Contents

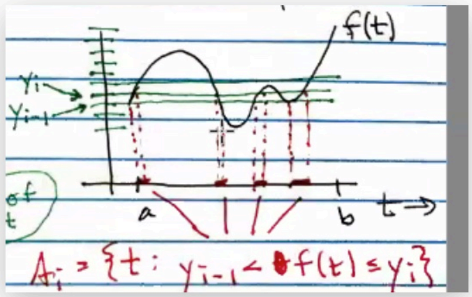
Beginning of Aug.23, 2021

0.1 Introduction

A primary goal of this course is to study *integration* in a general context. However, our current methods are not good enough. For example, we can have a sequence of continuous functions $\{f_n\}$ in $L_c^2([0,1])$ which converges (meaning that $d_2(f_n, f) \to 0$ for some f) whereas the limit is not in $L_c^2([0,1])$.

Riemann integration is also not good enough — the rational indicator function $\chi_{\mathbb{Q}\cap[0,1]}$ is not Riemann integrable as the upper sum is always 1 and the lower sum always 0.
Instead of partitioning the *domain* of the function, we instead consider partitioning the *range* of the function. Instead we obtain a partition $\{y_1, \ldots, y_k\}$ of the range of the function and sets

$$A_i = \{t : y_{i-1} < f(t) \leqslant y_i\}.$$



However, how do we define the "total length" $\mu(A_i)$ of A_i where A_i can look really weird? Some possibilities:

(1) $\mu(A)$ can stand for the *probability* of A for some random procedure.

(2) In physics, $\mu(A)$ can be the *mass* or *charge* in A (for charge, we may want to allow negative $\mu(A)$).

Basic Definitions and Intuitions

Now we will review some basic concepts from **set theory**.

(1) A **relation** on a set X is a set R of *ordered paris* (meaning a subset of $X \times X$), for example $\leqslant$ on $\mathbb{R}$.

(2) A **partial order** is a relation $\prec$ such that

  (i) (transitivity) if $x \prec y$ and $y \prec z$ then $x \prec z$,

  (ii) (symmetry) if $x \prec y$ and $y \prec x$ then $x = y$, and

  (iii) (reflexivity) $x \prec x$ for all x.

  An example of partial order can be defined on $\mathbb{R}^2$ by $u = (u_1, u_2) \prec v = (v_1, v_2)$ if $u_1 \leqslant v_1$ and $u_2 \leqslant v_2$. Note that points like $(1, 2)$ and $(2, 1)$ are not comparable. Partial orders need only be defined on a subset of $X \times X$.

(3) A **linear (total) order** is a partial order also satisfying

  (iv) any two elements are comparable, i.e., for all $x, y \in X$, either $x \prec y$ or $y \prec x$.

  For example we can consider $\mathbb{N}$ with the order $2, 4, 6, \ldots, 1, 3, 5, \ldots$. In this example there is no "last number before 1".

(4) A **maximal** element is an element x_0 such that $x_0 \prec y$ only if $y = x_0$. A **minimal** element is defined similarly. There can be multiple maximal elements.

(5) A set X is **well-ordered** by a relation $<$ if X is linearly ordered and every *nonempty* subset $E \subset X$ has a *unique minimal element*. For example, $\mathbb{N} = \{1, 2, \ldots\}$ is well-ordered, but $\mathbb{R}$ is not, as $(0, 1)$ does not have a smallest element. $\mathbb{N}$ written as $\{2, 4, 6, \ldots, 1, 3, 5, \ldots\}$ is well-ordered but $\{\ldots, 6, 4, 2, \ldots, 5, 3, 1\}$ is not.

(6) A **Cartesian product** of 2 sets A_1, A_2 is defined by

$$A_1 \times A_2 := \{(x_1, x_2) : x_i \in A_i\}$$

or equivalently

$$\{\text{all functions } f \text{ on } \{1,2\} \text{ with } f(i) \in A_i\}.$$

(7) A **general Cartesian product** of $\{A_\alpha, \alpha \in I\}$ (a collection of sets over the index set I) is

$$\prod_{\alpha \in I} A_\alpha \{\text{all functions } f \text{ on } I \text{ with } f(\alpha) \in A_\alpha \text{ for all } \alpha \in I\}.$$

(8) The **Axiom of Choice** says that for every nonempty collection $\{A_\alpha, \alpha \in I\}$, $\prod_{\alpha \in I} A_\alpha$ is nonempty. This says that we can always "select one element from each A_α".

(9) From the **Axiom of Choice** we can prove the **Hausdorff Maximal Principle**:

Every partially ordered set has a maximal linearly ordered subset.

(10) **Zorn's Lemma** (which is equivalent to the Hausdorff Maximal Principle):

Let X be a **poset** (partially ordered set) such that every linearly ordered $E \subset X$ has an upper bound (maybe not in E). Then X has a maximal subset.

Beginning of Aug.25, 2021

Hausdorff ⇒ Zorn's. let X be as in Zorn's lemma (i.e., every lienarly ordered subset has an upper bound). By Hausdorff, X has a maixmal linearly ordered E (which cannot be enlarged). Zorn's assumption says that E has an upper bound. Call it e. We claim that e is maximal.

Suppose for contradiction that e is not maximal, so there existss y such that $y \geqslant e$. But then the set $E \cup \{y\}$ is linearly ordered and bigger than E, contradicting the Hausdorff assumption. Therefore no such y exists, i.e., e is maximal.

For the converse, see p.5 of text. □

Example 0.1.1: Application of Zorn's Lemma. Let V be an infinite dimensional vector space, W a subspace, and $f : W \to \mathbb{R}$ a linear functional. Can we always extend f to all of V (still linear)? Note that if V is finite dimensional this is obvious.

Solution. Yes; first notice that we can always extend by one dimension. Take some $x \notin W$ and choose any value $f(x) = b$. For every $y \in \text{span}\{W \cup \{x\}\}$ has form $y = w + cx$ with $w \in W$. Then we can extend f by

$$f(y) = f(w + cx) := f(w) + cb.$$

It is easy to verify that f has been extended and is still linear.
Now we generalize. A definition first. If W_1, W_2 are subspace of V and $g_i : W_i \to \mathbb{R}$, we say (W_2, g_2) is an **extension** of (W_1, g_1) if

$$W_2 \supset W_1 \qquad \text{and} \qquad g_2 = g_1 \text{ on } W_1.$$

(We write $(W_2, g_2) \geqslant (W_1, g_1)$.) This is a partial order on

$$\mathcal{W} := \{\text{all linear functionals on subspaces of } V\}.$$

Now we check conditions for Zorn's lemma. Does every linearly ordered subset of V has an upper bound? Suppose $\mathcal{E} = \{(W_\alpha, g_\alpha) : \alpha \in A\}$ is a linearly ordered subset, meaning that any two W_α, W_β, one always contains the other. We consider

$$W_{\text{up}} := \bigcup_{\alpha \in A} W_\alpha \qquad \text{and} \qquad g_{\text{up}} := g_\alpha \text{ on } W_\alpha.$$

Then $(W_{\text{up}}, g_{\text{up}})$ is an upper bound for all for $\mathcal{E}$. Thus the assumption of Zorn's lemma is satisfied.
Therefore by Zorn's lemma $\mathcal{W}$ contains a maximal $(W_{\max}, g_{\max})$. But $W_{\max}$ must be all of W, otherwise we could extend it by one dimension as mentioned above. This proves the claim!

□

Theorem 0.1.2: The Well-Ordering Principle

Every nonempty set X can be well ordered.

