] &\iff x \in \left[\frac{m}{3^{k-1}}, \frac{m+1}{3^{k-1}} \right] \quad \text{and} \quad a_k = 1 \\
&\iff x \in \left[\frac{m}{3^{k-1}}, \frac{m+1}{3^{k-1}} \right] \quad \text{and} \quad a_k = 1
\end{aligned}$$

by construction. This proves the claim.

□

Exterior Measure. -**Wednesday, 6.20.2018**

We now start developing the notion of measures. The two ingredients are exterior measures (which are defined for *all* subsets of $\mathbb{R}^n$) and measurable sets (which we will develop in the next subsection).

Basic Examples.

The exterior measure assigns to *any* subset of $\mathbb{R}^n$ a notion of size by approximating from the outside by rectangles. However, this notion is not strong enough since it only has σ-subadditivity. (The notion of measurable sets is essential to upgrading this to σ-additivity.)

Definition 25. Let E be *any* subset of $\mathbb{R}^n$. The **exterior measure** (or **outer measure**) of E is the number[11]

$$m_*(E) := \inf \left\{ \sum_{j=1}^{\infty} |Q_j| : E \subseteq \bigcup_{j=1}^{\infty} Q_j,\ Q_j \text{ are closed cubes} \right\} \tag{22}$$

Notice that the exterior measure is in $[0, \infty]$. This definition is equivalent to the following.

[10] Note that this representation is not unique since $\frac{1}{3} = \sum_{k=2}^{\infty} \left(\frac{2}{3}\right)^k$.

[11] We call coverings with closed cubes a **cube covering** in these notes (and likewise, rectangle covering, sphere covering, etc.)

Proposition 26 (ϵ-definition of exterior measure.). For all $\epsilon > 0$, there exists a covering of cubes $E \subset \bigcup_{j=1}^{\infty} Q_j$ such that

$$\sum_{j=1}^{\infty} m_*(Q_j) \leq m_*(E) + \epsilon \tag{23}$$

Intuitively, this just means that there is a cover small enough that the volume of the cover is slightly smaller than the exterior measure plus an ϵ error.

The proof is just the definition of the infimum.
We address some subtleties of this definition. The first point is that we do require an infinite sum (rather than a finite sum) in the definition of exterior measure.[12]

Definition 27. For any $E \subseteq \mathbb{R}$, the **outer Jordan content** is

$$J_*(E) := \inf \sum_{j=1}^{N} |I_j| \tag{24}$$

where the infimum is taken over every finite covering $E \subseteq \bigcup_{j=1}^{N} I_j$ by closed intervals I_j.

Proposition 28. For all $E \subset \mathbb{R}$,

$$J_*(E) = J_*(\overline{E}) \tag{25}$$

PROOF 29. Come back to this. □

Remark 30. Notice that $E := [0,1] \cap \mathbb{Q}$ satisfies $J_*(E) = 1$ and $m_*(E) = 0$. On the one hand, if $I := \bigcup_{j=1}^{N} I_j \neq [0,1]$ for any covering I, then I is a finite union of closed sets, so it is closed, hence $[0,1] \setminus I$ is open, hence there are rational points not in I. (The finite union is the difference between the J_* and m_*). Thus, $J_*(E) = 1$.

For m_*, we can simply cover $[0,1] \cap \mathbb{Q}$ with the rational points (which are closed intervals of length 0), and by the countability of the rationals and via the subadditivity of the outer measure, $m_*(E) = 0$.

The second point is that we can replace the cubes by rectangles or balls in the covering. We will show this for the rectangle here.

Proposition 31 (Outer measure can equivalently be defined in terms of rectangles.). Let

$$m_*^{\mathcal{R}}(E) := \inf \left\{ \sum_{j=1}^{\infty} |R_j| : E \subseteq \bigcup_{j=1}^{\infty} R_j,\ R_j \text{ are closed rectangles} \right\} \tag{26}$$

Then $m_*^{\mathcal{R}}(E) = m_*(E)$ for every $E \subset \mathbb{R}^n$.

PROOF 32. Come back to this. □

Let's observe some very basic examples of exterior measure. As we will see, it is a mix of topological and measure theoretic arguments.

Example 33 (Point.). $m_*(\{x\}) = 0$ since a point is a rectangle of side length 0.

[12] The following definition and proposition are from exercise 1.14.

Example 34 (Closed cube.). We claim that the exterior measure of a closed cube is equal to its volume, i.e. $m_*(Q) = |Q|$.

A common strategy for proving that an (exterior) measure is equal to some quantity is to show both sides of the inequality. One side is obvious: since Q overs itself, $m_*(Q) \le |Q|$.

Thus, we need to show $m_*(Q) \ge |Q|$. Consider an arbitrary covering $Q \subset \bigcup_{j=1}^{\infty} Q_j$ by cubes, and note that (by definition of infimum), it suffices to show that

$$|Q| \le \sum_{j=1}^{\infty} |Q_j| \tag{27}$$

Here is another common strategy. Fix $\epsilon > 0$, and choose an open cube S_j which contains Q_j so that $|Q_j| \le |S_j| \le (1+\epsilon)|Q_j|$. We will show the above equality with $|Q_j|$ replaced by $(1+\epsilon)|Q_j|$ and then passing to the limit.

Since Q_j cover Q and S_j contain Q_j, S_j is an open cover of Q. By compactness of Q, we can reduce this to a finite open cover: $j = 1, ..., N$ (with some relabeling of indices). (This part is a topological argument.) Now, by proposition 10 from the previous subsection, we can bound the volume of Q by S_j and hence by $(1+\epsilon)|Q_j|$ (this part is measure theoretic):

$$|Q| \le \sum_{j=1}^{N} |S_j| \le \sum_{j=1}^{N} (1+\epsilon)|Q_j| \le \sum_{j=1}^{\infty} (1+\epsilon)|Q_j| \tag{28}$$

Passing to the limit $\epsilon \to 0$, we get

$$|Q| \le \sum_{j=1}^{\infty} |Q_j| \tag{29}$$

as required.

Example 35 (The exterior measure of an open cube is its volume.). If Q is instead an open cube, the result $m_*(Q) = |Q|$ still holds. Once again, we prove both sides of the inequality. $m_*(Q) \le |Q| = |\overline{Q}|$ is immediate by definition of the volume of an open rectangle.

The other direction follows from a topological fact. Take an increasing sequence of closed cubes Q_n so that $Q = \bigcup_n Q_n$. Then we can see by inspection[13] that $\lim_{n\to\infty} |Q_n| = |Q|$. But if Q_n is a closed cube contained in Q, then $|Q_n| = m_*(Q_n) \le m_*(Q)$ since any covering of Q cover Q_n. Therefore, $|Q| \le m_*(Q)$.

Example 36 (Exterior measure of a rectangle is equal to its volume.). The exterior measure of a rectangle is equal to its volume. This follows immediately by applying example 34 but now instead using $m_*^{\mathcal{R}}$ from proposition 31 and replacing cubes with rectangles.

Alternatively, we can show this directly. We can replace the cube Q in example 34 with a rectangle and the proof works verbatim to get $|R| \le m_*(R)$.

To show the reverse inequality, fix $k \in \mathbb{N}$ and partition $\mathbb{R}^n$ into cubes of side length $\frac{1}{k}$. The idea is to show that the collection $\mathcal{Q}$ of all cubes contained completely inside R[14] contributes all of the volume of R in the limit $k \to \infty$, and so, we need to bound the approximation (from the above) of $|R|$ by the volumes of the cubes.

---

[13] We can make this perfectly rigorous. Wlog, take $|Q|$ be a cube with side lengths 1, and let Q_n be cubes with side lengths $1 - \frac{1}{n}$. Then the volume of Q_n is $\left(1 - \frac{1}{n}\right)^2 \to 1$ as $n \to \infty$.

[14] To be precise, $\mathcal{Q} := \{I \in C : I \subset R\}$ where I is the collection of all cubes of side length $\frac{1}{k}$.

Let $\mathcal{Q}'$ be the cubes that intersect both R and R^c.[15] Now, by this definition, R is covered by $\mathcal{Q} \cup \mathcal{Q}'$. On the other hand, since the elements of $\mathcal{Q}$ are almost disjoint[16], and the elements are all contained inside R,

$$\sum_{Q \in \mathcal{Q}} |Q| \leq |R| \tag{30}$$

Moreover, there are $O(k^{n-1})$ cubes in $\mathcal{Q}'$ since the boundary of R is an $n-1$ dimensional surface. Each of these cubes have volume k^{-n}, so

$$\sum_{Q \in \mathcal{Q}'} |Q| = O(1/k) \tag{31}$$

Therefore,

$$m_*(R) \leq \sum_{Q \in \mathcal{Q} \cup \mathcal{Q}'} |Q| \leq |R| + O(1/k) \tag{32}$$

and taking the limit $k \to \infty$ gives the desired result.

Example 37 (Exterior measure of $\mathbb{R}^n$.). The exterior measure of $\mathbb{R}^n$ is infinite. We just show that $\mathbb{R}^n$ contains arbitrarily large cubes.

If we have a covering for $\mathbb{R}^n$ by cubes, then this is a covering for any $Q \subset \mathbb{R}^n$. Thus, $|Q| \leq m_*(\mathbb{R}^n)$. But if we take Q to be a cube of side length $N \in \mathbb{N}$, then $N^n \leq m_*(\mathbb{R}^n)$, so $m_*(\mathbb{R}^n) = \infty$.

Example 38 (Cantor set has exterior measure 0.). The Cantor set has exterior measure 0. Often a good strategy for tackling the Cantor set is to look at the steps in its construction C_k. We know from construction that C_k is a disjoint union of 2^k closed intervals each of length 3^{-k}. Therefore, $\left|C^k\right| = \left(\frac{2}{3}\right)^k$. But since the Cantor set is contained in C_k,

$$m_*(C) \leq \left(\frac{2}{3}\right)^k \tag{33}$$

for all k. So, $m_*(C) = 0$.

Properties of the Exterior Measure.

We present some basic properties of exterior measure. We see here how useful the ϵ-definition of exterior measure is.

The following proposition becomes essential in the next subsection since the Lebesgue measure is defined in terms of the exterior measure.

Proposition 39 (Properties of Exterior Measures.). Let E, E_j be sets in $\mathbb{R}^n$.

1. (**Monotonicity.**) If $E_1 \subset E_2$, then $m_*(E_1) \leq m_*(E_2)$.
2. (**σ-subadditivity.**) If $E = \bigcup_{j=1}^{\infty} E_j$, then $m_*(E) \leq \sum_{j=1} 6\infty m_*(E_j)$.
3. $$m_*(E) = \inf_{\substack{E \subseteq \mathcal{O} \\ \mathcal{O} \text{ open}}} m_*(\mathcal{O}) \tag{34}$$

[15] i.e., $\mathcal{Q}' := \{I \in C : I \cap R \neq \emptyset,\ I \cap R^c \neq \emptyset\}$

[16] This is the nice feature of the partition-into-cubes arguments; the partitions are always almost disjoint, so we can simply add up the volume.

4. If $E = E_1 \cup E_2$ and $d(E_1, E_2) > 0$, then

$$m_*(E) = m_*(E_1) + m_*(E_2) \tag{35}$$

5. If $E = \bigcup_{j=1}^{\infty} Q_j$, Q_j almost disjoint, then

$$m_*(E) = \sum_{j=1}^{\infty} |Q_j| \tag{36}$$

Here is an obvious consequence of the the first statement.

Proposition 40 (Every bounded subset of $\mathbb{R}^d$ has finite exterior measure.). If E is a bounded set, say $E \subset B_r(0)$, then $m_*(E) \leq m_*(B_r(0))$.

Note that the converse of this fails.

Example 41. Consider the set

$$\left\{(x, y) \in \mathbb{R}^2 : x \geq 0,\ y \leq \frac{1}{\lceil x \rceil^2}\right\} \tag{37}$$

i.e., the region under $y = \frac{1}{n^2}$, $\lfloor x \rfloor = n$. The (exterior) measure of this is clearly finite ($\frac{\pi^2}{6}$), but the set is certainly not bounded.

Here is an important consequence of the fifth statement.

Proposition 42 (The exterior measure of an open set equals the sum of the volumes of the cubes in the decomposition.). Let $\mathcal{O} = \bigcup_{j=1}^{\infty} Q_j$ be an open set in $\mathbb{R}^n$ where Q_j are almost disjoint cubes. Then

$$m_*(\mathcal{O}) = \sum_{j=1}^{\infty} m_*(Q_j) \tag{38}$$

Moreover, the sum is independent of the decomposition.

PROOF 43. We know that such a decomposition exists from proposition 17 from the previous subsection. Such a decomposition satisfies the hypothesis of the fifth part of the above proposition, so we do indeed get σ-additivity in this case. The independence of the sum is immediate because regardless of the decomposition, the sum are all equal to $m_*(\mathcal{O})$ and in particular, to each other.

□

The following proofs are very similar (except the first which is fairly easy to begin with). They all go back to using the ϵ-definition of exterior measure and then proving some bounds using point-set topology-esque arguments. The later three are stronger versions of the first two in the sense that one direction of the inequality follows from them.

PROOF 44 (Proof of 1.). Let Q_j be a cube covering of E_2. Then E_2 contains E_1, so it is a covering of E_1 as well. So $m_*(E_1)$ is an infimum over a larger set of coverings, so this gives $m_*(E_1) \leq m_*(E_2)$.

□

PROOF 45 (Proof of 2.). There is nothing to prove if $m_*(E_j) = \infty$, so assume finite exterior measure.

Fix $\epsilon > 0$. Then the ϵ-definition of exterior measure gives a cube covering $E_j \subset \bigcup_{k=1}^{\infty} Q_{k,j}$ satisfying

$$\sum_{k,j} |Q_{k,j}| \leq m_*(E_j) + \frac{\epsilon}{2^j} \tag{39}$$

The error $\frac{\epsilon}{2^j}$ may at first seem strange.[17] However, this is a common trick one uses when we ultimately have to sum $m_*(E_j)$ (in which case, by geometric series, we get back an error of ϵ).

Since $Q_{k,j}$ is a cube covering of E, we have

$$\begin{aligned}
m_*(E) &\leq \sum_{k,j} |Q_{k,j}| \\
&= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} |Q_{k,j}| \\
&\leq \sum_{j=1}^{\infty} m_*(E_j) + \frac{\epsilon}{2^j} \\
&\leq \epsilon + \sum_{j=1}^{\infty} m_*(E_j)
\end{aligned}$$

and taking $\epsilon \to 0$, we are done.

□

PROOF 46 (Proof of 3.). We prove both directions of the inequality. $m_*(E) \leq \inf m_*(\mathcal{O})$ is immediate from monotonicity. For the other direction, once again we use the ϵ-definition of exterior measure. Half of the error comes from the cube covering of E and the other half comes from approximating each cube covering by an open set.

We take cube coverings Q_j of E so that

$$\sum_{j=1}^{\infty} |Q_j| \leq m_*(E) + \frac{\epsilon}{2} \tag{40}$$

The other half of the error comes from approximating Q_j with open cubes (via weighted error trick). Let Q_j^0 be an open cube containing Q_j so that $\left|Q_j^0\right| \leq |Q_j| + \frac{\epsilon}{2^{j+1}}$. We adjust the index in the 2^{j+1} so that the error sums up to $\frac{\epsilon}{2}$.

Thus $\mathcal{O} = \bigcup_{j=1}^{\infty} Q_j^0$ is open, so by σ-subadditivity,

$$\begin{aligned}
m_*(\mathcal{O}) &\leq \sum_{j=1}^{\infty} m_*(Q_j^0) \\
&= \sum_{j=1}^{\infty} |Q_j^0| \\
&= \sum_{j=1}^{\infty} \left(|Q_j| + \frac{\epsilon}{2^{j+1}}\right) \\
&\leq \frac{\epsilon}{2} + \sum_{j=1}^{\infty} |Q_j| \\
&\leq m_*(E) + \epsilon
\end{aligned}$$

Thus, $\inf m_*(\mathcal{O}) \leq m_*(E)$ as desired.

[17] We will call this trick the **weighted error trick** in these notes.

$\square$

PROOF 47 (Proof of 4.). Once again, we prove both directions of the inequality. By σ-subadditivity, we have

$$m_*(E) \leq m_*(E_1) + m_*(E_2) \tag{41}$$

For the other direction, the idea is to get a cube covering for all of E, but then split the covering into the disjoint parts covering E_1, E_2 respectively. From there, we can use the ϵ-definition of exterior measure.

Fix $\epsilon > 0$ and $d(E_1, E_2) > \delta > 0$. By ϵ-definition of exterior measure, choose cube coverings of E so that

$$\sum_{j=1}^{\infty} |Q_j| \leq m_*(E) + \epsilon \tag{42}$$

WLOG, we can assume that the diameter of Q_j are less than δ (since we can subdivide each Q_j to make them smaller). Since $d(E_1, E_2) > \delta$, Q_j can intersect at most one of E_1, E_2. In other words, if J_1, J_2 denote the sets of indices of Q_j which intersect E_1, E_2 respectively, then $J_1 \cap J_2 = \emptyset$. Then of course, Q_j cover E_1 for $j \in J_1$ and likewise for E_2, J_2.

Therefore,

$$\begin{aligned} m_*(E_1) + m_*(E_2) &\leq \sum_{j \in J_1} |Q_j| + \sum_{j \in J_2} |Q_j| \\ &\leq \sum_{j=1}^{\infty} |Q_j| \\ &\leq m_*(E) + \epsilon \end{aligned}$$

and so we are done. Note that we must have inequality in the second line since there may be cubes that do not intersect either E_1 nor E_2. $\square$

PROOF 48 (Proof of 5.). Of course, σ-additivity gives one direction:

$$m_*(E) \leq \sum_{j=1}^{\infty} |Q_j| \tag{43}$$

Let $\tilde{Q}_j$ be a cube strictly contained in the interior of Q_j such that $|Q_j| \leq |Q_j| + \frac{\epsilon}{2^j}$ (which exists from proposition 19 from the last subsection). Now since Q_j are almost disjoint, for every N, the cubes $\tilde{Q}_1, \tilde{Q}_2, ..., \tilde{Q}_N$ are disjoint, hence there is a finite distance between one another. (Notice we use here the fact that there are only finitely many $\tilde{Q}_j$.) This satisfies the condition of the part 4 of the proposition, and so by repeated application of part 4,

$$m_*(E) \geq m_* \left(\bigcup_{j=1}^{N} \tilde{Q}_j \right) = \sum_{j=1}^{N} \left| \tilde{Q}_j \right| \geq \sum_{j=1}^{N} \left(|Q_j| - \frac{\epsilon}{2^j} \right) \tag{44}$$

and passing to the limit $N \to \infty$ gives

$$m_*(E) \geq \sum_{j=1}^{\infty} |Q_j| - \epsilon \tag{45}$$

and so,

$$m_*(E) \geq \sum_{j=1}^{\infty} |Q_j| \tag{46}$$

$\square$

Measurable Sets and Lebesgue Measure. -Friday, 6.22.2018

In the previous subsection, we looked at some basic properties of the exterior measure. In particular, from parts 4 and 5 of proposition 39 (henceforth referred to as "property of exterior measure"), we suspect that perhaps we can get

$$m_*(E_1 \cup E_2) = m_*(E_1) + m_*(E_2) \tag{47}$$

for disjoint sets E_1, E_2 in $\mathbb{R}^n$. In fact, this is *not* the case, and this happens for "sufficiently nice" (i.e. measurable) sets in the sense we will now define.

Basic Properties of the Lebesgue Measurable Sets.

There are many different ways of defining measurability, but they all turn out to be equivalent.[18]

Definition 49. A set $E \subset \mathbb{R}^n$ is **(Lebesgue) measurable** if for all $\epsilon > 0$, there is an open set $\mathcal{O}$ containing E such that

$$m_*(\mathcal{O} \setminus E) \leq \epsilon \tag{48}$$

So, loosely speaking, a Lebesgue measurable sets are subsets of $\mathbb{R}^n$ which can be well-approximated by open sets. Notice that open sets are then Lebesgue measurable. (We can take $\mathcal{O}$ to be the set itself.)

Definition 50. Let E be a Lebesgue measurable set. We define the **(Lebesgue) measure of** E to be the number

$$m(E) := m_*(E) \tag{49}$$

We call a set E a **(Lebesgue) null set** if $m_*(E) = m(E) = 0$.[19]

Notice that the Lebesgue measure satisfies the properties in proposition 39 (monotonicity, σ-subadditivity, etc.) from the previous subsection since we *defined* it to be the exterior measure for a restricted collection of subsets $\mathbb{R}^n$.

Remark 51. We will see this later, but the notion of measure does not simply exist for the sake of developing the integral. More specifically, there are many theorems in the theory of the Lebesgue integral (Tonelli, Lebesgue density, etc.) which are statements about integrals which says a lot about measurability of sets, often done by the trick

$$m(E) = \int \chi_E \tag{50}$$

This may be a phenomena we do not quite see in classical calculus in which we never talk about rectangles again after talking about upper and lower sums. This may be the reflection of the sophistication that measurable sets have over plain old rectangles.

We now find means of constructing measurable sets. We take $\mathcal{O}$ to be the open set in the definition of Lebesgue measurable sets.

Proposition 52 (Open sets are measurable.).

PROOF 53. Take $\mathcal{O}$ to be the set itself.

□

[18] For instance, in Bass, *Real Analysis for Graduate Students*, we define measurable sets to be elements of a σ-algebra. Here, we will prove the properties of a σ-algebra as *theorems* instead of them being parts of the definition of a σ-algebra.

[19] As we will see later, if $m_*(E) = 0$ implies that E is measurable, so what we have here is redundant.

Proposition 54 (Sets of exterior measure 0 are measurable.). Moreover, if F is a subset of a set of exterior measure 0, then it is measurable.

PROOF 55. Let E have exterior measure 0. Then for any $\epsilon > 0$, by part 3 of the property of exterior measure, there is $\mathcal{O}$ containing E such that $m_*(\mathcal{O}) \leq \epsilon$, and by monotonicity,

$$m_*(\mathcal{O} \setminus E) \leq m_*(\mathcal{O}) \leq \epsilon \tag{51}$$

which by definition of measurable sets implies that E is measurable.

If $F \subset E$, then by monotonicity of exterior measure, F has exterior measure 0, and so it too is measurable.

□

Here is an immediate consequence.

Proposition 56 (Cantor set is measurable and is a Lebesgue null set.).

Exercise 1.3, 1.4

Proposition 57 (Countable union of measurable sets is measurable.).

PROOF 58. We use the weighted error trick. Suppose $E = \bigcup_{j=1}^{\infty} E_j$ for measurable sets E_j. By definition of measurability, we can then take open set $\mathcal{O}_j$ such that

$$m_*(\mathcal{O}_j \setminus E_j) \leq \frac{\epsilon}{2^j} \tag{52}$$

Take $\mathcal{O} := \bigcup_{j=1}^{\infty} \mathcal{O}_j$. Then $\mathcal{O} \setminus E \subseteq \bigcup_{j=1}^{\infty}(\mathcal{O}_j \setminus E_j)$, so by monotonicity of Lebesgue measure,

$$m_*(\mathcal{O} \setminus E) \leq \sum_{j=1}^{\infty} m_*(\mathcal{O}_j \setminus E_j) \leq \epsilon \tag{53}$$

So, E is measurable.

□

Now we want to show that closed sets are measurable. But first, we need a lemma from point set topology.

Proposition 59. If F is closed, K is compact, and these sets are disjoint, then $d(F, K) > 0$.

PROOF 60.

Construct an original proof of this.

□

An immediate corollary of this is the following.

Proposition 61 (If disjoint sets are closed and compact respectively, then the exterior measure is additive.). If E_1, E_2 are respectively closed and compact, and they are disjoint, then

$$m_*(E_1 \cup E_2) = m_*(E_1) + m_*(E_2) \tag{54}$$

PROOF 62. This combines the previous proposition with property 4 of exterior measure.

□

Remark 63. The above statement automatically holds for measures because closed sets and compact sets are always always measurable, as we will see.

Add exercise 1.5

Proposition 64 (Closed sets are measurable.). In particular, compact sets are measurable.

PROOF 65. Since closed sets are countable unions of compact sets[20], it suffices to show that compact sets are measurable.

So suppose F is compact.[21] Fix $\epsilon > 0$. From property 3 of exterior measure, there is some open set $\mathcal{O}$ containing F such that $m_*(\mathcal{O}) \le m_*(F) + \epsilon$. Since F is closed, $\mathcal{O} \setminus F$ is open, and so, from proposition 17, there are almost disjoint cubes Q_j such that

$$\mathcal{O} \setminus F = \bigcup_{j=1}^{\infty} Q_j \tag{55}$$

For a fixed N, $K := \bigcup_{j=1}^{N} Q_j$ is closed and bounded, hence compact. So, from the previous proposition, $d(K, F) > 0$. Now, $\mathcal{O}$ covers $K \cup F$, so

$$\begin{aligned} m_*(\mathcal{O}) &\ge m_*(F) + m_*(K) \\ &= m_*(F) + \sum_{j=1}^{N} m_*\left(Q_j\right) \end{aligned}$$

where in the second line we used fourth and fifth properties of the exterior measure.

Rearranging this gives

$$\sum_{j=1}^{N} m_*\left(Q_j\right) \le m_*(\mathcal{O}) - m_*(F) \le \epsilon \tag{56}$$

Passing to the limit $N \to \infty$, and by subadditivity,

$$m_*(\mathcal{O} - F) \le \sum_{j=1}^{N} m_*\left(Q_j\right) \le \epsilon \tag{57}$$

as desired.

□

Proposition 66 (Complement of measurable set is measurable.).

PROOF 67. Let E be a measurable set. The idea is to write E^c as a union of two measurable sets. First by definition of measurability, for all $n \in \mathbb{N}$, there is an open set $\mathcal{O}_n$ containing E so that

$$m_*(\mathcal{O}_n \setminus E) \le \frac{1}{n} \tag{58}$$

Take $S := \bigcup_{n=1}^{\infty} \mathcal{O}_n^c$. This is a countable union of closed (hence measurable) sets, so it is measurable. Now, by De Morgan's law,

$$E^c \setminus S = E^c \cap \left(\bigcup_{n=1}^{\infty} \mathcal{O}_n^c \right)^c = E^c \cap \bigcap_{n=1}^{\infty} \mathcal{O}_n \subseteq \mathcal{O}_n \setminus E \tag{59}$$

and so, by monotonicity of the exterior measure,

$$m_*(E^c \setminus S) \le m_*(\mathcal{O}_n \setminus E) \le \frac{1}{n} \tag{60}$$

for all n. So, $m_*(E^c \setminus S) = 0$, and sets with exterior measure 0 are measurable, so $E^c \setminus S$ is measurable. Thus, E^c is a union of two measurable sets, so it is measurable.

[20] For a closed set F, we can take $F = \bigcup_{n=1}^{\infty} F \cap B_n(0)$ where $F \cap B_n(0)$ is compact since finite intersection of closed sets are closed and closed subsets of compact sets are compact.

[21] Note that compact sets have finite exterior measure since they are bounded (i.e. contained in some ball around the origin) and by subadditivity of the exterior measure.

□

Proposition 68 (Countable intersection of measurable sets is measurable.).

PROOF 69. This follows from De Morgan's law, and what we have proved in the above. If E_i are measurable, then

$$\bigcap_{i=1}^{\infty} E_i = \left(\bigcup_{i=1}^{\infty} E_i^c\right)^c \tag{61}$$

and the countable union as well as complement of measurable sets are measurable, so the above is measurable.

□

In conclusion, the family of measurable sets is closed under the usual set theoretic operations. Even better, we have shown that they are closed under *countable* intersections and unions. This is very important when we do analysis. We will introduce here a definition which is very important in the context of abstract measure theory, and sums up what we have shown above.

Definition 70. Let X be a set (in our case $X = \mathbb{R}^n$). A set $\mathcal{A} \subseteq \mathcal{P}(X)$ is a σ**-algebra over** X if

1.) $\emptyset, X \in \mathcal{A}$
2.) $\mathcal{A}$ is closed under compliments
3.) $\mathcal{A}$ is closed under finite and countable unions and intersections.

If we excluded the axiom for countable unions and intersections, we simply call $\mathcal{A}$ an algebra. We call the pair $(X, \mathcal{A})$ a **measurable space**.

In this subsection, we have shown the following.

Proposition 71. $(\mathbb{R}^n, \mathcal{L})$ is a measurable space (where $\mathcal{L}$ denotes the Lebesgue σ-algebra, i.e. the collection of all Lebesgue measurable sets). In other words, the collection of Lebesgue measurable sets form a σ-algebra.

For the remainder of this subsection (and until we get to abstract measure theory), we will restrict our discussion to $(\mathbb{R}^n, \mathcal{L})$.

Basic Properties of the Lebesgue Measure.

We are now ready to establish some basic properties of the Lebesgue measure.

Proposition 72 (Lebesgue measure is σ-additive). Let $E_1, E_2, ...$ be disjoint measurable sets, and let $E = \bigcup_{j=1}^{\infty} E_j$. Then

$$m(E) = \sum_{j=1}^{\infty} m(E_j) \tag{62}$$

PROOF 73. We first consider the case when E_j is bounded. We prove both direction of the inequality. One direction is obvious by the subadditivity of the measure:

$$m(E) \le \sum_{j=1}^{\infty} m(E_j) \tag{63}$$

Thus, we must prove the other direction.

By definition of measurability, for each E_j^c, there is a closed set $F_j \subseteq E_j$ such that $m_*(E_j \setminus F_j) \le \epsilon/2^j$. Notice that for all N,

$$
\begin{aligned}
m_*(E_j) &\leq m_*(F_j) + m_*(E_j \setminus F_j) \leq m_*(F_j) + \epsilon/2^j \\
\sum_{j=1}^{N} m_*(E_j) &\leq \sum_{j=1}^{N} m_*(F_j) + \sum_{j=1}^{N} \epsilon/2^j \leq \sum_{j=1}^{N} m_*(F_j) + \epsilon
\end{aligned}
$$

by subadditivity of the exterior measure.

For each fixed N, the sets $F_1, ..., F_N$ are closed and bounded (this is where we use boundedness of E_j), hence compact. They are also disjoint. Therefore, from proposition 61, we get

$$
m\left(\bigcup_{j=1}^{N} F_j\right) = \sum_{j=1}^{N} m(F_j) \tag{64}
$$

for fixed N. Now since $\bigcup_{j=1}^{N} F_j \subseteq E$, we have

$$
m(E) \geq \sum_{j=1}^{N} m\left(F_j\right) \geq \sum_{j=1}^{N} m\left(E_j\right) - \epsilon \tag{65}
$$

Passing to the limit $N \to \infty$, and $\epsilon \to 0$, we get

$$
m(E) \geq \sum_{j=1}^{\infty} m\left(E_j\right) \tag{66}
$$

as desired.

In the general case, the only idea is to break up the set E into bounded pieces. Namely, we consider an increasing sequence of cubes Q_k such that their union is all of $\mathbb{R}^n$. Then we can take disjoint shells by defining

$$
\begin{cases}
S_1 := Q_1 & k = 1 \\
S_k := Q_k \setminus Q_{k-1} & k \geq 2
\end{cases} \tag{67}
$$

Then the set $E_{j,k} := E_j \cap S_k$ is bounded, and

$$
E = \bigcup_{j=1}^{\infty} \bigcup_{k=1}^{\infty} E_{j,k} \tag{68}
$$

But this is a countable disjoint union, so

$$
\begin{aligned}
m(E) &= \sum_{j,k} m(E_{j,k}) \\
&= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} m(E_{j,k}) \\
&= \sum_{j=1}^{\infty} m(E_j)
\end{aligned}
$$

The first line uses the bounded part of the proposition.

□

Remark 74. Note that the above also implies the additivity of Lebesgue measure since we can just take $E_k = \emptyset$ for $k \geq N$ for some N.

The above proposition establishes the connection between
1.) Our primitive notion of a "volume" given by exterior measure
2.) the more refined notion of measurable sets
3.) the countably infinite operations allowed on these sets

Here is an immediate corollary of the previous proposition.

Proposition 75. If $E_2 \subseteq E_1$ are measurable subsets of $\mathbb{R}^n$, then

$$m(E_1 \setminus E_2) = m(E_1) - m(E_2) \tag{69}$$

PROOF 76. Observe that $G_1 := E_1 \setminus E_2, G_2 := E_2$ are disjoint measurable subsets of $\mathbb{R}^n$. Then $E_1 = G_1 \cup G_2$. By additivity of measure,

$$m(E_1) = m(G_1) + m(G_2) \tag{70}$$

which implies the claim.

□

We now talk about limit theorems concerning increasing and decreasing sequences of measurable sets.

Definition 77. A sequence of subsets $E_1, E_2, ...$ in $\mathbb{R}^n$ **increases to** E if $E_k \subseteq E_{k+1}$ for all k and $E = \bigcup_{k=1}^{\infty} E_k$. We denote this by $E_k \nearrow E$.

Similarly, we say $E_1, E_2, ...$ in $\mathbb{R}^n$ **decreases to** E if $E_k \supseteq E_{k+1}$ for all k and $E = \bigcap_{k=1}^{\infty} E_k$. We write this as $E_k \searrow E$.

Proposition 78 (Limit Theorem for Increasing/Decreasing Measurable Sets). Suppose E_k are measurable subsets of $\mathbb{R}^k$.
1.) If E_k increases to E, then
$$m(E) = \lim_{N \to \infty} m(E_N) \tag{71}$$
2.) If E_k decreases to E and $m(E_k) < \infty$ for some k, then
$$m(E) = \lim_{N \to \infty} m(E_N) \tag{72}$$

PROOF 79 (Proof of first part.). The main idea is to decompose E_k into disjoint measurable sets and then use σ-additivity.

Let

$$\begin{cases} G_1 := E_1 & k = 1 \\ G_k := E_k \setminus E_{k-1} & k \geq 2 \end{cases} \tag{73}$$

Then $E = \bigcup_{k=1}^{\infty} G_k$ is a disjoint countable union of measurable sets, so

$$
\begin{aligned}
m(E) &= m\left(\bigcup_{k=1}^{\infty} G_k\right) \\
&= \sum_{k=1}^{\infty} m\left(G_k\right) \\
&= \lim_{N\to\infty} \sum_{k=1}^{N} m\left(G_k\right) \\
&= \lim_{N\to\infty} m\left(E_k\right)
\end{aligned}
$$

as desired.

□

PROOF 80 (Proof of second part.). For the second part, since the sequence is decreasing, we can assume that $m(E_1) < \infty$. The idea here is similar; we split up E_1 into a countable disjoint union of measurable sets. We need $m(E_1) < \infty$ because in the end, we need to cancel $m(E_1)$ from both sides of the equation. We cannot do that if it is infinite.

Take $G_k := E_k \setminus E_{k+1}$ for each k. Then

$$
E_1 = E \cup \bigcup_{k=1}^{\infty} G_k \tag{74}
$$

Thus,

$$
\begin{aligned}
m(E_1) &= m(E) + m\left(\bigcup_{k=1}^{\infty} G_k\right) \\
&= m(E) + \lim_{N\to\infty} \sum_{k=1}^{N-1} m\left(E_k \setminus E_{k+1}\right) \\
&= m(E) + \lim_{N\to\infty} \sum_{k=1}^{N-1} m(E_k) - m(E_{k+1}) \\
&= m(E) + m(E_1) - \lim_{N\to\infty} m(E_N)
\end{aligned}
$$

where in the third line, we used proposition 75, and the last line follows from telescoping.

Canceling $m(E_1)$ gives

$$
m(E) = \lim_{N\to\infty} m(E_N) \tag{75}
$$

as desired.

□

Remark 81. If $m(E_k) = \infty$, the second statement fails. Take for instance $E_n := (n, \infty)$. Then $\lim_{N\to\infty} m(E_N) = \infty$ but $E = \emptyset$, so $m(E) = 0$.

We now use the above limit theorem to prove the following proposition which is an important geometric and analytic insight into the nature of measurable sets in terms of their relation to open and closed sets. The point is, an arbitrary measurable set can be well-approximated by the open sets that contain it and alternatively, by the closed sets that it contains.

But first, a definition.

Definition 82. The **symmetric difference between sets** E, F is the set

$$E \triangle F := (E \setminus F) \cup (F \setminus E) \tag{76}$$

Proposition 83 (Approximation of measurable sets by open and closed sets.). Suppose E is measurable subset of $\mathbb{R}^n$. Then for every $\epsilon > 0$,

1.) There exists an open set $\mathcal{O}$ containing E such that $m(\mathcal{O} \setminus E) \leq \epsilon$.
2.) There exists an closed set F contained in E such that $m(E \setminus F) \leq \epsilon$.
3.) If $m(E) < \infty$, there exists a compact set K contained in E such that $m(E \setminus K) \leq \epsilon$.
4.) (**Littlewood's First Principle.**) If $m(E) < \infty$, there exists a finite union $F = \bigcup_{j=1}^{N} Q_j$ of closed cubes such that $m(E \triangle F) \leq \epsilon$. In other words, sets of finite measure can be approximated as unions of cubes arbitrarily well (in such a way that the part hanging from the overlap (i.e. the symmetric difference) can be made arbitrarily small). Even more concisely, **every set is a finite union of intervals**.