Proof. Let $\mathcal{W}$ be the set of all subsets $W \subset E \times E \subset X \times X$ where E some subset of X and W is an ordering on E. (In other words we are considering an ordering of orderings here, and soon we will attempt to find a "maxiamal" ordering that hopefully applies to all of X, thereby making it a well-ordered set.) We can order $\mathcal{W}$ (partially) via $W_1 \leqslant W_2$ if W_1 well-orders some E_i $(i = 1, 2)$ and W_2 extends the ordering "upward":

(1) $E_1 \subset E_2$ and the 2 orders agree on E_1 (meaning $W_1 \subset W_2$ as sets), and

(2) In the ordering W_2, all elements of e_1 are $\leqslant$ all elements in $E_2 - E_1$.

We want to apply Zorn's lemma. Check hypothesis: suppose that some collection $\mathcal{W}^* = \{W_\alpha : \alpha \in A\} \subset \mathcal{W}$ is linearly ordered. Namely, any two orderings in $\mathcal{W}^*$ satisfy (1) and (2). Then the union of all these sets, $\bigcup_\alpha E_\alpha$, is our candidate for upper bound. Pick $x, y \in \bigcup_\alpha E_\alpha$. Is $x \leqslant y$? Pick $x \in E_\alpha$ and $y \in E_\beta$. One of E_α, E_β is bigger, say the latter. Then both $x, y \in E_\beta$ and the ordering of E_β will determine either $x \leqslant y$ or $y \leqslant x$.
A similar argument shows that it is well-ordered. Hence the ordering on $\bigcup_{\alpha \in A} E_\alpha$ is an upper bound for the subset $\mathcal{W}^*$ of $\mathcal{W}$. Therefore, by Zorn's lemma, $\mathcal{W}^*$ has a maximal element (maximal well-ordered subset of X). Call it W, an well-ordering of some $E \subset X$. Clearly $E = X$. Otherwise, we can enlarge the ordering by setting some $y \notin E$ at the top ($x \leqslant y$ for all $x \in E$), contradicting the maximality. □

Beginning of Aug.27, 2021

Corollary 0.1.3: Well-Ordering Principle ⇒ AC

Proof: We well-order $\bigcup_{\alpha\in A} X_\alpha$ and let $f(\alpha)$ denote the mimimal element in X_α. Then $f \in \prod_{\alpha\in A} X_\alpha$.

Cardinality

Definition 0.1.4: Cardinality

Let E, F be sets. We say $\text{card}(E) \leqslant \text{card}(F)$ if there exists an injection $f : E \to F$ or, equivalently, there exists a surjection $g : F \to E$. (This equivalence requires AC.)

Remark. Any two sets E, F are comparable in this way. Let d be the set of *all injections* to F defined on subsets of E. We can order these injections (one being an extension of the other) and use Zorn's lemma to deduce that there exists a maximal injection $f : E_{\text{max}} \to F$ (with largest possible domain). Now consider any two points x, y such that

$$x \in E - E_{\text{max}} \qquad \text{and} \qquad y \in F - f(E_{\text{max}}).$$

Since f corresponds to the "maximal" injection, either x or y does not exist (or we can find a "larger" injection). If x does not exist, i.e., E_{max} is all of E, then $f : E \to F$ is injective. Otherwise, if $f(E_{\text{max}})$ is all of F, then $f : E_{\text{max}} \to F$ is surjective.

If we have injections both ways, is there necessarily a bijection between the sets?

Theorem 0.1.5: Schröder-Bernstein Theorem

If $\text{card}(X) \leqslant \text{card}(Y)$ and $\text{card}(Y) \leqslant \text{card}(X)$, then $\text{card}(X) = \text{card}(Y)$ (bijection exists!).

Proof. Let $f : X \to Y$ and $g : Y \to X$ be the injections.
We start from some $x_0 \in X$. If $x_0 \in g(Y)$, we can apply g^{-1} to get $^{-1}(x_0) = x_1 \in Y$. If $x_0 \notin g(Y)$ then we simply stop.
We then repeat: if $x_1 \in f(X)$, we apply f^{-1} to get $^{-1}(x_1) = x_2 \in X$. If $x_1 \notin f(X)$, the process simply stops from x_1.
We continue doing this as long as we can. Now we define

$$X_X = \{x_0 \in X : \text{process eventually stops in } X\}$$
$$X_Y = \{x_0 \in X : \text{process eventually stops in } Y\}$$
$$X_\infty = \{x_0 \in X : \text{process never stops}\}.$$

(Similarly we could define Y_X, Y_Y, and Y_∞.)

For example, along the trajectory $x_0 \overset{g^{-1}}{\to} x_1 \overset{f^{-1}}{\to} x_2 \overset{g^{-1}}{\to} x_3$, all 4 points (x_0, x_1, x_2, and x_3) will end up at x_3. Further notice that each element can only appear in exactly one sequence.
Thus, g maps Y_Y *bijectively* to X_Y (e.g. $(x_1) = x_0, g(x_3) = x_2$), and similarly f maps X_X bijectively to Y_X. Also,

f maps X_∞ *bijectively* to Y_∞. Therefore, we can combine them and obtain a bijection $h : X \to Y$:

$$h(x_0) := \begin{cases} g^{-1}(x_0) & \text{if } x_0 \in X_Y \\ f(x_0) & \text{if } x_0 \in X_X \\ f(x_0) & \text{if } x_0 \in X_\infty. \end{cases}$$

□

Definition 0.1.6: Countability, etc.

(1) We say X is **countable** if $\text{card}(X) \leqslant \text{card}(\mathbb{N})$.

(2) We say X has the cardinality of the **continuum**, denoted $\mathfrak{c}$, if $\text{card}(X) = \text{card}(\mathbb{R})$.

(3) The following sets all have cardinality $\mathfrak{c}$:

$$\mathbb{R} \qquad \mathcal{P}(\mathbb{Z}) \qquad [0,1] \qquad \{0,1\}^{\mathbb{N}} \qquad [0,1]^{\mathbb{N}}.$$

Theorem 0.1.7: $\mathbb{R}$ is Uncountable

Proof. The famous Cantor diagonalization. Omitted.

Theorem 0.1.8: Continuum Hypothesis

$\mathfrak{c}$ is the smallest cardinality $> \text{card}(\mathbb{N})$.
This has been proven to be *undecidable* using standard set theory. There is no answer as of now.

Chapter 1

Measures

We want to assign sizes $\mu(E)$ to sets E. The most intuitive, special case is by defining $\mu(E)$ as the volume of E in $\mathbb{R}^n$. We want to also extend it to other sets. A natural question arises – what criterion should we choose when defining such notion?

(1) $\mu(Q) = 1$ for the unit cube Q. $\mu(E) \geqslant 0$ for all E.

(2) μ could be translation-invariant. If F can be obtained by translation E then $\mu(F)$ should be equal to $\mu(E)$.

(3) μ should be countably additive. If $\{E_n\}$ is a finite or countable collection of sets, then μ of their union should be sun of $\mu(E_n)$.

Some other properties which are consequences from above:

(4) If $A \subset B$ then $\mu(A) \leqslant \mu(B)$. Indeed, we can write B as $(B - A) \sqcup A$.

(5) μ of a singleton is 0. Indeed, $[0,1]$ contains infinitely many points, so by (3),

$$\mu([0,1]) = \sum_{x \in [0,1]} \mu(\{x\}).$$

Example 1.0.1. $\mu([1/5, 25)) = 1/5$ because

$$1 = \mu([0,1]) = \mu([0,1)) = \mu([0,1/5)) + \ldots + \mu([4/5,1)).$$

Similarly, $\mu([0.23, 0.43)) = 1/5$.