The first two and the last two proofs are similar. The first two are just the definition of measurability whereas the latter two are more constructive. (We get existence of objects via definitions, and then we construct our desired objects out of them.)

PROOF 84 (Proof of first statement.). This is the definition of measurability.

□

PROOF 85 (Proof of second statement.). For this, since E^c is measurable, by definition of measurability, there is an open set $\mathcal{O}$ containing E^c and $m(\mathcal{O} \setminus E^c) \leq \epsilon$. Taking $F := \mathcal{O}^c$, then F is closed, contained in E (since $\mathcal{O}$ covers E^c), and $E \setminus F = \mathcal{O} \setminus E^c$ (by set theory). Thus, $m(E \setminus F) \leq \epsilon$.

□

PROOF 86 (Proof of third statement.). We use the usual trick of intersecting closed sets with compact balls. From the second statement, there is a closed set F in E such that $m(E \setminus F) \leq \epsilon/2$. For each n, consider the compact (hence measurable) sets $K_n := F \cap B_n(0)$. Then $E \setminus K_n$ is (a difference of measurable sets, hence) measurable, and it decreases to $E \setminus F$ by definition of K_n. Now since $m(E) < \infty$, from the second part of proposition 78,

$$m(E \setminus K) = \lim_{n \to \infty} m(E \setminus K_n) \tag{77}$$

and so, for large enough n, $m(E \setminus K_n) \leq \epsilon$.

□

PROOF 87 (Proof of fourth statement.). Fix $\epsilon > 0$. By definition of exterior measure, there exists a family of closed cubes Q_j such that E is covered by Q_j, and

$$\sum_{j=1}^{\infty} |Q_j| \leq m(E) + \epsilon/2 \tag{78}$$

Since E has finite measure[22], the series on the LHS converges, and so there exists some N such that the tail of the series is small, i.e.

$$\sum_{j=N+1}^{\infty} |Q_j| < \epsilon/2 \tag{79}$$

[22] Notice that we are *not* using proposition 78 here.

If $F := \sum_{j=1}^{N} |Q_j|$ is the concatenation, then

$$\begin{aligned} m(E \triangle F) &= m(E \setminus F) + m(F \setminus E) \\ &\leq m\left(\bigcup_{j=N+1}^{\infty} Q_j\right) + m\left(\bigcup_{j=1}^{\infty} Q_j \setminus E\right) \\ &\leq \sum_{j=N+1}^{\infty} |Q_j| + \sum_{j=1}^{\infty} |Q_j| - m(E) \\ &\leq \epsilon \end{aligned}$$

as desired. □

Here is a nice corollary to the third statement. (This is a key step in the proof of 291).

Proposition 88. A measurable set of finite measure differs from an (countable) union of compact sets by a null set.

PROOF 89. Here is the proof that follows from the previous proposition.
Proof 1. Take a sequence of compact sets K_n such that $m(E \setminus K_n) \leq \frac{\epsilon}{2^n}$ (given by the third part of the previous proposition). Then

$$\begin{aligned} m\left(E \setminus \bigcup_{n=1}^{\infty} K_n\right) &= m\left(E \cap \bigcap_{n=1}^{\infty} K_n^c\right) \\ &= m\left(\bigcap_{n=1}^{\infty} (E \setminus K_n)\right) \\ &\leq m\left(\bigcup_{n=1}^{\infty} (E \setminus K_n)\right) \\ &\leq \sum_{n=1}^{\infty} m\left(E \setminus K_n\right) \\ &\leq \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = \epsilon \end{aligned}$$

Thus, $E \setminus \bigcup_{n=1}^{\infty} K_n$ is a null set. □

Proof 2. If in addition we assume that E is bounded, then we already know that measurable sets are countable unions of closed sets (i.e. F_σ) up to a null set, and so, from the boundedness, the closed sets are compact, and we get out conclusion. □

Invariance Properties of Lebesgue Measure.

We prove some invariance properties of Lebesgue measure. We need these properties if we were to have any reasonable notion of volume. The proofs for translation, dilation, and reflection invariance are identical.

Proposition 90 (Translation invariance of Lebesgue Measure.). If E is a measurable set and $h \in \mathbb{R}^n$, then let

$$E_h := E + h := \{x + h : x \in E\} \tag{80}$$

Then E_h is also measurable and $m(E + h) = m(E)$.

PROOF 91. Let's first consider the easy case when E is a cube (say wlog[23], the cube $[-1,1]^n$). Then E_h is also a cube because

$$E_h = \prod_{i=1}^{n} [-1+h_i, 1+h_i] \tag{81}$$

where $h = (h_1, ..., h_n)$. Since cubes are closed sets, they are measurable. Also, from the above formula, $m(E_h) = 2^n = m(E)$. So, we have the theorem for this special case.

Now let's establish $m_*(E+h) = m_*(E)$ for arbitrary sets $E \subseteq \mathbb{R}^n$ from which, in the case when E is measurable, we get $m(E+h) = m(E)$. We know by definition of exterior measure that

$$\begin{aligned} m_*(E) &= \inf \sum_{j=1}^{\infty} |Q_j| \\ &= \inf \sum_{j=1}^{\infty} |Q_j + h| \\ &= m_*(E+h) \end{aligned}$$

Notice that the first infimum is over all coverings of E whereas the second is over the coverings of $E+h$. Passing from first to second line is a bit fishy. We know that $|Q_j| = |Q_j + h|$ because as we showed above, volumes of cubes can be computed explicitly, and they are clearly translation invariant. It is also clear that $Q_j + h$ is a covering for $E+h$ by simply set theoretic considerations; all elements of $E+h$ are of the form $x+h$ for $x \in E$, but all x belongs in some Q_j, and so, all $x+h$ belongs in some $Q_j + h$.

For measurability of E_h, we take $E \subseteq \mathcal{O}$ open such that $m_*(\mathcal{O} \setminus E) < \epsilon$. Then $\mathcal{O}_h$ is open (because for any $x+h \in \mathcal{O}_h$, we can take $B_\delta(x+h)$ which is contained in $\mathcal{O}_h$ via an easy set theoretic argument) and contains E_h. We know that $m_*(\mathcal{O}_h \setminus E_h)$ since we already showed above the translation invariance for exterior measures.

□

Proposition 92 (Dilation Invariance of Lebesgue Measure.). Suppose $\delta > 0$, and let

$$\delta E := \{\delta x : x \in E\} \tag{82}$$

If E is measurable, then δE is measurable, and $m(\delta E) = \delta^n m(E)$.

Again, this is to be expected from a reasonable notion of volume.

PROOF 93. The idea is as same as before. First, if E is a cube (wlog $[-1,1]^n$), then the result holds since δE would still be a closed cube (i.e., $[-\delta, \delta]^n$) and

$$m(\delta E) = \delta^n \tag{83}$$

As before, this implies the dilation invariance of the exterior measure, and so the formula $m(\delta E) = \delta^n m(E)$. As for measurability, consider $\mathcal{O}$ containing E such that $m(\mathcal{O} - E) < \frac{\epsilon}{\delta^n}$ for a fixed $\epsilon > 0$. We then see that $\mathcal{O}$ is open (by the same argument as before, only replace translation with dilation), and we already established the invariance of the measure, so $m(\delta\mathcal{O} - \delta E) < \epsilon$.

□

Proposition 94 (Reflection invariance of Lebesgue measure.). If E is measurable, then $-E := \{-x : x \in E\}$ is measurable, and $m(-E) = m(E)$.

[23] This reduction helps us compute the volume more concretely. We can do this reduction because we can replace ± 1 with some other number accounting for dilation and translation, and we can carry out the exact same computation.

The proof of can be done in exactly the same manner as before.

Here are some additional invariance theorems. The first is a more general version of the dilation invariance.

Proposition 95. [24] If $\delta = (\delta_1, ..., \delta_n) \in (\mathbb{R}_{>0})^n$, and $E \subset \mathbb{R}^n$, let

$$\delta E := \{(\delta_1 x_1, ..., \delta_n x_n) : x \in E\} \tag{84}$$

If E is measurable, then δE is measurable, and $m(\delta E) = \delta_1 ... \delta_n m(E)$

PROOF 96. The idea of the proof is identical to the previous theorems. First, the theorem holds for a cube since if we take $E = [-1, 1]^n$, then

$$\delta E = \prod_{i=1}^{n} [-\delta_i, \delta_i] \tag{85}$$

which is measurable (since it is closed) and has volume

$$m(\delta E) = \prod_{i=1}^{n} (2\delta_i) = \delta_1 ... \delta_n 2^n \tag{86}$$

Now, if E is a general subset of $\mathbb{R}^n$, then

$$\begin{aligned} m_*(\delta E) &= \inf \sum_{i=1}^{n} |\delta Q_j| \\ &= \inf \sum_{i=1}^{n} \delta_1 ... \delta_n |Q_j| \\ &= \delta_1 ... \delta_n m_*(E) \end{aligned}$$

We showed above that $|\delta Q_j| = \delta_1 ... \delta_n |Q_j|$. We can also see that if Q_j are cube coverings of E, then δQ_j are coverings for δE. (Just observe that all $\delta x \in \delta E$ must be of the form $(\delta_1 x_1, ..., \delta x_n)$ for $(x_1, ..., x_n) \in E$ which in turn must belong to one of the Q_j's.) If E is measurable and assuming that δE is measurable (which we will show in a moment), then this gives

$$m(\delta E) = \delta_1 ... \delta_n m(E) \tag{87}$$

Now, we must show that δE is measurable. Since E is measurable, take open set $\mathcal{O}$ that contains E such that

$$m_*(\mathcal{O} \setminus E) \leq \frac{\epsilon}{\delta_1 ... \delta_n} \tag{88}$$

Then $\delta \mathcal{O}$ is an open set (since every element $y := \delta x \in \delta \mathcal{O}$, we can take an open ball $B_d(x) \subset \mathcal{O}$, then $\delta x \in \delta B_d(x) \subset \delta \mathcal{O}$). Additionally,

$$m_*(\delta \mathcal{O} \setminus \delta E) \leq \epsilon \tag{89}$$

by the invariance we showed for exterior measures we showed above.

□

Proposition 97. Suppose L is a linear transformation of $\mathbb{R}^n$. If E is measurable, then so is $L(E)$. (In fact, $m(L(E)) = |\det L| \, m(E)$, as we will see in the next section.)

[24] Exercise 1.7

PROOF 98. The main idea is the equivalence: a set is measurable iff it differed from a F_σ set by a Lebesgue null set. We show below that $L(E)$ can thus broken into a F_σ part and a null set part.

Claim 1. If E is compact, then so is $L(E)$. Hence if E is F_σ, then so is $L(E)$.

A soft argument would be that a linear transformation is continuous, and that a continuous image of a compact set is compact.

Is there a hard argument to show this, using the boundedness of L?

This implies that the image $L(F)$ of closed set F is also measurable since $F = \bigcup_{n=1}^\infty F \cap B_n(0)$, and so,

$$L(F) = \bigcup_{n=1}^{\infty} L(F \cap B_n(0)) \tag{90}$$

and we already showed that an image $L(F \cap B_n(0))$ of a compact set is already measurable. Measurable sets are closed under countable unions, so $L(F)$ is measurable.

But now, measurable sets are closed under countable unions, so F_σ sets are measurable. □

Claim 2. If E is a null set, then $L(E)$ is a null set.

Since linear operators are bounded (or equivalently, they are continuous),

$$|L(x) - L(x')| \leq M\,|x - x'| \tag{91}$$

for some M. If we consider a cube $Q := [-1,1]^n$, then the furthest points contained in Q are the ones at the furthest corners (for instance, for $n = 2, 3$, they are the opposing corners of the cube). The distance between those two points, is just $c_n := 2\sqrt{n}$ (again, consider the case $n = 2, 3$). So, an upper bound to the distance between points in $L(Q)$ is simply $c_n M$.

(Via scaling) in general, a cube of side length ℓ is mapped into a cube of side length $c_n M \ell$ where $c_n := 2\sqrt{n}$. Now, by definition of exterior measure, this implies that

$$m_*(L(Q)) \leq c_n^n M^n\,|Q| \tag{92}$$

where Q is a cube of side length ℓ.

Now if E is a Lebesgue null set, then there is a cube covering Q_j such that $\sum_{j=1}^\infty |Q_j| < \epsilon$ for a fixed $\epsilon > 0$. Then from what we have above,

$$\begin{aligned}
m_*(L(E)) &\leq m_*\left(L\left(\bigcup_{j=1}^{\infty} Q_j\right)\right) \\
&= m_*\left(\bigcup_{j=1}^{\infty} L\left(Q_j\right)\right) \\
&\leq \sum_{j=1}^{\infty} m_*\left(L\left(Q_j\right)\right) \\
&\leq \sum_{j=1}^{\infty} c_n^n M^n\,|Q_j| \\
&\leq c_n^n M^n \epsilon
\end{aligned}$$

Since $c_n^n M^n$ is just a constant that depends on the dimension n, we can take $\epsilon \to 0$ and conclude that $L(E)$ is a Lebesgue null set. This proves claim 2.

For the final step, we use a proposition that we will prove later. Namely, a set is measurable iff it differed from a F_σ set by a Lebesgue null set. Therefore, E differs from an F_σ set by a Lebesgue null set, i.e. $E = F \cup N$ where F is F_σ, N is null, and the union is disjoint. Then $L(E) = L(F) \cup L(N)$. From claim 1, $L(F)$ is also F_σ hence measurable. From claim 2, $L(N)$ is a null set, hence measurable. So, $L(E)$ is a union of measurable sets, hence it is also measurable.

□

The claims in the above proof deserve to be a theorem.

Proposition 99. The image of a compact set under a linear transformation is a compact set. The image of a F_σ set under a linear transformation is a F_σ set.

Proposition 100. Image of a Lebesgue null set under a linear transformation is a Lebesgue null set.

σ-algebras and Borel Sets.

We will now move on to some general discussions on σ-algebras and Borel Sets. Recall that

Definition 101. Let X be a set (in our case $X = \mathbb{R}^n$). A set $\mathcal{A} \subseteq \mathcal{P}(X)$ is a **σ-algebra over** X if
1.) $\emptyset, X \in \mathcal{A}$
2.) $\mathcal{A}$ is closed under compliments
3.) $\mathcal{A}$ is closed under finite and countable unions and intersections.

Definition 102. The **Borel σ-algebra** $\mathcal{B}_{\mathbb{R}^n} \subseteq \mathcal{P}(\mathbb{R}^n)$ is the smallest σ-algebra containing all open sets in $\mathbb{R}^n$, i.e.

$$\mathcal{B}_{\mathbb{R}^n} := \bigcup \mathcal{A} \tag{93}$$

where the intersection is taken over all σ-algebras containing the open sets of $\mathbb{R}^n$. The elements of $\mathcal{B}_{\mathbb{R}^n}$ are called **Borel sets**.

Remark 103. Any (arbitrary) intersection of σ-algebras is again a σ-algebra, so indeed, $\mathcal{B}_{\mathbb{R}^n}$ is a σ-algebra under the above definition. Notice that the definition 93 guarantees of existence and uniqueness of the Borel σ-algebra (since it is given as a formula).

Let's verify the above statement.

Proposition 104. (Arbitrary) intersections of σ-algebras are a σ-algebra.

PROOF 105. Let $\{\mathcal{A}_\alpha\}_{\alpha \in A}$ be a family of σ-algebras. We claim that

$$\mathcal{A} := \bigcup_{\alpha \in A} \mathcal{A}_\alpha \tag{94}$$

satisfies the axioms of σ-algebras. The arguments are identical for unions, intersections, and complements. (We just observe that if a set is in $\mathcal{A}$, then it is in $\mathcal{A}_\alpha$ for all α, and so, indeed, $\mathcal{A}$ inherits the properties of each $\mathcal{A}_\alpha$.) We will just prove this for unions.

Let $U_1, U_2, \ldots \in \mathcal{A}$. Then in particular, $U_1, U_2, \ldots \in \mathcal{A}_\alpha$ for all $\alpha \in A$. But since each $\mathcal{A}_\alpha$ is a σ-algebra, $\bigcup_{i=1}^\infty A_i \in \mathcal{A}_\alpha$ for all α, and so, $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$. So, indeed it is closed under countable unions.

□

We also showed before that

Proposition 106. The collection $\mathcal{L}$ of Lebesgue measurable sets form a σ-algebra.

Here is an important connection between Lebesgue and Borel σ-algebras.

Proposition 107. The Borel σ-algebra is *strictly* contained in the Lebesgue σ-algebra.

PROOF 108. Of course, there is two parts to this proof. First, we must show inclusion. Then we must show an element that is Lebesgue measurable but not a Borel set.

The inclusion follows immediately from the definition of a Borel σ-algebra. We showed before that all open sets are measurable, so it is contained in the Lebesgue σ-algebra. But by definition, the Borel σ-algebra is the smallest σ-algebra containing the open sets, and so, indeed, the Borel sets are all in the Lebesgue σ-algebra.

The construction of a measurable non-Borel set is a bit tricky.[25] $\square$

Come back to this after talking about measurable functions.

Definition 109. Let $C \subset \mathcal{P}(\mathbb{R}^n)$. The **$\sigma$-algebra generated by** C is the set

$$\sigma(C) := \bigcap_{\substack{A_\alpha \ \sigma\text{-alg.} \\ C \subset A_\alpha}} A_\alpha \tag{95}$$

Remark 110. In other words, when we say a σ-algebra $\mathcal{A}$ is generated by the set C, we mean $\sigma(C) = \mathcal{A}$.

Proposition 111. σ-algebra generated by a set is indeed a σ-algebra.

PROOF 112. This follows immediately from the above definition and proposition 104. $\square$

Proposition 113. The Borel σ-algebra on $\mathbb{R}^n$ is generated by open sets, open balls, or closed sets. on $\mathbb{R}$, it is generated by any of the following:

1.) $A := \{(-\infty, a) : a \in \mathbb{R}\}$
2.) $B := \{(-\infty, a] : a \in \mathbb{R}\}$
3.) $C := \{(a, \infty) : a \in \mathbb{R}\}$
4.) $D := \{[a, \infty) : a \in \mathbb{R}\}$

PROOF 114. The open sets generate the Borel sets by definition of Borel sets. All open sets are countable unions of open balls, so the σ-algebra generated by open balls contain open sets, and hence all Borel sets.

From this, it suffices to show that the generated σ-algebra contains all open sets or all open balls (or simply open intervals for $\mathbb{R}$). The closed sets generate the Borel sets because any σ-algebra containing all closed sets must contain all open sets, and so, it must contain all Borel sets.

On $\mathbb{R}$, observe that

$$(a, b) = (-\infty, b) \setminus (-\infty, a] = (-\infty, b) \setminus \bigcap_{k=1}^{\infty} (-\infty, a + 1/k) \tag{96}$$

and so, all open intervals are contained in the σ-algebra generated by A. Since $(-\infty, a)^c = [a, \infty)$, this implies that D generates the Borel σ-algebra too. By the same reason, if we show the claim for B, then that implies the conclusion for C.

For C, observe that

$$(a, b) = (-\infty, b) \setminus (-\infty, a] = \bigcup_{k=1}^{\infty} (-\infty, b - 1/k] \setminus (-\infty, a] \tag{97}$$

and so, we are done. $\square$

[25] Exercise 1.35

A naive approach to Borel sets is to try listing the possible sets one can construct from open (gebiete) and closed (fermé) sets which are the simplest Borel sets. The next simplest are the countable intersections (durschnitt) of open sets and countable unions (somme) of closed sets.[26]

Definition 115. A subset of $\mathbb{R}^n$ is a G_δ **set** if it is a countable intersection of open sets. Likewise, a set is a F_σ **set** if it is a countable union of closed sets.

Here is a very important theorem. (We use this to show that the Lebesgue measurable sets are closed under linear transformations over $\mathbb{R}^n$, see proposition 97. Recall that in that proof, we use both directions of the equivalence; we first decompose E into an F_σ set and a null set, and then apply the linear transformation. The work is showing that the linear transformation of a null set is a null set which we did via a Lipshitz bound of the linear transformation and a geometric argument on the cubes.)

Proposition 116 (Measurable iff differs from G_δ and F_σ by null set.). A subset E of $\mathbb{R}^n$ is measurable
1.) iff E differs from a G_δ set by a Lebesgue null set.
2.) iff E differs from a F_σ set by a Lebesgue null set.

PROOF 117. For both, backward direction is obvious: since G_δ, F_σ, and null sets are all measurable, E is a union of measurable sets, hence measurable.

The idea for the forward direction for both statements is just proposition 83 (this gives the open/closed sets) and taking the respective set theoretic operations (this gives the δ and σ).

For the forward direction of the first statement, we do the usual constructive proof where we approximate the set E by open sets (which is suggested by the fact that G_δ sets are requires us to obtain open sets). Namely, for each $n \geq 1$, there is an open set $\mathcal{O}_n$ containing E such that $m(\mathcal{O}_n \setminus E) \leq 1/n$. Then $S = \bigcup_{n=1}^{\infty} \mathcal{O}_n$ which is a G_δ set containing E, and $(S \setminus E) \subseteq (\mathcal{O}_n \setminus E)$ for each n. Therefore, by monotonicity of the measure, $m(S \setminus E) \leq 1/n$ for all n. Hence $S \setminus E$ is exterior measure 0, and so, it is a Lebesgue null set.

For the forward direction of the second statement, we do the analogous procedure for closed sets, i.e. for each $n \geq 1$, take closed sets F_n contained in E such that $m(E \setminus F_n) \leq 1/n$. Then $U := \bigcup_{n=1}^{\infty} F_n$ is a F_σ set contained in E such that $(E \setminus U) \subseteq (E \setminus F_n)$. Therefore, once again by the monotonicity of the measure, $m(E \setminus U) \leq 1/n$ for all n. Therefore, $E \setminus U$ has exterior measure 0, and so, it is a Lebesgue null set as desired. □

Here is a very important connection between Lebesgue σ-algebra and Borel σ-algebra.

Proposition 118. The Lebesgue σ-algebra is a **completion** of the Borel σ-algebra, i.e. $\mathcal{L}$ is the smallest σ-algebra containing $\mathcal{B}_{\mathbb{R}^n}$ and the Lebesgue null sets $\mathcal{N}$.

PROOF 119. Denote by $\mathcal{S}$ the smallest σ-algebra containing the Borel sets and the Lebesgue null sets. We already know that $\mathcal{L} \supseteq \mathcal{S}$ from before. So, we need to show the other inclusion. But this follows immediately from the previous proposition: from the above, if $E \in \mathcal{L}$, then $E = G \cup N$ where G is a G_δ (hence Borel) set and $N, N' \in \mathcal{N}$ are null sets. Therefore, $\mathcal{L} \subseteq \mathcal{S}$ as desired. □

<u>Construction of Non-measurable Sets.</u>

We now want to talk about some pathologies in this theory, namely the subsets of $\mathbb{R}^n$ which are nonmeasurable. Here, we focus on constructing a subset in $\mathbb{R}$ (which as a consequence of a theorem in the next section gives a nonmeasurable set in $\mathbb{R}^n$).

For this construction, we need to recall some very basic concepts.

[26] Note that it is pointless to consider countable unions of open sets since such a set is open (and likewise for countable intersections of closed sets).

Definition 120. An **equivalence relation** $\sim$ is a relation which is symmetric, reflexive, and transitive.

And as with many, many constructions of pathologies in mathematics, our construction requires the axiom of choice.

Proposition 121 (Axiom of Choice). [27] Suppose E is a set and $\{E_\alpha\}_{\alpha \in A}$ is a family of nonempty subsets of E (where A is not assumed to be countable). Then there exists a function $\alpha \mapsto x_\alpha$ such that $x_\alpha \in E_\alpha$ for all α.

Proposition 122 (Vitali's Theorem.). Let $\sim$ be an equivalence relation[28] given by

$$x \sim y \iff x - y \in \mathbb{Q} \tag{98}$$

and let $\mathcal{E}_\alpha, \alpha \in A$ be the equivalence classes given by this equivalence relation. Also let $\mathcal{N} := \{x_\alpha\}_{\alpha \in A}$ be the collection of the representatives of the equivalence classes. Then $\mathcal{N}$ is not measurable.

Notice that we used the Axiom of Choice in constructing $\mathcal{N}$ because we needed to pick out the representatives from each equivalence class.

Remark 123. The set N given by the above construction is the **Vitali set**.

PROOF 124. The idea is to consider the unions of the shifts of $\mathcal{N}$ (where the shits are confined within a compact interval, say $[-1,2]$). Via σ-additivity and monotonicity, we can bound from the top and bottom of the measure of this union. (The disjointness required for the σ-additivity is our first claim. The second claim is that it contains $[0,1]$ and is contained in the compact interval.) But by translation invariance of Lebesgue measure, this bounds the infinite sum of the measure of $m(\mathcal{N})$ which forces it to not be nonnegative, and hence get a contradiction.
Suppose for contradiction that $\mathcal{N}$ is measurable. Let $\{r_k\}_{k=1}^{\infty} = \mathbb{Q} \cap [-1,1]$. The consider

$$\mathcal{N}_k := \mathcal{N} + r_k \tag{99}$$

i.e., the shifts of $\mathcal{N}$ by the rationals in $[-1,1]$.

Claim 1. $\mathcal{N}_k$ are disjoint.

Suppose for contradiction that $\mathcal{N}_k \cap \mathcal{N}_{k'} \neq \emptyset$. Then there exists rationals $r_k \neq r_{k'}$ and $\alpha, \beta \in A$ such that $x_\alpha + r_k = x_\beta + r_{k'}$. Hence,

$$x_\alpha - x_\beta = r_{k'} - r_k \in \mathbb{Q} \setminus \{0\} \tag{100}$$

But since $x_\alpha - x_\beta \neq 0$, we have $\alpha \neq \beta$. Thus, $x_\alpha \sim x_\beta$ which contradicts the fact that $\mathcal{N}$ contains one representative from each equivalence class.

Claim 2.

$$[0,1] \subseteq \bigcup_{k=1}^{\infty} \mathcal{N}_k \subseteq [-1,2] \tag{101}$$

The second inclusion is immediate since $\mathcal{N}_k \subseteq [-1,2]$, by choice of r_k. For the first inclusion, since $\sim$ partitions $[0,1]$ into equivalence classes, for all $x \in [0,1]$, there is a unique $x_\alpha \in \mathcal{N}$ such that $x \sim x_\alpha$. Then $x \in \mathcal{N}_k$ where k is given by $r_k = x - x_\alpha$.

[27] Of course, the Axiom of Choice can be considered either as an axiom or a consequence of the well-ordering principle (or any other statement equivalent to AofC).

[28] It is immediate that $\sim$ is an equivalence relation. One way to see this is to note that $\mathbb{Q}$ is an additive group (i.e., $0 \in \mathbb{Q}$, $x - y \in \mathbb{Q} \implies -(x-y) \in \mathbb{Q}$, and $x - y, y - z \in \mathbb{Q} \implies (x-y)+(y-z) \in \mathbb{Q}$).

Since $\mathcal{N}$ is measurable, $\mathcal{N}_k$ are measurable, and since $\mathcal{N}_k$ are disjoint, by monotonicity and σ-additivity, claim 2 implies that

$$1 \leq \sum_{k=1}^{\infty} m(\mathcal{N}_k) \leq 3 \tag{102}$$

But by translation invariance of Lebesgue measure, we must have $m(\mathcal{N}_k) = m(\mathcal{N})$ for each k, so

$$1 \leq \sum_{k=1}^{\infty} m(\mathcal{N}) \leq 3 \tag{103}$$

This is impossible since either $m(\mathcal{N}) = 0$ or $m(\mathcal{N}) > 0$. In the former case, the sum is vanishes, and so the above inequality is violated. In the latter case, the sum blows up, and again, the inequality is violated. Thus, we have our contradiction.

□

Remark 125. Observe that the only part where we used the fact that m is a measure (rather than an exterior measure) is when we get the upper bound

$$\sum_{k=1}^{\infty} m(\mathcal{N}) \leq 3 \tag{104}$$

from which we deduced that $\mathcal{N}$ is a null set. To show the other inequality, we just needed monotonicity which exterior measures also satisfy. This implies that Vitali sets have positive exterior measure.

Measurable Functions. -Monday, 6.25.2018

We are now ready to tackle the fundamental object we deal with in the theory of Lebesgue integration: measurable functions. Let's compare Riemann and Lebesgue's theories of integration.

Definition 126. A **characteristic function χ_E of a set** $E \subset X$ is the function $\chi : X \to \{0, 1\}$ such that

$$\chi_E(x) = \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases} \tag{105}$$

The building blocks for Riemann integration are step functions whereas for Lebesgue integration, they are simple functions.

Definition 127. A **step function** is a function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f := \sum_{k=1}^{\mathbb{N}} a_k \chi_{R_k} \tag{106}$$

where $a_k \in \mathbb{R}$, $N < \infty$, and R_k are rectangles in $\mathbb{R}^n$.

Definition 128. A **simple function** is a function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f := \sum_{k=1}^{\mathbb{N}} a_k \chi_{E_k} \tag{107}$$

where $a_k \in \mathbb{R}$, $N < \infty$, and E_k are measurable sets in $\mathbb{R}^n$ of finite measure.

Notice that the generalization only entails changing rectangles to measurable sets of finite measure.

Definitions and Basic Properties of Measurable Functions.

Definition 129. A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ is **finite-valued** if for all $x \in \mathbb{R}^n$, $f(x) \in \mathbb{R}$.

Remark 130. We allow f to take $\pm\infty$ because in many situations, we have functions that take on infinite values on at most a set of measure zero.

Example 131. Continuous functions $f : \mathbb{R}^n \to \mathbb{R}$ are finite-valued since on a compact ball $B_n(0),\ n \in \mathbb{N}$, f is bounded (extreme value theorem), and in particular takes on values in $\mathbb{R}$.

Here is an obvious proposition.

Proposition 132. Composition of finite-valued maps is finite valued.

PROOF 133. Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be finite-valued. By definition, $g(x) \in \mathbb{R}$ for all $x \in \mathbb{R}$. But $f(\mathbb{R}^n) \subseteq \mathbb{R}$, so $(g \circ f)(x) \in \mathbb{R}$ for all $x \in \mathbb{R}$, and so $g \circ f$ is finite-valued.

□

Definition 134. Let $E \subset \mathbb{R}^n$ be measurable. A function $f : E \subset \mathbb{R}^n \to \mathbb{R}$ is **measurable** if for all $a \in \mathbb{R}$

$$\{f < a\} := f^{-1}([-\infty, a)) = \{x \in E : f(x) < a\} \tag{108}$$

is measurable.

Remark 135. Proving that a function is measurable usually amounts to writing $\{f < a\}$ as a countable union or intersection of measurable sets, based on the hypothesis. The arguments are fairly point-set topological.

We will prove some theorems on the behavior of the sum, product, powers, etc. of measurable functions. For now, we will start with more basic facts. Let's start with the equivalent definitions of measurable functions.

Proposition 136 (Equivalent Characterizations of Measurability.). Let $E \subset \mathbb{R}^n$ be measurable, and let $f : E \subset \mathbb{R}^n \to \mathbb{R}$. The following are equivalent:

1.) $\{f < a\}$ is measurable, i.e. f is a measurable function.
2.) $\{f \leq a\}$ is measurable.
3.) $\{f > a\}$ is measurable.
4.) $\{f \geq a\}$ is measurable.

PROOF 137 (Proof: 1 and 2 are equivalent.). The proof for both directions are essentially the same.

(1 $\implies$ 2.) Observe that

$$\{f \leq a\} = \bigcap_{k=1}^{\infty} \{f < a + 1/k\} \tag{109}$$

Since countable intersections of measurable sets are measurable, we are done.

□

(1 $\impliedby$ 2.) Observe that

$$\{f < a\} = \bigcup_{k=1}^{\infty} \{f \leq a - 1/k\} \tag{110}$$

Since countable unions of measurable sets are measurable, we are done.

□

3 and 4 are equivalent by the exact same proof.

PROOF 138 (Proof: 1 and 4 are equivalent.). Observe that $\{f < a\} = \{f \geq a\}^c$, and complements of measurable sets are measurable.

□

Proposition 139 (Equivalent Characterizations of Measurability for Finite-Valued Functions.). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a finite-valued function. The following are equivalent.

1.) f is measurable.
2.) $f^{-1}(\mathcal{O})$ is measurable for every open set $\mathcal{O} \subseteq \mathbb{R}$. (Equivalently, replace $\mathcal{O}$ with (a, b) for $a, b \in \mathbb{R}$.)
3.) $f^{-1}(F)$ is measurable for every closed set $F \subseteq \mathbb{R}$.

PROOF 140. (1 $\implies$ 2.) Observe that

$$(a, b) = [-\infty, b) \setminus [-\infty, a] = [-\infty, b) \setminus \bigcap_{k=1}^{\infty} [-\infty, a + 1/k) \tag{111}$$

and so,

$$f^{-1}((a, b)) = f^{-1}([-\infty, b)) \setminus \bigcap_{k=1}^{\infty} f^{-1}([-\infty, a + 1/k)) \tag{112}$$

But from the previous proposition, f is measurable, so $f^{-1}([-\infty, b)), f^{-1}([-\infty, a + 1/k))$ is measurable. Thus, $f^{-1}((a, b))$ is measurable, so $f^{-1}(\mathcal{O})$ is measurable. □

□

Come back to this.

An immediate corollary to this is the following.

Proposition 141. If $f : \mathbb{R}^n \to \mathbb{R}$ is continuous, then f is measurable. If f is measurable and finite-valued, and $\Phi : \mathbb{R} \to \mathbb{R}$ is continuous, then $\Phi \circ f$ is measurable.

PROOF 142. For the first statement, since f is continuous, f is finite-valued. So, from the previous proposition, f is measurable iff the preimage (with respect to f) of all open sets is measurable. But this is clearly the case since f is continuous, and the preimage of open sets under continuous maps are open.

For the second part, since Φ is continuous, $\mathcal{O} := \Phi^{-1}((-\infty, a)),\ a \in \mathbb{R}$ is open, and since f is finite-valued and measurable, by the previous proposition,

$$(\Phi \circ f)^{-1}((-\infty, a)) = f^{-1}(\mathcal{O}) \tag{113}$$

is measurable. Therefore, $\Phi \circ f$ is measurable.

□

Here are some additional theorems on measurable functions.

Proposition 143 (Sup, inf, limsup, liminf of measurable is measurable). Let $f_n : \mathbb{R}^n \to \mathbb{R}$ be measurable functions. Then

$$\sup_n f_n(x) \quad \inf_n f_n(x) \quad \limsup_{n\to\infty} f_n(x) \quad \liminf_{n\to\infty} f_n(x) \tag{114}$$

are measurable.

PROOF 144. The first two just follows from the usual trick of writing the set $\{f > a\}$ as countable unions and intersections of measurable sets. For the third and fourth, one needs to recall the relation between limsup/liminf and sup/inf.

For the first, observe that

$$\left\{\sup_{n\in\mathbb{N}} f_n > a\right\} = \bigcup_{n=1}^{\infty} \{f_n > a\} \tag{115}$$

which is a countable union of measurable sets. Since $\inf_n f_n(x) = -\sup_n(-f_n(x))$, this shows that the first two functions are measurable.

The third and fourth follows from the first two:

$$\limsup_{n\to\infty} f_n(x) = \inf_{k\in\mathbb{N}} \sup_{n\geq k} f_n \qquad \liminf_{n\to\infty} f_n(x) = \sup_{k\in\mathbb{N}} \inf_{n\geq k} f_n \tag{116}$$

We already showed that sup/inf of measurable functions are measurable, and so, $\sup_{n\geq k} f_n, \inf_{n\geq k} f_n$. Therefore, from the above, the limsup, liminf are just sup/inf of measurable functions.

□

Here is an immediate corollary.

Proposition 145 (Limit of a sequence of measurable functions is measurable). If $f_n : \mathbb{R}^n \to \mathbb{R}$ are measurable functions, and

$$f(x) := \lim_{n\to\infty} f_n(x) \tag{117}$$

then f is measurable.

PROOF 146. By hypothesis, the limit $\lim_{n\to\infty} f_n(x)$ exists for each x, so

$$f(x) := \limsup_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x) \tag{118}$$

which we showed in the previous proposition is measurable.

□

Proposition 147 (Sums, Products, and Powers of Measurable Functions.). Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be measurable functions. Then the integer powers $f^k, k \geq 1$ is measurable. Additionally, the sum $f+g$ and product fg are measurable provided f, g are both finite-valued.

PROOF 148. The first statement is straightforward. As usual, for powers, we need to consider the odd and even powers separately. If k is odd, then

$$\left\{f^k > a\right\} = \left\{f > a^{1/k}\right\} \tag{119}$$

which is measurable since f is measurable. If k is even, and $a \geq 0$,

$$\left\{f^k > a\right\} = \left\{f > a^{1/k}\right\} \cup \left\{f < -a^{1/k}\right\} \tag{120}$$

For the sum, observe that

$$\{f+g > a\} = \bigcup_{r\in\mathbb{Q}} \{f > a-r\} \cap \{g > r\} \tag{121}$$

which of course is a countable union of measurable sets since f, g are measurable.

Where do we use the fact that f, g are finite valued?????

For product, we just use polarization[29]:

$$fg = \frac{1}{4}((f+g)^2 - (f-g)^2) \tag{122}$$

which we know is a multinomial of the measurable function $f+g$, $f-g$ (based on what we proved above). □

Add remark on composition of measurable functions. (They are NOT measurable. See stackexchange)

Proposition 149. If f, g are measurable functions, then

$$h_1 := \min(f, g), \; h_2 := \max(f, g) \tag{123}$$

are measurable.

[29] This is in the sense of inner product spaces; we are considering the vector space of all functions from $\mathbb{R}^n \to \mathbb{R}$, and the inner product can be defined to be simply $\langle f, g\rangle := fg$.

PROOF 150. We just prove that h_1 is measurable, and the same proof applies to h_2. For $a \in \mathbb{R}$, it suffices to show $\{h_1 > a\}$ is a measurable set. Observe that

$$\{h_1 > a\} = \{f > a\} \cap \{g > a\} \tag{124}$$

and so, it is an intersection of measurable sets, and so, it is measurable.

□

Definition 151. Let $f, g : E \subseteq \mathbb{R}^n \to \mathbb{R}$. f, g **are equal almost everywhere on** E if the set $\{x \in E : f(x) \neq g(x)\}$ is a null set, and we write $f(x) = g(x)$ a.e. $x \in E$.

More generally, let $P(x)$ be a property of f at $x \in E$. Then$f$ **satisfies** $P(x)$ **almost everywhere on** E if the set $\{x \in E : P(x)$ is false$\}$ is a null set.

Remark 152. We can relax the conditions we proved above to conditions holding almost everywhere. For instance, in proposition 145, we can weaken the hypothesis to $f(x) = \lim_{n\to\infty} f_n(x)$ almost everywhere. The proof of this proposition hinges on the proof of 143. There, we can just intersect all the sets we are working with with a null set since f is given by the limit of f_n only on a null set.

Another example: for the second part of proposition 147, we just need to take f, g to be finite-valued almost everywhere. In which case, $f+g$, fg will both be defined almost everywhere. But this means $f+g$, fg agrees with a measurable function almost everywhere, and so, (from the following proposition) this implies that they are measurable.

Proposition 153. If $f : E \subset \mathbb{R}^n \to \mathbb{R}$ is measurable and $f = g$ a.e. then g is measurable.

PROOF 154. Observe that

$$\{g < a\} \setminus \{f < a\} \subseteq \{x \in E : f(x) \neq g(x)\} \tag{125}$$

is a null set. On the other hand,

$$\{f < a\} \cap \{g < a\} = \{f < a\} \cap (\{f < a\} \setminus \{g < a\})^c \tag{126}$$

is measurable since $\{f < a\} \setminus \{g < a\}$ is a null set (by the same argument as above). Therefore, $\{g < a\}$ is a union of measurable sets, and so, it is measurable.

□

Approximation by simple functions or step functions.

We now look at how we can approximate measurable functions using simple and step functions (i.e. write measurable functions in terms of these functions which are easier to think about). Note that the three theorems are in order to dependence (i.e., the second theorem depends on the first, and the third depends on the second).

Proposition 155 (Nonnegative measurable function is approximated pointwise by increasing sequence of nonnegative simple functions.). Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a nonnegative measurable function. Then there exits an increasing sequence of nonnegative simple functions φ_k which converge to f pointwise, i.e.

$$\varphi_k(x) \leq \varphi_{k+1}(x) \qquad \lim_{k\to\infty} \varphi_k(x) = f(x) \tag{127}$$

for all x.

PROOF 156. The proof is by constructing φ_k. The logic of the construction is as follows. We break the construction down into two steps. We first consider the truncation of f which has compact support. Compact support is essential since we need a sequence of measurable sets with finite measure in order to construct a simple function. The measurable sets should be such that as k grows, the partition gets finer. To construct the simple function, we need to make sure that it monotonically increases to the truncation (so that the final function φ_k monotonically increases). Finally, to obtain φ_k, we just take the two variables N (for the truncation) and M (for the partition of the range) to be 2^k.

We first look at the truncation. For $N \geq 1$, let $Q_N := [-N, N]^n$. Then

$$F_N(x) := \begin{cases} f(x) & x \in Q_N,\ f(x) \leq N \\ N & x \in Q_N,\ f(x) > N \\ 0 & \text{otherwise} \end{cases} \tag{128}$$

In other words, inside the cube Q_N, F_N agrees with f if f is below a certain threshold, and is constant if it is above the threshold. Outside the cube, F_N vanishes. Compact support of F_N is crucial for constructing the simple functions (as we will see later).

From this construction, $\lim_{N\to\infty} F_N(x) = f(x)$ pointwise (since for any $x \in \mathbb{R}^n$, $x \in Q_{N_1}$ for some N_1 and $f(x) \leq N_2$ for some N_2, from which we can just take $N := \max(N_1, N_2)$ to get $F_N(x) = f(x)$).

Now we partition Q_N based on the range of F_N (namely $[0, N]$) as follows. Fix natural numbers $N, M \geq 1$, then we define

$$E_{\ell,M} := \left\{ x \in Q_N : \frac{\ell}{M} < F_N(x) \leq \frac{\ell+1}{M} \right\} \tag{129}$$

for $\ell = 1, ..., NM - 1$ and

$$E_{0,M} := \left\{ x \in Q_N : 0 \leq F_N(x) \leq \frac{1}{M} \right\} \tag{130}$$

for $\ell = 0$. Notice that

$$E_{\ell,M} = \begin{cases} Q_N \cap \left\{ \frac{\ell}{M} < f \leq \frac{\ell+1}{M} \right\} & \ell = 0, ..., NM-2 \\ Q_N \cap \left(\left\{ \frac{NM-1}{M} < f \leq N \right\} \cup \{ f > N \} \right) & \ell = NM-1 \end{cases} \tag{131}$$

and $E_{\ell,m}$ is an intersection of measurable sets (this is where we use the measurability of f). Also, $E_{\ell,M}$ has finite measure because $E_{\ell,M} \subseteq Q_N$. (This is why we need to introduce the cubes Q_N in the definition of $E_{\ell,M}$.)

Therefore, we can take the simple function[30]

$$F_{N,M}(x) := \sum_{\ell=1}^{NM-1} \frac{\ell}{M} \chi_{E_{\ell,M}}(x) \tag{132}$$

Now,

$$0 \leq F_{N,M}(x) \leq \frac{NM-1}{M} \tag{133}$$

since $E_{\ell,M}$ is disjoint. Therefore, $F_N - F_{N,M} \geq 0$. Also, by definition of $E_{\ell,M}$,

$$F_N(x) - F_{N,M}(x) \leq \frac{\ell+1}{M} - \frac{\ell}{M} = \frac{1}{M} \tag{134}$$

for all x. (So in particular, $F_{N,M} \to F_N$ as $M \to \infty$.)

Now if we choose $N = M = 2^k$ with $k \in \mathbb{N}$, we can let $\varphi_k := F_{2^k,2^k}$. Then from the above inequality

$$0 \leq F_{2^k}(x) - \varphi_k(x) \leq \frac{1}{2^k} \tag{135}$$

[30] Recall that the sets E_k by which we define simple functions must satisfy two conditions: measurability and finite measure.

for all x. Therefore, $\lim_{k\to\infty} \varphi_k(x) = \lim_{k\to\infty} F_{2^k}(x) = f(x)$.

Now we want to show that φ_k is an increasing sequence. Let $x \in \mathbb{R}^n$. Then

$$E_{\ell,2^k} = \left\{ x \in Q_{2^k} : \frac{\ell}{2^k} < F_{2^k}(x) \leq \frac{\ell+1}{2^k} \right\} \tag{136}$$

and

$$E_{m,2^{k+1}} = \left\{ x \in Q_{2^{k+1}} : \frac{m}{2^{k+1}} < F_{2^{k+1}}(x) \leq \frac{m+1}{2^{k+1}} \right\} \tag{137}$$

But now, (just by the definition of F_N) $F_{2^k} \leq F_{2^k+1}$, so

$$\frac{m}{2^{k+1}} < F_{2^{k+1}}(x) \leq \frac{m+1}{2^{k+1}} \implies \frac{\ell}{2^k} < F_{2^k}(x) \leq \frac{\ell+1}{2^k} \tag{138}$$

for $m = 2\ell, 2\ell+1$. In other words,

$$E_{2\ell,2^{k+1}} \cup E_{2\ell+1,2^{k+1}} = E_{\ell,2^k} \tag{139}$$

for all $\ell = 0, ..., 2^k - 1$. i.e., at each k, the $E_{\ell,m}$ gets partitioned into exactly two parts.

Now, for $x \in E_{\ell,2^k}$,

$$\varphi_k(x) = \frac{\ell}{2^k} = \frac{2\ell}{2^{k+1}} \leq \varphi_{k+1}(x) \tag{140}$$

since $\varphi_{k+1}(x) = \frac{2\ell}{2^{k+1}}, \frac{2\ell+1}{2^{k+1}}$. Therefore, φ_k is an increasing sequence. Thus, φ_k satisfies the desired properties.

□

Remark 157. This approximation theorem for nonnegative measurable functions are useful for passing from a property of a simple function to a property about general nonnegative measurable functions. In doing so, we can simple use the monotone convergence theorem.

Here is another theorem of a similar nature. We drop the assumption that f is nonnegative, and we allow the extended limit $-\infty$.

Figure out the remarks on the bottom of page 31: "Note that the result holds for nonnegative functions that are extended-valued, if the limit $+\infty$ is allowed." This is worthy of a "remark."

Proposition 158 (Measurable functions are pointwise limit of (absolutely increasing) simple functions.). Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is measurable. Then there exists a sequence of simple functions ϕ_k such that

$$|\varphi_k(x)| \leq |\varphi_{k+1}(x)| \quad \text{and} \quad \lim_{k\to\infty} \varphi_k(x) = f(x) \tag{141}$$

for all $x \in \mathbb{R}^n$. In particular $|\phi_k(x)| \leq |f(x)|$ for all $x \in \mathbb{R}^n$ and $k \geq 1$.

PROOF 159. Most of the work for this proof has already been done in the previous proposition. As is typically the case, when passing from the previous "nonnegative theorem" to this general theorem, all we really have to do is decompose the given function into a linear combination of nonnegative functions. All of the power of the previous theorem then caries over. We then get a sequence of simple functions converging to the nonnegative and nonpositive parts from which we can construct our ϕ in the obvious way.

Take the nonnegative and nonpositive part decomposition of f, i.e.

$$f^+(x) := \max(f(x), 0),\ f^-(x) := \max(-f(x), 0) \tag{142}$$

Then $f = f^+ - f^-$. Since f^+, f^- are nonnegative functions, we can apply the previous proposition to them to get two increasing sequences of nonnegative simple functions $\varphi_k^{(1)}(x), \varphi_k^{(2)}(x)$ which converge pointwise to f^+, f^-, respectively. Then take

$$\varphi_k(x) := \varphi_k^{(1)}(x) - \varphi_k^{(2)}(x) \tag{143}$$

But sum of limits is limits of the sum, so $\varphi_k(x) \to f(x)$ as $k \to \infty$, pointwise.

Now we claim that the sequence $|\varphi_k|$ is increasing.

Come back to this. The argument is not quite precise.

We now approximate by step functions instead of simple functions. Before we do this, we need a lemma.

Proposition 160. Let E be a measurable set with finite measure, and let Q_j be closed cubes such that $m\left(E \triangle \bigcup_{j=1}^N Q_j\right) \leq \epsilon$[31], for any $\epsilon > 0$. Then

1.) There exists almost disjoint rectangles $\tilde{R}_1, ..., \tilde{R}_M$ such that $\bigcup_{j=1}^N Q_j = \bigcup_{m=1}^M \tilde{R}_m$.

2.) There exists a collection of disjoint rectangles $R_m \subseteq \tilde{R}_m$ such that $m\left(E \triangle \bigcup_{m=1}^M R_m\right) \leq 2\epsilon$

If one draws pictures, these statements are obvious; for the first, we can extend the edges of the cubes and use these lines to partition the union. For the second, we take slightly smaller rectangles. Let's try to let this intuition be the guiding points of our proof.

PROOF 161. For the first statement, we can proceed by induction on N. For $N = 1$, there is nothing to prove since a cube is a rectangle. So, let's consider the induction step. Let $S := \bigcup_{j=1}^N Q_j = \bigcup_{j=1}^M \tilde{R}_j$ be a partition of the first N cubes into rectangles (given by the induction hypothesis). If the $N+1$st cube (wlog, let it be $Q_{N+1} := [-1,1]^n$) is disjoint or almost disjoint from S, then we can just take $\tilde{R}_{M+1} = Q_{N+1}$, and we would be done. So, let Q_{N+1} intersect S. Now consider the $n-1$-dimensional planes given by extending all the sides of $Q_j, 1 \leq j \leq N$. [32]

We claim that these $2nN$ planes (given by $x_d = p_{j,d} \pm \ell_d,\ 1 \leq d \leq n,\ 1 \leq j \leq N$) partition Q_{N+1} into almost disjoint rectangles. This is immediate since the planes intersecting Q_{N+1} call them $P_{1,d}, ..., P_{\nu_d,d},\ d = 1, ..., n$ partitions $[-1,1]$, so we observe that $\tilde{R}_m := \prod_{d=1}^n [P_{i,d}, P_{i+1,d}]$ gives the partition of Q_{N+1} (where we take here the convention $P_{0,d} := -1, P_{\nu_d+1,d} := 1$). □

Finish the proof.

Now we prove the theorem.

Proposition 162 (Measurable functions are pointwise limits of step functions almost everywhere). Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is measurable. Then there exists a sequence of step functions ψ_k that converges pointwise to $f(x)$ for almost every x.

PROOF 163. The idea is to restrict our attention to indicator function χ_E (via the previous theorem), and then approximate the measurable set E with finite measure using Littlewood's first principle. This is the key idea which ultimately gives the "almost everywhere" of the conclusion. This gives us rectangles from which we can construct step functions.

First, we claim that WLOG, we can take $f = \chi_E$ for $m(E) < \infty$. From the previous theorem, there exists a sequence of simple functions such that

$$|\varphi_k(x)| \leq |\varphi_{k+1}(x)| \quad \text{and} \quad \lim_{k\to\infty} \varphi_k(x) = f(x) \tag{146}$$

[31] The cubes are given by Littlewood's first principle in the following proposition.

[32] This is a perfectly rigorous procedure. In 2 dimensions, a cube $[-2,2]^2$ can extend its sides to give the four lines $x_1 = \pm 2, x_2 = \pm 2$. Similarly, for 3 dimensions, the sides of $[-2,2]^3$ can be extended to the six planes $x_1 = \pm 2, x_2 = \pm 2, x_3 = \pm 2$. Continuing this pattern, in n-dimensions, we just get the set of $2n$ planes $x_1 = \pm 2, x_2 = \pm 2, ..., x_n = \pm 2$. The collection of all the planes given by extending the sides of

$$Q_j := \prod_{i=1}^n [p_{j,d} - \ell_j, p_{j,d} + \ell_j], \qquad j = 1, ..., N \tag{144}$$

(where the cube is centered around the point $p_j := (p_{j,1}, p_{j,2}, ..., p_{j,n}) \in \mathbb{R}^n$ and has side length $2\ell_j$) are the $2nN$ planes given by

$$x_1 = p_{1,1} \pm \ell_1, p_{2,1} \pm \ell_2, ..., p_{n,1} \pm \ell_n; ...; x_n = p_{1,N} \pm \ell_1, p_{2,N} \pm \ell_2, ..., p_{n,N} \pm \ell_N \tag{145}$$

or simply $x_d = p_{j,d} \pm \ell_d,\ 1 \leq d \leq n,\ 1 \leq j \leq N$. ($d$ is for dimension.)

for all $x \in \mathbb{R}^n$. Each φ_k is a simple function, i.e. it can be written as

$$\varphi_k(x) = \sum_{j=1}^{N} a_{k,j} \chi_{E_j}(x) \tag{147}$$

for measurable sets E_j with finite measure. Therefore, it suffices to establish the claim for characteristic functions χ_{E_j} over sets of finite measure.

Fix $\epsilon > 0$. Since E has finite measure, by Littlewood's First Principle we can approximate E with finitely many cubes, i.e.

$$m\left(E \triangle \bigcup_{j=1}^{N} Q_j\right) \leq \epsilon \tag{148}$$

Now, by the usual trick of partitioning $\bigcup_{j=1}^{N} Q_j$ into rectangles by extending the sides of the cubes, we get a family of almost disjoint rectangles $\tilde{R}_i$ such that

$$m\left(E \triangle \bigcup_{i=1}^{M} \tilde{R}_i\right) \leq \epsilon \tag{149}$$

We can then get a slightly smaller disjoint [33] rectangle R_i inside $\tilde{R}_i$ such that [34]

$$m\left(E \triangle \bigcup_{i=1}^{M} R_i\right) \leq 2\epsilon \tag{150}$$

Since $f = \chi_E$, we have

$$f(x) = \sum_{j=1}^{M} \chi_{R_j} \tag{151}$$

except for a set of measure $\leq 2\epsilon$.

Now notice that the RHS of the above is a step function. We can thus construct a sequence of step functions ψ_k such that the sets

$$E_k := \{x \in \mathbb{R}^n : f(x) \neq \psi_k(x)\} \tag{152}$$

is decreases rapidly, i.e. $m(E_k) \leq 2^{-k}$.[35] Then taking

$$F = \limsup_{j\to\infty} E_j = \bigcap_{K=1}^{\infty} \bigcup_{j=K+1}^{\infty} E_j \tag{153}$$

[33] We need this so that the resulting step function $\psi := \sum \chi_{R_j}$ does indeed approximate $f = \chi_E$. If for instance R_1, R_2 are not disjoint, then for a point $x \in R_1 \cap R_2$ in their intersection, $\psi(x) = \sum \chi_{R_j}(x) = 2$ which is not in the range of f which means that f and ψ cannot agree at this point. This is fatal to the construction since we need the union of rectangles to be the good set on which f agrees with the step function.

[34] First, note that by construction, the family of rectangles $\tilde{R}_i$ is almost disjoint. Therefore, each edge of rectangles can be perturbed by a small distance δ so that adjacent rectangles are not sharing edges. To put this more rigorously, it suffices to consider the unit cube partitioned into rectangles by partitioning the n-axis into points $p_{1,1}, ..., p_{1,N_1}, ..., p_{n,1}, ..., p_{n,N_n}$ and taking their products. (We can make this reduction since constructing rectangles by extending the edges of the cubes amounts to this structure.) Let $\delta_{1,1}$ be the perturbation at the point $p_{1,1}$ to make the rectangle almost disjoint but have sufficiently large volume. In order to make the union of the rectangles have volume more than $1-\epsilon$ for some $\epsilon > 0$, then we must do the following. Fix $\epsilon' > 0$ depending on ϵ, which we determine later. Take the total perturbation along the axis to be $\sum_{i=1}^{N_1} \delta_{1,i} = \frac{\epsilon'}{2^n}$. Then by elementary geometry, the volume of the union of the rectangles becomes $\left(1 - \frac{\epsilon'}{2^n}\right)^n = 1 - \frac{n}{2^n}\epsilon' + o(\epsilon')$. Therefore, for small enough ϵ', we can get $\frac{n}{2^n}\epsilon' + o(\epsilon') \leq \epsilon$, and so, the union of the rectangles is larger than $1-\epsilon$, as desired.

[35] We can do this by the construction above. For each k, we can run the above construction again with $\epsilon = 2^{-(k+1)}$ to get $\psi_k(x) := \sum_{j=1}^{M} \chi_{R_j}(x)$ which agrees with f except for a set of measure $2\epsilon = 2^{-k}$. If we take the intersection F of these bad sets, and take its complement, then on this set $\psi_k \to f$ on this set.

i.e. the collection of elements that appear infinitely many times in E_j, or in other words, the points at which ψ_k disagree with f at arbitrarily large k. Now notice that by construction,

$$m(F) \leq m\left(\bigcup_{j=K+1}^{\infty} E_j\right) \leq \sum_{k=K+1}^{\infty} 2^{-k} = 2^{-K} \tag{154}$$

so, $m(F) = 0$, and $\psi_k \to f$ pointwise on F^c as desired.

□

Littlewood's Three Principles.

Littlewood's three principles illustrate the intuitive connection between measure theory and basic analysis. We have already done the first principle.

Proposition 164 (Littlewood's First Principle). Every measurable set is can be approximated arbitrarily well with finite union of intervals. More precisely: Suppose E is measurable subset of $\mathbb{R}^n$ with finite measure. Let $\epsilon > 0$. Then there exists a finite union $F = \bigcup_{j=1}^{N} Q_j$ of closed cubes such that $m(E \triangle F) \leq \epsilon$.

Proposition 165 (Littlewood's Second Principle: Lusin's Theorem). For every measurable finite valued function, the domain on can be approximated arbitrarily well by a set on which the function is continuous. Suppose f is measurable and finite valued on measurable set E with finite measure. Then for every $\epsilon > 0$, there exists a closed set F_ϵ contained in E and $m(E \setminus F_\epsilon) \leq \epsilon$ such that $f|_{F_\epsilon}$ is continuous.

Proposition 166 (Littlewood's Third Principle: Egorov's Theorem). For every pointwise convergent sequence of measurable functions, the domain can be approximated arbitrarily well by a closed set on which the function converges uniformly. More precisely: Suppose f_k is a sequence of measurable functions defined on a measurable set E with finite measure such that it converges pointwise almost everywhere on E. Given $\epsilon > 0$, there exists a closed set $A_\epsilon \subset E$ such that $m(E - A_\epsilon) \leq \epsilon$ and f_k converges uniformly on A_ϵ.

Remark 167. Notice the similarity between the three Littlewood's principles.

1.) The set E in the three theorems all have finite measure.
2.) All three theorems involve approximating this set E with a closed set with nice properties. This uses the fact that E has finite measure since we often need to subtract from the measure of E. In doing so, we do not want to calculate with something with infinite measure.
3.) The set E has a priori some nice properties. (For the first principle, this is finite measure (although this is also required of the other two theorems, we see from the proof that having finite measure allows us to approximate E by cubes from the outside). For the second theorem, this is finite value of f (infinite value is clearly pathological for continuous functions). For the third theorem, this is pointwise convergence. The three theorems all upgrade this nice property on the nice closed set.

Remark 168. Note that for the second principle, we mean that f viewed as a a function only on F_ϵ is continuous, rather than that f, as a function defined on E, is continuous on F_ϵ.

To illustrate the distinction, consider the step function $g(x) := \chi_{\{x \geq 0\}}(x)$. This function is discontinuous (at the point 0) as a function defined on all of $\mathbb{R}$, but it is continuous as a function defined on the set $\mathbb{R}_{\geq 0}$.

For completeness, recall the proof of Littlewood's first principle. We get the cubes via the definition of exterior measure. (Recall that we just defined the measure to be equal to the exterior measure restricted to the class of measurable sets.) This approximation contributes half of the error. The finite measure of E shows that the sum of the volumes of the cubes must converge, and so, we can concatenate this series. This gives us the other half of the error.

We use third principle to prove the second principle, so we will prove the third principle first.

PROOF 169 (Proof of Littlewood's Third Principle). Fix $\epsilon > 0$. The construction can be done by first principles. The idea is to approximate E with a measurable set $\tilde{A}_\epsilon$ which has the required uniform convergence conditions (which incurs an error of $\frac{\epsilon}{2}$) and then approximate this by a closed set A_ϵ (which incurs the rest of the error). The construction of the first set can be done by general techniques in measure theory (and this is the bulk of the work of the proof) and the second approximation can be done in a single line. The finite measureness of E is crucial to this first approximation.

We assume without loss of generality that $f_k \to f$ in all of E (since if B is the null set on which the function does not converge pointwise, then we can simply consider the set $E \setminus B$ which is still a measurable set with finite measure). Now let

$$E_k^n := \left\{ x \in E : |f_j(x) - f(x)| < \frac{1}{n}, \quad \text{for all } j > k \right\} \tag{155}$$

i.e. n controls the error of the approximation of f_j, and we choose only the function which are further along the sequence (and k controls how far we choose them). Now for fixed n, if $f_j(x)$ for $j > k$ is less than $\frac{1}{n}$ away from $f(x)$, then surely $f_j(x)$ for $j > k+1$ is less than $\frac{1}{n}$ away from $f(x)$. Therefore $E_k^n \subseteq E_{k+1}^n$, and so E_k^n is an increasing sequence for fixed n. Furthermore, f_j converge pointwise on all of E, so for fixed n (again, this controls the error), for each $x \in E$ and, there must be some k for which $|f_j(x) - f(x)| < \frac{1}{n}$ for all $j > k$. In other words, $E_k^n \nearrow E$ as $k \to \infty$.

Under these conditions, it is natural to use the limit theorem for increasing measurable sets. So, for each n (which again gives the error), there is some k_n such that

$$m(E \setminus E_{k_n}^n) = m(E) - m(E_{k_n}^n) < \frac{1}{2^n} \tag{156}$$

(Notice that this step requires that E has finite measure. If not, then we can simply get $m(E_{k_n}^n) \to \infty$ without the measure between $E, E_{k_n}^n$ to shrink.) We can then choose $N \in \mathbb{N}$ such that the tail of the geometric series is small, i.e.

$$\sum_{n=N}^{\infty} \frac{1}{2^n} < \frac{\epsilon}{2} \tag{157}$$

and we let

$$\tilde{A}_\epsilon := \bigcap_{n \geq N} E_{k_n}^n \tag{158}$$

which by monotonicity of the measure gives[36],

$$\begin{aligned} m\left(E \setminus \tilde{A}_\epsilon\right) &= m\left(E \cap \left(\bigcup_{n \geq N} (E_{k_n}^n)^c\right)\right) \\ &\leq \sum_{n=N}^{\infty} m\left(E \setminus E_{k_n}^n\right) < \frac{\epsilon}{2} \end{aligned}$$

This gives the set $\tilde{A}_\epsilon$. We now want to show that this indeed satisfies the uniform convergence property that we wanted. This is almost immediate from our construction of $\tilde{A}_\epsilon$.

Fix $\delta > 0$, and take $n \geq N$ such that $\frac{1}{n} < \delta$. But $x \in \tilde{A}_\epsilon$ implies that $x \in E_{k_n}^n$ which implies that

$$|f_j(x) - f(x)| < \delta \tag{159}$$

[36] Note that we cannot use σ-additivity because $E_{k_n}^n$ is not disjoint (in fact, as we noted before, it is an increasing sequence of sets)

for $j > k_n$. So, f_k converges uniformly on $\tilde{A}_\epsilon$.

Now, we need to approximate $\tilde{A}_\epsilon$ with a closed set. But we know that a measurable set can be approximated arbitrarily well by a closed set contained in the set, i.e. there exists a closed set $A_\epsilon \subseteq \tilde{A}_\epsilon$ with $m(\tilde{A}_\epsilon \setminus A_\epsilon) < \epsilon/2$. Therefore,

$$m(E \setminus A_\epsilon) = m(E \setminus \tilde{A}_\epsilon) + m(\tilde{A}_\epsilon \setminus A_\epsilon) < \epsilon \tag{160}$$

by additivity of measure.

□

Now we have the necessary tools to prove the second principle.

PROOF 170 (Littlewood's Second Principle.). The spirit of this proof is very similar to the proof of previous theorem. Fix $\epsilon > 0$. The idea is to first approximate E by a set $A_{\epsilon/3}$ on which step functions approximate f uniformly well (via Egorov's theorem) which incurs the error $\epsilon/3$. We then take away a subset of $A_{\epsilon/3}$ so that the step functions are continuous on this new set (which gives another error of $\epsilon/3$) from which f is a uniform limit of continuous functions, so f is continuous on this set. Finally, we take a subset of this which is closed (which incurs the rest of the error). Notice that the proof ends in the same way as Egorov's theorem does. Also note that the finite measureness of E comes from using Egorov's theorem.

We recall from general theory (proposition 162) that the measurable function f can be pointwise approximated by step functions f_n almost everywhere on E. But by Egorov's theorem, there is a closed set $A_{\epsilon/3}$ such that f_n converge uniformly to f and $m(E \setminus A_{\epsilon/3})$. This is the first step.

Now since step functions are just finite sums of weighted indicator functions on rectangles [37], so the points of discontinuity are included in the union of all the boundaries of the rectangles, which are Lebesgue null sets. So, in particular we can find measurable sets $E_n \subseteq E$ such that $m(E_n) < 1/2^n$ and f_n is continuous outside E_n (in other words, E_n contains all of the bad points). Then let

$$F' := A_{\epsilon/3} \setminus \bigcup_{n \geq N} E_n \tag{161}$$

for N large enough so that the tail of the geometric series is small, i.e. $\sum_{n=N}^{\infty} 1/2^n < \epsilon/3$. Now by construction of E_n, all f_n is continuous on F', and so f (being the uniform limit of continuous functions f_n) is continuous on F'. (Notice that f_n, f are being considered only as functions on F' rather than E. This is how we get that f is continuous as a function on F' but the points of F' are not continuous points when f is viewed as a function of the entire domain f.)

Now to finish, since F' is measurable (which is immediate since $A_{\epsilon/3}$ and E_n are measurable), so it contains a closed set F_ϵ such that $m(F' \setminus F_\epsilon) < \epsilon/3$. Of course, F_ϵ inherits the properties of F' (since the only property we are considering is continuity of f).

□

Exercises and Problems for Chapter 1: Measure Theory -Monday, 7.9.2018

Problem (2. Cantor-Lebesgue Function). [38]
a.) We assume that every number in $[0, 1]$ has a ternary expansion.[39]

Observe that

[37] Note that the points of discontinuity for f_n in the entire domain can be infinite. It is only finite for the one dimensional case.
[38] Also see problem 21 for an application of this example.
[39] Note that this representation is not unique since $\frac{1}{3} = \sum_{k=2}^{\infty} \left(\frac{2}{3}\right)^k$.

$$
\begin{aligned}
x \in C &\iff \forall k \geq 0 \qquad x \in C_k \\
&\iff \forall k \geq 0 \qquad x \in \left(\left[\frac{3m}{3^k}, \frac{3m+1}{3^k} \right] \cup \left[\frac{3m+2}{3^k}, \frac{3m+3}{3^k} \right] \right) \text{ for } m = 0, ..., 3^{k-1} - 1 \\
&\iff \forall k \geq 0 \qquad x \notin \left[\frac{3m+1}{3^k}, \frac{3m+2}{3^k} \right] \text{for } m = 0, ..., 3^{k-1} - 1 \\
&\iff \forall k \geq 0 \qquad a_k \neq 1
\end{aligned}
$$

In the last step, we observed that

$$
\begin{aligned}
x \in \left[\frac{3m+1}{3^k}, \frac{3m+2}{3^k} \right] &\iff x \in \left[\frac{m}{3^{k-1}}, \frac{m+1}{3^{k-1}} \right] \quad \text{and} \quad a_k = 1 \\
&\iff x \in \left[\frac{m}{3^{k-1}}, \frac{m+1}{3^{k-1}} \right] \quad \text{and} \quad a_k = 1
\end{aligned}
$$

by construction. This proves the claim. □

b.) Let's first show well-definedness of F. Since we are specifying the expansion of x, the choice of the representation is not the problem. The only thing we must verify is that the series we define F by does indeed converge. But this is immediate since $b_k = 0, 1$, and so,

$$F(x) \leq \sum_{k=1}^{\infty} \frac{1}{2^k} = 1 \tag{162}$$

so, $F(x)$ converges for all $x \in [0, 1]$.

Now let's show continuity. Fix $\epsilon > 0$. Denote by b_k^x the term in the series of $F(x)$, and likewise b_k^y for $F(y)$. We first observe that by definition of the ternary expansion,

$$|x - y| \leq \frac{1}{3^k} \tag{163}$$

(i.e. lies in the same connected component in the kth step of the construction of the Cantor set) iff $b_k^x = b_k^y$ (iff $a_k^x = a_k^y$, where we used the same convention for a_k^x as for b_k^x).

Let $N \in \mathbb{N}$ be the smallest number such that $\frac{1}{2^N} < \epsilon$. Take $\delta \leq \frac{1}{3^N}$, and let $|x - y| < \delta$. Then $b_k^x = b_k^y$ for all $1 \leq k \leq N$. So,

$$\begin{aligned}
|F(x) - F(y)| &\leq \sum_{k=1}^{\infty} \frac{|b_k^x - b_k^y|}{2^k} \\
&= \sum_{k=1}^{N} \frac{|b_k^x - b_k^y|}{2^k} + \sum_{k=N+1}^{\infty} \frac{|b_k^x - b_k^y|}{2^k} \\
&= \sum_{k=N+1}^{\infty} \frac{|b_k^x - b_k^y|}{2^k} \\
&\leq \sum_{k=N+1}^{\infty} \frac{2}{2^k} \\
&\leq 2 \sum_{k=N+1}^{\infty} \frac{1}{2^k} = \frac{1}{2^k} < \epsilon
\end{aligned}$$

where in the fourth line, we used the triangle inequality and the fact that $b_k^x, b_k^y = 0, 1$. This proves continuity.

Now let's show how F behaves at 0 and 1. This is immediate once we look at their ternary expansion. Observe that $a_k = 0$ for all k for 0, and $a_k = 2$ for all k for 1. For the former, we thus get $b_k = 0$ for all k, and so, $F(0) = 0$. For the latter, we get $b_k = 1$ for all k, and thus,

$$F(1) = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1 \tag{164}$$

as desired. □

c.) Proof 1. Via Soft Argument. Assume the result of part d[40], i.e. that the Cantor-Lebesgue function F is a continuous map from $[0, 1]$ to $[0, 1]$. Now, by the intermediate value theorem, any value between 0 and 1 is achieved by the Cantor-Lebesgue function on $[0, 1]$. Now, if a value y is achieved on the complement of the Cantor set $\mathcal{C}$, say on the interval (a, b), then the endpoints a, b are on $\mathcal{C}$, and F is constant on the intervals contained in $\mathcal{C}^c$. Therefore, $F(a) = F(b) = y$, and so, indeed $F : \mathcal{C} \to [0, 1]$ is surjective. □

Proof 2. Via Ternary Expansion. Based on the characterization from part a., it suffices to show that for all $y \in [0, 1]$, there exists a sequence b_k taking its values in the set $\{0, 1\}$ such that

$$y = \sum_{k=1}^{\infty} \frac{b_k}{2^k} \tag{165}$$

In other words, this just asks us to show that there is a binary representation for every element in $[0, 1]$. We construct the appropriate sequence b_k, and claim that this gives the representation of y.

For k, consider the partition $t_i = \frac{i}{2^k}, i = 0, 1, ..., 2^k$ of the interval $[0, 1]$. Then let

$$b_k = \begin{cases} 1 & \frac{2m+1}{2^k} \leq y < \frac{2m+2}{2^k},\ m \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases} \tag{166}$$

[40] Note that we cannot just use the continuity on the Cantor set from part b since the intermediate value theorem requires the *connectedness* of the entire domain.

(This is best seen by drawing the picture of the partition.) We now claim that the series $s := \sum_{k=1}^{\infty} \frac{b_k}{2^k}$ obtained by this converges to y. We can show by induction on k that the partial sum

$$s_k := \sum_{j=1}^{k} \frac{b_j}{2^j} \tag{167}$$

is in the same partition as y in the kth partition of $[0,1]$ (i.e., $[0,1]$ is split into intervals of length $\frac{1}{2^k}$). For $k = 1$, we see by inspection that this is true. For the induction step, we know by the induction hypothesis that s_k is in the same partition as y for the kth partition. So, we see by inspection that if

$$s_k = \frac{i}{2^k} = \frac{2i}{2^{k+1}} \leq y, \frac{2i+1}{2^{k+1}} < \frac{i+1}{2^k} = \frac{2i+2}{2^{k+1}} \tag{168}$$

then by inspection, s_{k+1} is in the same partition as y in the $k+1$th step (depending on whether y is greater or lesser than $\frac{2i+1}{2^{k+1}}$).