Beginning of Aug.30, 2021

However, in fact we can find **nonmeasurable sets** where no μ works satisfying all the axioms above.
Suppose we can find some subset $A \subset [0,1)$ such that there are *infinitely many* disjoint translates $A_1, A_2, \ldots$ whose union in $[0,1)$. Then

$$1 = \mu([0,1)) = \sum_{i=1}^{\infty} \mu(A_i)$$

which is impossible!!

Example of Non-Measurable Sets. We define an equivalence relation on $[0, 1)$ by $a \sim b$ if $a - b \in \mathbb{Q}$. There are *uncountably many* such equivalence classes, where class is countable (since $\mathbb{Q}$ is). Using AC, we may pick *one* point from *each* equivalence class and form a set A. (There is no formula that describes what A looks like, but we know we can.)
Then

(1) the difference between any two elements in A is irrational (since they below to different equivalence classes), and

(2) each $x \in [0, 1)$ can only be in *exactly one* equivalence class.

Therefore (1) implies that the translate $A + r$ for some $r \in \mathbb{Q}$ forms a new set that is disjoint from A, and (2) implies that the countable union of sets of form $A + r, r \in \mathbb{Q}$ is the entire $[0, 1)$.
It follows that A cannot be in the domain of μ. □

Domain of Measure

The question arises — if not the entire power set, **what should the domain of a measure be**, then?
Clearly, given X, the domain of a measure μ on X needs to be a collection $\mathfrak{F} \subset \mathbb{P}(X)$. Using a weakened version (compared to above), we tentatively require the following:

(i) $\mathfrak{F}$ contains X and $\varnothing$.

(ii) $\mathfrak{F}$ is closed under complementation.

(iii) $\mathfrak{F}$ is closed under finite / countable unions. (Their intersection counterparts follow from De Morgan's law and (ii).)

Definition 1.0.2: σ-Algebra

We say $\mathfrak{F}$ is an **algebra** (or **field**) if (iii) holds for finite unions. We say it's a **σ-algebra** (or **σ-field**) if (iii) holds for countable unions.

Proposition 1.0.3

The intersection of two σ-algebras is still a σ-algebra. *One can simply verify this via the axioms.*
More generally, let $\mathcal{E}$ be a *collection* of subsets of X, and let

$$\sigma(\mathcal{E}) = \text{intersection of all } \sigma\text{-algebras containing } \mathcal{E}.$$

Then:

(1) $\sigma(\mathcal{E})$ is a σ-algebra, and

(2) $\sigma(\mathcal{E})$ is the *smallest* σ-algebra containing $\mathcal{E}$: that is, for all σ-algebra $\mathfrak{F} \supset \mathcal{E}$, we have $\mathfrak{F} \supset \sigma(\mathcal{E})$.

We say $\sigma(\mathcal{E})$ is the σ-algebra **generated** by $\mathcal{E}$.

Definition 1.0.4: Borel σ-Algebra

A connection to topology — let X be a topological space, i.e., it has a family $\mathfrak{J}$ of **open sets** (closed under arbitrary union and finite intersections). The **Borel σ-algebra** is the σ-algebra generated by $\mathfrak{J}$ (by all open sets). We write $\mathcal{B}_X = \sigma(\mathfrak{J})$.

Note that open sets are NOT closed under arbitrary intersections, but $\mathcal{B}_X$ is, so (nearly in all cases) $\mathcal{B}_X$ is bigger. Open sets do not constitute a σ-algebra in general.

Definition 1.0.5: G_δ and F_σ Sets

A $\boldsymbol{G_\delta}$ set is a countable intersection of open sets. An $\boldsymbol{F_\sigma}$ set is a countable union of closed sets. Similarly, a $\boldsymbol{G_{\delta\sigma}}$ set is a countable union of G_δ sets, and so on...

Product Spaces

Recall that $X = \prod_{\alpha \in A} X_\alpha$ is a Cartesian product, and $\pi_\alpha : X \to X_\alpha$ is the **projection** onto the α^{th} coordinate. For $E_\alpha \subset X_\alpha$, the inverse image $\pi_\alpha^{-1}(E_\alpha)$ is almost the same product $\prod_{\alpha \in A} X_\alpha$, except X_α is now replaced by E_α (all other X_i's remain the same; they do not bring changes to our projection).

Definition 1.0.6: Unrestricted Coordiante

An **unrestricted coordinate** in A is an α such that if $x_\beta = x_\beta$ for all $\beta \neq \alpha$, then $x_\beta \in A$ if and only if $y_\beta \in A$. (*Changing α's coordinate never changes the membership in A.*) For example consider a vertical cylinder in $\mathbb{R}^3$. The z-coordinate is unrestricted.

Definition 1.0.7: Product Algebra

Suppose each X_a corresponds to a σ-algebra $\mathfrak{M}_\alpha$. We define the **product σ-algebra** in $X = \prod_{\alpha \in A} X_\alpha$ by

$$\bigotimes_{\alpha \in A} \mathfrak{M}_\alpha = \text{the } \sigma\text{-alpha generated by} \{\pi_\alpha^{-1}(E_\alpha) : \alpha \in A, E_\alpha \in \mathfrak{M}_\alpha\}.$$

In other weeds, $\bigotimes_{\alpha \in A} \mathfrak{M}_\alpha$ is generated by all the preimages of all elements in $\mathfrak{M}_\alpha$'s.

Note that this product σ-algebra is closed under intersection, so it also contains sets of form

$$\bigcap_{i=1}^{\infty} \pi_{\alpha_i}^{-1}(E_{\alpha_i}) \text{ with } E_{\alpha_i} \in \mathfrak{M}_{\alpha_i}.$$

There is another σ-algebra in $X = \prod_\alpha X_\alpha$:

$$\mathfrak{F} := \text{the } \sigma\text{-algebra generated by } \{\text{all abstract rectangles } \prod_{\alpha \in A} E_\alpha \text{ with } E_\alpha \in \mathfrak{M}_\alpha\}.$$

The first way is to get a collection of abstract rectangles with restriction on one coordinate, and the second starts with completely unrestricted rectangles. Then we generate two σ-algebras from these two.

Notice that $\bigotimes_{\alpha \in A} \mathfrak{M}_\alpha \subset \mathfrak{F}$: indeed, the latter has a larger generator.

Proposition 1.0.8

If $\mathfrak{F}, \mathfrak{G}$ are σ-algebras and $\mathfrak{F}$ contains all generators of $\mathfrak{G}$, then $\mathfrak{F} \supset \mathfrak{G}$.
Indeed, $\mathfrak{G}$ is the smallest σ-algebra containing its generators, so $\mathfrak{F}$, another σ-algebra, cannot be smaller.

Proposition 1.0.9

If A is countable, then the aformentioned $\bigotimes_{\alpha \in A} \mathfrak{M}_\alpha = \mathfrak{F}$.

Proof. It suffices to prove the $\supset$ direction.
When A is countable, every abstract rectangle is simply the *countable* intersection of rectangles in which we each impose one restriction, so by σ-algebra's closure under *countable* intersection, each

$$\prod_{\alpha \in A} E_\alpha = \bigcap_{\alpha \in A} \underbrace{\pi_\alpha^{-1}(E_\alpha)}_{\in \bigotimes_{\alpha \in A} \mathfrak{M}_\alpha} \in \bigotimes_{\alpha \in A} \mathfrak{M}_\alpha .$$

□

Remark: (General Principle). Suppose that

(1) Y is a collection of sets each having some property P.