Now the result is immediate. Fix $\epsilon > 0$. Take the smallest $N \in \mathbb{N}$ such that $\frac{1}{2^N} < \epsilon$. Then from the above, s_N and y are in the same partition in the Nth step, so $|s_N - y| \leq \frac{1}{2^N} < \epsilon$. Therefore, $s = \lim_{N\to\infty} s_N = y$. $\square$

d.) Since constant functions are obviously continuous, and we already have continuity on the Cantor set from part b, we just need to show that $F(a) = F(b)$ when (a,b) is an open interval in the complement of the Cantor set. From the ternary expansion characterization of the Cantor set from part a, our claim is equivalent to showing that for any N,

$$\sum_{k=N+1}^{\infty} \frac{1}{2^k} = \frac{1}{2^N} \tag{169}$$

But this is just follows from the formula for geometric series:

$$\sum_{k=N+1}^{\infty} \frac{1}{2^k} = \frac{1}{1-1/2} \cdot \frac{1}{2^{N+1}} = \frac{1}{2^N} \tag{170}$$

as desired. $\square$

Uniform Convergence Construction of the Cantor-Lebesgue Function.

The Stein introduces the following construction as a part of his exposition of the theory of differentiability (in particular, of the theory of bounded variation functions). Since the construction only requires calculus, this is best presented here.

Consider the standard Cantor set $\mathcal{C} \subseteq [0,1]$, i.e.

$$\mathcal{C} := \bigcup_{k=0}^{\infty} C_k \tag{171}$$

where C_k is a disjoint union of 2^k closed intervals. Let $F_1(x)$ be the continuous increasing function on $[0,1]$ such that

$$F_1(x) := \begin{cases} \frac{3}{2}x & x \in [0,1/3] \\ 1/2 & x \in C_1^c = [1/3,2/3] \\ \frac{3}{2}x - \frac{1}{2} & x \in [2/3,1] \end{cases} \tag{172}$$

i.e., constant on C_1^c and linear on C_1 so that $F(0) = 0$ and $F(1) = 1$. We proceed recursively to construct a sequence of continuous increasing functions in this manner:

$$F_n(x) := \begin{cases} \frac{1}{2}F_{n-1}(3x) & x \in [0, 1/3] \\ 1/2 & x \in [1/3, 2/3] \\ \frac{1}{2}F_{n-1}(3x) + 1/2 & x \in [2/3, 1] \end{cases} \tag{173}$$

An explicit formula formula is given by the following[41]:

$$F_n(x) = \begin{cases} \sum_{i=1}^{k_0} \frac{a_i}{2^i} & a_{k_0} = a_{k_0}(x) = 1 \\ \left(\frac{3}{2}\right)^n \left(x - \sum_{i=1}^{k} \frac{a_i}{3^i}\right) + \sum_{i=1}^{k} \frac{a_i}{2^i} & a_k = a_k(x) = 0, 2 \quad \forall k \end{cases} \qquad 1 \le k \le n \tag{174}$$

where $x = \sum_{k=1}^{\infty} \frac{a_k(x)}{3^k}$ and $k_0 := k_{0,N}(x) := \{k : a_k(x) = 1,\ 1 \le k \le n\}$. The $\frac{1}{2}$ factor in the sums just come from the fact that $F_1(x) = \frac{3}{2}x$ for $x \in [0, 1/3]$.

In the limit $n \to \infty$, the first term for $a_k = 0$ decays (since $x - \sum_{i=k}^{\infty} \frac{a_i}{3^i}$ is the tail of the series), and so, the F_n certainly converges the Cantor-Lebesgue function given by ternary expansion.

Verify that the above formula is indeed correct....

As we see here, although geometrically the natural approach to the Cantor-Lebesgue function is to think of it as the uniform limit of simpler functions, as a formula, it is much cleaner to write it in terms of the dyadic expansion if we want an explicit formula.

But now, (without relying on the explicit formula above) we see that $|F_{k+1} - F_k| \le 2^{-k-1}$ by the construction (since the only points at which the consecutive functions can disagree on is on the Cantor set, but the distance between the constant parts of the F_{k+1} is 2^{-k-1}). So,

$$|F_N - F_M| \le \sum_{k=M}^{N-1} |F_{k+1} - F_k| \le 2^{-k-1} \le \frac{1}{2^M} \cdot \frac{1}{2} \tag{175}$$

so, the sequence of functions is Cauchy in the sup norm, and so, we have uniform convergence. The limit is then a continuous increasing function, and we call this function the Cantor-Lebesgue function. We notice that this function definitely agrees with the ternary expansion definition of the Cantor-Lebesgue function on the complement of the Cantor set. On the Cantor set, we see that the sequence of functions converge to $\sum_{k=1}^{\infty} \frac{a_k}{2^{k+1}}$ where a_k is the coefficient of 3^{-k} in the ternary expansion of x.

Problem (4. Cantor-like Sets.). **a.)** In the construction of $\hat{\mathcal{C}}$, let $I_{k,i}, 1 \le i \le 2^{k-1}$ be the interval of length ℓ_k centrally situated, i.e. at the midpoints of each interval (for instance, for $k = 1$, the point $\frac{1}{2}$, at $k = 2$, the point $\frac{1}{4}, \frac{3}{4}$, at $k = 3$, the point $\frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}$, so in other words, points $\frac{2m+1}{2^k}$ for $m = 1, 2, ..., 2^{k-1} - 1$). Then

$$\hat{\mathcal{C}} = [0, 1] \setminus \left(\bigcup_{k=1}^{\infty} \bigcup_{i=1}^{2^{k-1}} I_{k,i} \right) \tag{176}$$

and by hypothesis and by σ-subadditivity,

$$m\left(\bigcup_{k=1}^{\infty} \bigcup_{i=1}^{2^{k-1}} I_{k,i} \right) \le \sum_{k=1}^{\infty} 2^{k-1} \ell_k < 1 \tag{177}$$

so,

$$m\left(\hat{\mathcal{C}}\right) = m\left([0, 1]\right) - m\left(\bigcup_{k=1}^{\infty} \bigcup_{i=1}^{2^{k-1}} I_{k,i} \right) = 1 - \sum_{k=1}^{\infty} 2^{k-1} \ell_k > 0 \tag{178}$$

[41] We can deduce this from recursive rule although it is certainly easier to see this if we knew about the ternary expansion approach *a priori*.

as desired. □

b.) We know from problem 2 part b that every $x \in [0,1]$ has a sequence of half points (i.e., points of the form $s_N := \sum_{k=1}^{N} \frac{b_k}{2^k}$, $b_k = 0, 1$) such that $s_N \to x$ as $N \to \infty$. Since the open intervals of the k th construction of the Cantor set is centrally situated, the center of the intervals are these points s_N for all the 2^N possibilities for the values of b_k. (Or in other words, "centrally situated" simply means that the center of the interval is at the points $\frac{2m+1}{2^k}$ for $m = 1, 2, ..., 2^{k-1} - 1$, and so, from $k = 1, ..., N$, the collection of central points are just $\frac{i}{2^N}$, $1 \le i \le 2^N - 1$.)

Come back to this. This solution* is a bit fishy.

Thus, if $x_n := s_n$ are the points given from problem 2 b. converging to x, then we can take $I_n := \left(x_n - \frac{1}{n}, x_n + \frac{1}{n}\right) \cap \hat{\mathcal{C}}^c$ (which is an open subinterval in the complement of $\hat{\mathcal{C}}^c$). Then $x_n \to x$, $x_n \in I_n$, and $m(I_n) \le \frac{2}{n} \to 0$ as $n \to \infty$. □

c.) Choose the same $x \in \hat{\mathcal{C}}$ as in part b. We know that by construction, there is a point $y_n \in \hat{\mathcal{C}}^c$ such that $|x_n - y_n| < \frac{1}{2^{n-1}}$ (since we can just consider the point y_n lying in the same half interval[42] as x_n in the nth stage of the construction of $\hat{\mathcal{C}}$). But

$$|x - y_n| \le |x_n - x| + |x_n - y_n| \to 0 \tag{179}$$

as $n \to \infty$ from part a. and the condition on y_n. So, x is a limit point of $\hat{\mathcal{C}}$. Since $\hat{\mathcal{C}}$ is also closed, it is perfect.

We also see from part b that $\hat{\mathcal{C}}$ does not contain an open interval. If it did, consider for any $x \in \hat{\mathcal{C}}$ the open interval $B_\epsilon(x) \subseteq \hat{\mathcal{C}}$. Then in particular, there are no points $y \in \hat{\mathcal{C}}^c$ in $B_\epsilon(x)$. But this violates what we saw in part b which is that for any ϵ-ball around x, there is the point $x_n \in B_\epsilon(x)^c$ such that x_n is in this ϵ-ball. □

d.) For small enough ℓ_j are chosen (so that we are in the situation of part a.), then uncountability is immediate. If $\hat{\mathcal{C}}$ is a countable set of points, each singleton is a null set, so the entire set $\hat{\mathcal{C}}$ will be a countable disjoint union of null sets, and so a null set. But we know that we have positive measure, so we see that $\hat{\mathcal{C}}$ is uncountable. So, the case when the Cantor-like set is a null set (e.g. the standard Cantor set) is the tough case.

Proof 1. Easy Proof. Nonempty perfect sets are uncountable (Rudin theorem 2.43), and we've shown in part c that the Cantor-like set is perfect, so it is uncoutable. □

Problem (5.). a.) The only idea is the limit theorem for decreasing sequence of measurable sets. First we see that $\mathcal{O}_n \supseteq \mathcal{O}_{n+1}$ just by definition of $\mathcal{O}_n$. Also, we should also have

$$E = \bigcap_{n=1}^{\infty} \mathcal{O}_n \tag{180}$$

$\subseteq$ is immediate from the definition of $\mathcal{O}_n$. For the other direction, suppose z is a point in the RHS. Then $d(z, E) < \frac{1}{n}$ for all n, so $d(z, E) = 0$ which is equivalent to $z \in E$, and so we have $\supseteq$.

In addition, $\mathcal{O}_1$ is bounded because E is compact, hence contained in some open ball $B_N(0)$, and so, we claim that $\mathcal{O}_1 \subseteq B_{N+1}(0)$.

This should follow immediately from triangle inequality. If $x \in \mathcal{O}_1$, then for any $\epsilon > 0$, there is some $y_0 \in E$ such that

$$|x - y| < 1 + \epsilon \tag{181}$$

[42] By half interval, we mean the subinterval given by $\left[\frac{i}{2^k}, \frac{i+1}{2^k}\right]$, $0 \le i \le 2^k - 1$ on the kth stage of the construction of $\hat{\mathcal{C}}$.

and on the other hand, $y_0 \in B_N(0)$, so

$$|y - 0| < N \tag{182}$$

and so, putting the two together from triangle inequality,

$$|x - 0| < 1 + N + \epsilon \tag{183}$$

and since ϵ is arbitrary, this implies that $x \in B_{N+1}(0)$. Thus, $\mathcal{O}_1$ is bounded, hence has finite measure.

We now have the necessary hypothesis for the limit theorem for decreasing sequence of measurable sets:

$$m(E) = \lim_{n\to\infty} \mathcal{O}_n \tag{184}$$

□

b.)

Problem (10. Outer Jordan Content.). The **outer Jordan content of** $E \subseteq \mathbb{R}$ is a number given by

$$J_*(E) := \inf \sum_{j=1}^{N} |I_j| \tag{185}$$

where infimum is taken over all *finite* coverings of E by closed intervals.

a.) To show the equality, we do the usual approach with measures, namely show both directions of the inequality. One direction is trivial; we must get $J_*(E) \leq J_*(\overline{E})$ since a covering of $\overline{E}$ is a covering of E.

For the other direction, we use the other typical technique for measures, namely for $\epsilon > 0$, take closed intervals I_j such that

$$J_*(E) + \epsilon \geq \sum_{j=1}^{N} |I_j| \tag{186}$$

But since $\overline{E}$ is by definition, the intersection of all closed coverings of E, so in particular, it is contained in $\bigcup I_j$. Thus,

$$J_*(E) + \epsilon \geq J_*(E) \tag{187}$$

and taking $\epsilon \to 0$, we get the claim. □

b.) The set

$$E := \mathbb{Q} \cap [0, 1] \tag{188}$$

has exterior Lebesgue measure (hence Lebesgue measure) 0 but outer Jordan content 1. The former is immediate by σ-additivity of measure. For the latter, observe that if I_j is a covering of E does not cover all of E, then

$$[0, 1] \setminus \left(\bigcup_{j=1}^{N} I_j \right)^c \tag{189}$$

is open in the subspace topology of $[0, 1]$. (Notice we are using the fact that there are finitely many intervals since in order for this to be open, we need $\bigcup_{j=1}^{N} E_j$ to be closed.) But now, E is dense in $[0, 1]$, so E intersects this new set which violates the fact that I_j covers E. So, there are no finite coverings of E by closed intervals, and so, the outer Jordan content is 1.

Problem (13.). **a.)** Let F be a closed set. Then consider

$$\mathcal{O}_n := \left\{ x \in \mathbb{R}^n : d(x, F) < \frac{1}{n} \right\} \tag{190}$$

Then

$$F = \bigcap_{n=1}^{\infty} \mathcal{O}_n \tag{191}$$

since $\subseteq$ is immediate from the definition of $\mathcal{O}_n$, and $\supseteq$ is also immediate since if x is an element not in F, then (since $\{x\}$ is a compact set and F is closed, we have $d(x, F) > 0$, so) there exists some n such that $d(x, F) > \frac{1}{n}$, and thus, $x \notin \mathcal{O}_n$, so x is not an element of the RHS. This shows equality.

Now, we need to show that $\mathcal{O}_n$ is an open set. This is straightforward. Take $x \in \mathcal{O}_n$. Then let $d := d(x, F)$, then we must have $B_{\frac{1}{n}-d}(x) \subseteq \mathcal{O}_n$. Let $z \in B_{\frac{1}{n}-d}(x)$, and for fixed $\epsilon > 0$, there is $y_0 \in F$ such that $|x - y_0| \leq d + \epsilon$. Then

$$|z - y_0| \leq |z - x| + |x - y_0| \leq \frac{1}{n} - d + d + \epsilon = \frac{1}{n} + \epsilon \tag{192}$$

and since ϵ is arbitrary, we have $d(z, F) < \frac{1}{n}$, so indeed $B_{\frac{1}{n}-d}(x) \subseteq \mathcal{O}_n$. Thus, $\mathcal{O}_n$ is open. Therefore, F is indeed a G_δ-set.

On the other hand, consider the open set $G = F^c$. Then

$$G = F^c = \left(\bigcap_{n=1}^{\infty} \mathcal{O}_n \right)^c = \bigcup_{n=1}^{\infty} \mathcal{O}_n^c \tag{193}$$

and $\mathcal{O}_n^c$ is closed, so G is an F_σ-set. □

b.) We claim that $\mathbb{Q}$ is F_σ but not G_δ. First note that $F := \mathbb{Q}$ is an F_σ-set since

$$F = \bigcup_{q \in \mathbb{Q}} \{q\} \tag{194}$$

On the other hand, suppose for contradiction that

$$F = \bigcap_{k=1}^{\infty} I_k \tag{195}$$

where I_k is an open set. Then $\mathbb{Q} \subseteq I_k$ for each k. Fix k. Then every $q \in \mathbb{Q}$ is an interior point in the open set I_k, so for each q, take ϵ_q such that $(q + \epsilon_q, q - \epsilon_q) \subseteq I_k$. We claim that for any choice of ϵ_q,

$$\mathbb{R} = \bigcup_{q \in \mathbb{Q}} (q + \epsilon_q, q - \epsilon_q) \subseteq I_k \tag{196}$$

(and since the choice of ϵ_q is arbitrary, we see that $I_k = \mathbb{R}$ for all k from which we have a contradiction because $\bigcap I_k = \mathbb{R} \neq \mathbb{Q}$).

The inclusion $\supseteq$ is immediate. For the other direction, take $\alpha \in \mathbb{R}$.

Problem (16. Borel-Cantelli Lemma.). Suppose E_k is a sequence of measurable sets in $\mathbb{R}^d$ such that the sum of their measures is finite, i.e.

$$\sum_{k=1}^{\infty} m(E_k) < \infty \tag{197}$$

Then $E := \limsup_{k\to\infty}(E_k)$ is measurable, in particular, it is a null set.

a.) To show that E is measurable, we just need to recall what a limsup of a sequence of sets is. By definition,

$$E = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} E_k \tag{198}$$

Now, $\bigcup_{k\geq n} E_k$ is a countable union of measurable sets, so it is measurable. So, E is a countable intersection of measurable sets which is thus measurable. $\square$

b.) This follows immediately from the fact that the sum of the measures is finite. By hypothesis, for all $\epsilon > 0$, there exists some N such that

$$\sum_{k=N}^{\infty} m(E_k) < \epsilon \tag{199}$$

Therefore, by monotonicity of the measure,

$$m\left(\bigcup_{k\geq N} E_k\right) \leq \sum_{k=N}^{\infty} m(E_k) < \epsilon \tag{200}$$

But since $E \subseteq \bigcup_{k\geq N} E_k$ for all N, again by monotonicity of measure,

$$m(E) < \epsilon \tag{201}$$

Since ϵ is arbitrary $m(E) = 0$, as desired. $\square$

Problem (21. Continuous function mapping measurable to nonmeasurable set.). Consider the Cantor-Lebesgue function F (from problem 2), let $\mathcal{C}$ denote the Cantor set, and let $\mathcal{N}$ be the nonmeasurable function constructed in the book (i.e. the collection of all representatives the equivalence classes in $[0, 1]$ where elements are equivalent iff they differ by a rational).

From problem 2, $F : \mathcal{C} \to [0, 1]$ is a surjective map, so $F^{-1}(\mathcal{N}) \subseteq \mathcal{C}$. But since the Cantor set is a null set, by monotonicity of measure, this implies that $F^{-1}(\mathcal{N})$ is a null set, and thus measurable. This proves the claim. $\square$

Chapter 2. Integration Theory

The Lebesgue Integral: Basic Properties and Convergence Theorems. -Sunday, 7.1.2018

We now take the machinery from the previous section and develop the Lebesgue integral. We proceed in four stages:

1.) simple functions
2.) bounded functions with support on finite measure
3.) nonnegative functions
4.) integrable functions

Additionally, we assume that the functions are measurable and real valued. (We consider the case when it is complex valued later.) Simultaneously, we also prove some basic properties of the Lebesgue integral such as linearity, monotonicity, and convergence theorems (which is one of the big advantages of the Lebesgue integral).

Stage One: Simple Functions.

Recall that a simple function is a function of the form

$$\varphi(x) = \sum_{i=1}^{N} a_i \chi_{E_i} \tag{202}$$

where $a_i \in \mathbb{R}$ and E_i are measurable sets with finite measure.

Remark 171. A useful way to think about of a simple function is as a list of values a_i. So, for instance we can deduce from $\varphi \geq 0$ that $a_i \geq 0$ for each i.

Definition 172. Let

$$\varphi(x) = \sum_{k=1}^{M} c_k \chi_{F_k}(x) \tag{203}$$

where c_k are nonzero and distinct, and F_k are disjoint measurable sets with finite measure, then the **Lebesgue integral of the simple function** φ **on** $\mathbb{R}^n$ is the number

$$\int_{\mathbb{R}^n} \varphi(x) dx := \sum_{k=1}^{M} c_k m(F_k) \tag{204}$$

Additionally, if E is a measurable set with finite measure, then **Lebesgue integral of the simple function** φ **on** E is the number

$$\int_E \varphi(x) dx := \int_{\mathbb{R}^n} \varphi(x) \chi_E(x) dx \tag{205}$$

We sometimes denote the Lebesgue integral of π over E by

$$\int_E \varphi(x) dm \tag{206}$$

to emphasize the measure m.

Remark 173. We refer to the representation of φ in the definition as the **canonical form** of φ, and F_k are the **canonical sets**. How does one find c_k, F_k given an *arbitrary* simple function? (This also shows the uniqueness of the canonical form of φ.) By definition, the simple function 202 can take at most N distinct nonzero values, say $c_1, ..., c_M$, and so, we can just set

$$F_k := \{x : \varphi(x) = c_k\} = \varphi_k^{-1}(\{c_k\}) \tag{207}$$

The sets F_k must be disjoint because φ is a function. F_k is measurable since it is a preimage of the closed set $\{c_k\}$, and preimages of closed sets under measurable functions are measurable. Also,

$$\bigcup_{k=1}^{M} F_k \subseteq \bigcup_{i=1}^{N} E_i \tag{208}$$

and each E_i has finite measure, so by monotonicity of measure, each F_k must have finite measure. Therefore, indeed this choice of F_k gives a valid simple function $\varphi(x)$.

Now we want to show that in fact we do not need to rely on the canonical form to define the Lebesgue integral. But first, we need an elementary lemma.

Proposition 174. [43] Given the collection of sets $F_1, ..., F_n$, there exists sets $F_1^*, ..., F_N^*$with $N := 2^n - 1$, so that

1.) $\bigcup_{k=1}^n F_k = \bigcup_{j=1}^n F_j^*$
2.) F_j^* is disjoint
3.) For every k, $F_k = \bigcup_{F_j^* \subseteq F_k} F_j^*$

PROOF 175. Let $F_j^*, j = 1, ..., 2^n - 1$ be the sets given by

$$F_j^* := F_1' \cap F_2' \cap ... \cap F_n' \tag{209}$$

where F_1' is either F_1 or F_1^c. We cannot take the set $F_1^c \cap F_2^c \cap ... \cap F_n^c = \left(\bigcup_{k=1}^n F_k\right)^c$ because of the first condition (and so, we chose $N = 2^n - 1$). Let's verify that indeed F_j^* satisfy the desired hypothesis.

Condition 1. If $x \in \bigcup_{k=1}^n F_k$, then for some l,

$$x \in F_l = F_l \cap \left(\bigcup_{F_i' = F_i, F_i^c} F_1' \cap ... F_{l-1}' \cap F_{l+1} \cap ... \cap F_n' \right) \subseteq \bigcup_{j=1}^{n} F_j^* \tag{210}$$

So, we've show the inclusion $\subseteq$. For the other direction, Note that $F_j^* \subseteq F_k$ for some k since we did not allow all F_j^* to be F_j^c. This proves the first condition.

Condition 2. Suppose $F_i^* \cap F_j^* \neq \emptyset$. Suppose that for some k, $F_i^* \subseteq F_k, F_j^* \subseteq F_k^c$. Then they would be disjoint which is not the case. So for all k, F_i^*, F_j^* are in the same sets. But by how we defined the indices, this implies that $i = j$. This proves the second condition.

Condition 3. For all l, the $F_j^* \subseteq F_l$ iff $F_l' = F_l$ in the representation of F_j^*. On the other hand, observe that

$$F_l = F_l \cap \left(\bigcup_{F_i' = F_i, F_i^c} F_1' \cap ... F_{l-1}' \cap F_{l+1} \cap ... \cap F_n' \right) \tag{211}$$

This proves the third condition.

□

Now we are ready to prove the theorem.

Proposition 176 (Independence of Representation.). If $\varphi = \sum_{k=1}^N a_k E_k$ is any representation of φ, then

$$\int \varphi = \sum_{k=1}^{N} a_k m(E_k) \tag{212}$$

[43] Exercise 2.1

PROOF 177. We do this in two steps. We first consider the case when E_k is disjoint, and then we consider the general case. For the first case, the idea is to construct the canonical sets F_m by taking unions of the disjoint sets. From here, we can easily pass to the definition of the Lebesgue integral. For the general case, we can consider the refinements for the partition of $\bigcup_{k=1}^N E_k$ so that the sets become disjoint. We can then go back to the first case.

Let's consider the disjoint case. The sets E_k in the decomposition

$$\varphi = \sum_{k=1}^{N} a_k \chi_{E_k} \tag{213}$$

are disjoint, but a_k are not necessarily distinct or nonzero. So, for each distinct nonzero value a among a_k, take

$$\varphi^{-1}(a) = E_a' := \bigcup_{\{k:\ a_k = a\}} E_k \tag{214}$$

But φ is a function, so E_a' are disjoint, and by additivity of the measure,

$$m(E_a') = \sum_{\{k:\ a_k = a\}} m(E_k) < \infty \tag{215}$$

Therefore,

$$\varphi = \sum_{a \in \varphi(\mathbb{R}^n) \setminus 0} a \chi_{E_a'} \tag{216}$$

Since this is in canonical form, we have

$$\int_{\mathbb{R}^n} \varphi = \sum_{a \in \varphi(\mathbb{R}^n) \setminus 0} a m(E_a') = \sum_{k=1}^{N} a_k m(E_k) \tag{217}$$

as desired. This proves the disjoint case.

For the general case, let

$$\varphi = \sum_{k=1}^{N} a_k \chi_{E_k} \tag{218}$$

where E_k are not necessarily disjoint. By the previous proposition, there exists a refined partition $E_1^*, ..., E_n^*$ such that

1.) E_j^* partitions $\bigcup_{k=1}^N E_k$, i.e. $\bigcup_{k=1}^N E_k = \bigcup_{j=1}^n E_j$
2.) E_j^* are disjoint
3.) For each k, $E_k = \bigcup_{E_j^* \subseteq E_k} E_j^*$

Now, for each j, let $a_j^* = \sum_{\left\{k : E_j^* \subseteq E_k\right\}} a_k$. Then we get

$$\varphi = \sum_{j=1}^{n} a_j^* \chi_{E_j^*} \tag{219}$$

and this goes back to the disjoint case. So,

$$\int_{\mathbb{R}^n} \varphi = \sum_{j=1}^{n} a_j^* m(E_j^*) = \sum_{k=1}^{N} \sum_{E_j^* \subseteq E_k} a_k m(E_j^*) = \sum_{k=1}^{N} a_k m(E_k) \tag{220}$$

where the last equality used the additivity of the measure.

$\square$

We will now prove some basic properties the Lebesgue integral on simple functions which is useful for computations.

Proposition 178 (Basic Properties of Lebesgue Integral for Simple Functions.). For the following, φ, ψ are simple functions.

1.) **Linearity.** The Lebesgue integral is (real)-linear, i.e. for any $a, b \in \mathbb{R}$,

$$\int (a\varphi + b\psi) = a \int \varphi + b \int \psi \tag{221}$$

2.) **Additivity.** If E, F are disjoint measurable sets with finite measure, then

$$\int_{E \cup F} \varphi = \int_E \varphi + \int_F \varphi \tag{222}$$

3.) **Monotonicity.** If $\varphi \leq \psi$, then

$$\int \varphi \leq \int \psi \tag{223}$$

4.) **Triangle inequality.** If φ is simple, then $|\varphi|$ is also simple. Additionally,

$$\left| \int \varphi \right| \leq \int |\varphi| \tag{224}$$

Remark 179. The power of the definition of the Lebesgue integral is apparent from how easy the following proofs are. (In particular, properties ii through v follows right out of the definition.)

PROOF 180. All of them follow immediately from the definition.

Proof of i.) This follows immediately by definition (although it is a notational mess if we write it out explicitly.) If

$$\varphi = \sum_{k=1}^{N} a_k \chi_{E_k} \tag{225}$$

and

$$\psi = \sum_{l=1}^{M} b_l \chi_{G_l} \tag{226}$$

then we just take linear combinations of a_k, b_l on the regions on which E_k, G_l intersect. □

Proof of ii.) Directly from the definition of characteristic functions, we see that

$$\chi_{E \cup F} = \chi_E + \chi_F \tag{227}$$

for disjoint sets E, F. Now,

$$\begin{aligned}
\int_{E \cup F} \varphi &= \int_{\mathbb{R}^n} \varphi \chi_{E \cup F} \\
&= \int_{\mathbb{R}^n} \varphi (\chi_E + \chi_F) \\
&= \int_{\mathbb{R}^n} \varphi \chi_E + \varphi \chi_F \\
&= \int_E \varphi + \int_F \varphi
\end{aligned}$$

where last line follows from the linearity. □

Proof of iii.) Consider $\eta := \psi - \varphi \geq 0$, and we can show that

$$\int_{\mathbb{R}^n} \eta \geq 0 \tag{228}$$

But since $\eta \geq 0$, we must have $c_k \geq 0$ for

$$\eta = \sum_{k=1}^{N} c_k \chi_{F_k} \tag{229}$$

and so,

$$\int_{\mathbb{R}^n} \eta = \sum_{k=1}^{N} c_k m(F_k) \geq 0 \tag{230}$$

□

Proof of iv.) Given

$$\varphi = \sum_{k=1}^{N} a_k \chi_{E_k} \tag{231}$$

from the definition of simple functions,

$$|\varphi| = \sum_{k=1}^{N} |a_k| \chi_{E_k} \tag{232}$$

and so, by the standard triangle inequality (i.e. for sums),

$$\left| \int_{\mathbb{R}^n} \varphi \right| = \left| \sum_{k=1}^{N} a_k m(E_k) \right| \leq \sum_{k=1}^{N} |a_k| m(E_k) = \int \mathbb{R}^n \varphi \tag{233}$$

□

We also have the following very important idea.

Proposition 181. If f, g are simple functions such that $f = g$ a.e., then $\int_{\mathbb{R}^n} f = \int_{\mathbb{R}^n} g$.

PROOF 182. It suffices to show that $h := f - g$ which vanishes a.e. has integral $\int_{\mathbb{R}^n} = 0$. But this is immediate from definition since

$$\int_{\mathbb{R}^n} h = \sum_{k=1}^{N} a_k m(E_k) = 0 \tag{234}$$

where h is supported on $\bigcup_{k=1}^{N} E_k$, and $m(E_k) = 0$ for each k.

□

Stage Two: Bounded Functions Supported on a Set of Finite Measure.

Definition 183. The **support of a measurable function** f is the set on which f does not vanish, i.e.

$$\text{supp}(f) := \{x : f(x) \neq 0\} \tag{235}$$

f **is supported on a set** E if $\text{supp}(f) \subseteq E$.

Proposition 184. $\text{supp}(f)$ is measurable.

PROOF 185. Note that $\text{supp}(f) = f^{-1}(\mathbb{R}^*)$, and since $\mathbb{R}^*$ is open and f is measurable, we have the conclusion.

□

The following lemma is crucial for defining the integral for the class of bounded functions supported on sets of finite measure. (Specifically, this theorem is the content of the proof that the definition is well-defined.)[44]

Proposition 186. Let f be a bounded function supported on a set E of finite measure. If φ_n is a sequence of simple functions bounded by M, supported on E, converging pointwise to $f(x)$ for a.e. x, then

1.) $I_n := \int_{\mathbb{R}^d} \varphi_n$ converges
2.) If $f = 0$ a.e., then the limit $\lim_{n\to\infty} \int_{\mathbb{R}^d} \varphi_n = 0$.

PROOF 187. The conclusion is obvious if φ_n converges uniformly to f on E. Since f is measurable, we instead use Egorov's theorem (this is where we use finite measure of E) which gives us uniform convergence on a closed measurable set which approximates E arbitrarily well. This is really the only idea. Other than this, it is just a computation.

We claim that I_n is a Cauchy sequence. Fix $\epsilon > 0$. Since E has finite measure, Egorov's theorem gives a closed measurable set A_ϵ contained in E such that $m(E \setminus A_\epsilon) \leq \epsilon$, and φ_n converges uniformly to f on A_ϵ. Then

$$\begin{aligned}
|I_n - I_m| &\leq \int_E |\varphi_n(x) - \varphi_m(x)|\, dx \\
&= \int_{A_\epsilon} |\varphi_n(x) - \varphi_m(x)|\, dx + \int_{E\setminus A_\epsilon} |\varphi_n(x) - \varphi_m(x)|\, dx \\
&\leq \int_{A_\epsilon} |\varphi_n(x) - \varphi_m(x)|\, dx + 2Mm(E \setminus A_\epsilon) \\
&\leq \int_{A_\epsilon} |\varphi_n(x) - \varphi_m(x)|\, dx + 2M\epsilon
\end{aligned}$$

By uniform convergence, we can also bound the first term by ϵ; specifically, for large m, n, $|\varphi_n(x) - \varphi_m(x)| < \epsilon$, so

$$|I_n - I_m| < m(A_\epsilon)\epsilon + 2M\epsilon \leq (m(E) + 2M)\epsilon \tag{236}$$

Since ϵ is arbitrary and $m(E) + 2M < \infty$ (we use finite measure of E here again), and so, I_n is Cauchy. This proves the first claim.

For the second, we can pass to the limit $m \to \infty$ in the above bound to get

$$|I_n| = |I_n - f| < (m(E) + 2M)\epsilon \tag{237}$$

so, $I_n \to 0$ as $n \to \infty$.

□

Definition 188. The **Lebesgue integral of a bounded measurable function supported on a set of finite measure** f is the number

$$\int_{\mathbb{R}^d} f(x)dx = \lim_{n\to\infty} \int_{\mathbb{R}^d} \varphi_n(x)dx \tag{238}$$

[44] This theorem can be thought of as a special case of the bounded convergence theorem when $f = 0$. The bounded convergence theorem gives more regularity to f as the conclusion. We will prove the bounded convergence theorem later (rather than now) since we need the Lebesgue integral of bounded measurable function supported on a set of finite measure in order to state the theorem.

where φ_n is any sequence of bounded simple functions supported on supp(f) converging pointwise to f almost everywhere.

If E has finite measure, and f is bounded with $m(\text{supp}(f)) < \infty$, then the **Lebesgue integral of a bounded measurable function supported on a set of finite measure** f **over the set** E is the number

$$\int_E f(x)dx := \int_{\mathbb{R}^d} f(x)\chi_E(x)dx \tag{239}$$

Remark 189. We know by proposition 158 from the previous section that φ_n exists. The limit exists by the first part of the preceding proposition. We will check below that the definition is well-defined.

Proposition 190. The above definition is well-defined, i.e. if φ_n, ψ_n are both sequences of simple functions bounded by $M > 0$ supported on supp(f) converging pointwise to f almost everywhere, then

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} \varphi_n(x)dx = \lim_{n\to\infty} \int_{\mathbb{R}^d} \varphi_n(x)dx \tag{240}$$

PROOF 191. This follows immediately from the second part of the preceding proposition.

Consider $\eta_n := \varphi_n - \psi_n$, then this sequence is bounded by $2M$, supported on supp(f), and $\eta_n \to f - f = 0$ a.e. as $n \to \infty$. Then by the second part of the preceding proposition,

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} \eta_n = 0 \tag{241}$$

which is equivalent to the claim.

□

We once again proof the basic properties that the Lebesgue integral should satisfy.

Proposition 192. Let f, g be bounded measurable functions supported on a set of finite measure. Then

1.) **Linearity.** For $a, b \in \mathbb{R}$,

$$\int_{\mathbb{R}^d} (af + bg) = a \int_{\mathbb{R}^d} f + b \int_{\mathbb{R}^d} g \tag{242}$$

2.) **Additivity.** If E, F are disjoint subsets of $\mathbb{R}^d$, then

$$\int_{E\cup F} f = \int_E f + \int_F f \tag{243}$$

3.) **Monotonicity.** If $f \leq g$, then

$$\int_{\mathbb{R}^d} f \leq \int_{\mathbb{R}^d} g \tag{244}$$

4.) **Triangle Inequality.** $|f|$ is also bounded, supported on a set of finite measure, and

$$\left| \int_{\mathbb{R}^d} f \right| \leq \int_{\mathbb{R}^d} |f| \tag{245}$$

PROOF 193. The proof is trivial: we approximate f, g by simple functions φ_n, ψ_n and then use this to upgrade the properties of the Lebesgue integrals for simple functions.

Proof of linearity. Since limit of the sum is sum of the limit, and since linear combinations of simple functions are again simple functions, $a\varphi_n + b\psi_n$ is a simple function approximating $af + bg$.

Now, by definition,

$$\begin{aligned}
\int_{\mathbb{R}^d}(af+bg) &= \lim_{n\to\infty}\int_{\mathbb{R}^d}(a\varphi_n+b\psi_n)\\
&= \lim_{n\to\infty} a\int_{\mathbb{R}^d}\varphi_n+b\int_{\mathbb{R}^d}\psi_n)\\
&= a\lim_{n\to\infty}\int_{\mathbb{R}^d}\varphi_n+b\lim_{n\to\infty}\int_{\mathbb{R}^d}\psi_n)\\
&= a\int_{\mathbb{R}^d}f+b\int_{\mathbb{R}^d}g
\end{aligned}$$

□

Proof of additivity.

$$\begin{aligned}
\int_{E\cup F} f &= \int_{\mathbb{R}^d} f(x)\chi_{E\cup F}(x)dx\\
&= \lim_{n\to\infty}\int_{\mathbb{R}^d}\varphi_n(x)\chi_{E\cup F}(x)dx\\
&= \lim_{n\to\infty}\left(\int_{\mathbb{R}^d}\varphi_n(x)\chi_E(x)dx+\int_{\mathbb{R}^d}\varphi_n(x)\chi_E(x)dx\right)\\
&= \lim_{n\to\infty}\int_{\mathbb{R}^d}\varphi_n(x)\chi_E(x)dx+\lim_{n\to\infty}\int_{\mathbb{R}^d}\varphi_n(x)\chi_E(x)dx\\
&= \int_E f+\int_F f
\end{aligned}$$

□

Proof of monotonicity.
Since $f\geq g$, for large enough n, $\pi_n\geq\psi_n$ (this is standard calculus trick). So,

$$\begin{aligned}
\int_{\mathbb{R}^d} f &= \lim_{n\to\infty}\int_{\mathbb{R}^d}\varphi_n(x)dx\\
&\leq \lim_{n\to\infty}\int_{\mathbb{R}^d}\psi_n(x)dx\\
&= \int_{\mathbb{R}^d} g
\end{aligned}$$

□

Proof of triangle inequality. If f is bounded $|f|\leq M$ for some M, and so, $|f|$ is bounded if f is bounded. Also, $|f|=0$ iff $f=0$, so supp(f) = supp$(|f|)$. Thus, $m(\text{supp}(f))<\infty$. Now,

$$\begin{aligned}
\left|\int_{\mathbb{R}^d} f\right| &= \left|\lim_{n\to\infty}\int_{\mathbb{R}^d}\varphi_n\right|\\
&= \lim_{n\to\infty}\left|\int_{\mathbb{R}^d}\varphi_n\right|\\
&\leq \lim_{n\to\infty}\int_{\mathbb{R}^d}|\varphi_n|\\
&= \int_{\mathbb{R}^d}|f|
\end{aligned}$$

□

Here is the first important convergence theorem (i.e. a theorem about swapping an integral and a limit).

Proposition 194 (Bounded Convergence Theorem.). Suppose f_n is a sequence of measurable functions all bounded by M a.e., are supported on a set E of finite measure a.e. and converges to f pointwise a.e. as $n \to \infty$. Then f is measurable, bounded, supported on E for a.e. x[45], and

$$\int_E |f_n - f| \to 0 \qquad \text{as } n \to \infty \tag{246}$$

Consequently,

$$\int_E f_n \to \int_E f \qquad \text{as } n \to \infty \tag{247}$$

PROOF 195. The only part of this proof that requires any work is the first limit. The second limit follows from the first by triangle inequality:

$$\left| \int_E f - \int_E f_n \right| \leq \int_E |f_n - f| \to 0 \tag{248}$$

as $n \to \infty$.[46] f is measurable since it is the limit of measurable functions[47]. f is bounded by M a.e. since

$$|f| = \left| \lim_{n\to\infty} f_n \right| = \lim_{n\to\infty} |f_n| \leq M \qquad \text{a.e.} \tag{249}$$

f is supported on E a.e. since f_n all vanish outside E except for a set of measure 0.

Now we need to prove the first limit. The argument is identical to the proof of lemma 186 we proved for the definition of the Lebesgue integral for bounded measurable functions with finite measures. The key idea is Egorov's theorem.

Fix $\epsilon > 0$. By Egorov's theorem, there is a closed set A_ϵ contained in E such that $m(E \setminus A_\epsilon) \leq \epsilon$ and f_n converges uniformly to f on A_ϵ. Then for sufficiently large n, we have a uniform bound, i.e. $|f_n(x) - f(x)| \leq \epsilon$ for all $x \in A_\epsilon$. Therefore,

$$\begin{aligned} \int_E |f_n(x) - f(x)|\, dx &\leq \int_{A_\epsilon} |f_n(x) - f(x)|\, dx + \int_{E\setminus A_\epsilon} |f_n(x) - f(x)|\, dx \\ &\leq \epsilon m(E) + 2M\epsilon \end{aligned}$$

for large n. Since ϵ is arbitrary, we have the claim.

□

Proposition 196. If f is nonnegative bounded measurable function supported on a set E with finite measure, and $\int_E f = 0$, then $f = 0$ almost everywhere.

PROOF 197. Partition supp(f) into the sets

$$E_k := \left\{ x \in E : f(x) \geq \frac{1}{k} \right\}, \qquad k \in \mathbb{N} \tag{250}$$

Then by monotonicity of the integral,

[45] In other words, the supp(f) is contained in E except for a null set.
[46] This applies to all convergence theorems; we start with a limit of the form $\lim_{n\to\infty} \int_{\mathbb{R}^d} |f_n - f| = 0$, and then conclude that we can swap limit and integral.
[47] Recall that this followed from the fact that limsup and liminf of measurable functions are measurable

$$k^{-1}\chi_{E_k} \leq f(x)$$
$$k^{-1}m(E_k) \leq \int_E f$$

So, $m(E_k) = 0$ for all k, and since $\text{supp}(f) = \bigcup_{k=1}^{\infty} E_k$, f vanishes almost everywhere.

□

Connection Between Riemann and Lebesgue Integral.

We are now ready to show that Riemann integrable functions are Lebesgue integrable as well as address the question: when can we swap limits and integrals? We can talk about this problem at this point (rather than waiting until later) because the definite Riemann integral is defined for a bounded function on a bounded set. So, to our theory of Lebesgue integrals up until this point is adequate for comparing with this.

Proposition 198 (Riemann and Lebesgue agree on an interval.). Suppose f is Riemann integrable on $[a,b] \subset \mathbb{R}$. Then f is Lebesgue measurable, and

$$\int_{[a,b]}^{\mathcal{R}} f(x)dx = \int_{[a,b]}^{\mathcal{L}} f(x)dx \tag{251}$$

where the integral on the left is the Riemann integral and the one on the right is the Lebesgue integral.

PROOF 199. The idea is to instead consider step functions (which arise from the definition of the Riemann integral) for which the result is obvious. We can then pass to the limit (via bounded convergence theorem) in order to go back to the original function, and inherit the nice behavior.

Since a Riemann integrable function is bounded, let $|f(x)| \leq M$. We also have from the definition of Riemann integrability (specifically, the upper/lower sums) the sequence of step functions φ_k (the lower sum) and ψ_k (the upper sum) that satisfy $|\varphi_k|, |\psi_k| \leq M$ for all $x \in [a,b]$ and $k \in \mathbb{N}$, and

$$\varphi_1 \leq \varphi_2 \leq ... \leq f \leq ... \leq \psi_2 \leq \psi_1 \tag{252}$$

and

$$\lim_{k\to\infty} \int_{[a,b]}^{\mathcal{R}} \varphi_k(x)dx = \lim_{k\to\infty} \int_{[a,b]}^{\mathcal{R}} \psi_k(x)dx =: \int_{[a,b]}^{\mathcal{R}} f(x)dx \tag{253}$$

Now recall that the definitions of the Riemann and Lebesgue integrals of step functions are exactly the same. For Riemann integrals, (since the integral is the supremum of the lower sums and infimum of the upper sums) the upper and lower sum will equal the integral itself:

$$\int_{[a,b]}^{\mathcal{R}} \varphi_k(x)dx = \sum_{s_i \in \mathcal{P}} c_i(s_{i+1} - s_i) \tag{254}$$

where $\mathcal{P}$ is the partition of $[a,b]$ associated with the step function $\varphi_k(x) = \sum_i c_i \chi_{[s_i, s_{i+1}]}$. For the Lebesgue integral (since step functions are simple functions), recall that

$$\int_{[a,b]}^{\mathcal{L}} \varphi_k(x)dx = \sum_{s_i \in \mathcal{P}} c_i(s_{i+1} - s_i) \tag{255}$$

Therefore, indeed we have

$$\int_{[a,b]}^{\mathcal{R}} \varphi_k(x)dx = \int_{[a,b]}^{\mathcal{L}} \varphi_k(x)dx \tag{256}$$

and

$$\int_{[a,b]}^{\mathcal{R}} \psi_k(x)dx = \int_{[a,b]}^{\mathcal{L}} \psi_k(x)dx \tag{257}$$

for all k.

We now have the main ingredients to complete the proof. Now, since we are approximating f from above and below by φ_k, ψ_k, we can consider their limit:

$$\tilde{\varphi}(x) := \lim_{k\to\infty} \varphi_k(x),\ \tilde{\psi}(x) := \lim_{k\to\infty} \psi_k(x) \tag{258}$$

and since $\varphi_k \leq f \leq \psi_k$, we get $\tilde{\varphi} \leq \phi \leq \tilde{\psi}$.

Now, since the step functions φ_k, ψ_k are all bounded by M, supported on $[a,b]$, and converge to $\tilde{\varphi}, \tilde{\psi}$ everywhere, we can use the bounded convergence theorem to get

$$\lim_{k\to\infty} \int_{[a,b]}^{\mathcal{L}} \varphi_k(x)dx = \int_{[a,b]}^{\mathcal{L}} \tilde{\varphi}_k(x)dx \tag{259}$$

and

$$\lim_{k\to\infty} \int_{[a,b]}^{\mathcal{L}} \psi_k(x)dx = \int_{[a,b]}^{\mathcal{L}} \tilde{\psi}_k(x)dx \tag{260}$$

But from equations 253, 256, 257, we get

$$\int_{[a,b]}^{\mathcal{L}} (\tilde{\varphi}_k(x) - \tilde{\psi}_k(x))dx = 0 \tag{261}$$

Now, since $\psi_k - \varphi_k \geq 0$, we must have $\tilde{\varphi}_k(x) - \tilde{\psi}_k(x) \geq 0$. But a nonnegative function whose integral vanishes must be 0 a.e. So, $\tilde{\varphi}_k(x) = \tilde{\psi}_k(x) = f$ a.e. As a consequence, f must also be measurable by proposition 153 from the previous section.

Finally, $\varphi_k \to f$ a.e., so by definition of Lebesgue integral of f,

$$\int_{[a,b]}^{\mathcal{L}} f(x)dx = \lim_{k\to\infty} \int_{[a,b]}^{\mathcal{L}} \varphi_k(x)dx \tag{262}$$

and so, by equations 253, 256, 257, we get

$$\int_{[a,b]}^{\mathcal{L}} f(x)dx = \int_{[a,b]}^{\mathcal{R}} f(x)dx \tag{263}$$

as desired.

□

Stage Three: Nonnegative Functions.

Definition 200. The **(extended) Lebesgue integral of measurable nonnegative function** $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$[48] is the number

$$\int_{\mathbb{R}^n} f(x)dx := \sup_g \int_{\mathbb{R}^n} g(x)dx \tag{264}$$

where the supremum is taken over all measurable functions g such that $0 \leq g \leq f$, and it is the only class of functions we know how to integrate thus far, i.e. bounded functions supported on sets with finite measure.

[48] In particular, f is not necessarily bounded.

Definition 201. A **measurable nonnegative function** f **is (Lebesgue) integrable** if

$$\int_{\mathbb{R}^n} f(x)dx < \infty \tag{265}$$

We define integrals on subsets of $\mathbb{R}^n$ using the usual way using characteristic functions.

Example 202. Let $f_a : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$f_a := \begin{cases} |x|^{-a} & |x| \leq 1 \\ 0 & |x| > 1 \end{cases} \tag{266}$$

This is integrable exactly when $a < n$

Justify this.

Example 203. Let $F_a : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$F_a := \frac{1}{1 + |x|^a} \tag{267}$$

This is integrable exactly when $a > n$.

Justify this.

We again justify the following essential properties.

Proposition 204. Let f, g be nonnegative measurable functions.

1.) **Linearity.** If $f, g \geq 0$, and $a, b \in \mathbb{R}_{>0}$, then

$$\int_{\mathbb{R}^n} (af + bg) = a \int_{\mathbb{R}^n} f + b \int_{\mathbb{R}^n} g \tag{268}$$

2.) **Additivity.** If E, F are disjoint sets, then

$$\int_{E \cup F} f = \int_E f + \int_F f \tag{269}$$

3.) **Monotonicity.** If $f \leq g$, then

$$\int_{\mathbb{R}^n} f \leq \int_{\mathbb{R}^n} g \tag{270}$$

4.) If g is integrable, and $f \leq g$, then f is integrable.

5.) f is integrable, then f is finite a.e., i.e.$f(x) < \infty$ a.e.

6.) If $\int_{\mathbb{R}^n} f = 0$, then f vanishes a.e.

PROOF 205. The only proofs that we need to do any work are 1, 5, and 6 (and the proof for 6 is identical to one we provide before).

Proof of 1.) First note that for $c \in \mathbb{R}$, since we can pull constants out of supremums, and we know linearity for Lebesgue integrals of bounded measurable functions supported on sets of finite measure, we have

$$\int_{\mathbb{R}^n} cf = c \int_{\mathbb{R}^n} f$$

Thus, take wlog $a = b = 1$.
We need to prove both sides of the inequality (since the integral is defined in terms of a supremum which is inherently an inequality).

Lets prove $\leq$ direction. Take $\varphi \leq f, \psi \leq g$ as required by definition of the integral. Then

$$\varphi + \psi \leq f + g \tag{271}$$

But since $\varphi + \psi$ is a bounded measurable function supported on a finite measure,
we have

$$\int_{\mathbb{R}^n} \varphi + \int_{\mathbb{R}^n} \psi = \int_{\mathbb{R}^n} \varphi + \psi \leq \int_{\mathbb{R}^n} (f + g) \tag{272}$$

So, $\int_{\mathbb{R}^n} (f+g)$ is an upper bound to the sum of integrals of any two elements φ, ψ. Therefore,

$$\int_{\mathbb{R}^n} f + \int_{\mathbb{R}^n} g \leq \int_{\mathbb{R}^n} (f + g) \tag{273}$$

which proves the $\leq$ direction.

For the $\geq$ direction, take η bounded measurable function with support on a set of finite measure and $\eta \leq f + g$. Take $\eta_1 := \min(f, \eta)$, then

$$\eta_1 \leq f,\ \eta_2 \leq g \tag{274}$$

Also, η_1, η_2 must be bounded measurable function with support on a set of finite measure. Therefore,

$$\int \eta = \int (\eta_1 + \eta_2) = \int \eta_1 + \int \eta_2 \leq \int f + g \tag{275}$$

where in the last inequality, we used monotonicity. Taking the supremum over η gives

$$\int_{\mathbb{R}^n} f + \int_{\mathbb{R}^n} g \geq \int_{\mathbb{R}^n} (f + g) \tag{276}$$

as desired. $\square$

Proof of 2.) Observe that

$$\begin{aligned}
\int_{E \cup F} f &= \int_{\mathbb{R}^n} f \chi_{E \cup F} \\
&= \int_{\mathbb{R}^n} f(\chi_E + \chi_F) \\
&= \int_E f + \int_F f
\end{aligned}$$

by linearity. $\square$

Proof of 3.)
Since f is a measurable nonnegative bounded function such that $f \leq g$ and is supported on a set with finite measure, by definition of supremum,

$$\begin{aligned}
\int_{\mathbb{R}^n} g &= \sup_h \int_{\mathbb{R}^n} h \\
&\geq \int_{\mathbb{R}^n} f
\end{aligned}$$

$\square$

Proof of 4.)
Observe that from monotonicity,

$$\int_{\mathbb{R}^n} f \leq \int_{\mathbb{R}^n} g < \infty \tag{277}$$

so f is integrable. □

Proof of 5.) Let

$$E_k := \{x : f(x) \geq k\} \tag{278}$$

and

$$E_\infty := \{x : f(x) = \infty\} \tag{279}$$

Note that $E_k \searrow E_\infty$. We want to use the limit of decreasing sets theorem.

Then

$$\int_{\mathbb{R}^n} f \geq \int_{\mathbb{R}^n} f\chi_{E_k} \geq km(E_k) \tag{280}$$

by definition of E_k. So,

$$\frac{\int_{\mathbb{R}^n} f}{k} \geq m(E_k) \tag{281}$$

and so, $m(E_k) \to 0$ as $k \to \infty$. (Notice that we are using the integrability of f here since the argument will not work if the LHS in the above is ∞.) Therefore, in particular, for some k, the measure $m(E_k) < \infty$, and so, we can pass to the limit:

$$m(E_\infty) = \lim_{k\to\infty} m(E_k) = 0 \tag{282}$$

as desired. □

Proof of 6.) The corresponding proof (proposition 196) for bounded measurable functions supported on sets with finite measure works verbatim.

□

We now want to show a very important convergence theorem.

Proposition 206 (Fatou's Lemma). Suppose f_n is a sequence of nonnegative measurable functions. If $\lim_{n\to\infty} f_n = f$ pointwise a.e. (so, in particular, f is nonnegative measurable), then

$$\int f \leq \liminf_{n\to\infty} \int f_n \tag{283}$$

This also holds for when RHS, LHS are both $+\infty$.

Remark 207. Note that in general, we do not have equality. Consider

$$f_n(x) = \begin{cases} n & 0 < x < \frac{1}{n} \\ 0 & \text{otherwise} \end{cases} \tag{284}$$

Then $f_n \to 0$ for all x, but $\int f_n = 1$ for all n. However, this still satisfies Fatou's lemma.[49]

[49] In this view, this example is all we need to remember for the details of Fatou's lemma.

Remark 208 (How to remember Fatou's lemma.). The story up to now is all about nonnegative measurable functions, so that's a given. We also definitely want f_n to be bounded below by some number, at the very least if we remember that Fatou bounds the integral of f from above with the integral of the sequence; otherwise, the integral of the sequence could go off to $-\infty$, and the bound would not work.

The whole point of Fatou's lemma is that for a pointwise a.e. converging sequence of functions, we want to naïvely swap the limit and the integral. We certainly cannot do this (by the previous remark or from calculus). We further do not know if the limit of the integral will converge. But we always know that liminf exists, so we want to create an inequality using this.

PROOF 209. The only big theorem we use is the bounded convergence theorem. We do the usual approach for Lebesgue integral of nonnegative measurable function f: consider instead the nonnegative bounded measurable function g less than f supported on a set of finite measure. But this is not quite enough since we want these approximations to be less than f_n (so when we pass to the limit, we get the desired inequality). So, instead, take the minimum of f_n and g which has nice properties. We then pass to the limit g via bounded convergence, and we are done.

Let g be nonnegative, measurable, supported on a set of finite measure, and $0 \leq g \leq f$. Take

$$g_n := \min(g, f_n) \tag{285}$$

The minimum of two measurable functions is measurable (proposition 149 from last section), so g_n is measurable. The support of g_n is contained in the support of f_n, g, so g_n has support of finite measure. $g_n \to g$ a.e. since $g \leq f$, so for large n, $g \leq f_n$. (The a.e. comes from the fact that $f_n \to f$ a.e.) Therefore, by bounded convergence theorem,

$$\int g_n \to \int g \tag{286}$$

But now, $g_n \leq f_n$ by construction, so $\int g_n \leq \int f_n$, and so,

$$\int g \leq \liminf_{n\to\infty} \int f_n \tag{287}$$

(We must use liminf here since we have no reason to believe that the sequence $\int f_n$ converges. Recall that liminf always exists.)

□

We just need a tiny adjustment to get a full limit.

Proposition 210 (Fatou with Full Limit or Strong Monotone Convergence). Suppose f_n is a sequence of nonnegative measurable functions $f_n \to f$ pointwise a.e. (so, in particular, f is nonnegative measurable). Additionally, let $f_n \leq f$ a.e. Then

$$\lim_{n\to\infty} \int f_n = \int f \tag{288}$$

PROOF 211. (Since we already have Fatou,) we just need to prove the inequality for limsup to get the claim. But this is immediate since $f_n \leq f$ a.e., so by monotonicity of the integral,

$$\limsup_{n\to\infty} \int f_n \leq \int f \tag{289}$$

and combined with Fatou, we are done.

□

A special case of this is the monotone convergence theorem.

Proposition 212 (Monotone Convergence Theorem). Suppose f_n is a sequence of nonnegative measurable functions which increase to f, i.e. $f_n \nearrow f$. Then

$$\lim_{n\to\infty}\int f_n = \int f \tag{290}$$

Remark 213 (How to remember MCT.). As we see in the proof, the whole point why monotone convergence works is because we have a condition *in addition to Fatou* that makes the liminf a full limit. So, what would be the condition that makes a $\leq \liminf$ into a full limit? We certainly want $\geq \limsup$. So, we take the easiest possible condition to make this happen, namely to take $f_n \leq f$. This is strong MCT. For the standard MCT, we can just take increasing.

In short, we want nonnegative measurable monotone increasing because *Fatou, Fatou, and limsup*.

PROOF 214. Observe that the hypothesis of the theorem satisfies the hypothesis of the previous theorem. The result then follows immediately.

□

Here is a very useful consequence that is important for practical applications[50] of measure theory.

Proposition 215 (Swapping sums and integrals). Consider the series $\sum_{k=1}^{\infty} a_k(x)$ whose terms are nonnegative measurable. Then we can swap the sum and the integral[51], i.e.

$$\int \sum_{k=1}^{\infty} a_k(x)dx = \sum_{k=1}^{\infty}\int a_k(x)dx \tag{291}$$

If the sum on the RHS is finite, i.e. $\sum_{k=1}^{\infty}\int a_k(x)dx < \infty$, then the series $\sum_{k=1}^{\infty} a_k(x)$ converges a.e.

Remark 216. This theorem is a version of the Tonelli theorem and Fubini theorem which says that if a function $f : \mathbb{R}^{d_1}\times\mathbb{R}^{d_2} \to [0,\infty]$ is nonnegative measurable, then the slice and integral of the slice are measurable, and the integral $\int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}} f$ is given by an iterated integral. The comment "if the sum on the RHS is finite..." says that if $f \in L^1(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})$, then the slices of f are integrable, i.e. $f^y \in L^1(\mathbb{R}^{d_1})$ and $f_x \in L^1(\mathbb{R}^{d_2})$ which is the statement of the Fubini theorem.

Using the Fubini-Tonelli theorem[52], we can generalize the above proposition to:

Proposition 217. Consider the series $\sum_{k=1}^{\infty} a_k(x)$ where $a_k(x)$ is measurable. If either

$$\int \sum_{k=1}^{\infty} |a_k(x)|\, dx$$

$$\sum_{k=1}^{\infty}\int |a_k(x)|\, dx$$

are finite ($a_k \in L^1(\mathbb{R}^{d_1}\times\mathbb{N})$), then

$$\int \sum_{k=1}^{\infty} a_k(x)dx = \sum_{k=1}^{\infty}\int a_k(x)dx \tag{292}$$

[50] Practical applications always refers to applications within mathematics which occurs frequently.
[51] Note that the integral is over $\mathbb{R}^n$, so the integral can be either a definite integral over a compact interval or an improper integral over an unbounded region.
[52] Technically, we don't prove the theorem in full generality in these notes, but this application is still valid since the additional hypothesis is that the measure space is σ-finite.

PROOF 218 (Proof to swapping sums and integrals.). The only key idea is monotone convergence, and it is fairly obvious to what sequence we must apply the theorem to. The convergence follows immediately from the fact that integrability implies finiteness a.e.

Let f_n be the partial sums and f be the series, i.e.

$$f_n(x) := \sum_{k=1}^{n} a_k(x), \; f(x) := \sum_{k=1}^{\infty} a_k(x) \tag{293}$$

Then f_n is a sum of measurable functions, so it is measurable. Each term a_k is nonnegative, so $f_k \nearrow f$. So, by monotone convergence theorem,

$$\begin{aligned}
\sum_{k=1}^{\infty} \int a_k(x)dx &= \lim_{n\to\infty} \sum_{k=1}^{n} \int a_k(x)dx \\
&= \lim_{n\to\infty} \int \sum_{k=1}^{n} a_k(x)dx \\
&= \lim_{n\to\infty} \int f_n \\
&= \int f \qquad \text{(monotone convergence)} \\
&= \int \sum_{k=1}^{\infty} a_k(x)
\end{aligned}$$

In other words, we swap the sum and integral for a finite sum, and then we use monotone convergence to pass to the limit. This proves the first claim.

Now, if

$$\sum_{k=1}^{\infty} \int a_k(x)dx < \infty \tag{294}$$

then from the above equality,

$$\int \sum_{k=1}^{\infty} a_k(x) < \infty \tag{295}$$

and so, the series $\sum_{k=1}^{\infty} a_k(x)$ is integrable. But integrable functions are finite a.e., so we have the second claim.

□

Come back to the two observations on p.63 once I have done exercise 16 of chapter 1.

Stage Four: General Case

Here is the most general case (for real valued functions).

Definition 219. The **measurable function** $f : \mathbb{R}^n \to \mathbb{R}$ **is (Lebesgue) integrable (or** L^1**)** if the function $|f|$ is integrable (in the sense of nonnegative measurable functions), i.e. the L^1**-norm** is finite:

$$\|f\|_{L^1} := \int_{\mathbb{R}^n} |f| < \infty \tag{296}$$

The space of all integrable functions on $\mathbb{R}^n$ is the the space $L^1(\mathbb{R}^n)$.[53]

Definition 220. Let $f \in L^1(\mathbb{R}^n)$. The **Lebesgue integral of an L^1 function** is the number

$$\int f := \int f^+ - \int f^- \tag{298}$$

where the integrals on the RHS are in the sense of Lebesgue integrals of nonnegative functions, and f_+, f^- are the positive and negative parts of f, i.e.

$$f^+ := \max(f, 0),\ f^- := \min(f, 0) \tag{299}$$

(which are clearly nonnegative).

Remark 221. Since $f^\pm \leq |f|$, $f^\pm$ is integrable whenever f is.

Proposition 222 (The definition is independent of the decomposition of f.). The definition of the Lebesgue integral is independent of the decomposition of f, i.e. if $f = f_1 - f_2 = g_1 - g_2$ for nonnegative integrable functions f_1, f_2, g_1, g_2, then

$$\int f_1 - \int f_2 = \int g_1 - \int g_2 \tag{300}$$

PROOF 223. This follows immediately from the linearity of the Lebesgue integral for nonnegative measurable functions. By hypothesis,

$$f_1 + g_2 = g_1 + f_2 \tag{301}$$

and integrating both sides gives

$$\int f_1 + \int g_2 = \int (f_1 + g_2) = \int (g_1 + f_2) = \int g_1 + \int f_2 \tag{302}$$

and so,

$$\int f_1 - \int f_2 = \int g_1 - \int g_2 \tag{303}$$

as desired.

□

Remark 224. The integrability and the value of the integral of a function f is unchanged if we modified the function f on a null set. So, it is reasonable to adopt the convention (in the context of integration) that functions are undefined on null sets. (This justifies why we let L^1 be the equivalence classes of functions equivalent up to null sets.) As a consequence of this convention, we know unambiguously what the sums of integrable functions are; integrable functions are finite a.e., and so, in particular, the ambiguity of adding extended value $\pm\infty$ only occurs on null sets (which by the above convention, we choose to ignore).

Proposition 225 (Basic Properties of Lebesgue Integrals of L^1 functions.). The Lebesgue integral of L^1 functions are linear, additive, monotonic, and satisfies the triangle inequality.

[53] Technically, $L^1(\mathbb{R}^n)$ is the space of all *equivalence classes* of integrable functions, but we do the usual simplification of only looking at the representative of those equivalence classes. The equivalence is defined to be

$$f \sim g \iff f = g \text{ a.e.} \tag{297}$$

This is a sensible definition since we have treated functions agreeing a.e. as the same. We continue this discussion in the next subsection.

PROOF 226. The idea for all of them is to just pass to nonnegative measurable functions via the definition of the integral. Let f, g be L^1 functions.

Proof of linearity. Let $a, b \in \mathbb{R}$. Then

$$(af+bg)^+ = af^+ + bg^+,\ (af+bg)^- = af^- + bg^- \tag{304}$$

so,

$$\begin{aligned}
\int (af+bg) &= \int (af^+ + bg^+) - \int (af^- + bg^-) \\
&= a\int f^+ + b\int g^+ - a\int f^- - b\int g^- \qquad \text{(linearity)} \\
&= a\left(\int f^+ - \int f^-\right) + b\left(\int g^+ - \int g^-\right) \\
&= a\int f + b\int g
\end{aligned}$$

□

Proof of additivity. If E, F are disjoint sets,

$$\begin{aligned}
\int_{E\cup F} f &= \int_{E\cup F} f^+ - \int_{E\cup F} f^- \\
&= \left(\int_E f^+ + \int_F f^+\right) - \left(\int_E f^- + \int_F f^-\right) \\
&= \int_E (f^+ - f^-) + \int_F (f^+ - f^-) \\
&= \int_E f + \int_F f
\end{aligned}$$

□

Proof of monotonicity. If $f \leq g$, then

$$f^+ + g^- \leq g^+ + f^-$$

and so, by monotonicity for nonnegative measurable functions,

$$\begin{aligned}
\int (f^+ + g^-) &\leq \int (g^+ + f^-) \\
\int (f^+ - f^-) &\leq \int (g^+ - g^-) \\
\int f &\leq \int g
\end{aligned}$$

where we used linearity. □

Proof of triangle inequality. Since $f^+, f^- \leq |f|$, and disjoint union of the supports of f^+, f^- is the support of $|f|$, so

$$\begin{aligned}\left|\int f\right| &= \left|\int f^+ - \int f^-\right| \\ &= \int f^+ + \int f^- \\ &\leq \int |f|\end{aligned}$$

□

We now more towards the proof of the most important convergence theorem: the dominated convergence theorem. But first, we prove some easier theorems which use similar proof techniques.

The main theorem for the proofs of the following two theorems is the monotone convergence theorem. For both parts, wlog we can assume that $f \geq 0$ (by replacing f by $|f|$)

Proposition 227 (L^1 functions vanish at infinity.). Let f be L^1 on $\mathbb{R}^n$ and $\epsilon > 0$. Then there exists a ball B of finite measure such that

$$\int_{B^c} |f| < \epsilon \tag{305}$$

PROOF 228. The trick is to realize that

$$\int_{B^c} f = \int f(1 - \chi_B) = \int f - \int_B f \tag{306}$$

Otherwise, proof is just by first principles.

Consider the sequence of balls $B_N := B_N(0)$, and if we take $f_N := f\chi_{B_N}$, then f_N is nonnegative, measurable[54], increasing, and pointwise converge to f. So, f_N satisfies the hypotheses of monotone convergence theorem, and so,

$$\lim_{N\to\infty} \int f_N = \int f \tag{307}$$

So, for some large N

$$0 \leq \int_{B_N^c} f = \int f - \int f\chi_{B_N} < \epsilon \tag{308}$$

as desired.

□

Proposition 229 (Absolute Continuity of the Lebesgue Integral.). Let f be L^1 on $\mathbb{R}^n$, and let $\epsilon > 0$. Then there is some $\delta > 0$ such that $m(E) < \delta$ implies

$$\int_E |f| < \epsilon \tag{309}$$

PROOF 230. The idea is to consider the part of f from a set where f takes on low values and from the set where f takes on high values. The mass contributed by the latter set can be bounded via the technique from the first part (i.e. by monotone convergence theorem). So, our job is to bound the second part.

[54] Recall that the product of measurable functions is measurable since $\{fg < a\} = \bigcup_{r\in\mathbb{Q}}(\{f < r\} \cap \{g < \frac{a}{r}\})$.

Fix $\epsilon > 0$. We use the idea from the proof of the first part; take $f_N := f\chi_{E_N}$ where

$$E_N := \{x : f(x) \leq N\} \tag{310}$$

i.e. points for which $f(x)$ lies fairly low. But f_N is nonnegative, measurable, and also increasing to f. So by monotone convergence theorem, there is some N such that[55]

$$\int (f - f_N) < \frac{\epsilon}{2} \tag{311}$$

Let $\delta > 0$ which we shall determine later. If $m(E) < \delta$, then

$$\begin{aligned}
\int_E f &= \int_E (f - f_N) + \int_E f_N \\
&= \int_E (f - f_N) + \int_E f_N \\
&\leq \int (f - f_N) + \int_E f_N \\
&\leq \int (f - f_N) + m(E)N \\
&< \int (f - f_N) + N\delta
\end{aligned}$$

We want this to be less than ϵ, so we just take $N\delta < \frac{\epsilon}{2}$. □

Remark 231. The idea for the first part of the proof is often used (e.g. in the second part of the proof, the following proof of DCT) to show that the mass of a function is small outside some set. What are the conditions on the sequence of sets E_N can we do this? The key lies in the fact that we use the monotone convergence theorem. Mainly, we need E to be 1.) measurable (so that χ_{E_N} is a measurable function[56]) and 2.) increasing to $\mathbb{R}^n$ (so that f_N increases to f). If the sequence satisfies these properties, then we have the mass vanishing. This is worth generalizing in the following corollary.

Proposition 232. Let E_N be an increasing sequence of measurable sets such that $E_N \nearrow \mathbb{R}^n$. Then for $\epsilon > 0$, there exists some N such that

$$\int_{E_N^c} f < \epsilon \tag{312}$$

for any $f \in L^1(\mathbb{R}^n)$.

PROOF 233. Do the same as in the first part of the previous proposition. □

We are now ready for the big theorem.

Proposition 234 (Dominated Convergence Theorem). Suppose f_n is a sequence of measurable functions pointwise converging a.e. to f. If additionally, they are all bounded absolutely by a nonnegative integrable function g (i.e. $|f_n| \leq g$), then

$$\int |f_n - f| \to 0 \qquad \text{as } n \to \infty \tag{313}$$

which implies that

$$\int f_n \to \int f \tag{314}$$

as $n \to \infty$.

[55] For intuition, notice that $f - f_N = f\chi_{E_N^c}$, i.e. the values of f lying outside of the low set E_N.

[56] Note that a characteristic function χ_A is measurable iff A is measurable since $A = \chi_A^{-1}(1)$ and $A^c = \chi_A^{-1}(0)$.

PROOF 235. The use both bounded convergence and monotone convergence. We use the same method of proof as the absolute continuity of Lebesgue integrals: bound the integral inside and outside a good set. Insider the good set, we bound using bounded convergence[57].

For $N \in \mathbb{N}_0$, let[58]

$$E_N := \{|x| \leq N, g(x) \leq N\} \tag{315}$$

For $\epsilon > 0$, we can apply the exact same proof as proposition 227. Namely, take $g_N(x) = g(x)\chi_{E_N}$ and apply the same proof verbatim. Observe that g_N is nonnegative, measurable, and also increasing to g. So, by monotone convergence theorem, we get

$$\lim_{N\to\infty} \int g_N = \int g \tag{316}$$

and so,

$$\int_{E_N^c} g < \epsilon \tag{317}$$

for large N.

On the other hand, $f_n\chi_{E_N}$ is bounded by N (by choice of E_N) and supported on a set of finite measure[59]. This satisfies the hypothesis of bounded convergence theorem, so for large n

$$\int_{E_N} |f_n - f| < \epsilon \tag{318}$$

Now we are done. Combining the above two estimates, we get

$$\begin{aligned}
\int |f_n - f| &= \int_{E_N} |f_n - f| + \int_{E_N^c} |f_n - f| \\
&= \int_{E_N} |f_n - f| + \int_{E_N^c} g \qquad \text{(usual triangle inequality)} \\
&\leq \epsilon + 2\epsilon = 3\epsilon
\end{aligned}$$

as desired.

□

Remark 236. Notice that the dominated convergence theorem gives convergence of the functions f_n in the L^1-norm which in turn implies convergence of the (sequence of) integrals of f_n (via the triangle inequality). Note that by the same argument, convergence in the norm is stronger than convergence of the integral, by triangle inequality.

Complex-valued Functions.

This is the final stage to our generalization of the Lebesgue integral (although this step is equally unsophisticated as step four).

57 Recall that for absolute continuity, we got the bound for δ from this step. For the proof of the vanishing of L^1 functions at infinity, we did not have to do this step

58 Notice that this is the intersection of the sets B_N from the vanishing of L^1 functions at infinity and E_N from the proof of absolute continuity the proof. Indeed, even at the level of the sets we deal with, the two proofs are very similar to this proof of DCT.

59 Note that we cook up the set E_N for this step to work (in other words E_N is rigged so that we can apply bounded convergence theorem to $f_n\chi_{E_N}$.)

Definition 237. A **measurable complex valued function** $f : \mathbb{R}^n \to \mathbb{C}$ is a function $f = u+iv,\ u, v : \mathbb{R}^n \to \mathbb{R}$ is such that the real and imaginary parts u, v are measurable. The **Lebesgue integrable complex valued function** $f : \mathbb{R}^n \to \mathbb{C}$ is a function such that its modulus is integrable, i.e. the nonnegative function

$$|f| = \sqrt{u^2 + v^2} \tag{319}$$

is integrable. The The **Lebesgue integral of a complex valued function** $f : \mathbb{R}^n \to \mathbb{C}$ is the number given by

$$\int f(x)dx := \int u(x)dx + i \int v(x)dx \tag{320}$$

where the integrals on the RHS are in the sense of the integrals of measurable functions.

We define integrals of complex valued functions on measurable sets $E \subseteq R^n$ in the usual way.

Proposition 238. The complex-valued integrable functions on a measurable set $E \subseteq \mathbb{R}^n$ forms a $\mathbb{C}$-vector space.

PROOF 239. The niceness is just inherited from the real valued counterpart. The set of complex-valued integrable functions on a measurable set forms an abelian group because real valued integrable functions on a measurable set forms an abelian group and addition is defined componentwise. Similarly for scalar multiplication. □

Remark 240. On a similar note, the additivity, linearity, and monotonicity of the integral is inherited directly from real valued function.

Proposition 241 (Triangle Inequality.). The triangle inequality holds for complex-valued integrable functions on a measurable set $E \subseteq \mathbb{R}^n$.