(2) $\mathcal{P}$, the collection of all sets with such property, is a σ-algebra.

Then every set in $\sigma(Y)$ has this property.
Proof: since $\mathcal{P}$ is a σ-algebra and $\mathcal{P} \supset Y$, we have $P \supset \sigma(Y)$.
Upshot: to show that a σ-algebra has some property, we only need to verify its "generators".

Beginning of Sept.1, 2021

Proposition 1.0.10

If A is uncountable (and $\mathfrak{M}_\alpha$ is assumed to be nontrivial, i.e., not just containing A_α and $\varnothing$), then

$$\bigotimes_{\alpha \in A} \mathfrak{M}_\alpha \neq \mathfrak{F}.$$

Proof. Consider the property P of a set B defined by "B has only countably many restricted coordinates." For example, the generators $\pi_\alpha^{-1}(E_\alpha)$ all have one restricted coordinates so they each have P. It follows that P is closed under complements and countable unions[!] so it is a σ-algebra. Therefore, the general principle says that *every* set in $\bigotimes_{\alpha \in A} \mathfrak{M}_\alpha$ has this property. However, in $\mathfrak{F}$ we have sets with *uncountably* many restricted coordiantes. Therefore $\mathfrak{F}$ is strictly bigger. □

Proposition 1.0.11

If each $\mathfrak{M}_\alpha$ is generated by some collection $\mathcal{E}_\alpha$, then

$$\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathcal{E}_\alpha, \alpha \in A\}$$

generates $\prod_{\alpha \in A} \mathfrak{M}_\alpha$.

Example 1.0.12. Let $\mathcal{B}_\mathbb{R}$ be the Borel sets in $\mathbb{R}$, generated by the colleciton of open sets who are in turn generated by the open intervals. By definition $\mathcal{B}_\mathbb{R} \times \mathcal{B}_\mathbb{R}$ is generated by

$$\mathcal{E} = \{\text{all sets } G \times \mathbb{R} \text{ or } \mathbb{R} \times G \text{ with } G \text{ open}\},$$

but this example states that it is also generated by

$$\mathcal{D} = \{\text{all sets } I \times \mathbb{R} \text{ or } \mathbb{R} \times I \text{ with } I \text{ open}\}.$$

Proof. Every set in $\mathcal{E}$ is a countable union of sets in $\mathcal{D}$ so one direction $\mathcal{E} \subset \sigma(\mathcal{D})$ is trivial. Also, $\mathcal{D} \subset \mathcal{E}$, so $\sigma(\mathcal{D}) \subset \sigma(\mathcal{E})$, and the two sets are indeed equal. □

Example 1.0.13: Special Example: Metric Spaces. Let $X_1, \dots, X_n$ be metric spaces with metrics d_i in X_i. Then there are two ways to make a σ-algebra in $\prod_{j=1}^n X_i$:

(1) Get a σ-algebra in X_i first, for example $\mathcal{B}_{X_i}$. Then take $\bigotimes_{i=1}^n \mathcal{B}_{X_i}$.

(2) Define the product metric $d(x, y) = \max_{i \leqslant n} d_i(x_i, y_i)$ on $\prod_{i=1}^n X_i$. This creates open sets in the product space and generates Borel sets $\mathcal{B}_X$ from them.

Are these two the same? It relates to separability.

Proposition 1.0.14

Suppose $X_1, \dots, X_n$ are metric spaces and $X = \prod_{i=1}^n X_i$ with the product metric. Then $\bigotimes_{i=1}^n \mathcal{B}_{X_i} \subset \mathcal{B}_X$. If in addition each X_i is separable, the two are equal.

Proof. Note that $\mathcal{E}_i := \{\text{open sets in } X_i\}$ generates $\mathcal{B}_{X_i}$, so

$$\{\pi_i^{-1}(u_i) \mid u_i \in \mathcal{E}_i\}$$

generates $\bigotimes_{i=1}^n \mathcal{B}_{X_i}$. All of these $\pi_i^{-1}(u_i)$'s are *open* so

$$\bigotimes_{i=1}^n \mathcal{B}_{X_i} \subset \sigma(\text{all open sets in } X) = \mathcal{B}_X.$$

Now for the separable case: each X_i contains a countable dense subset Y_i. Define

$$\mathcal{E}_i := \{\text{all "special" balls in } X_i \text{ with centers in } Y_i \text{ with rational radius}\}$$

which is countable. Then, if $G \subset X_i$ is open and $p \in G$, there exists a ball $B(r,q) \in \mathcal{E}_j$ with $p \in B(r,q) \subset G$. Therefore, every $p \in G$ is in such a ball, and G is a *countable* union of special balls from $\mathcal{E}_i$.

Now we generalize this into the product space. If $U \subset X$ is open and $x \in U$, then each coordinate of x is contained in a "special open ball". Analogously, U is a union of products of special balls. In particular, since $\mathcal{E}_i$'s are countable, U is a *countable* union of such balls. Therefore, the product σ-algebra contains *all* open sets and therefore the Borel algebra, i.e.,

$$\mathcal{B}_X = \sigma(\text{all of } U) \subset \bigotimes_{i=1}^{n} \mathcal{B}_{X_i}.$$

□

Example 1.0.15: What goes wrong in the non-separable case?. For the countable case, for example $\mathbb{R} \times \mathbb{R}$, we can let $\mathcal{E} = \{\text{open intervals in } \mathbb{R} \text{ with rational endpoints}\}$. Then $\mathcal{E}$ is countable and every open interval is a union of sets in E. Then, given $U \subset \mathbb{R} \times \mathbb{R}$ open, for all $p \in U$, there exists $I_1, I_2 \in \mathcal{E}$ with $p \in I_1 \times I_2 \subset U$. The set

$$\bigcup \{I_1 \times I_2 \mid I_1, I_2 \in \mathcal{E}, I_1 \times I_2 \subset U\}$$

is *all* of U, so U is a countable union of sets $I_1 \times I_2$.

However, for a non-separable metric space X, there exists *uncountably* many disjoint open sets, say $\{G_\alpha, \alpha \in A\}$. Then

$$U = \bigcup_{\alpha \in A} G_\alpha \times G_\alpha$$

is open in $X \times X$. We can show that this is *not* in the product σ-algebra $\mathcal{B}_\alpha \times \mathcal{B}_\alpha$, but it's open and therefore in $\mathcal{B}_{X \times X}$. In this case $\mathcal{B}_{X \times X} \neq \mathcal{B}_\alpha \times \mathcal{B}_\alpha$.

Definition 1.0.16: Elementary Family

An **elementary family** is an $\mathcal{E} \subset \mathbb{P}(X)$ such that

(1) $\varnothing \in \mathcal{E}$,

(2) $\mathcal{E}$ is closed under *finite* intersection, and

(3) for all $E \in \mathcal{E}, E^c$ is a *finite* disjoint unions of members of $\mathcal{E}$.

For example, the collection of all abstract rectangles (product of two intervals, possibly infinite and possibly degenerate) in $\mathbb{R}^2$ consist of an elementary family. (1) is clear. (3) is clear for "basic" rectangles: $\mathbb{R}^2 - [a,b] \times [c,d]$ is the disjoint union of 8 "infinite" rectangles. (Indeed, $x = c, x = d, y = a$, and $y = b$ divide $\mathbb{R}^2$ into 9 disjoint "rectangles" and we took away one of them.) It follows that finite intersections also preserve this property, and hence (2) holds.