PROOF 242. By triangle inequality for complex modulus,

$$|f + g| \leq |f| + |g| \tag{321}$$

and so by monotonicity of the integral, we get the conclusion. □

The Space L^1 of Integrable Functions . -Monday, 7.9.2018

Topological Properties of L^1.

We discuss here some topological properties of the space $L^1(\mathbb{R}^d)$ relevant to analysis.

Definition 243. The **L^1-norm of a function f on R^d** is the number

$$\|f\| := \|f\|_{L^1} := \int_{\mathbb{R}^d} |f(x)|\, dx \tag{322}$$

Definition 244 (Working Definition of $L^1(\mathbb{R}^d)$.). The space $L^1(\mathbb{R}^d)$ is the collection of all functions $f : \mathbb{R}^d \to \mathbb{R}$ such that

$$\|f\| = \int_{\mathbb{R}^d} |f(x)|\, dx < \infty \tag{323}$$

As we noted before, the formal definition of $L^1(\mathbb{R}^d)$ is the collection of all equivalence classes of functions f such that $\|f\| < \infty$ where equivalence is defined to be

$$f \sim g \iff f = g \qquad \text{a.e.} \tag{324}$$

Notice that since the norm is defined in terms of an integral, it is well defined up to the representative of the equivalence class (i.e. if two functions f, g are equivalent, then their norms agree).

Proposition 245 (L^1-norm is a norm.). Let $f, g \in L^1(\mathbb{R}^d)$.

1.) $\|af\| = |a| \, \|f\|$ for all $a \in \mathbb{C}$
2.) $\|f + g\| \leq \|f\| + \|g\|$
3.) $\|f\| = 0$ iff $f = 0$ a.e.

In particular, $d(f, g) := \|f - g\|$ defines a metric on $L^1(\mathbb{R}^d)$.

Note that the final comment about the metric holds true for any vector norm.

PROOF 246. We don't need to do anything. The first property follows immediately from the linearity of the integral. The second property follows from integrating (over $\mathbb{R}^d$) the usual triangle inequality

$$|f + g| \leq |f| + |g| \tag{325}$$

and then using linearity of the integral. We already showed before the nontrivial direction of the third property (sixth property in proposition 204 from last subsection). So, we are done.

$\square$

Not only is $L^1(\mathbb{R}^d)$ a normed vector space, it is complete, i.e. it is a Banach space.

Proposition 247 (Riesz-Fischer). $L^1(\mathbb{R}^d)$ is a Banach space, i.e. it is complete under its metric.

PROOF 248. Take a sequence $\{f_n\}$ which is Cauchy (with respect to the L^1-norm). We claim that this sequence is convergent to some $f \in L^1(\mathbb{R}^d)$. The first question is: what is this f?

We take a subsequence $\{f_{n_k}\}$ which converges quickly, i.e.

$$\left\| f_{n_{k+1}} - f_{n_k} \right\| \leq 2^{-k} \tag{326}$$

for all $k \in \mathbb{N}$. (We of course obtain this from the Cauchyness of the sequence.) This then allows us to define a function

$$f(x) := f_{n_1}(x) + \sum_{k=1}^{\infty} (f_{n_{k+1}}(x) - f_{n_k}(x)) \tag{327}$$

i.e., a series whose partial series (up to K) telescopes to f_{K+1}. Note that at this point, we do not know is the series converges. (Showing $f \in L^1(\mathbb{R}^d)$ will solve this issue. Why?)

In order to show that f is integrable, we can bound its absolute value from above by the function

$$g(x) := |f_{n_1}(x)| + \sum_{k=1}^{\infty} \left| f_{n_{k+1}}(x) - f_{n_k}(x) \right| \tag{328}$$

and show it is integrable. We rigged our initial choice of the subsequence for this step:

$$\begin{aligned} \int g &= \int \left(|f_{n_1}(x)| + \sum_{k=1}^{\infty} \left| f_{n_{k+1}}(x) - f_{n_k}(x) \right| \right) \\ &= \int |f_{n_1}(x)| + \sum_{k=1}^{\infty} \left\| \left| f_{n_{k+1}}(x) - f_{n_k}(x) \right| \right\| \qquad \text{(monotone convergence)} \\ &\leq \int |f_{n_1}(x)| + \sum_{k=1}^{\infty} 2^{-k} < \infty \end{aligned}$$

where in the second line, we can use monotone convergence to swap sum and integral because the Kth telescoping sum gives $f_{n_{K+1}}$ which is integrable. So, $g \in L^1(\mathbb{R}^d)$.

But since $|f| \leq g$ (by triangle inequality), we have $f \in L^1(\mathbb{R}^d)$. But since L^1 functions are finite a.e., this implies that the series defining f converges a.e. This proves the existence of f.

Now we need to show convergence in the norm. First, since the partial sums defining f are just $f_{n_{K+1}}$, we have the pointwise convergence

$$f_{n_{K+1}} \to f \tag{329}$$

almost everywhere. In order to get convergence in the norm, we need to use dominated convergence; specifically, since $|f - f_{n_k}| \leq g$ for all k, we have

$$\|f - f_{n_k}\|_1 \to 0 \tag{330}$$

as $k \to \infty$, as required.

Finally, to show that the entire sequence (rather than just the subsequence) converges to f, we just need to go far enough down the sequence that the entire sequence is very close to the good subsequence. Fix $\epsilon > 0$. Choose N large enough so that $n, m > N$ implies $\|f_n - f_m\| < \frac{\epsilon}{2}$. For $n_k > N$, we must have $\|f_{n_k} - f\| < \epsilon/2$ which by triangle inequality gives

$$\|f_n - f\| \leq \|f_n - f_{n_k}\| + \|f_{n_k} - f\| < \epsilon \tag{331}$$

for $n > N$. This shows that f_n converges to f in the L^1 norm as required.

□

If we stop in the middle of the above argument, we have the following corollary.

Proposition 249. A Cauchy sequence in the L^1 norm has a pointwise convergent subsequence.

We proved this a an intermediate step in Riesz-Fischer.

Families Dense in L^1.

There are a few families of functions dense in L^1. This is in general very useful for proving properties of integrable functions because it is often easier to prove it for the following class of functions, and then we can approximate the integrable function using this class of functions to obtain the general result.[60]

Proposition 250 (Families dense in L^1.). The following families of functions are dense in $L^1(\mathbb{R}^n)$:

1.) Simple functions
2.) Step functions
3.) Continuous functions with compact support

PROOF 251. The proofs are deceptively easy. For the first two statements, we use the ideas from the approximation theorems for measurable functions that we proved in the last section, and in that sense, we do not need any new ideas for the first two proofs. The third proof is easy for one dimensions, and higher dimensions falls right out from the one dimensional case.

WLOG, it suffices to consider $f \in L^1(\mathbb{R}^n)$ such that $f \geq 0$[61]. For each of the following, our goal is to show that there is an element from each of the family which is arbitrarily close to f in the norm. We also need to

[60] Note that we have taken a similar approach to this before when we were defining the Lebesgue integral; we first explored the properties using simple functions and then passed to the limit to obtain properties of bounded functions with support on a set with finite measure.

[61] Firstly, we can assume that f is real valued since we can consider the real and imaginary parts separately. In which case, we can also consider the positive and negative parts of this function separately, and both of these functions are nonnegative, so it suffices to consider a nonnegative function.

prove the three statements in this order because they preceding statements simplify the later ones.

Proof. Simple Functions. We need to do nothing here. We know from the previous section that a nonnegative measurable function has an increasing sequence of nonnegative simple functions pointwise converging to it (proposition 155). We then use dominated convergence to get convergence in the norm:

$$|f - \varphi_k| \to 0 \tag{332}$$

as $k \to \infty$. Therefore, there are simple functions arbitrarily close to f in the L^1-norm. $\square$

Proof. Step Functions. The idea of the proof is very similar to the proof that measurable functions are pointwise limits of step functions a.e. (proposition 162 from last section).
From the first statement, an L^1 function can be approximated by a simple function, and so it suffices to show that indicator functions χ_E of set E of finite measure can be approximated well by a step function, i.e. $\|\chi_E - \psi\|_{L^1}$ is small. But we have already provided a procedure to construct such a step function in the proof of proposition 162 from last section.
Namely, by Littlewood's First Principle, approximate E by finitely many cubes which we can partition to get almost disjoint rectangles. We can then make adjustments to the edges of the rectangles so that we get smaller disjoint rectangles $R_j, j = 1, ..., M$ contained in the almost disjoint rectangles:

$$m\left(E \triangle \bigcup_{j=1}^{M} R_j\right) \leq 2\epsilon \tag{333}$$

for fixed $\epsilon > 0$. Therefore, χ_E and $\psi = \sum_{j=1}^{M} \chi_{R_j}$ differ on a set of measure at most 2ϵ. Therefore,

$$\|\chi_E - \psi\|_{L^1} = \int_{E \triangle \bigcup_{j=1}^{M} R_j} (\chi_E - \psi) \leq 2m\left(E \triangle \bigcup_{j=1}^{M} R_j\right) < 2\epsilon \tag{334}$$

since $|\chi_E - \psi| = 0, 1$[62]. $\square$

Proof. Continuous functions with compact support. From the second statement, it suffices to approximate a characteristic function of a rectangle $f = \chi_R$. In the one dimensional case, we can approximate this by a smooth bump function (such as the ones we use for constructing partitions of unity) or even more elementary, the function

$$g(x) = \begin{cases} 1 & a \leq x \leq b \\ 0 & x \leq a - \epsilon, \ x \geq b + \epsilon \end{cases} \tag{335}$$

and g is linear on $[a-\epsilon, a]$ and $[b, b+\epsilon]$, i.e. $\frac{1}{\epsilon}(x - (a-\epsilon))$ and $-\frac{1}{\epsilon}(x - b)$ respectively. Then g does not agree with f only on the intervals $[a - \epsilon, a]$ and $[b, b + \epsilon]$, so

$$\|f - g\|_{L^1} = \int_{[a-\epsilon,a]\cup[b,b+\epsilon]} (f - g) \leq 2m([a - \epsilon, a] \cup [b, b + \epsilon]) \leq 4\epsilon \tag{336}$$

Now in n-dimensions, we can do this same construction for each component of the approximating function. In other words, characteristic functions of a rectangle is just a product of characteristic functions of intervals, so the desired continuous function of compact support is simply the product of functions like g that we just constructed.

$\square$

[62] Notice that $\chi_E, \psi = 0, 1$ since χ_E is a characteristic function, and ψ is the sum of characteristic functions over disjoint rectangles, so the only way the two functions agree are when they are equal (so $|\chi_E - \psi| = 0$), and they disagree if one takes the value 1 and the other takes the value 0 (in which case $|\chi_E - \psi| = 1$).

Invariance Properties of Lebesgue Integral.

Here we discuss the invariance properties of the Lebesgue integral with respect to translation, dilation, and reflection. Together, this gives the linear change of variables formula of the Lebesgue integral.

Proposition 252. Let $f \in L^1(\mathbb{R}^d)$. If $c \in \mathbb{R}, y \in \mathbb{R}^d$, then

$$\int_{\mathbb{R}^d} f(cx+y)dx = \frac{1}{c}\int_{\mathbb{R}^d} f(x)dx \tag{337}$$

PROOF 253. This follows immediately from the following theorems.

□

Proposition 254 (Lebesgue integral is translation invariant.). If $f \in L^1(\mathbb{R}^d)$, then $f(x-h) \in L^1(\mathbb{R}^d)$ for $h \in \mathbb{R}^d$ and

$$\int_{\mathbb{R}^d} f(x-h)dx = \int_{\mathbb{R}^d} f(x)dx \tag{338}$$

PROOF 255. We check for classes of functions which are easier to deal with, namely characteristic, simple, nonnegative measurable, and finally integrable/complex valued.

For characteristic functions, the result follows immediately from the translation invariance of the Lebesgue measure which we have already shown:

$$\int \chi_E(x)dx = m(E) = m(E_h) = \int \chi_E(x-h)dx \tag{339}$$

Now the Lebesgue function is linear, so

$$\int \sum_{n=1}^N a_n\chi_{E_n}(x)dx = \sum_{n=1}^N a_n \int \chi_{E_n}(x)dx = \sum_{n=1}^N a_n \int \chi_{E_n}(x-h) = \int \sum_{n=1}^N a_n\chi_{E_n}(x-h)dx \tag{340}$$

so the assertion holds for simple functions.

For nonnegative measurable function f, there is a sequence of simple functions ϕ_n converging pointwise a.e. to f, then since the simple functions are translation invariant, we can pass to f by monotone convergence.

So now for integrable and complex-valued functions, we can just observe that $\|f_h\| = \|f\|$ for $f_h(x) := f(x-h)$, and so, f_h is integrable if f is. The invariance holds immediately from the definition of Lebesgue integral of integrable and complex valued functions since it is written as a linear combination of the integral of the nonnegative function.

□

Proposition 256 (Lebesgue integral is invariant under dilation and translation.). Let $f \in L^1(\mathbb{R}^d)$. Then for $\delta > 0$,

$$\delta^d \int_{\mathbb{R}^d} f(\delta x)dx = \int_{\mathbb{R}^d} f(x)dx \tag{341}$$

and

$$\int_{\mathbb{R}^d} f(-x)dx = \int_{\mathbb{R}^d} f(x)dx \tag{342}$$

PROOF 257. The idea of the proof is the same as in the previous proposition. We can inherit the invariance of Lebesgue measure under dilation and reflection by first considering the characteristic function. We pass to simple functions using linearity of the Lebesgue integral. Then pass to nonnegative measurable functions using increasing sequence of nonnegative measurable functions and using the monotone convergence theorem. Finally, we pass to integrable and complex valued functions using the integral for these two cases.

$\square$

Remark 258. When we say the Lebesgue integral is "invariant" under dilation, we just mean we need to tack on a scaling factor to get the correct value.

Here is a very useful corollary.

Proposition 259 (Convolution is commutative.). If $f, g \in L^1(\mathbb{R}^d)$, then $h_x(y) := f(x-y)g(y) \in L^1(\mathbb{R}^d)$, and the **convolution** $f * g$ is equal to $g * f$, i.e.

$$(f*g)(x) := \int_{\mathbb{R}^d} f(x-y)g(y)dy = \int_{\mathbb{R}^d} f(y)g(x-y)dy =: (g*f)(x) \tag{343}$$

PROOF 260. We observe that the above change of variables is given by a translation and a reflection. Taking $x-y$, we translate this by $-x$, and then we reflect this to get y. From the previous propositions, the integral should be invariant throughout this transformation. The integrability of $f(x-y)g(y)$ with respect to y also follows from the previous theorems.

$\square$

Here is another corollary.

Proposition 261. For all $\epsilon > 0$,

$$\int_{|x|\geq\epsilon} \frac{dx}{|x|^a} = \epsilon^{-a+d} \int_{|x|\geq 1} \frac{dx}{|x|^a} \tag{344}$$

for $a > d$, and

$$\int_{|x|\leq\epsilon} \frac{dx}{|x|^a} = \epsilon^{-a+d} \int_{|x|\leq 1} \frac{dx}{|x|^a} \tag{345}$$

for $a < d$.

PROOF 262. This follows immediately by the dilation invariance.

$\square$

Translations and Continuity.

Definition 263. Let f be a function on $\mathbb{R}^d$. For $h \in \mathbb{R}^d$, the **translation of the function** f **by the vector** h is $f_h(x) := f(x-h)$.

We notice that if the translation f_h converges pointwise to f as $h \to 0$, then this is equivalent to the continuity of f. However, a general integrable function can be discontinuous everywhere even if corrected on a null set. . Nevertheless, there is an overall continuity that is satisfied by an arbitrary integrable function given by the norm.

Insert exercise 2.15

Proposition 264. Suppose $f \in L^1(\mathbb{R}^d)$. Then

$$\|f_h - f\| \to 0 \qquad \text{as } h \to 0 \tag{346}$$

PROOF 265. Fix $\epsilon > 0$. First we note that the proposition holds for a continuous function with compact support $g \in L^1(\mathbb{R}^d)$. Since g is continuous, we can take $\delta > 0$ such that $|h| < \delta$ implies $|g(x-h) - g(x)| < \epsilon/m(\text{supp} g)$, so

$$\|g_h - g\| = \int_{\mathbb{R}^d} |g(x-h) - g(x)|\, dx \leq \epsilon/m(\text{supp} g) \cdot m(\text{supp} g) = \epsilon \tag{347}$$

The proof for the general case is just a simple 3ϵ-argument using the density of the continuous functions with compact support in $L^1(\mathbb{R}^d)$.

We can find a continuous functions with compact support $g \in L^1(\mathbb{R}^d)$ with $\|f - g\| < \epsilon$. Now,

$$\begin{aligned}\|f_h - f\| &\leq \|f_h - g_h\| + \|g_h - g\| + \|f - g\| \\ &\leq 3\epsilon\end{aligned}$$

The first and third term we bound by hypothesis and translation invariance. The second follows from proposition for continuous functions with compact support.

□

Fubini's Theorem. -Monday, 7.16.2018

Fubini's Theorem.

Definition 266. Let f be a function on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. The **slice of the function f corresponding to** $y \in \mathbb{R}^{d_2}$ is the function f^y of the variable $x \in \mathbb{R}^{d_1}$ given by

$$f^y(x) := f(x, y) \tag{348}$$

and likewise, $f_x(y) := f(x, y)$.

The **slices of a set** $E \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ are the sets

$$E^y := \left\{x \in \mathbb{R}^{d_1} : (x, y) \in E\right\},\ E_x := \left\{y \in \mathbb{R}^{d_2} : (x, y) \in E\right\} \tag{349}$$

Remark 267. Note the convention: the superscript indicates fixing a point in $\mathbb{R}^{d_2}$ and subscript indicates fixing a point in $\mathbb{R}^{d_1}$.

Definition 268. A **product set** of sets $E_j \subseteq \mathbb{R}^{d_j}, j = 1, ..., N$ are sets of the form

$$E = \prod_{j=1}^{N} E_j \tag{350}$$

Example 269. We need to be careful in dealing with slices. A slice of a measurable set or function may not be measurable.

For instance, if a set A in $\mathbb{R}^2$ is a product of a nonmeasurable set (such as the one we constructed using the axiom of choice in the last section) and a null set (say, the Cantor set), then the resulting set is a null set in $\mathbb{R}^2$ whose slice is not measurable. If a measurable function f is such that $f^{-1}(a) = A$, then its slice f_x is is not measurable since $f_x^{-1}(a)$ is the nonmeasurable set.

Proposition 270 (Fubini's Theorem). Suppose $f \in L^1(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$. Then for almost every $y \in \mathbb{R}^{d_2}$, $f^y \in L^1(\mathbb{R}^{d_1})$ and $\int_{\mathbb{R}^{d_1}} f^y(x)dx \in L^1(\mathbb{R}^{d_2})$.

Moreover, we have the formula

$$\int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f(x, y)dx\right) dy = \int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}} f \tag{351}$$

and hence

$$\int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f(x, y)dx\right) dy = \int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} f(x, y)dy\right) dx \tag{352}$$

Obviously, the first statement is symmetric.

Remark 271. Fubini's theorem says that an integral in. $\mathbb{R}^d$ can be computed by iterating lower-dimensional integrals, and the iteration can be taken in any order.

A closely related theorem is by Tonelli:

Proposition 272 (Tonelli's Theorem). For a nonnegative measurable function $f : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to [0, \infty]$, for almost every $y \in \mathbb{R}^{d_2}$, the slice $f^y \in \mathbb{R}^{d_2}$ is measurable and the integral $\int_{\mathbb{R}^{d_1}} f^y(x)dx$ is measurable on $\mathbb{R}^{d_2}$.

Moreover,

$$\int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f(x,y)dx \right) dy = \int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}} f \tag{353}$$

in the extended sense, i.e. if the LHS is infinity, then RHS is infinity.

Remark 273. Notice the difference between Fubini and Tonelli. Fubini gives the *integrability* of the slice and integral of the slice whereas Tonelli gives *measurability*.

However, we often use the theorems in conjunction. Starting with a measurable function f on $\mathbb{R}^d$, we apply Tonelli's theorem to the nonnegative measurable function $|f|$, and compute the integral over $\mathbb{R}^d$ via iterated integration. If the integral is finite, then we see that $f \in L^1(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$ which is the hypothesis for Fubini's theorem. We can then use Fubini to calculate the original integral. These two are used together often enough that it is often called a single theorem.

Proposition 274 (Fubini-Tonelli Theorem). If f is measurable, and if any of the integrals

$$\int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} |f(x,y)|\, dy \right) dx$$
$$\int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} |f(x,y)|\, dx \right) dy$$
$$\int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}} |f(x,y)|\, dxdy$$

is finite (i.e., $f \in L^1(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$), then

$$\int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} f(x,y)dy \right) dx = \int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f(x,y)dx \right) dy = \int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}} f \tag{354}$$

Or, concisely: if a measurable function is L^1, then we can compute the integral as iterated integral.

We recall that we proved a version of this as a consequence of the monotone convergence theorem (proposition 215). Recall that that was just a special case of the Tonelli and Fubini theorems for the counting and Lebesgue measure. At the end of the day, the necessary hypothesis for swapping integrals is that the function is integrable which can be obtained by Tonelli's theorem by considering the absolute value of the function.

PROOF 275 (Proof of Tonelli's Theorem). The main idea is to apply Fubini's theorem on truncations and pass to the limit using monotone convergence theorem. Note that we use all three parts of Fubini's theorem.

We post-pone the proof of Fubini's theorem until we discuss general measure spaces since there is a much cleaner proof on p.86 of Bass.

We use the usual truncation

$$f_k(x,y) := \begin{cases} f(x,y) & (x,y) \in B_k(0),\ f(x,y) < k \\ 0 & \text{otherwise} \end{cases} \tag{355}$$

Since f_k is bounded with compact support, it is integrable. By Fubini's theorem, almost every slice is integrable, hence measurable. So, there exists a null set $E_k \subset \mathbb{R}^{d_2}$ such that f_k^y is measurable for $y \in E_k^c$. If we then take $E = \bigcup_k E_k$, then f^y is measurable for all $y \in E^c$ and all k, and E is a null set. Now, a truncation f_k^y increases to f^y, so by monotone convergence, if $y \notin E$,

$$\int_{\mathbb{R}^{d_1}} f_k(x,y)dx \nearrow \int_{\mathbb{R}^{d_1}} f(x,ydx) \tag{356}$$

as $k \to \infty$.

Now again by Fubini's theorem, $\int_{\mathbb{R}^{d_1}} f_k(x,y)dy$ is measurable for all $y \in E^c$, and by the above limit, the limit of a measurable function is measurable, so $\int_{\mathbb{R}^{d_1}} f(x,y)dy$ is measurable.

Now we can apply monotone convergence again to get

$$\int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}} f_k = \int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f_k(x,y)dx\right) dy \nearrow \int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f(x,y)dx\right) dy \tag{357}$$

where the LHS follows from Fubini's theorem. So, by monotone convergence on f_k, we conclude that

$$\int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}} f_k \to \int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}} f \tag{358}$$

which gives the conclusion.

$\square$

Applications of Tonelli's Theorem.

We explore some consequences of Tonelli's theorem. We get quite a few useful result on measurability of sets and functions. They will be quite useful for constructing examples.

Here is an immediate consequence of Tonelli's theorem.

Proposition 276 (Almost every slice of a measurable set is measurable.). If $E \subseteq \mathbb{R}^{d_1} \times \mathbb{R}^{d^2}$ is measurable, then for almost every $y \in \mathbb{R}^{d_2}$, the slice $E^y \subseteq \mathbb{R}^{d_1}$ is measurable. Moreover, $m(E^y)$ is a measurable function of y and

$$m(E) = \int_{\mathbb{R}^{d_2}} m(E^y)dy \tag{359}$$

PROOF 277. Apply Tonelli's theorem to the function $f(x,y) := \chi_E$. Then the slice

$$E^y = (f^y)^{-1}(\mathbb{R}^{d_1}) \tag{360}$$

is measurable since f^y is measurable (by first part of Tonelli's theorem). Moreover, the function

$$m(E^y) = \int_{\mathbb{R}^{d_1}} f^y(x)dx \tag{361}$$

is measurable (by second part of Tonelli's theorem), and (by third part of Tonelli's theorem)

$$m(E) = \int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}} f(x,y)dxdy = \int_{\mathbb{R}^{d_2}} f^y(x)dx = \int_{\mathbb{R}^{d_2}} m(E^y)dy \tag{362}$$

as desired.

$\square$

Example 278. The converse to the above assertion doesn't hold, i.e. there is a nonmeasurable set E for which almost every slice E^y is measurable.

Let $\mathcal{N} \subseteq \mathbb{R}$ be a nonmeasurable set, and consider

$$E := [0,1] \times \mathcal{N} \subset \mathbb{R}^2 \tag{363}$$

Then $E^y = [0,1], \emptyset$ depending on y, so it is measurable for every y. If E was measurable, then the previous proposition implies that almost every slice $E_x = \mathcal{N}$ is measurable which is certainly not true.

Example 279. Here is another counterexample which is even more striking. This time we find a non-measurable set for which all the x and y slices are measurable. This construction relies on the continuum hypothesis.

We assume that there exists a partial ordering $\prec$ of $\mathbb{R}$ such that $\{x : x \prec y\}$ is countable[63] for each $y \in \mathbb{R}$. Do Problem 5 in chapter 2

Consider

$$E = \{(x,y) \in [0,1]^2,\ x \prec y\} \tag{364}$$

For all $y \in [0,1]$, E^y is countable (by construction of $\prec$), so E^y is a null set, so in particular, it is measurable. Meanwhile, E_x is countable since it is a complement of a countable set. Now, E cannot be measurable by the previous proposition because if it was, then the formula

$$\int_{\mathbb{R}^{d_2}} m(E^y)dy = \int_{\mathbb{R}^{d_1}} m(E_x)dx \tag{365}$$

gives a contradiction.

We now want to look at the relation between sets and their product sets. Notice that in the following discussion, we need to use exterior measure instead of measure since we do not a priori that the sets are measurable.

We first prove a lemma.

Proposition 280. Let E_1, E_2 be any sets in $\mathbb{R}^{d_1}, \mathbb{R}^{d_2}$ respectively. Then

$$m_*(E_1 \times E_2) \le m_*(E_1)_*(E_2) + O(\epsilon) \tag{366}$$

PROOF 281. This is a routine exercise. Fix $\epsilon > 0$. Then by definition of exterior measure, there exists countably many cubes Q_k, Q_l' that cover E_1, E_2 and

$$\sum_{k=1}^{\infty} |Q_k| \le m_*(E_1) + \epsilon,\ \sum_{l=1}^{\infty} |Q_l'| \le m_*(E_2) + \epsilon \tag{367}$$

Now $Q_k \times Q_l'$ cover E_1, E_2, so by subadditivity,

$$\begin{aligned}
m_*(E_1 \times E_2) &\le \sum_{k,l=1}^{\infty} |Q_k \times Q_l'| \\
&= \left(\sum_{k=1}^{\infty} |Q_k|\right)\left(\sum_{l=1}^{\infty} |Q_l'|\right) \\
&\le (m_*(E_1) + \epsilon)(m_*(E_2) + \epsilon) = m_*(E_1)_*(E_2) + O(\epsilon)
\end{aligned}$$

and since ϵ is arbitrary, we get the desired result.

[63] Note how powerful cardinality is. Countability implies measure zero, and for a finite measure space, this means the complement is full measure. This is all we need in this construction.

□

Proposition 282 (Measurability of product sets.). Let $E := E_1 \times E_2$. If E is measurable and E_2 has positive exterior measure, then E_1 is measurable. Conversely, the product of measurable sets is measurable, i.e. if E_1, E_2 are measurable, then $E_1 \times E_2$ is measurable.

Moreover, the measure of a product is the product of measures:

$$m(E_1 \times E_2) = m(E_1)m(E_2) \tag{368}$$

PROOF 283. Let's first prove the first statement. By Tonelli's theorem, for almost every $y \in \mathbb{R}^{d_2}$, the slice

$$(\chi_{E_1 \times E_2})^y(x) = \chi_{E_1}(x)\chi_{E_2}(y) \tag{369}$$

is measurable. Our claim is that there exists some $y \in E_2$ such that $\chi_{E_1 \times E_2}(x, y) = \chi_{E_1}(x)$ is measurable (from which we get that $E_1 = \chi^{-1}_{E_1 \times E_2}(\{1\})$ is measurable).

The existence of such a y requires the hypothesis $m_*(E_2) > 0$. Let F be the set of $y \in \mathbb{R}^{d_2}$ such that E^y is measurable. But the slice E^y is measurable for a.e. y by the previous proposition, so F^c is a null set. But $E_2 \cap F \neq \emptyset$ since it has positive exterior measure:

$$0 < m_*(E_2) \leq m_*(E_2 \cap F) + m_*(E_2 \cap F^c) = m_*(E_2 \cap F) \tag{370}$$

since F^c is a null set. So, we can take $y \in E_2 \cap F$, and this proves the claim. □

For the second assertion, we just need to show that E is measurable. This second assertion is certainly true for G_δ-sets since the product of G_δ-sets are G_δ-sets which are then measurable.[64]

The idea is then to extend to all measurable sets: recall that that a set is measurable iff it differs from a G_δ-set by a null set (proposition 116). So, for each E_j, there is a G_δ-set G_j containing E_j and $m_*(G_j - E_j) = 0$. Now, $G = G_1 \times G_2$ is measurable, so

$$G \setminus E = (G_1 \times G_2) \setminus (E_1 \times E_2) \subseteq ((G_1 \setminus E_1) \times G_2) \cup (G_1 \times (G_2 \setminus E_2)) \tag{371}$$

But now, by the previous proposition, this implies that

$$m(G \setminus E) \leq \epsilon(m(G_1) + m(G_2)) \tag{372}$$

so, $G \setminus E$ is a null set, hence E is measurable. □

The third assertion about $m(E)$ is an immediate consequence of proposition 276:

$$\begin{aligned} m(E) &= \int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} \chi_{E_1 \times E_2}(x, y) dx \right) dy \\ &= \left(\int_{\mathbb{R}^{d_1}} \chi_{E_1}(x) dx \right) \left(\int_{\mathbb{R}^{d_2}} \chi_{E_2}(y) dy \right) \\ &= m(E_1)m(E_2) \end{aligned}$$

□

Here is an immediate consequence.

Proposition 284. Suppose f is measurable on $\mathbb{R}^{d_1}$. Then $\tilde{f}(x, y) := f(x)$ is measurable on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$.

[64] Let $U_1 = \bigcap_{i=1}^\infty V_i, U_2 = \bigcap_{j=1}^\infty W_j$ be G_δ-sets. Then $U_1 \times U_2 = \bigcap_{i=1}^\infty (V_i \times W_i)$ which we can verify easily from set theory.

PROOF 285. This is a straightforward application of the previous proposition. WLOG, we can assume f is real-valued (since we can consider the real and imaginary parts). Since f is measurable, $\{x \in \mathbb{R}^{d_1} : f(x) < a\}$ is measurable for all $a \in \mathbb{R}$, and so,

$$\left\{(x,y) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} : \tilde{f}(x,y) < a\right\} = \left\{x \in \mathbb{R}^{d_1} : f(x) < a\right\} \times \mathbb{R}^{d_2} \tag{373}$$

is measurable by the previous proposition. Thus, $\tilde{f}$ is measurable.

□

We are now ready to go back to the classical interpretation of the integral, namely as the area under the graph of the function.

Proposition 286 (Lebesgue integral as area under the graph.). Let f be a nonnegative function on $\mathbb{R}^d$, and let

$$\mathcal{A} := \left\{(x,y) \in \mathbb{R}^d \times \mathbb{R} : 0 \leq y \leq f(x)\right\} \tag{374}$$

i.e. the region under the graph of f. Then f is measurable iff $\mathcal{A}$ is measurable in $\mathbb{R}^{d+1}$.

If f is measurable, then

$$\int_{\mathbb{R}^d} f(x)dx = m(\mathcal{A}) \tag{375}$$

PROOF 287. For the first statement, if f is measurable on $\mathbb{R}^d$, then by the previous proposition

$$F(x,y) := y - f(x) \tag{376}$$

is a sum of measurable functions, so it is measurable. Thus, $\mathcal{A} = \{y \geq 0\} \cap \{F \leq 0\} = \{F = 0\}$ is measurable. ($\{y \geq 0\}$ is measurable since $h(x,y) = y$ is measurable.) This gives the forward direction.

The backward direction follows immediate from proposition 276. The slice

$$\mathcal{A}_x = \{y \in \mathbb{R} : (x,y) \in \mathcal{A}\} = [0, f(x)] \tag{377}$$

for any $x \in \mathbb{R}^d$. So, from proposition 276, $f(x) = m(\mathcal{A}_x)$ is measurable, and

$$m(\mathcal{A}) = \int m(\mathcal{A}_x)dx = \int_{\mathbb{R}^d} f(x)dx \tag{378}$$

□

Chapter 2: Integration Theory -Wednesday, 7.11.2018

Problem (3.). First observe that if $I \subseteq (k\pi, (k+4)\pi)$, then[65]

[65] Here, we can equivalently take the linear change of variables $y = x - (k+1)\pi$, but we have not proven the change of variables formula yet.

$$
\begin{aligned}
\int_I f(x)dx &= \int_{\mathbb{R}} f(x)\chi_I(x)dx \\
&= \int_{\mathbb{R}} f(x-(k+1)\pi)\chi_I(x)dx \qquad \text{(periodicity of } f\text{)} \\
&= \int_{\mathbb{R}} f(y)\chi_I(y+(k+1)\pi)dy \qquad \text{(let } y = x-(k+1)\pi\text{)} \\
&= \int_{\mathbb{R}} f(y)\chi_{I-(k+1)\pi}(y)dy \\
&= \int_{\mathbb{R}} f(y)\chi_{[-\pi,\pi]}(y)dy \\
&= \int_{-\pi}^{\pi} f(y)dy
\end{aligned}
$$

where we used the notation

$$
I-(k+1)\pi := [k\pi-(k+1)\pi, (k+2)\pi-(k+1)\pi] = [-\pi,\pi] \tag{379}
$$

The parts which are a bit fishy are from lines 2 to 3 and 3 to 4. From line 2 to 3 is simply a substitution, and so, nothing has really happened. Line 3 to 4 follows from the definition of the indicator function:

$$
\begin{aligned}
\chi_I(y+(k+1)\pi) &= \begin{cases} 1 & k\pi \le y+(k+1)\pi \le (k+2)\pi \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} 1 & -\pi \le y \le \pi \\ 0 & \text{otherwise} \end{cases} \\
&= \chi_{I-(k+1)\pi}(y)
\end{aligned}
$$

From this reduction, we can assume wlog that $I := (a,b) \subseteq [-\pi, 3\pi]$. Then

$$
\begin{aligned}
\int_I f(x)dx &= \int_a^{\pi} f(x)dx + \int_{\pi}^{b} f(x)dx \\
&= \int_a^{\pi} f(x)dx + \int_{-\pi}^{b-2\pi} f(x+2\pi)dx \\
&= \int_a^{\pi} f(x)dx + \int_{-\pi}^{a} f(x)dx \qquad \text{(periodicity)} \\
&= \int_a^{b} f(x)dx
\end{aligned}
$$

where in the third line, we used the fact that $b-a = 2\pi$, given as hypothesis. □

Problem (6. Integrability of f on $\mathbb{R}$ does not imply $f(x) \to 0$ at $\pm\infty$). a.) Take the continuous function which takes on the values n on $\left[n, n+\frac{1}{n^3}\right)$. The easy was is to take smooth bump functions (which decay on an interval of, say length ϵ) on each of these intervals, multiply them by n, and then take a piecewise function of this.

To be more concrete, we can construct a similar function as follows. Take

$$f(x) := \begin{cases} 2n^4\left(x - \left(n - \frac{1}{2n^3}\right)\right) & x \in \left[n - \frac{1}{2n^3}, n\right), n = 2, 3, 4, ... \\ -2n^4\left(x - \left(n + \frac{1}{2n^3}\right)\right) & x \in \left[n, n + \frac{1}{2n^3}\right), n = 2, 3, 4, ... \\ 0 & \text{otherwise} \end{cases} \tag{380}$$

i.e., a sequence of spikes at $n = 2, 3, 4, ...$ which become taller ($f(n) = n$) and sharper (slope $2n^4$ for the nth spike) as $n \to \infty$. We see that this is continuous at the $n - \frac{1}{2n^3}, n, n + \frac{1}{2n^3}$, and so, it is continuous everywhere. Notice that

$$\limsup_{n\to\infty} f(n) = \infty \tag{381}$$

but

$$\int f \leq \sum_{n=2}^{\infty} n \cdot \frac{1}{n^3} < \infty \tag{382}$$

so, $f \in L^1(\mathbb{R})$. □

b.) We assume WLOG[66] that $f \geq 0$. Suppose for contradiction that $f \not\to 0$ as $x \to \infty$[67], i.e. we can take

$$\limsup_{x\to\infty} f(x) \geq M \tag{383}$$

for some M. Then for all $N > 0$, there is some $x \geq N$ such that $f(x) \geq M$. So, take a sequence of points x_n such that $f(x_n) \geq M$. Then by the uniform continuity, for a fixed $\epsilon > 0$, there is some δ such that $x \in B_\delta(x)$ implies $|f(x) - f(y)| < \epsilon$. So,

$$f(y) > M - \epsilon \tag{384}$$

for all $y \in B_\delta(x_n)$. So, in particular,

$$\int f \geq \sum_{n=1}^{\infty} 2\delta \cdot (M - \epsilon) = \infty \tag{385}$$

which violates the integrability of f. Therefore, f must decay at infinity.