Beginning of Sept.3, 2021

1.1 Measures

Definition 1.1.1: Measures, etc.

Let X be a set with a σ-algebra $\mathfrak{M}$, the collection of **measurable sets**. We call $(X, \mathfrak{M})$ a **measure space**. We say a function $\mu : \mathfrak{M} \to [0, \infty]$ is a **measure** if

(1) $\mu(\varnothing) = 0$ and

(2) μ is *countably additive*: if $\{E_i\}$ are disjoint in $\mathfrak{M}$ then

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

Note that if we let the tail of $\{E_i\}$ to be $\varnothing$, we obtain finite additivity of μ as well.

Proposition 1.1.2: Monotonicity of Measure

If $E, F \in \mathfrak{M}$ and $E \subset F$, then $\mu(E) \leqslant \mu(F)$.
Proof: write $F = E \sqcup (F - E)$.

Definition 1.1.3: Finite, σ-finite, & Semifinite Measures

If $\mu(X) < \infty$ then $\mu(E) < \infty$ for all $E \in \mathfrak{M}$. If so we call μ a **finite measure**.
We say μ is **σ-finite** if X can be written as a countable union of sets, each of which has finite measure.
We say μ is **semifinite** if every E with $\mu(E) = \infty$ has a subset $F \subset E$ with finite measure.

Example 1.1.4: Lebesgue Measure (later). The Lebesgue measure m is defined on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ such that $\mu(I)$ is the length of I for all interval I. Then $\mu(\mathbb{R}) = \infty$ but $\mathbb{R}$ is the countable union of $[n-1, n)$ so m is σ-finite but not finite.

Example 1.1.5. If we let $X = \mathbb{R}$, $\mathfrak{M} = \mathbb{P}(\mathbb{R})$ and define

$$\mu(E) = \begin{cases} 0 & E \text{ countable} \\ \infty & E \text{ uncountable}, \end{cases}$$

then μ is neither σ-finite nor semifinite.

Example 1.1.6: Point Mass. Let $A \subset \mathfrak{M}$ and define $\mu(A) = 1$ if $x \in A$ and 0 otherwise.

Theorem 1.1.7: More Properties of Measures

μ is subadditive: if $E_i \in \mathfrak{M}$ then

$$\mu(\bigcup_{i \geqslant 1} E_i) \leqslant \sum_{i \geqslant 1} \mu(E_i).$$

Proof. We define $F_i := E_i - (\bigcup_{i<j} E_i)$. Then the F_i's are disjoint and the claim follows from the additivity of disjoint sets and monotonicity. □

Proposition 1.1.8: Continuity from Below

(1) Suppose $E_1 \subset E_2 \subset \ldots$ in $\mathfrak{M}$. Then

$$\mu(\bigcup_{i\geqslant 1} E_i) = \lim_{i\to\infty} \mu(E_i).$$

(2) Suppose $E_1 \supset E_2 \supset \ldots$ in $\mathfrak{M}$ *and some E_n has finite measure*. Then

$$\mu(\bigcap_{i\geqslant 1} E_i) = \lim_{i\to\infty} \mu(E_i).$$

Proof.

(1) Let $E_0 = \varnothing$. Then

$$\mu(\bigcup_{i\geqslant 1} E_i) = \mu(\bigcup_{i\geqslant 1}(E_i - E_{i-1})) = \sum_{i=1}^{\infty} \mu(E_i - E_{i-1}) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(E_i - E_{i-1}) = \lim_{n\to\infty} \mu(E_n).$$

(2) Analogous — if $\{E_i\}$ is decreasing, then $\{E_i^c\}$ is increasing, and we can apply (1).

More formally, fix n and let $F_i = E_n - E_i$ for $i \geqslant n$. It follows that $F_{n+1} \subset F_{n+2} \subset \ldots$, and

$$E_n = (\bigcup_{i\geqslant n} F_i) \cup (\bigcap_{i\geqslant 1} E_i).$$

Therefore,

$$\mu(E_n) - \mu(\bigcap_{i\geqslant 1} E_i) = \mu(\bigcup_{i\geqslant n} F_i) = \lim_{i\to\infty} \mu(F_i) = \lim_{i\to\infty} (\mu(E_n) - \mu(E_i)),$$

and since $\mu(E_n) < \infty$, subtracting makes sense and we are done.

Note that $\mu(E_1) < \infty$ is important: if we let $E_n := [n, \infty)$ then the infinite intersection is $\varnothing$ whereas the limit of $\mu(E_n) = \infty$.

□

Definition 1.1.9: Null Set

A **null set** is a set E with $\mu(E) = 0$. (What's null depends on μ.)

We say "something is true **almost everywhere** (a.e)" if something is true on all $x \notin E$ where E is a null set.

Remark. If $u(E) = 0$, there might be "bad" sets $F \subset E$ with $F \notin \mathfrak{M}$. But we want to assign a measure 0 to them as well. However, we cannot add *one* set to $\mathfrak{M}$ as $\mathfrak{M} \cup \{F\}$ is not a σ-algebra anymore. This leads to the following notion called completion.

Definition 1.1.10

A measure μ is called **complete** if for all $F \subset E$ where $E \in \mathfrak{M}$ with $u(E) = 0$, we have $\mu(F) = 0$.

Example 1.1.11. Let μ be the Lebesgue number on $[0, 1]$ and let $\mathcal{C}$ be the standard middle-thirds Cantor set. It turns out that $\mathcal{C}$ is a null set.
$\mathcal{C}$ is "isomorphic to $[0, 1]$" in the sense that there exists a monotone bijection $\varphi : [0, 1] \to \mathcal{C}$ by

$$\text{binary to "ternary (left or right)"} : (0.00101...) \mapsto (\text{LLRLR}...)$$

for example. One can show that if A is non-Borel in $[0, 1]$ then $\varphi(A)$ is a non-Borel set in $\mathcal{C}$.

Theorem 1.1.12: Completion

Let $(X, \mathfrak{M}, \mu)$ be a measure space, and define

$$\mathfrak{N} := \{\text{all null sets of } \mu \text{ in } \mathfrak{M}\}$$

and

$$\overline{\mathfrak{M}} := \{E \cup F \mid E \in \mathfrak{M} \text{ and } F \subset N \text{ for some } N \in \mathfrak{N}\}.$$

Then $\overline{\mathfrak{M}}$ is a σ-algebra and there is a *unique* way to extend μ to $\overline{\mu}$ which is a *complete* measure on $\mathfrak{M}$.