Problem (9. Tchebychev Inequality). Observe that

$$\alpha\chi_{E_\alpha} \leq f\chi_{E_\alpha} \leq f \tag{386}$$

and so, by monotonicity of the integral,

$$\alpha m(E_\alpha) = \alpha \int \chi_{E_\alpha} \leq \int f \tag{387}$$

so,

$$m(E_\alpha) \leq \frac{1}{\alpha} \int f \tag{388}$$

We need to take $\alpha > 0$ so that we can divide by α, and so that the inequality does not flip when we do so. We need to f nonnegative to get the first inequality $f\chi_{E_\alpha} \leq f$.

[66] We can do this since we can replace f by $|f|$; if f is uniformly continuous and integrable, then so is $|f|$. If $|f|$ converges to 0, then so will f.

[67] We can apply the same argument to $x \to -\infty$

Problem (10. Some Equivalent Conditions for Integrability.). The first equivalence is immediate. For the forward direction, by monotonicity of the integral,

$$2^k m(F_k) < \int_{F_k} f(x)dx \tag{389}$$

and since F_k partitions the entire space, we get

$$\sum_{k=-\infty}^{\infty} 2^k m(F_k) < \sum_{k=-\infty}^{\infty} \int_{F_k} f(x)dx = \int f < \infty \tag{390}$$

and so, if f is integrable, the sum converges. For the converse, observe that

$$\int_{F_k} f(x)dx \leq 2^{k+1} m(F_k) \tag{391}$$

and so, summing again,

$$\int f = \sum_{k=-\infty}^{\infty} \int_{F_k} f(x)dx \leq 2 \sum_{k=-\infty}^{\infty} 2^k m(F_k) < \infty \tag{392}$$

so, we have the first equivalence.

Chapter 3. Differentiation and Integration

Differentiation of the Integral. -Wednesday, 7.18.2018

The goal of this chapter is to prove the first and second fundamental theorem of calculus. Namely, we address the following two questions:

1.) If $f \in L^1([a,b])$, and $F(x) := \int_a^x f(y)dy$, then does this imply that F is differentiable a.e. and $F' = f$?
2.) For $F : [a,b] \to \mathbb{R}$, when does F' exist, is integrable, and

$$F(b) - F(a) = \int_a^b F'(x)dx \tag{393}$$

Let's start with the first problem (i.e. first fundamental theorem of calculus). Naturally, the only thing we can do to take a derivative of the indefinite integral F is to take its difference quotient:

$$\begin{aligned} \frac{F(x+h) - F(x)}{h} &= \frac{1}{h}\int_x^{x+h} f(y)dy \\ &= \frac{1}{|I|}\int_I f(y)dy \end{aligned}$$

which is an average over the set $I := (x, x+h)$. Therefore, the question of the differentiability of F becomes the question of whether

$$\lim_{|I| \to 0} \frac{1}{|I|}\int_I f(y)dy = f(x) \tag{394}$$

holds for suitable $x \in I$. We can also consider higher dimensional versions of this question for which f map from a set in $\mathbb{R}^d$, and $I \subseteq \mathbb{R}^d$. In doing this, we restrict our attention to when the sets are balls. For the sake of our discussion, we call the above problem the **average problem**.

Proposition 288. If $f : \mathbb{R}^d \to \mathbb{R}$ is continuous, then the average problem has an affirmative answer, i.e. for all $x \in \mathbb{R}^d$,

$$\lim_{m(B) \to 0} \frac{1}{m(B)}\int_B f(y)dy = f(x) \tag{395}$$

for balls $B \subset \mathbb{R}^d$.

PROOF 289. Since f is continuous, for all $\epsilon > 0$, there is some $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$. So, if $\operatorname{diam}(B) < \delta$,

$$\begin{aligned} \left| f(x) - \frac{1}{m(B)}\int_B f(y)dy \right| &= \left| \frac{1}{m(B)}\int_B (f(x) - f(y))dy \right| \\ &\leq \frac{1}{m(B)}\int_B |f(x) - f(y)|\, dy < \epsilon \end{aligned}$$

$\square$

Hardy-Littlewood Maximal Function.

In considering the average problem, we consider the following auxiliary function.

Definition 290. If $f \in L^1(\mathbb{R}^d)$, then the **(Hardy-Littlewood) maximal function of** f is the function $f^* : \mathbb{R}^d \to \mathbb{R}$ given by

$$f^*(x) := \sup_{x \in B} \int_B |f(y)|\, dy \tag{396}$$

Proposition 291 (Main Properties of the Hardy-Littlewood Maximal Function.). Let $f \in L^1(\mathbb{R}^d)$. Then the maximal function f^* is measurable, finite-valued a.e., and satisfies the **weak-type bound**

$$m(\{x \in \mathbb{R}^d : f^*(x) > \alpha\}) \le \frac{3^d}{\alpha} \|f\|_{L^1} \tag{397}$$

for all $\alpha > 0$.

Remark 292. The constant 3^d comes from the Vitali covering lemma, as we shall see. We can make this bound tighter via geometric measure theory, but we do not need to do this since we just need to know that it is independent of α, f.

Remark 293. We define precisely what is meant by weak and strong bounds.

Let T be an operator (i.e. a map) from measurable functions on one measure space to another measure space. The **operator** T **is weak type** (p,p) if there exists some $C > 0$ such that

$$m(\{|Tf| > \alpha\}) \le C\alpha^{-1/p} \|f\|_p \tag{398}$$

(and we call the inequality a **weak** (p,p)**-bound**). The **operator** T **is strong type** (p,p) if there exists some $C > 0$ such that

$$\|Tf\|_p \le C \|f\|_p \tag{399}$$

Strong (p,p)-bounds imply the weak (p,p) bounds.

Work out the details in Strong type inequality (p,q) implies weak type inequality (p,q)?. Also, come back to this after doing the Marcinkiewicz interpolation theorem. See What does a "weak type" inequality mean? for the comment: "The Marcinkiewicz interpolation theorem shows how to obtain strong type inequalities from weak type

Vitali Covering Lemma.

The proofs of the first two parts of the property of the Hardy-Littlewood maximal function is fairly straightforward (given that the second part relies on the inequality). The inequality relies on a finite version of Vitali covering lemma. This lemma is core in the theory of differentiation. Since the general Vitali covering also relies on the fintie Vitali covering, we will also prove this here.

Proposition 294 (Finite Vitali Covering Lemma.). Let $\mathcal{B} = \{B_i\}_{i=1}^N$ be a finite collection of balls in $\mathbb{R}^d$. Then there exists a finite subcollection $B_{i_1}, ..., B_{i_k}$ such that

$$m\left(\bigcup_{l=1}^N B_l\right) \le 3^d \sum_{j=1}^k m(B_{i_j}) \tag{400}$$

i.e., $3B_{i_j}$ cover the original balls.

Remark 295. In the two applications of the finite Vitali covering lemma (the general Vitali covering lemma and weak-type inequality of the maximal function), the key idea is to approximate the set using *compact sets* to extract a finite subcovering.

We use the notation $3B := B_{3r}(p)$ where $B := B_r(p)$.

PROOF 296. The proof is a constructive proof. We have a rule for choosing the finite subcollection of balls, and in the end, we prove that it satisfies the appropriate bound.

The key geometric notion at the heart of the proof is the following. This is where the number 3 comes from.

Geometric principle. Suppose B and B' are balls that intersect, and such that B is bigger than B' (in the sense of radius). Then $B' \subseteq 3B$.

This is immediate when one draws a picture. Now here is the rule for selecting the balls:

1.) Choose a maximal ball B_{i_1} from $\mathcal{B}_1 := \mathcal{B}$ (where maximal means one with largest radius; note that we use the fact that $\mathcal{B}$ is finite here). Take $\mathcal{B}_2$ to be the elements of $\mathcal{B}_1$ which are not B_{i_1} nor the balls that intersect B_{i_1}.
2.) Iterate for at most N steps.

The balls that are eliminated in the nth step are the balls which are covered by $3B_{i_n}$ (by the geometric principle).

We now show that the subcollection obtained by the above algorithm satisfies the inequality. By construction $B \in \mathcal{B}$ intersects some element B_{i_j} in the subcollection, and $B \subseteq 3B_{i_j}$. So,

$$\begin{aligned} m\left(\bigcup_{l=1}^{N} B_l\right) &\le m\left(\bigcup_{j=1}^{k} 3B_{i_j}\right) \\ &\le \sum_{j=1}^{k} m\left(3B_{i_j}\right) = 3^d \sum_{j=1}^{k} m\left(B_{i_j}\right) \end{aligned}$$

□

We now prove the

Definition 297. A collection of balls $\mathcal{B}$ is a Vitali covering of a set E [68] if all points of E have a ball containing it with arbitrarily small measure, i.e., if for every $x \in E$ and $\epsilon > 0$, there is a ball B in the collection which contains B and the measure is small, i.e $m(B) < \epsilon$.

Proposition 298 ((General) Vitali Covering Lemma). Let E be a set of finite measure, and let $\mathcal{B}$ be a Vitali covering of E. For any $\delta > 0$, there exists finitely many disjoint balls $B_1, ..., B_N$ in $\mathcal{B}$ which cover "most" of E, i.e.

$$\sum_{i=1}^{N} m(B_i) \ge m(E) - \delta \tag{401}$$

PROOF 299. The idea is to apply Vitali covering to portions of the region E so that it covers at least a constant amount of area A[69]. Then for large k, the amount covered must be at least $m(E) - \delta$.

The proof is again, very algorithmic. Take A small, say less than $m(E)$. We first claim that there is some collection of disjoint balls $B_1, ..., B_N$ in $\mathcal{B}$ which cover a portion of E, i.e.[70]

$$\sum_{i=1}^{N_1} m(B_i) \ge \frac{A}{3^d} \tag{402}$$

[68] Note that we do not impose any condition on the cardinality of $\mathcal{B}$.
[69] In Stein, we take this constant amount to be δ which is the same letter as what appears in the hypothesis. However, the only part in which the δ from the hypothesis appears is in the final line, and so we choose to make these two letters distinct.
[70] Of course, the 3^d comes from the Vitali covering lemma. As we see here, A is the $m(\bigcup_{l=1}^{N} B_i)$ in the finite Vitali covering lemma i.e., the area of a region in E which are being covered by the finite subcollection of balls.

First, since $E_1 := E$ is measurable, there is a compact set $E_1' \subseteq E_1$ such that $m(E_1') \geq A$ (proposition 83). [71] But by compactness, we can cover E_1' with finitely many balls from $\mathcal{B}$ to which we can use finite Vitali covering lemma and obtain disjoint balls $B_1, ..., B_{N_1}$ satisfying the inequality.

If this subcollection satisfies the desired inequality, we stop here. Otherwise, we iterate. In the latter case, we consider

$$E_2 := E_1 \setminus \bigcup_{i=1}^{N_1} \overline{B_i} \tag{403}$$

We then repeat the previous argument, i.e. choose a compact set $E_2' \subseteq E_2$ with $m(E_2') \geq A$. Take $\mathcal{B}_2$ to be the elements in $\mathcal{B}_1$ which are disjoint from $\bigcup_{i=1}^{N_1} \overline{B_i}$.

Now $\mathcal{B}_2$ certainly cover E_2 since the union of $\mathcal{B}_2$ and $\{B_i\}_{i=1}^{N_1}$ must cover E_1. Additionally, it is a *Vitali* covering. Indeed, since $\mathcal{B}$ is a Vitali covering of E_1, for $x \in E_2$, there is some ball in $\mathcal{B}$ containing x. Also, $\bigcup_{i=1}^{N_1} \overline{B_i}$ is a compact set[72], so dist $\left(x, \bigcup_{i=1}^{N_1} \overline{B_i}\right) = R > 0$. But now, by definition of Vitali covering, we can take a ball $B_r(p)$ containing x such that $m(B_r(p)) \leq Cr^d < CR^d$. (Or in other words, since the volume of a ball containing x is controlled by its radius, we can take one with radius smaller than R.)

Now, we are allowed to iterate the step from before to obtain balls $B_i, N_1 + 1 \leq i \leq N_2$ such that there total volume is at least $\frac{A}{3^d}$, and so,

$$\sum_{i=1}^{N_2} m(B_i) \geq \frac{2A}{3^d} \tag{404}$$

After k iterations of the above, we get balls $B_i,\ 1 \leq i \leq N_k$ with

$$\sum_{i=1}^{N_k} m(B_i) \geq \frac{kA}{3^d} \tag{405}$$

But we are done since there exists some $k \in \mathbb{N}$ such that

$$k \geq \frac{m(E) - \delta}{3^d A} \tag{406}$$

□

We are now ready to prove the property of maximal functions.

PROOF 300 (Proof of the property of the Hardy-Littlewood maximal function.). As noted before, the second assertion follows immediately from the third since

$$m(\left\{x \in \mathbb{R}^d : f^(x) = \infty\right\}) \leq m(\left\{x \in \mathbb{R}^d : f^(x) \geq \alpha\right\}) \leq \frac{3^d}{\alpha} \|f\|_{L^1} \to 0 \tag{407}$$

as $\alpha \to \infty$.

---

[71] The statement of the proposition says that for $\epsilon > 0$, there is a compact set $E_1' \subseteq E_1$ such that $m(E_1 \setminus E_1') < \epsilon$. But now, if we take $m(E_1 \setminus E_1') = m(E_1) - a$ for $m(E_1) > a \geq A$, then

$$\begin{aligned} m(E_1') &= m(E_1) - m(E_1 \setminus E_1') \\ &= a \geq A \end{aligned}$$

as required.

[72] This is why we need to take *closure* when we define E_2.

The first assertion is also fairly simple; the set $\{f^* > \alpha\}$ is open since if y is in this set, then by definition of f^*, and in particular the supremum, there is a ball B containing y so that

$$\frac{1}{m(B)} \int_B |f(y)|\, dy > \alpha \tag{408}$$

and so, for $x \in B$, we can take a smaller ball $B' \subseteq B$ around y so that the above inequality holds. Therefore, $x \in \{f^* > \alpha\}$.

For the third assertion, we prove the assertion for an arbitrary compact subset of $\{f^* > \alpha\}$. But measurable sets are approximated arbitrarily well by compact sets (proposition 88), so this gives us the desired result[73].

Take $x \in \{f^* > \alpha\}$. Again, by definition of the maximal function, and in particular the supremum, there is a ball B_x containing x such that

$$\frac{1}{m(B_x)} \int_{B_x} |f(y)|\, dy > \alpha \tag{410}$$

so,

$$m(B_x) < \frac{1}{\alpha} \int_{B_x} |f(y)|\, dy \tag{411}$$

Take a compact set $K \subseteq \{f^* > \alpha\}$[74] which is covered by $B_x, x \in \{f^* > \alpha\}$, so extracting a finite subcover $B_l, l = 1, ..., N$ and applying finite Vitali covering lemma, we get a disjoint subcollection $B_{l_j}, j = 1, ..., k$ such that

$$m\left(\bigcup_{l=1}^{N} B_l\right) \leq 3^d \sum_{j=1}^{k} m(B_{i_j}) \tag{412}$$

So,

$$\begin{aligned} m(K) &\leq m\left(\bigcup_{l=1}^{N} B_l\right) \\ &\leq 3^d \sum_{j=1}^{k} m\left(B_{i_j}\right) \\ &\leq \frac{3^d}{\alpha} \sum_{j=1}^{k} \int_{B_{i_j}} |f(y)|\, dy \\ &\leq \frac{3^d}{\alpha} \int_{\bigcup_{j=1}^{k} B_{i_j}} |f(y)|\, dy \\ &\leq \frac{3^d}{\alpha} \|f\|_1 \end{aligned}$$

and this holds for all compact $K \subseteq E$, we are done.

$\square$

[73] To be more precise, for all compact $K \subseteq E$ we have $m(K) \leq M$, then since we can choose $\epsilon > 0$ such that $m(E \setminus K) < \epsilon$, so

$$m(E) = m(E \setminus K) + m(K) \leq M + \epsilon \tag{409}$$

and we can take $\epsilon \to 0$ to get desired result.

[74] Recall that measurable sets contain compact sets.

Lebesgue Differentiation Theorem

We now have a solution to the averaging problem. This is the key idea behind (both) Fundamental Theorems of Calculus. (We will spell this out more explicitly later.)

Proposition 301 (Lebesgue Differentiation Theorem). If $f \in L^1(\mathbb{R}^n)$, then

$$f(x) = \lim_{m(B)\to 0, x\in B} \frac{1}{m(B)} \int_B f(y)dy \tag{413}$$

for a.e. x.

PROOF 302. As usual, the goal is to show that the set of bad points have measure 0. The key idea is to approximate f with continuous functions for which the statement trivially holds. As we have seen a number of times before, this is the key strategy in analysis: to approximate an arbitrary function with good functions and then use some bounding results to get the desired claim. The bounding results which are useful here are Tchebychev and weak-type estimate for the maximal function; this is a natural guess since they are the only results we know thus far that relates an upper bound of the function to the measure of the set on which the function is bounded.

We claim that for $\alpha > 0$, the set

$$E_\alpha = \left\{ x \in \mathbb{R}^d : \limsup_{m(B)\to 0, x\in B} \left| \frac{1}{m(B)} \int_B f(y)dy \right| > 2\alpha \right\} \tag{414}$$

has measure zero so that $E = \bigcup_{n=1}^\infty E_{1/n}$ has measure zero, hence we get the claim on E^c.

Fix $\alpha > 0$. Then continuous functions with compact support are dense in $L^1(\mathbb{R}^d)$, so there is such a g with $\|f - g\|_1 < \epsilon$. Therefore,

$$\begin{aligned}\limsup_{m(B)\to 0, x\in B} \left| \frac{1}{m(B)} \int_B f(y)dy \right| &\le \limsup_{m(B)\to 0, x\in B} \frac{1}{m(B)} \int_B |f(y) - g(y)|\, dy + \frac{1}{m(B)} \int_B |g(y) - g(x)|\, dy + |g(x) - f(x)| \\ &\le (f-g)^*(x) + |g(x) - f(x)|\end{aligned}$$

where $*$ indicates the maximal function. Therefore, $E_\alpha \subseteq F_\alpha \cup G_\alpha$ for [75]

$$F_\alpha := \{x : (f-g)^* > \alpha\}, \; G_\alpha := \{x : |g(x) - f(x)| > \alpha\} \tag{415}$$

so we just need to bound these sets. But now we know how to do this. For the former, by weak-type inequality,

$$m(F_\alpha) \le \frac{3^d}{\alpha} \|f - g\|^1 \tag{416}$$

and for the latter, by Tchebychev,

$$m(G_\alpha) \le \frac{1}{\alpha} \|f - g\|^1 \tag{417}$$

But we chose g so we can control these norms, so we can choose g so that

$$m(E_\alpha) \le \frac{3^d + 1}{\alpha} \|f - g\|^1 \le \frac{3^d + 1}{\alpha} \epsilon \tag{418}$$

and thus, E_α is a null set as required.

[75] We chose to use 2α instead of α in the claim precisely for this step. If $u_1 + u_2 > 2\alpha$ for positive numbers u_1, u_2, then this occurs only if $u_i > \alpha$ for either $i = 1, 2$.

□

Here is an immediate consequence:

Proposition 303. For $f \in L^1(\mathbb{R}^d)$, $f^* \geq |f|$ a.e.

PROOF 304. By applying Lebesgue differentiation theorem to $|f|$, we have the result.

□

We note here that in the Lebesgue differentiation theorem, since we are taking a sequence of smaller balls, the theorem is a local one. Thus, it makes sense to introduce the following definition.

Definition 305. A measurable function f is locally integrable if $f\chi_B$ is integrable for all $B \subseteq \mathbb{R}^n$. The set of all locally integrable functions is denoted by $L^1_{\text{loc}}(\mathbb{R}^d)$.

Remark 306. Note that we require that it is integrable over *all* balls. So, we just ignore the behavior at infinity.

Example 307. Here are some examples: $e^{|x|}, |x|^{-1/2} \in L^1_{\text{loc}}(\mathbb{R})$. We see this since there are closed formulas for their antiderivatives (for $x > 0$ and $x < 0$ separately).

We can weaken the hypothesis of the Lebesgue differentiation theorem by our earlier comment.

Proposition 308 (Lebesgue Differentiation Theorem for Locally Integrable Functions). If $f \in L^1_{\text{loc}}(\mathbb{R}^n)$, then

$$f(x) = \lim_{m(B)\to 0, x\in B} \frac{1}{m(B)} \int_B f(y)dy \tag{419}$$

for a.e. x.

Here are some additional notions that result from the Lebesgue differentiation theorem. Since the theorem is about how much mass the function f has on certain sets, it is now natural to talk about the *density* of a given set (where the terminology is inspired by the obvious intuition).

Definition 309. A point $x \in \mathbb{R}^d$ is a point of Lebesgue density of a measurable set E if

$$\lim_{m(B)\to 0, x\in B} \frac{m(B\cap E)}{m(B)} = 1 \tag{420}$$

The **density of the set E at x** is the number

$$d(x, E) := \lim_{m(B)\to 0, x\in B} \frac{m(B\cap E)}{m(B)} \tag{421}$$

if the limit exists.

Remark 310. It should be clear why the Lebesgue differentiation theorem is useful for this notion. (As we will see in the proof that follows,) taking $f = \chi_E$, the RHS of the Lebesgue differentiation theorem is just the density of E.

Intuitively, the small balls around x are almost entirely covered by E.

Example 311. An interior point of a measurable set (by definition of openness) is a density point of the set.

Proposition 312. Let $E \subseteq \mathbb{R}^n$ be measurable. Then almost every point x in E is a point of density of E, and almost every point x not in E has density 0 (i.e. not a point of density).

PROOF 313. Take $f = \chi_E$ for the Lebesgue differentiation theorem. Then

$$\begin{aligned}\chi_E(x) &= \lim_{m(B)\to 0, x\in B} \frac{1}{m(B)} \int_B \chi_E(y)dy \\ &= \lim_{m(B)\to 0, x\in B} \frac{m(B\cap E)}{m(B)}\end{aligned}$$

and the result follows from the definition of χ_E.

□

Definition 314. The Lebesgue set of a locally integrable function f on $\mathbb{R}^d$ are points $x \in \mathbb{R}^d$ such that $f(x) < \infty$ and

$$\lim_{m(B)\to 0, x\in B} \frac{1}{m(B)} \int_B |f(y) - f(x)| = 0 \tag{422}$$

The points the Lebesgue set are called **Lebesgue points**.

Here are a few simple observations.

Proposition 315. Let f be locally integrable on $\mathbb{R}^d$.

1.) If f is continuous at x, then x is in the Lebesgue set of f.
2.) If x is in the Lebesgue set of f, then the statement of the Lebesgue differentiation theorem holds for this point, i.e.

$$f(x) = \lim_{m(B)\to 0} \frac{1}{m(B)} \int_B f(y)dy \tag{423}$$

3.) Almost all points in $\mathbb{R}^d$ are Lebesgue points.

PROOF 316. The first statement is immediate. Fix $\epsilon > 0$. Since f is continuous at x, take the radius small enough so that the integrand is less than $\epsilon > 0$. This gives the claim.

The second statement is immediate by triangle inequality for integrals:

$$\lim_{m(B)\to 0} \frac{1}{m(B)} \left| \int_B f(y) - f(x)dy \right| \leq \lim_{m(B)\to 0} \frac{1}{m(B)} \int_B |f(y) - f(x)|\, dy \tag{424}$$

For the third statement, we use the typical strategy of approximating the target $(f(x) - f(y))$ with "nice functions" for which we know the claim is true. Fix $\epsilon > 0$. By Lebesgue differentiation theorem, we know that the "nice functions" are $|f(y) - r|$:

$$\lim_{m(B)\to 0, x\in B} \frac{1}{m(B)} \int_B |f(y) - r|\, dy = |f(x) - r| \tag{425}$$

for $x \notin E_r$ for some null set E_r. So, take $x \notin E := \bigcup_{r\in\mathbb{Q}} E_r$ where by countability, E is still a null set. We can then make the above small by taking r close to $f(x)$, i.e. $|f(x) - r| < \epsilon$. Then

$$\begin{aligned}\lim_{m(B)\to 0} \frac{1}{m(B)} \int_B |f(y) - f(x)| &\leq \frac{1}{m(B)} \int_B |f(y) - r| + \frac{1}{m(B)} \int_B |f(x) - r| \\ &= 2\,|f(x) - r| \\ &\leq 2\epsilon\end{aligned}$$

as desired.

□

p.107 remarks that the density points depend on the functions.

Definition 317. **A collection of sets $\{U_\alpha\}$ shrink regularly to** $x \in \mathbb{R}^d$ (or **has bounded eccentricity at** x) if there exists a constant $c > 0$ such that for all α, there exists a ball B containing x, U_α, and $m(U_\alpha) \geq cm(B)$ (i.e., the measure of U_α is comparable to the measure of the ball).

Example 318. The set of all open cubes $\{Q_\alpha\}$ containing x shrink regularly to x.

Prove this using the definition.

this?

Example 319. In $\mathbb{R}^d$, $d \geq 2$, the collection of all rectangles containing x does not shrink regularly to x.

Consider very thin rectangles. Again, prove this.

ball vs. rectangle affects the analysis; solve problem 3.8

Proposition 320. Suppose f is locally integrable on $\mathbb{R}^d$, and x is in the Lebesgue set of f. If U_α shrink regularly to x, then the Lebesgue differentiation theorem holds at x on U_α, i.e.

$$\lim_{m(U_\alpha)\to 0} \frac{1}{m(U_\alpha)} \int_{U_\alpha} f(y)dy = f(x) \tag{426}$$

PROOF 321. Unravel definitions:

$$\lim_{m(U_\alpha)\to 0} \frac{1}{m(U_\alpha)} \int_{U_\alpha} |f(y) - f(x)| \leq \lim_{m(B_\alpha)\to 0, x\in B_\alpha} \frac{1}{cm(B_\alpha)} \int_B |f(y) - f(x)| \tag{427}$$

where B_α is the ball for which U_α is comparable. □

Good Kernels and Approximations to the Identity. -Saturday, 7.21.2018

Insert some basic properties of convolutions. Also, review standard proof of Weierstrass theorem from Lang, *Undergraduate Analysis.*

Differentiability of Functions. -Saturday, 7.21.2018

Go back to content after corollary 1.10 of chapter 2

We want to formulate the second fundamental theorem of calculus in the measure theoretic setting, i.e. find the condition on the function F such that

$$F(b) - F(a) = \int_a^b F'(x)dx \tag{428}$$

But of course, there are some functions which definitely does not satisfy this identity. First, there are non-differentiable functions (e.g. the Weierstrass nowhere differentiable function). Then there are functions which are differentiable, but the derivative is non (Lebesgue) integrable, such as in the following example.

Example 322. Here is a classic example from calculus. We claim that the function

$$F(x) := \begin{cases} x^2 \sin \frac{1}{x^2} & x \neq 0 \\ 0 & x = 0 \end{cases} \tag{429}$$

is differentiable everywhere on $\mathbb{R}$, but F' is not integrable (or, more specifically, it is differentiable but not $C^1(\mathbb{R})$).

From the usual algorithm, we see that

$$F'(x) = 2x \sin \frac{1}{x^2} - \frac{2}{x} \cos \frac{1}{x^2} \tag{430}$$

for $x \neq 0$. On the other hand,

$$F'(0) = \lim_{h\to 0} \frac{F(h) - F(0)}{h} = \lim_{h\to 0} h \sin \frac{1}{h^2} = 0 \tag{431}$$

since in the last line, we see that $\left|h\sin\frac{1}{h^2}\right| \leq |h|$ and h certainly vanishes. So, F' exists everywhere. Notice here that $\lim_{x\to 0} F'(x)$ does not exist because in short, the term $\frac{2}{x}\cos\frac{1}{x^2}$ dominates, oscillates, and blows up as $x \to 0$. ($2x\sin\frac{1}{x^2}$ vanishes by the same argument as $F'(0) = 0$.) More explicitly, for $x > 0$

$$\cos\frac{1}{x^2} = \begin{cases} 1 & x = \frac{1}{\sqrt{2n\pi}} \\ -1 & x = \frac{1}{\sqrt{(2n+1)\pi}} \end{cases} \tag{432}$$

where $n \in \mathbb{N}$. Meanwhile, $1/x > 0$ so for every neighborhood around 0, there are points at which $\frac{2}{x}\cos\frac{1}{x^2}$ is positive and negative respectively.

We exploit this blow up to show nonintegrability. Fix $\epsilon > 0$. Since $2x\sin\frac{1}{x^2} \to 0$, we can choose $\delta > 0$ such that $|x| < \delta$ then $\left|2x\sin\frac{1}{x^2}\right| < \epsilon$. Take N_0 large enough so that

$$\frac{1}{\sqrt{2N_0\pi - \pi/4}} < \delta \tag{433}$$

Also observe that

$$x \in \left(\frac{1}{\sqrt{2n\pi + \pi/4}}, \frac{1}{\sqrt{2n\pi - \pi/4}}\right), \qquad n \in \mathbb{N} \tag{434}$$

we have $\cos 1/x^2 \geq \frac{1}{2}$. So,

$$\begin{aligned}
\int_{[-1,1]} |F'(x)|\,dx &\geq \sum_{n=N_0}^{N} \int_{\frac{1}{\sqrt{2n\pi+\pi/4}}}^{\frac{1}{\sqrt{2n\pi-\pi/4}}} |F'(x)|\,dx \\
&\geq \sum_{n=N_0}^{N} \int_{\frac{1}{\sqrt{2n\pi+\pi/4}}}^{\frac{1}{\sqrt{2n\pi-\pi/4}}} \frac{2}{x}\cos\frac{1}{x^2} - \epsilon dx \\
&\geq \sum_{n=N_0}^{N} \int_{\frac{1}{\sqrt{2n\pi+\pi/4}}}^{\frac{1}{\sqrt{2n\pi-\pi/4}}} \frac{1}{x} - \epsilon dx \\
&= \sum_{n=N_0}^{N} \left(\log\left(\frac{1}{\sqrt{2n\pi+\pi/4}}\right) - \log\left(\frac{1}{\sqrt{2n\pi-\pi/4}}\right)\right) - \epsilon\left(\frac{1}{\sqrt{2n\pi-\pi/4}} - \frac{1}{\sqrt{2n\pi+\pi/4}}\right) \\
&= \log\left(\sqrt{\frac{2N\pi+\pi/4}{2N_0\pi-\pi/4}}\right) - \epsilon\left(\frac{1}{\sqrt{2N\pi-\pi/4}} - \frac{1}{\sqrt{2N_0\pi+\pi/4}}\right)
\end{aligned}$$

But now, the final sum blows up as we take $N \to \infty$, so indeed, the integral is unbounded. So, F' is not integrable.

Is there a cleaner way to show nonintegrability??

One way to address the above question is by taking F to be defined as an indefinite integral (of Lebesgue integrable functions). In this case, we definitely have both sides of the identity 428. But is there an alternate characterization for such functions? This leads us to the notion of bounded variation.

Come back to this to explain *why* this is the natural notion??

Functions of Bounded Variations.

Definition 323. Let $\gamma(t), t \in [a,b]$ be a continuous curve in $\mathbb{R}^2$. The **length (or total variation) of a curve** γ **on** $[a,x]$ is the number

$$L(\gamma) := \sup_{\{t_i\}_{i=1}^N} \sum_{i=1}^{N} |\gamma(t_i) - \gamma(t_{i-1})| \tag{435}$$

where the supremum is over all partitions $\{t_i\}$ of $[a, x]$.

A **curve** $\gamma(t) = (x(t), y(t)), t \in [a, b]$ **is rectifiable** or **has bounded variations** if the length is finite.

For the above definitions, we can instead have a function from $[a, b]$ to $\mathbb{C}$ (and hence functions from $[a, b]$ to $\mathbb{R}$) from the canonical linear isomorphism between $\mathbb{R}^2$ and $\mathbb{C}$. In this case, we allow the function to be discontinuous.

Definition 324. Let $F : [a, b] \to \mathbb{C}$ on $[a, x]$. The **positive variation of** F **on** $[a, x]$ is the function

$$P_F(a, x) = \sup \sum_{+} F(t_j) - F(t_{j-1}) \tag{436}$$

where the sum is over all partition and the sum is over all t_j such that $F(t_j) \geq F(t_{j-1})$. Likewise, the **negative variation of** F **on** $[a, x]$ is the function

$$N_F(a, x) = \sup \sum_{-} -(F(t_j) - F(t_{j-1})) \tag{437}$$

where the supremum is again over all partitions and the sum is over all t_j such that $F(t_j) \leq F(t_{j-1})$.

Of course, the total, positive, and negative variations are nonnegative, nondecreasing functions.

Remark 325. Notice the similarity with the total/positive/negative variation for measures (given by the Jordan decomposition). Here, we do not need a big decomposition theorem the way we do for measures because the notion of the positive and negative variations are relatively easier to define, at least when the domain of the function is one dimension.

Remark 326. The intuition for bounded variation is that the function does not oscillate to often with large amplitudes.

Example 327. A real valued, monotonic, bounded function has bounded variations by telescoping.

Example 328. If F is a complex valued differentiable function with bounded derivative, then F has bounded variation by mean value theorem.

Example 329. The function

$$F(x) = \begin{cases} x^a \sin(x^{-b}) & 0 < x \leq 1 \\ 0 & x = 0 \end{cases} \tag{438}$$

has bounded variation iff $a > b$.

Come back to this. Do exercise 3.12.

We know from curve theory that for C^1-curves,

$$L(\gamma) = \int_a^b |\gamma(t)| \, dt \tag{439}$$

We would like to generalize this condition here and ask under what conditions is γ rectifiable, and does the above Pythagorean formula hold? For the first, we need the class of curves such that the x, y components are functions of bounded variations which is actually the same condition we need for the differentiability a.e.

Here is a natural property of variations.

Proposition 330. Suppose F is real valued and has bounded variation on $[a, b]$. Then for all $x \in [a, b]$,

$$F(x) - F(a) = P_F(a, x) - N_F(a, x) \tag{440}$$

and

$$T_F(a, x) = P_F(a, x) + N_F(a, x) \tag{441}$$

PROOF 331. The idea is of course to just take a fine enough partition and manipulate the sums. Fix $\epsilon > 0$. Take the partition $\{t_i\}_{i=0}^N$ of $[a, x]$ such that

$$\left| P_F - \left(\sum_+ F(t_j) - F(t_{j-1}) \right) \right| < \epsilon, \; \left| N_F - \left(\sum_- -(F(t_j) - F(t_{j-1})) \right) \right| < \epsilon \tag{442}$$

We can do this simultaneously by taking refinements of partitions. Now, by telescoping,

$$F(x) - F(a) = \left(\sum_+ F(t_j) - F(t_{j-1}) \right) - \left(\sum_- -(F(t_j) - F(t_{j-1})) \right) \tag{443}$$

and by triangle inequality,

$$|(F(x) - F(a)) - (P_F - N_F)| < 2\epsilon \tag{444}$$

which gives the first identity.

For the second identity, we prove both sides of the inequality. Firstly, for any partition t_j,

$$\sum_{j=1}^N |F(t_j) - F(t_{j-1}| = \left(\sum_+ F(t_j) - F(t_{j-1}) \right) + \left(\sum_- -(F(t_j) - F(t_{j-1})) \right) \tag{445}$$

and so, $T_F \geq P_F + N_F$. Also,

$$\sum_{j=1}^N |F(t_j) - F(t_{j-1}| \leq \left(\sum_+ F(t_j) - F(t_{j-1}) \right) + N_F \leq P_F + N_F \tag{446}$$

so, $T_F \leq P_F + N_F$ as required.

$\square$

We are now ready to establish a better intuition for BV functions.

Proposition 332. A real valued function F on $[a, b]$ has bounded variations iff F is a difference of two increasing bounded functions.

Remark 333. This often reduces proofs regarding BV functions to just proofs on nondecreasing bounded functions on an interval.

PROOF 334. One direction is immediate. As we saw before, bounded increasing functions have bounded variation, then their difference also has bounded variation.