Proof. First we show that $\overline{\mathfrak{M}}$ is a σ-algebra. Complementation is given by

$$(E \cup F)^c = \underbrace{(E \cup N)^c}_{\in \mathfrak{M}} \cup \underbrace{(N - E \cup F)}_{\subset N} \in \overline{\mathfrak{M}}.$$

For a countable collection of $\{E_i \cup F_i\} \in \overline{\mathfrak{M}}$,

$$\bigcup_{n \geqslant 1} (E_n \cup F_n) = \underbrace{(\bigcup_{n \geqslant 1} E_n)}_{\in \mathfrak{M}} \cup \underbrace{(\bigcup_{n \geqslant 1} F_n)}_{\mu = 0} \in \overline{\mathfrak{M}}.$$

We define $\overline{\mu}$ on $\overline{\mathfrak{M}}$ by $\overline{\mu}(E \cup F) = \mu(E)$. To check that this is well-defined, if $E_1 \cup F_1 = E_2 \cup F_2$, by definition $F_2 \subset N_2$ for some null set N_2. Then

$$E_1 \subset E_2 \cup F_2 \subset E_2 \cup N_2$$

and so

$$\mu(E_1) \leqslant \mu(E_2) = \mu(N_2) = \mu(E_2)$$

and likewise $u(E_2) \leqslant \mu(E_1)$.
The rest of the problem is left as a homework problem... □

Beginning of Sept.8, 2021

1.2 Outer Measure

For example, consider A the set of finite unions of rectangles in $\mathbb{R}^2$. This an algebra but not a σ-algebra. We can define μ on A by assigning the area to it. It follows that μ is finitely additive. On the other hand we may sometimes want to extend μ to $\overline{\mu}$ on a σ-algebra, say $\sigma(A)$. Core idea: define an *outer measure* μ^*, an "upper bound" for our measure, and then show that the restriction of μ^* to $\sigma(A)$ is indeed a measure.

Definition 1.2.1: Outer Measure

A function $\mu^* : \mathbb{P}(X) \to [0, \infty]$ (containing ∞) is an **outer measure** if

(1) $\mu^*(\varnothing) = 0$,

(2) μ^* is **monotonous**: $A \subset B$ implies $\mu^*(A) \leqslant \mu^*(B)$, and

(3) μ^* is countably subadditive:

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

If we are able to define such outer measure, then for all countable collection $\{A_i\}$ that covers E, we have

$$\overline{\mu}(E) \leqslant \overline{\mu}(\bigcup_{i=1}^{\infty}(A_i) \leqslant \sum_{i=1}^{\infty} \mu(A_i).$$

More formally,

$$\mu^*(E) := \inf\{\sum_{i=1}^{\infty} \mu(A_i) : A_i \in A, E \subset \bigcup_{i=1}^{n} A_i\}.$$

However, it is not immediately clear whether μ^* is additive, in particular finitely additive. Our goal's to show that in general

(1) μ^* is an outer measure, and

(2) $\mu^* |_{\sigma(A)}$ *is* a measure, and $\mu^* = \mu$ on A.

Proposition 1.2.2

Let $E \subset \mathbb{P}(X)$ and $\rho : E \to [0, \infty]$ satisfying $\rho, X \in E$ and $\rho(\varnothing) = 0$. For all $A \subset X$, we define

$$\mu^*(A) := \inf\{\sum_{i=1}^{\infty} \rho(E_i) : E_i \in E, A \subset \bigcup_{i=1} E_i\}$$

is an outer measure. Note that we require no additional assumptions on E and ρ besides the very basic ones.

Example 1.2.3. Let $X = \mathbb{R}^2$, E be the set of rectangles, and ρ the area. The propositions tates that we can extend ρ to $\sigma(E)$, which is a σ-algebra rather than simply an algebra.

Proof. Notice that μ^* is well-defined (worse case sceneraio: X covers A so there's always a cover). Also, $\mu^*(\varnothing) = 0$ can be obtained by letting $E_i = \varnothing$. Monotonicity is clear. For subadditivity, suppose $A = \bigcup_{i=1}^{\infty} A_i$. Let $\epsilon > 0$ be given; it suffices to show that

$$\mu^*(A) \leqslant \sum_{i=1}^{\infty} \mu^*(A_i) + \epsilon.$$

Indeed, we can choose a "near-optimal" cover where

$$\{E_i^k, k \geqslant 1\} \text{ with } \sum_k^\infty \rho(E_i^k) \leqslant \mu^*(A_i) + \frac{\epsilon}{2^i}.$$

Therefore the errors add up to $\leqslant \epsilon$:

$$\bigcup_{i=1}^\infty \sum_{k=1}^\infty \mu(E_i^k) \leqslant \sum_{i=1}^\infty \mu^*(A_i) + \epsilon,$$

and it is clear that $\{E_i^k\}$ covers A. □

There are 3 "types" of outer measures:

(1) General — any function satisfying the axioms;

(2) Derived from sums over covers (taken infimum as above); and

(3) Derived from sums over covers of a *premeasure* (something we want to extend to an algebra) on an algebra.

What is a "nice" set? Any A splits X into A and A^c. The outer measure should *add* for "nice" A in the sense that cutting A into pieces and adding the outer measure of each part should result in the same as $\mu^*(A)$, namely

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \qquad \text{for all } E \subset X.$$

If so, we call A a $\boldsymbol{\mu^*}$**-measurable** set. If $\mu(E) = \infty$ the equation is clear. If $\mu(E) < \infty$, it is obvious from subadditivity that the LHS $\leqslant$ the RHS, so to have a "nice" A, it suffices to show

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

Theorem 1.2.4: Carathéodory's Theorem

Let μ^* be an outer measure on $\mathbb{P}(X)$ and

$$\mathfrak{M} := \{\text{all } \mu^*\text{-measurable subsets of } X\}.$$

Then $\mathfrak{M}$ is a σ-algebra and $\mu^* |_{\mathfrak{M}}$ is a complete measure.

Proof.

(1) It is clear that $\mathfrak{M}$ is closed under complements: the statement is symmetric.

(2) $\mathfrak{M}$ is closed under finite unions: if $A, B \in \mathfrak{M}$ then we want to show that $A \cup B \in \mathfrak{M}$. This can be done quite easily:

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

where $\geqslant$ is by subadditivity.

(3) $\mathfrak{M}$ is closed under differences $A - B$ because $A - B = (A^c \cup B)^c$.

(4) $\mathfrak{M}$ is closed under *countable* union. Let $\{A_i\}$ be a countable collection of sets and WLOG assume they are

disjoint. Since for all E_n and all n,

$$\mu^*(E \cap (\bigcup_{i=1}^{n} A_i)) = \sum_{i=1}^{n} \mu^*(E \cap A_i),$$

we have

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap (\bigcup_{i=1}^{n} A_i)) + \mu^*(E \cap (\bigcup_{i=1}^{n} A_i)^c) \\ &\geqslant \sum_{i=1}^{n} \mu^*(E \cap A_i) + \mu^*(E \cap (\bigcup_{i=1}^{\infty} A_i)^c).\end{aligned}$$

Since this holds for all n, letting $n \to \infty$ we have

$$\begin{aligned}\mu^*(E) &\geqslant \sum_{i=1}^{\infty} \mu^*(E \cap A_i) + \mu^*(E \cap (\bigcup_{i=1}^{\infty} A_i)^c) \\ &\geqslant \mu^*(E \cap (\bigcup_{i=1}^{\infty} A_i)) + \mu^*(E \cap (\bigcup_{i=1}^{\infty} A_i)^c)\end{aligned}$$

so again the infinite union shows $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{M}$, and so $\mathfrak{M}$ is closed under countable unions.

(5) μ^* is countably additive on $\mathfrak{M}$: in the above inequality, replacing E by $E \cap (\bigcup_{i=1}^{\infty} A_i)$ gives

$$\mu(E \cap (\bigcup_{i=1}^{\infty} A_i)) = \sum_{i=1}^{\infty} \mu^*(E \cap A_i).$$

Taking $E = X$ gives countable additivity.

(6) $\mu^*|_{\mathfrak{M}}$ is complete. If $\mu^*(A) = 0$ and $E \subset X$ then

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

so all $\leqslant$ are $=$, so $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$, so $A \in \mathfrak{M}$.

□

Premeasure

Now we try to extend μ on an algebra $\mathcal{A}$ to $\overline{\mu}$ on $\sigma(\mathcal{A})$. We have previously talked about generalizing “length” or “volume” to “Lebesgue measure” on all Borel sets. In general, questions include

(1) What μ can be extended?

(2) Do the notions of length or area work?

(3) How to extend?

Definition 1.2.5: Premeasure

Given $\mathcal{A} \subset \mathbb{P}(X)$ an algebra, $\mu : A \to [0, \infty]$ is a **premeasure** if

(1) $\mu(\varnothing) = 0$, and

(2) for $\{A_i\}$ a countable disjoint union in A *with union also in* A, $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$. (For example, a countable union of rectangles may be a rectangle, but they may as well form a weird shape, in which case (2) does not follow.)

A premeasure μ produces an outer measure μ^* via "sums over covers":

$$\mu^*(E) = \inf\{\sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, E \subset \bigcup_{i=1}^{\infty} A_i\}.$$

It follows that μ^* is a measure on all μ^*-measurable sets, but is this the extension we want?

(1) Are all sets in $\mathcal{A}$ μ^*-measurable?

(2) Do μ^* and μ agree on $\mathcal{A}$?

Theorem 1.2.6

Let μ be a premeasure on an algebra $\mathcal{A}$ and let μ^* be the corresponding "sums over squares" outer measure. Then indeed $\mu^*|_{\mathcal{A}} = \mu$ and all sets in $\mathcal{A}$ are μ^*-measurable.

Proof. We first show that $\mu \leqslant \mu^*$. Let $E \in \mathcal{A}$ and let $\{A_i\}$ be a cover of E (by sets in $\mathcal{A}$). We can make them disjoint by defining

$$B_n := (A_n - \bigcup_{i<n} A_i) \cap E.$$

Then $\bigcup_{n=1}^{\infty} B_n = E \in \mathcal{A}$ so

$$\mu(E) = \sum_{n=1}^{\infty} \mu(B_n) \leqslant \sum_{n=1}^{\infty} \mu(A_n).$$

Since the above inequality holds for any cover of E, taking infimum gives $\mu(E) \leqslant \mu^*(E)$.

Now we show that $\mu^* \leqslant \mu$: since $E, \varnothing, \varnothing, \ldots$ is a cover of E by sets in $\mathcal{A}$, we have

$$\mu^*(E) \leqslant \mu(E) + 0 + \ldots = \mu(E).$$

Hence $\mu^* = \mu$ on $\mathcal{A}$.

Now we show that all sets in $\sigma(\mathcal{A})$ is μ^*-measurable. Indeed, let $E \subset X, A \in \mathcal{A}$, and $\epsilon > 0$ be given. By definition of infimum there exists a cover $\{B_i\} \subset \mathcal{A}$ that is $< \epsilon$-optimal:

$$E \subset \bigcup_{i=1}^{\infty} B_i \quad \text{and} \quad \sum_{i=1}^{\infty} \mu(B_i) \leqslant \mu^*(E) + \epsilon.$$

Since each B_i can be split into $B_i \cap A$ and $B_i \cap A^c$, we obtain

$$\begin{aligned} \mu^*(E) + \epsilon &\geqslant \sum_{i=1}^{\infty} \mu(B_i) \\ &= \sum_{i=1}^{\infty} \mu(B_i \cap A) + \sum_{i=1}^{\infty} \mu(B_i \cap B^c). \end{aligned}$$

Note that the first term is a cover of $E \cap A$ and the second of $E \cap A^c$. Hence

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

Since ϵ is arbitrary, we are done — any set in $\mathcal{A}$ is μ^*-measurable. □

Theorem 1.2.7

Let $\mathcal{A}$ be an algebra and μ_0 a premeasure on $\mathcal{A}$. Let μ^* be the outer measure *obtained* from μ_0. Then:

(1) The restriction $\overline{\mu} := \mu^*\big|_{\sigma(\mathcal{A})}$ is a measure on $\sigma(\mathcal{A})$ which extends $\mu_0 := \overline{\mu}\big|_{\mathcal{A}}$.

(2) $\overline{\mu}$ is the "largest" extension: if v is any *other* extension of μ_0 to $\sigma(\mathcal{A})$ then

(i) $v \leqslant \overline{\mu}$ on $\sigma(\mathcal{A})$,

(ii) $v(E) = \overline{\mu}(E)$ if $\overline{\mu}(E) < \infty$, and

(iii) $v = \overline{\mu}$ on $\sigma(\mathcal{A})$ if μ_0 is σ-finite.

Note that the second statement says that $\overline{u}$ and v differ only when (1) $v(E) < \infty$ and $\overline{\mu}(E) = \infty$ for some E and (2) μ_0 is not σ-finite.

Beginning of Sept.13, 2021

Proof of "uniqueness". If v is an extension of μ_0 to a measure on any σ-algebra $\mathfrak{M}$ containing $\mathcal{A}$, then for all $E \in \mathfrak{M}$ there exists a cover $\{A_i\}$ of E by sets in $\mathcal{A}$.

$$v(E) =\leqslant \sum_{i=1}^{\infty} v(A_i) = \sum_{i=1}^{\infty} \mu_0(A_i).$$

Taking infimum over covers implies $\mu(E) \leqslant \mu^*(E)$ (and on σ-algebra $\mu^* = \overline{\mu}$).
On the other hand, suppose $\overline{\mu}(E) < \infty$ for some $E \in \sigma(\mathcal{A})$. We want to show that $\overline{\mu}(E) \leqslant v(E) + \epsilon$ for any $\epsilon > 0$. Indeed, choose a cover $\{A_i\}$ that is "optimal within ϵ":

$$\overline{\mu}(E) + \epsilon = \mu^*(E) \geqslant \sum_{i=1}^{\infty} \mu_0(A_i) = \sum_{i=1}^{\infty} \overline{\mu}(A_i) \geqslant \overline{\mu}(\bigcup_{i=1}^{\infty} A_i)$$

so $\overline{\mu}((\bigcup_{i=1}^{\infty} A_i) - E) \leqslant \epsilon$. Therefore

$$\begin{aligned}\overline{\mu}(E) \leqslant \overline{\mu}(\bigcup_{i=1} A_i) &= \lim_{n\to\infty} \overline{\mu}(\bigcup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} \mu(\bigcup_{i=1}^{n} A_i) \\ &= v(\bigcup_{i=1}^{\infty} A_i) = v(E) + v((\bigcup_{i=1}^{\infty} A_i) - E) \\ &\leqslant v(E) + \overline{\mu}((\bigcup_{i=1}^{\infty} A_i) - E) \leqslant v(E) + \epsilon.\end{aligned}$$

□

Proof of ν and $\overline{\mu}$. Suppose μ_0 is σ-finite. There exists $B_1 \subset B_2 \subset \mathcal{A}$ $\mu(B_i) < \infty$ such that $\bigcup_{i=1}^{\infty} B_i = X$. Then for all $E \in \sigma(\mathcal{A})$, we can write E as a limit:

$$\overline{\mu}(E) = \lim_{i\to\infty} \overline{\mu}(E \cap B_i)$$

by continuity from below. Each $\overline{\mu}(E \cap B_i)$ are finite, so $\overline{\mu}$ and ν agree on each one of them by the previous part.