This direction is also immediate from the previous proposition. Conversely, suppose F has bounded variation. Take

$$F_1(x) := P_F(a, x) + P(a), \; F_2(x) := N_F(a, x) \tag{447}$$

These two functions are both nondecreasing and bounded (hence has bounded variation). But by the lemma,

$$F(x) = F_1(x) - F_2(x) \tag{448}$$

as required.

□

Here is the regularity of total variations.

Proposition 335. Let F be a continuous complex valued function on $[a, b]$ with bounded variation. Then its total variation is also a continuous function.

PROOF 336. We will establish right and left continuity. Since $T_F(a, x)$ is a nondecreasing function, we only need to check values to the left of x when checking left continuity, and similarly for right continuity. Fix $\epsilon > 0$. Take the partition $\{t_i\}_{i=0}^N$ of $[a, x]$ so that the corresponding polygonal line well approximates F, i.e.

$$\sum_{i=1}^{N} |F(t_i) - F(t_{i-1})| \geq T_F(a, x) - \frac{\epsilon}{2} \tag{449}$$

But since F is continuous, we can find $0 < h < x - t_{N-1}$ such that

$$|F(x) - F(x - h)| < \frac{\epsilon}{2} \tag{450}$$

Therefore,

$$T_F(a, x - h) \geq T(x, h) - \epsilon \tag{451}$$

which gives left continuity.

The same argument gives right continuity.

□

We now go back to the theory of differentiation. This is what we have building up to.

Proposition 337 (Characterization of Differentiability a.e. for Continuous Functions). If F is a real valued continuous function with bounded variation, then F is differentiable a.e.

Since BV functions are a difference of bounded increasing functions, we can restrict our attention to functions which are increasing.
We need the following lemma.

Proposition 338 (Rising Sun Lemma.). Suppose $G : \mathbb{R} \to \mathbb{R}$ is continuous. Let

$$E := \{x \in \mathbb{R} : (\exists h := h_x > 0)(G(x + h) > G(x))\} \tag{452}$$

If E is nonempty, then it is open (hence a countable disjoint union of open intervals). If (a_k, b_k) is a finite interval in this union, then

$$G(b_k) = G(a_k) \tag{453}$$

Remark 339 (Intuition for Rising Sun Lemma.). The intuition for the above lemma is as follows. We can think of G as a mountain range (without any cliffs, i.e. discontinuities) in front of an ocean. Then E is the region which is in the shade as the sun rises above the ocean (and we assume that the sun's rays are all parallel to the ground). The theorem then says that we can split up region in the shade into smaller regions so that there is no abrupt change in the brightness (i.e. open intervals), and furthermore, each end of the small regions are at the same altitudes (i.e. $G(b_k) = G(a_k)$).

PROOF 340 (Rising Sun Lemma). The first part follows from general theory (of point set topology), namely

$$\begin{aligned} E &= \{x \in \mathbb{R} : (\exists h := h_x > 0)(G(x + h) > G(x))\} \\ &= \bigcup_{h>0} F_h^{-1}(\mathbb{R}_{>0}) \end{aligned}$$

(where $F_h(x) := G(x+h) - G(x)$) is a continuous image of an open set, so it must be open. Furthermore from the separability of $\mathbb{R}$, all open sets in $\mathbb{R}$ are countable disjoint unions of open intervals, and so, we can take $a_k < b_k$ such that

$$E = \bigcup_{k=1}^{\infty} (a_k, b_k) \tag{454}$$

For the second part, for an open interval (a_k, b_k), since a_k is an endpoint of the interval, $a_k \notin E$ so by definition of E, $G(a_k) \leq G(a_k)$. We will show that the other inequality is also impossible.

For this part of the proof, the geometric interpretation of the problem is a good guiding principle. [76]
We claim that b_k is a maximum of G on the interval $[a_k, b_k]$. Firstly, since $b_k \notin E$, for all $x \geq b_k$, $G(x) \leq G(b_k)$. Now for $y \in (a_k, b_k)$, if $G(y) \geq G(b_k)$, then this contradicts the fact that $y \in E$. Therefore, $G(y) < G(b_k)$ for all $y \in (a_k, b_k)$.

This immediately implies that $G(a_k) > G(b_k)$ is impossible. For, if this were the case, then by continuity of G, there exists a point $y \in (a_k, b_k)$ such that $G(a_k) > G(y) > G(b_k)$ which contradicts what we just proved. Therefore, $G(a_k) = G(b_k)$.

□

The form of the theorem that we need for our big proof is the following.

Proposition 341 (Modified Rising Sun Lemma.). Suppose H is a continuous real valued function on $[c, b]$. If E is the same set of points as in the rising sun lemma, then E is either empty or open. In the latter case, it is a disjoint union of intervals (a'_k, b'_k), and $H(a'_k) = G(b'_k)$ except possibly at $c = a'_k$ in which case we only have $G(a'_k) \leq G(b'_k)$.

PROOF 342. The proof for the rising sun lemma works verbatim for everything up to the last caveat about what happens at $a = a_k$. In this case, if H is a restriction of G (from the rising sun lemma) to the interval $[a, b]$, then the "except possibly" bit occurs precisely if c is chosen so that $a_k < c$, where the unprimed intervals a_k, b_k are from the rising sun lemma, and $a_k \leq a'_k \leq b_k = b'_k$. In short the caveat occurs when an interval is "cut short" by the restriction. (Or, going back to our mountain range picture, the mountain ends abruptly with a steep cliff rather than continuing indefinitely along the real line.)

□

Here is a lemma from basic measure theory which in turn requires a basic fact from real analysis.

Proposition 343. [77] Let F, G be continuous real valued functions on $[a, b]$.

1.) For any sequence $h_n \searrow 0$,
$$\limsup_{h \searrow 0} G(h) = \lim_{m \to \infty} \sup_{n \geq m} G(h_n) \tag{455}$$

2.) $D^+(F)(x) = \limsup_{h \searrow 0} \frac{F(x+h) - F(x)}{h}$ is measurable.

In the proof of the second part, we will take $G(h)$ to be the difference quotient $\frac{F(x+h)-F(x)}{h}$.

PROOF 344. The continuity is essential for the first statement. Since G is continuous at all $h \in [a, b]$,

$$\limsup_{h \searrow 0} G(h) = \liminf_{h \searrow 0} G(h) = \limsup_{h \nearrow 0} G(h) = \limsup_{h \nearrow 0} G(h) = G(0) \tag{456}$$

[76] Stein's proof is a little round about, so we prove an alternative proof which can be motivated by pictures. The picture we should consider is a decreasing sequence of mountain ranges where we zoom up on one valley. A point on the slope on the left is a_k and the peak on the right is b_k. The conclusion of the theorem says that we must have a_k, b_k be on the same altitudes.

[77] Exercise 3.14 a.).

(This is what it means for the limit $\lim_{h\to 0} G(h)$ to exist: both the right hand limit and left hand limit exist, and they are both equal to the value of G at the point.) We only use the first equality. For any sequence h_n,

$$\limsup_{h\searrow 0} G(h) \geq \lim_{m\to\infty} \sup_{n\geq m} G(h_n) \geq \liminf_{h\searrow 0} G(h) \tag{457}$$

just from the definition of supremums and infimums. So, from the equality from before, we get the desired conclusion. □

Now take

$$G(h) := \frac{F(x+h) - F(x)}{h} \tag{458}$$

Since F is continuous, this G is also continuous and so, we can use the first statement. But now, for any $M \in \mathbb{R}$

$$\begin{aligned}\left\{D^+(F)(x) > M\right\} &= \left\{\limsup_{h\searrow 0} \frac{F(x+h)-F(x)}{h} > M\right\} \\ &= \left\{\limsup_{n\to\infty} \frac{F(x+h_n)-F(x)}{h_n} > M\right\}\end{aligned}$$

for any $h_n \searrow 0$. But now, the supremum of a sequence of measurable functions $\frac{F(x+h_n)-F(x)}{h_n}$ is measurable, so this is a countable union of measurable sets. (Note that here again, we used the fact that F is continuous in order to get measurability of F.) Therefore, $D^+(F)(x)$ is measurable.

Notice that taking the sequence of h_n is essential in the above proof since we only have that the supremum of a *sequence* of measurable functions $\frac{F(x+h_n)-F(x)}{h_n}$ is measurable rather than an (uncountable) family of measurable functions $\frac{F(x+h)-F(x)}{h}$. This shows the critical role the first statement plays in the proof of the second statement.

□

PROOF 345 (Characterization of Differentiability a.e. - Continuous Case.). The key trick for this proof is using the auxiliary function $G(x) := -F(-x)$ to switch from a lower bound to an upper bound.

We first prove the case when F is real valued, has BV, and additionally is continuous. To prove the claim, we must show two things: the derivative exists a.e. and is finite. Since a derivative is a limit, we can consider the limsup and liminf of the right and left hand limits (hence four quantities). They are the **Dini numbers** [78] given by

$$\begin{aligned}D^+(F)(x) &:= \limsup_{h\searrow 0} \frac{F(x+h)-F(x)}{h} \\ D_+(F)(x) &:= \liminf_{h\searrow 0} \frac{F(x+h)-F(x)}{h} \\ D^-(F)(x) &:= \limsup_{h\nearrow 0} \frac{F(x+h)-F(x)}{h} \\ D_-(F)(x) &:= \liminf_{h\nearrow 0} \frac{F(x+h)-F(x)}{h}\end{aligned}$$

[78] The only benefit of introducing this new terminology is that we don't have to say the "limsup/inf of the right/left hand limits" each time; other than that, it is not a particularly special concept; it is totally natural considering that we are dealing with *limits*.

By definition of limsup/inf, we know immediately that

$$D_+(F)(x) \leq D^+(F)(x),\ D_-(F)(x) \leq D^-(F)(x) \tag{459}$$

To show that the limit exists, we can show that $D^+(F)(x) \leq D_-(F)(x)$ from which, a naive application will then give us

$$D_+(F)(x) \leq D^+(F)(x) \leq D_-(F)(x) \leq D^-(F)(x) \tag{460}$$

But also, (and this is the key trick) applying this to $G(x) = -F(-x)$ gives

$$\begin{aligned} D^+(G)(x) &= \limsup_{h \searrow 0} \frac{G(x+h) - G(x)}{h} \\ &= -\limsup_{h \searrow 0} \frac{F(-x-h) - F(-x)}{h} \\ &= -\limsup_{h' \searrow 0} \frac{F(-x+h') - F(-x)}{-h'} \qquad h = -h' \\ &= D^-(F)(-x) \end{aligned}$$

and likewise,

$$\begin{aligned} D_-(G)(x) &= \liminf_{h \nearrow 0} \frac{G(x+h) - G(x)}{h} \\ &= -\liminf_{h \nearrow 0} \frac{F(-x-h) - F(-x)}{h} \\ &= -\liminf_{h' \searrow 0} \frac{F(-x+h') - F(-x)}{-h'} \\ &= D_+(F)(-x) \end{aligned}$$

From which we get

$$D_+(F)(x) \leq D^+(F)(x) \leq D_-(F)(x) \leq D^-(F)(x) \leq D_+(F)(x) \tag{461}$$

So, it suffices to show that: $D^+(F)(x) \leq D_-(F)(x)$ and $D_+(F)(x) < \infty$ a.e. This will show the existence of $F'(x)$ a.e. We need the rising sun lemma in both claims.

(*Finiteness a.e.*.) We prove the finiteness of $D_+(F)(x)$ a.e. The goal (of course) is to contain the set on which F is big (say greater than γ), and then control it using γ – a statement similar to Chebyshev.

As we always do in measure theory, we prove that the set for which this set is large has small measure. That is, for fixed $\gamma > 0$, let

$$E_\gamma := \{x \in [a,b] : D_+(F)(x) > \gamma\} \tag{462}$$

We prove that this set is a null set. We first know measurability of this set from the previous proposition. We then know by the modified rising sun lemma [79] applied to the function $G(x) = F(x) - \gamma x$ [80] , we have

[79] We need the continuity of F in this step because the rising sun lemma requires G to be continuous. Stein remarks on p.121 that the rising sun lemma is a *covering argument*. We see what he means by that here; we cover the bad set (characterized by limsup) with intervals from which we can get a bound.

[80] We obtain this function by backtracking. The statement from the rising sun lemma talks about the form $F(x+h) - F(x) > 0$, so in order to get the difference quotient to be greater than γ, we must have γh on the RHS, which in turn can be rewritten as $\gamma(x_h - x)$.

$E_\gamma \subseteq \bigcup_{k=1}^\infty (a_k, b_k)$ [81] where $F(b_k) - F(a_k) \geq \gamma(b_k - a_k)$ (where the inequality comes the left most interval, and we have equality for everything else). We are now done: [82]

$$\begin{aligned} m(E_\gamma) &\leq m\left(\bigcup_{k=1}^\infty (a_k, b_k)\right) \\ &= \sum_{k=1}^\infty m\left((a_k, b_k)\right) \\ &= \sum_{k=1}^\infty (b_k - a_k) \\ &\leq \frac{1}{\gamma}\sum_{k=1}^\infty F(b_k) - F(a_k) \\ &\leq \frac{1}{\gamma}(F(b) - F(a)) \end{aligned}$$

which gives $m(E_\gamma) \to 0$ as $\gamma \to \infty$. This proves the first claim.

(*Inequality.*)

Come back to this.

$\square$

Here is what we have so far for the second fundamental theorem of calculus.

Proposition 346 (Continuous increasing bounded, then derivative is integrable.). If F is continuous and increasing, then F' exists a.e. Moreover, F' is measurable, nonnegative, and satisfies one direction of the inequality in the second fundamental theorem of calculus:

$$\int_a^b F'(x)dx \leq F(b) - F(a) \tag{464}$$

In particular, if F is bounded on $\mathbb{R}$, then F' is integrable on $\mathbb{R}$.

PROOF 347. (*Nonnegative, Measurable.*) We can show that F' is measurable and nonnegative by showing that it is the (sequential) limit of functions satisfying such properties. But we know from the previous proposition that

$$\lim_{n\to\infty} \frac{F\left(x + \frac{1}{n}\right) - F(x)}{1/n} = F'(x) \tag{465}$$

almost everywhere. But the difference quotient above is measurable (since it is continuous in x) and nonnegative (since F is increasing). Thus, we get the first claim.

(*Inequality.*) The key idea is Fatou's lemma. Recall that

$$\int_a^b F'(x)dx \leq \liminf_{n\to\infty} \int_a^b \frac{F\left(x + \frac{1}{n}\right) - F(x)}{1/n} dx \tag{466}$$

[81] The inclusion of course comes from the fact that for a fixed $\epsilon > 0$, there is $h_0 > 0$ such that

$$\limsup_{h \searrow 0} \frac{F(x+h) - F(x)}{h} + \epsilon \geq \frac{F(x+h_0) - F(x)}{h_0} \tag{463}$$

for any x.

[82] We actually don't need to use the measurability of E_γ in this step. We only need the measurability because we are taking $m(E_\gamma)$, but if we instead took the *exterior* measure, then the bounding argument will still work (since open intervals are the open balls in $\mathbb{R}$).

The strategy is to then exploit the continuity of F and to use the mean value theorem for integral. We can rewrite the above as

$$\begin{aligned}\frac{1}{n}\left(\int_a^b F\left(x+\frac{1}{n}\right)-F(x)dx\right) &= \frac{1}{n}\left(\int_{a+1/n}^{b+1/n} F\left(y\right)dy - \int_a^b F(x)dx\right) \\ &= \frac{1}{n}\left(\int_b^{b+1/n} F\left(y\right)dy - \int_a^{a+1/n} F(x)dx\right) \\ &\to F(b)-F(a)\end{aligned}$$

as desired.

□

Remark 348. The Cantor-Lebesgue function is an increasing, continuous functions such that $F(0) = 0, F(1) = 1$ but $F'(x) = 0$ a.e. So, indeed equality in the above proposition does not hold, and we need a stronger condition on F.

Absolutely Continuous Functions.

Definition 349. A **function** $F : I \subset \mathbb{R} \to \mathbb{R}$ **is absolutely continuous** if for any $\epsilon > 0$, there exists $\delta > 0$ such that $m\left(\bigcup_{k=1}^n b_k - a_k\right) < \delta$, (a_k, b_k) disjoint implies

$$\sum_{k=1}^N |F(b_k) - F(a_k)| < \epsilon \tag{467}$$

Remark 350. Note how this compares with absolute continuity of a measure; instead of a function F, we instead take a different measure ν and we say ν is absolutely continuous with respect to m (written $\nu \ll m$) if a set is ν-measure zero whenever it is m-measure zero.

Remark 351. Note the hierarchy:

$$\text{absolutely continuous} \implies \text{uniformly continuous} \implies \text{continuous} \tag{468}$$

Proposition 352. Absolutely continuity implies bounded variations for functions defined on compact intervals.

PROOF 353.

□

Proposition 354 (AC function is a difference of continuous monotonic functions.). The total, positive, and negative variations of an absolutely continuous function is (absolutely) continuous. In particular, an absolutely continuous function is the difference of two continuous monotonic functions.

PROOF 355.

□

Come back to this; fill this argument in at page 127

Remark 356. The function

$$F(x) := \int_a^x f(y)dy \tag{469}$$

for $f \in L^1([a,b])$ is absolutely continuous. We have already shown this before (proposition 229).

The following is why we are about AC functions.

Proposition 357 (AC then differentiable a.e.). If $F : [a, b] \to \mathbb{R}$ is absolutely continuous, then $F'(x)$ exists a.e.

PROOF 358. We don't need to do anything. Since an absolutely continuous function F is a difference of continuous monotonic functions (proposition 354). But monotonic functions have bounded variations, so F is a difference of a.e. differentiable functions (proposition 337). Therefore, F is differentiable a.e. $\square$

Here is another classic statement about derivatives. This part is where we need to do some work.

Proposition 359 (Derivative vanishes a.e. then constant.). $F'(x) = 0$ a.e. implies that $F \equiv c$ for some $c \in \mathbb{R}$.

Remark 360. Recall the Vitali covering lemma (proposition 298) which says: Suppose E is a set of finite measure with an (arbitrary) collection of balls such that every point in E has an arbitrarily small ball (in the sense of measure) in the collection. Then we can extract finitely many disjoint collection of balls which cover as much of E as we want.

PROOF 361. It suffices to show that $F(b) = F(a)$ (since we can then consider subintervals of $[a, b]$ to establish $F(x) = F(y)$ for any $x, y \in [a, b]$). The idea is that we consider the good set

$$E := \left\{ x \in (a, b) : F'(x) = \lim_{h \to 0} \frac{F(x + h) - F(x)}{h} = 0 \right\} \tag{470}$$

which by hypothesis is $m(E) = b - a$. On this set, we can bound $|F(b) - F(a)|$ using the vanishing derivative. On its complement (which is a small set), we need the absolute continuity. We want the Vitali covering lemma because in order to talk about the difference quotients, we want to have intervals. In effect, the Vitali covering lemma partitions $[a, b]$ into open and closed intervals on which we can estimate the difference quotient. [83]

Fix $\epsilon > 0$. Let's start with the good set E. For each $x \in E$ and $\eta > 0$, there are open intervals $I := I(x, \eta) := (a_x, b_x) \subseteq [a, b]$ containing x such that

$$|F(b_x) - F(a_x)| \leq \epsilon(b_x - a_x) \tag{471}$$

with $b_x - a_x < \eta$. The collection $\{I(x, \eta)\}$ is then a Vitali covering of E (by definition of $I(x, \eta)$). So, by the Vitali covering lemma, for $\delta > 0$, we can extract disjoint finite subcollection $\{I_i\}_{1 \leq i \leq N}$, $I_i := I(x_i, \eta_i)$ which cover a significant portion of E, i.e.

$$\sum_{i=1}^{N} m(I_i) \geq m(E) - \delta = (b - a) - \delta \tag{472}$$

and also, [84]

$$\sum_{i=1}^{N} |F(b_i) - F(a_i)| \leq \epsilon(b - a) \tag{473}$$

This establishes the estimate on the good set.

Now to tame the bad set, consider $\bigcap_{i=1}^{N} I_i^c$. Since this is a finite intersection of closed sets, it is closed, and so it is a finite union of intervals $\bigcup_{k=1}^{M} [\alpha_k, \beta_k]$ with total length $\leq \delta$. But by absolute continuity of F, we get

$$\sum_{k=1}^{M} |F(\beta_k) - F(\alpha_k)| \leq \epsilon \tag{474}$$

[83] The key usage of Vitali covering lemma seems to be that a.) it covers some good set (in this case the set E and in the proof of the weak (1-1) bound of Hardy-Littlewood maximal function, the compact set K), and then uses additivity (not even σ-additivity) to approximate this good set, and b.) as in this proof, we can split up the entire set (in this case $[a, b]$) into intervals on which we can establish some approximations, and we can add this all up using the disjoint finite additivity.

[84] Note that establishing this inequality requires that I_i are disjoint, and so, $\bigcup_{i=1}^{N} I_i \subseteq [a, b]$.

This is the bound for the bad set.

So, adding up the bounds for the good and bad sets, and in addition using the fact that the partitions are disjoint, we have

$$|F(b) - F(a)| \leq \sum_{k=1}^{N} |F(b_i) - F(a_i)| + \sum_{k=1}^{M} |F(\beta_k) - F(\alpha_k)| \leq \epsilon(m([a,b]) + 1) \tag{475}$$

which proves the claim.

□

This resolves the issue for second fundamental theorem of calculus.

Proposition 362 (Second Fundamental Theorem of Calculus for Measure Theory). If $F : [a,b] \to \mathbb{R}$ is AC, then it is differentiable a.e., the derivative is integrable, and the formula for second fundamental theorem of calculus holds, i.e.

$$F(x) - F(a) = \int_a^x F'(y)dy \tag{476}$$

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

Proposition 363 (First Fundamental Theorem of Calculus for Measure Theory.). If $f : [a,b] \to \mathbb{R}$ is integrable, then the function

$$F(x) = \int_a^x f(y)dy \tag{477}$$

is AC function and $F'(x) = f(x)$.

We presented the Lebesgue differentiation theorem before as being the measure theory version of (both) Fundamental Theorems of Calculus. Indeed, it is the key idea of the proof for both of the following theorems (and the proof is just a simple application).

PROOF 364 (Second Fundamental Theorem of Calculus for Measure Theory). We don't need to do anything for integrability of F'. Since AC functions are differences of continuous increasing functions f_1, f_2, and since F is on a compact set, it is also bounded, so the derivatives of f_1, f_2 are integrable (proposition 346), and thus, F is integrable.

For the second part, let

$$G(x) := \int_a^x F'(y)dy \tag{478}$$

We claim that $G - F$ is constant. First, by Lebesgue differentiation theorem,

$$\begin{aligned} G'(x) &= \lim_{h \to 0} \frac{1}{h} \int_{[x,x+h]} F'(y)dy \\ &= F'(x) \end{aligned}$$

for a.e. x. Since integral functions are always AC (remark 356), G is absolutely continuous, and so, $G - F$ is absolutely continuous [85]. But now $(F - G)' = 0$, so $F - G$ is constant, and $(F - G)(a) = F(a)$, so we get the desired result.

□

[85] We use the hypothesis in this step.

PROOF 365 (First Fundamental Theorem of Calculus for Measure Theory). F is an AC function because it is an integral function. Lebesgue differentiation theorem then gives

$$F'(x) = \lim_{h \to 0} \frac{1}{h} \int_{[x,x+h]} f(y)dy = f(x)$$

almost everywhere. □

Differentiability of Jump Functions.

Our goal is to prove the following:

Proposition 366 (Characterization of Differentiability a.e.). If F is a real valued function with bounded variation, then F is differentiable a.e.

In other words, we want to remove the continuity from proposition 337. Of course, we restrict our attention again to bounded increasing functions since we are dealing with BV functions. The key obstacle then are the discontinuities.

Definition 367. A **bounded increasing function** $F : [a, b] \to \mathbb{R}$ **has a jump discontinuity at** x if

$$\lim y \nearrow xF(y) < \lim y \searrow xF(y) \tag{479}$$

Of course, these are the only kinds of discontinuities bounded increasing functions can have since the left and right handed limits always exist. By inequality, we just mean $\neq$, but since F is increasing, we can upgrade to $<$. These discontinuities are easy to visualize and to deal with. For one, they are countable.

Proposition 368. A bounded increasing function $F : [a, b] \to \mathbb{R}$ has at most countably many discontinuities.

PROOF 369. If x is discontinuous at x, take a rational r_x such that

$$\lim y \nearrow xF(y) < r_x < \lim y \searrow xF(y) \tag{480}$$

For distinct continuities, we must have distinct rationals because if $x < z$, then

$$r_x < \lim y \searrow xF(y) \leq \lim y \nearrow zF(y) < r_z \tag{481}$$

So, there is an injection from the set of discontinuities to the set of rationals, and so, we have prove the claim. □

We now define a function that encodes all the discontinuities of BV functions.

Definition 370. The **jump function associated to a bounded increasing function** $F : [a, b] \to \mathbb{R}$ is

$$J_F(x) := \sum_{n=1}^{\infty} \alpha_n j_n \tag{482}$$

where $\{x_n\}_{n=1}^{N}$, $N \in \mathbb{N} \cup \{\infty\}$ are the discontinuities of F, $\alpha_n := F(x_n^+) - F(x_n^-)$,

$$\theta_n := \frac{F(x_n) - F(x_n^-)}{\alpha_n} \in [0, 1] \tag{483}$$

and

$$j_n(x) := \begin{cases} 0 & x < x_n \\ \theta_n & x = x_n \\ 1 & x > x_n \end{cases} \tag{484}$$

Proposition 371. Jump functions converge absolutely uniformly.

PROOF 372. This is immediate since

$$\sum_{n=1}^{\infty} |\alpha_n j_n(x)| \leq \sum_{n=1}^{\infty} \alpha_n \leq F(b) - F(a) < \infty \tag{485}$$

□

The following is why we defined the jump functions in the first place.

Proposition 373. If F is bounded and increasing on $[a, b]$, then $J := J_F$ has the same jump discontinuities at the discontinuities x_n of F, and has the same value of jumps. Moreover, J is all of the discontinuity of F, i.e. $F - J$ is continuous and increasing (and therefore, bounded variations).

PROOF 374. The tricky part for the discontinuities of J_F is that the sum is an infinite sum, so we need to be careful about continuity. The trick is then to concatenate the series so that the discontinuity comes from a finite sum which is much easier to handle.

j_n is continuous everywhere except at x_n, and the series for J converges uniformly, so J is continuous at $x \neq x_n$. Now for $x = x_N$ for some N,

$$J(x) = \sum_{n=1}^{N} \alpha_n j_n(x) + \sum_{N+1}^{\infty} \alpha_n j_n(x) \tag{486}$$

The second sum is continuous at x by the same reason that J is continuous on $x \neq x_n$. So, the finite sum gives us the jump, and the claim is then immediate.

$F - J$ is continuous since J has the same points of discontinuity as F and has the same jump values. Now $F - G$ is increasing since if $y > x$, then

$$J(y) - J(x) \leq \sum_{x < x_n \leq y} \alpha_n \leq F(y) - F(x) \tag{487}$$

where the last inequality follows since F is increasing. So,

$$F(x) - J(x) \leq F(y) - J(y) \tag{488}$$

as desired.

□

Remark 375. From the preceding proposition, we can write F as

$$F = (F - J) + J \tag{489}$$

Since $F - J$ has bounded variations, by proposition 337, it is differentiable a.e. So, we just need to show that J is differentiable a.e. in order to prove proposition 366.

Here is a lemma.

Proposition 376. [86] Let $G : [a, b] \to \mathbb{R}$ be a bounded increasing function, and let $J(x) = \sum_{n=1}^{\infty} \alpha_n j_n(x)$ be a jump function associated to G. Then

$$F(x) := \limsup_{h \to 0} \frac{J(x + h) - J(x)}{h} \tag{490}$$

is a measurable function.

PROOF 377.

□

Come back to this... How to show that $F^N_{k,m}$ is measurable??

[86] Exercise 3.14 b.

Chapter 4. Hilbert Spaces

The Hilbert Space L^2. -Friday, 8.3.2018

We start with the definition of a Hilbert space. We then proceed to motivate this with a bit more with $L^2(\mathbb{R}^d)$.

Definition 378. The **set** H **(or more formally, the pair** $(H, (\cdot,\cdot))$**) is a Hilbert space** if

1.) H is a real or complex vector space
2.) $(\cdot,\cdot)$ is a Hermitian inner product
3.) H is complete with respect to the norm; the **norm of** f **(induced by the inner product)** is the number $\|f\| := (f, f)$.
4.) H is **separable**, i.e. has a countable set whose span (in the sense of linear algebra) is dense in H. [87]

An easy consequence of the definition are the following fundamental inequalities.

Proposition 379. Let H be a Hilbert space. Then the Cauchy-Schwarz and triangle inequalities hold, i.e. for all $f, g \in H$,

$$(f, g) \leq \|f\| \, \|g\| \tag{491}$$

and

$$\|f+g\| \leq \|f\| + \|g\| \tag{492}$$

PROOF 380. These follow immediately from the definitions:

(*Cauchy-Schwarz.*)

Come back to this.

(*Triangle.*) For $f, g \in H$,

$$\begin{aligned}
\|f+g\|^2 &= (f+g, f+g) \\
&= \|f\|^2 + (f,g) + (g,f) + \|g\|^2 \\
&\leq \|f\|^2 + 2\|f\|\,\|g\| + \|g\|^2 \qquad \text{(Cauchy-Schwarz)} \\
&= (\|f\| + \|g\|)^2
\end{aligned}$$

So,

$$\|f+g\| \leq \|f\| + \|g\| \tag{493}$$

□

Let's now focus our attention on $L^2(\mathbb{R}^d)$.

Definition 381. Let f be a measurable function. The L^2**-norm of** f is the number

$$\|f\|_2 := \left(\int_{\mathbb{R}^d} |f(x)|^2 \right)^{1/2} \tag{494}$$

A **measurable function is square integrable** if $\|f\|_2 < \infty$. The set of all square integrable functions is denoted $L^2(\mathbb{R}^d)$.

Remark 382. Here is an immediate consequence of the definition. If $f \in L^2(\mathbb{R}^d)$, then by definition, $|f|^2 \in L^1(\mathbb{R}^d)$ so $|f|^2$ is finite a.e. Thus, L^2 functions are finite a.e.

[87] We are of course being very restrictive when we define the Hilbert space; we normally do not require separability to be one of the conditions.

Remark 383. [88] Note that neither of the spaces $L^1(\mathbb{R}^d), L^2(\mathbb{R}^d)$ contains the other.

For instance, the function $f(x) := \frac{1}{\prod_{i=1}^d x_i} \cdot \chi_{\mathbb{R}^d \setminus B(0,1)}$ is square integrable, but not integrable. [89] Another example is $g(x) = \frac{\sin x}{x}$. [90] (Get the convergence/divergence by comparing it with a series.) On the other hand, the function

$$h(x) = \frac{1}{\sqrt{\prod_{i=1}^d x_i}} \tag{496}$$

is $L^1(\mathbb{R}^d)$ but now $L^2(\mathbb{R}^d)$.
Note that $L_0^2(\mathbb{R}^d) \subseteq L_0^1(\mathbb{R}^d)$ (or more generally, supported on sets of finite measure E) since for $f \in L_0^2(\mathbb{R}^d)$

$$\begin{aligned} \|f\|_1 &= (f, \chi_E) = \int_{\mathbb{R}^d} f\chi_E \\ &\leq \|f\|_2 \|\chi_E\|_2 \qquad \text{(Cauchy-Schwarz)} \\ &= m(E)^{1/2} \|f\|_2 \end{aligned}$$

where $(\cdot, \cdot)$ is the inner product in $L^2(\mathbb{R}^d)$.

On the other hand, $L_{\text{bdd}}^1(\mathbb{R}^d) \subseteq L_{\text{bdd}}^2(\mathbb{R}^d)$ (where "bdd" indicates being bounded) since if $f \in L_{\text{bdd}}^1(\mathbb{R}^d)$ bounded by M, then

$$\begin{aligned} \|f\|_2^2 &= \int_{\mathbb{R}^d} f^2 \\ &\leq M \int_{\mathbb{R}^d} f \\ &= M \|f\|_1^2 \end{aligned}$$

Let's now show that $L^2(\mathbb{R}^d)$ is a Hilbert space.

Proposition 384. $L^2(\mathbb{R}^d)$ is a Hilbert space.

PROOF 385. It is immediate that $L^2(\mathbb{R}^d)$ is a real vector space. Also

$$(f, g) := \int_{\mathbb{R}^d} f\overline{g} \tag{497}$$

certainly defines a Hermitian inner product (i.e., it is linear in the first variable, antisymmetric, and the norm is positive except for at 0).

(*Completeness.*) The work is really in showing that it is a Banach space with respect to its norm. The idea of the proof is almost exactly the same as Riesz-Fischer for L^1 spaces, i.e. consider a Cauchy sequence, show that a suitable subsequence converges pointwise to an L^1 function, show that the subsequence converge to it in the L^1 norm, and finally, show that the entire sequence converge to it in the L^1 norm.

---

[88] Exercise 5 a.
[89] The main motivation for this example is the $d = 1$ case, i.e.

$$f(x) := \frac{1}{x} \cdot \chi_{\mathbb{R} \setminus (-1,1)} \tag{495}$$

The general case follows from Fubini-Tonelli. The motivation is the same for $h(x)$ as well.
[90] See my solutions to Boot Camp 2018, complex analysis problem set 1.

The only difference is in the step showing that the subsequence converges to the limit function in the norm. Here, we actually have to rely on the definition of the L^2-norm (namely, we use a bound on the integrands, and pass to a bound of the norms via monotonicity of integrals). The other three steps only rely on generic properties of the norm, so there is no difference between L^1 and L^2.

Since the rest of the proof is identical to the L^1 proof, let's just show the convergence in the norm part. We define f_{n_k}, f, g in the exact same way as the L^1 proof. If

$$S_K(f) := f_1 + \sum_{k=1}^{K} (f_{n_{k+1}} - f_{n_k}) \tag{498}$$

is the partial sum defining f, then by triangle inequality,

$$\left| f - f_{n_{K+1}} \right| = |f - S_K(f)| \leq 2g \tag{499}$$

and so, squaring and applying the dominated convergence theorem gives the convergence in norm.

(*Separability.*) Finally, we need to show separability of the space. (Again, remember that Stein's definition is a bit weaker than the standard one since he only requires a countable set whose span is dense in the space.) The idea is to approximate L^2 functions using L^1 functions and in turn exploit the dense sets (i.e. the step functions) in L^1 space.

Since we require countability, we will need to use the rationals. Also, since we want a dense set, we want a (very simple) family of functions which approximate measure theoretic functions very well. From this, a reasonable candidate is the family of functions of the form $r\chi_R$, where $r \in \mathbb{C}$ with rational real and imaginary part, and R is a rectangle with rational coordinates.

Take $f \in L^2(\mathbb{R}^d)$ and fix $\epsilon > 0$. What is a good L^1 function which approximates f well? The common trick is one which is bounded and has compact support: [91]

$$g_n(x) = \begin{cases} f(x) & |x| \leq n, |f(x)| \leq n \\ 0 & \text{otherwise} \end{cases} \tag{500}$$

Since by construction, g_n is bounded and has compact support so it is an L^1 function. Furthermore, it converges pointwise a.e. [92] and by triangle inequality, $|f - g_n| \leq 2|f|$, so integrating, we can use the dominated convergence theorem to get $\|f - g_n\|_2 \to 0$, i.e. convergence in the norm. So, in particular, for we can choose some large N such that

$$\|f - g_N\|_2 \leq \frac{\epsilon}{2} \tag{501}$$

This is half of the work; we now have to focus on approximating the L^1 function g_N. We can first approximate this by a step function φ bounded by N and so that $\|g_N - \varphi\|_1 < \frac{\epsilon^2}{16N}$. But now, (by density of the rationals) we can approximate φ arbitrarily well by linear combination ψ of functions of the form $r\chi_R$ which is also bounded by N and $\|g_N - \psi\|_1 < 2 \cdot \frac{\epsilon^2}{16N}$. To convert this into a bound on the L^2 norm,

$$\int |g_N - \psi|^2 \leq 2N \int |g - \psi| < \frac{\epsilon^2}{4} \tag{502}$$

where we used the bound $|g - \psi| \leq 2N$ (which follows from triangle inequality from the bounds on g_N, ψ) to get the first inequality.

[91] g_n is easy to visualize as a function from $\mathbb{R}^2$ to $\mathbb{R}$. g_n is just the graph of f in the cylinder of radius n and height n (rather than a *ball*!!!). We ignore everything outside of this compact region.

[92] g_n won't converge to f at points at which it is infinity, but we know from the property of L^1 functions that this only occurs on a null set.

Therefore, $\|g_N - \psi\|_2 < \frac{\epsilon}{2}$ and combining this with the bound on $f - g_N$, we are done.

We thus have that $L^2(\mathbb{R}^d)$ is a Hilbert space.

$\square$

Hilbert Spaces: Examples and Properties. -Saturday, 8.4.2018