Thus

$$\overline{\mu}(E) = \lim_{i\to\infty} \overline{\mu}(E \cap B_i) = \lim_{i\to\infty} \nu(E \cap B_i) = \nu(E).$$

□

1.3 Borel Measures on $\mathbb{R}$

Here we consider the **h-intervals** of form $(a, b]$ where $-\infty \leqslant a < b \leqslant \infty$. Note that the collection of these intervals form an elementary family, so the collection of finite unions of them form an algebra (not σ-algebra). Define

$$\mathfrak{A} := \{\text{all finite unions of h-intervals}\}.$$

We take a measure μ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ that is finite on bounded sets. Define its **distribution function**

$$F(x) := \begin{cases} -\mu((x, 0]) & x < 0 \\ 0 & x = 0 \\ \mu((0, x]) & x > 0. \end{cases}$$

Proposition 1.3.1: Properties of F

Clearly F is nondecreasing. It is right-continuous: for all $x > 0$ and $y_n \downarrow x$,

$$F(x^+) = \lim_{y_n\downarrow x} F(y_n) = \lim_{y_n\downarrow x} \mu((0, y_n]) = \mu(\bigcap_{n=1}^{\infty}(0, y_n]) = \mu((0, x]) = F(x),$$

and a similar argument proves the case for $x \leqslant 0$ and $y_n \downarrow x$ using continuity from below.

Also, $\mu((a, b]) = F(b) - F(a)$ for $a < b$. For $0 < a < b$, this is proven by

$$F(b) - F(a) = \mu((0, b]) - \mu((0, a]) = \mu((a, b]).$$

Conversely, if we have a nondecreasing, right continuous function F, then we can obtain a measure μ described by

$$\mu((a, b]) = F(b) - F(a).$$

Proposition 1.3.2

If $F : \mathbb{R} \to \mathbb{R}$ is nondecreasing and right-continuous, define μ on $\mathcal{A} := \{\text{finite union of h-intervals}\}$ by $\mu(\varnothing) = 0$ and

$$\mu(\bigcup_{i=1}^{n}(a_i, b_i]) := \sum_{i=1}^{n}(F(b_i) - F(a_i))$$

for disjoint h-intervals $(a_i, b_i]$. Then μ is a premeasure on $\mathcal{A}$ (i.e., from finite additivity we can deduce countable additivity).

Proof. We first note that μ is well-defined (for finite unions, even if there are more than one ways to represent a union, all middle terms cancel each other out).

Now we show that μ is countably additive "within $\mathcal{A}$", i.e., countably additive for h-intervals whose countable union is still an h-interval. To put formally, if $I = (a, b]$ is a countable disjoint unions $I_j := (a_j, b_j]$, then we need to show that $\mu(I) = \sum_{j=1}^{\infty} \mu(I_j)$.

The $\geqslant$ direction is easy: it suffices to show $\mu(I) \geqslant \sum_{j=1}^{n} \mu(I_j)$ and take limits in n. Note that $I - \bigcup_{j=1}^{n} I_j \in \mathcal{A}$, so finite additivity of μ implies

$$\mu(\bigcup_{i=1}^{n} 9_j) = \sum_{j=1}^{n} \mu(I_j) = \mu(I) - \mu(I - \bigcup_{j=1}^{n} I_j) \leqslant \mu(I).$$

The $\leqslant$ direction is harder. First suppose that $(a, b]$ is bounded. Let $\epsilon > 0$ be given.
Idea: we want to modify $\{I_i\}$ *to open covers that cover a compact set. To this end, we shrink* $(a, b]$ *to* $[a+\delta, b]$ *and* $(a_i, b_i + \delta_i)$ *and therefore obtain a finite subcover satisfying* $\mu(I) \leqslant \sum_{i=1}^{n} \mu(I_i) + \epsilon$ *for arbitrary* $\epsilon > 0$. For detailed construction, refer to Folland's book. □

Beginning of Sept.15, 2021

Remark. A follow up on the previous F: for $a > 0$ and any $a_n \uparrow a$, we have

$$\mu((0,a)) = \mu(\bigcup_{n\geqslant 1} (0, a_n]) = \lim_{n\to\infty} \mu((0, a_n]) = \lim_{n\to\infty} F(a_n)$$

so $\mu((0,a)) = F(a^-)$. (A similar argument holds for $a < 0$). Hence $F(a) - F(a^-)$ is the jump in F at a, and

$$\mu(\{a\}) = \mu((0,a]) - \mu((0,a)).$$

Remark. Back to the latest proposition — since μ is σ-finite, it can be *uniquely* extended to a measure on the Borel sets!

Theorem 1.3.3: Correspondence Between Measures and Distribution Functions

Given $F : \mathbb{R} \to \mathbb{R}$ that is increasing and right-continuous, there is a *unique* measure, which we call μ_F, on $\mathcal{B}_{\mathbb{R}}$ with the property stated in the previous proposition:

$$\mu((a,b]) = F(b) - F(a) \qquad \text{for all } a < b.$$

In addition, observe that $\mu_F = \mu_G$ if and only if $F - G \equiv$ a constant.

Conversely, given μ on $\mathcal{B}_{\mathbb{R}}$ finite on bounded sets, there is an increasing right-continuous F with $\mu = \mu_F$.

Definition 1.3.4: Lebesgue-Stieltjes Measure & Lebesgue Measure (on $\mathbb{R}$)

Recall we can completed the above μ_F to a unique complete measure (also called μ_F). This measure is called the **Lebesgue-Stieltjes measure** associated to F. Its domain is

$$\{D \cup E : D \text{ Borel}, E \subset \text{some null Borel set}\}.$$

(What is a null set depends on F.)
If $F(x) = x$, we obtain the **Lebesgue measure** on $\mathbb{R}$.

Example 1.3.5. Define

$$F(x) = \begin{cases} x & \text{if } x < 2 \\ x + 3 & \text{if } x \geqslant 2. \end{cases}$$

Let m be the Lebesgue measure. Then $\mu_F = m$ on $(-\infty, 2) \cup (2, \infty)$, and $\mu_F(E) = m(E) + 3$ if $2 \in E$.

In general, if $\mu(\mathbb{R}) < \infty$, we can write $F(a) = \mu((-\infty, a])$. Then $\lim_{a\to-\infty} F(a) = 0$ and $\lim_{a\to\infty} F(a) = \mu(\mathbb{R})$.

Lemma 1.3.6

Let μ be a σ-finite premeasure on an algebra $\mathcal{A}$, μ^* the outer measure (obtained by infimum covering), and

$$\mathfrak{M}^* = \{\text{all } \mu^*\text{-measurable sets}\}, \overline{\mathfrak{M}} = \{D \cup E : D \in \sigma(\mathcal{A}), E \subset \text{null}\}$$

(i.e., $\overline{\mathfrak{M}}$ is the domain of completion). Then $\mathfrak{M}^* = \overline{\mathfrak{M}}$, and $\mu^*|_{\mathfrak{M}^*}$ is the completion of μ.

Proof. Carathéodory's theorem implies that $\mu^*|_{\mathfrak{M}^*}$ is complete, so the domain must include $D \cup E$ and we must have $\mu^*(D \cup E) = \mu^*(D)$. Hene $\overline{\mathfrak{M}} \subset \mathfrak{M}^*$. The other inclusion is a Folland exercise. ☐

For the special case $\mu = \mu_F$ on $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, μ_F extends to $\overline{\mathfrak{M}} = \mathfrak{M}^*$, with

$$\mu_F(E) = \mu_F^*(E) = \inf\{\sum_{i=1}^{\infty} \mu_F((a_i, b_i]) : E \subset \bigcup_{i=1}^{\infty}(a_i, b_i]\}.$$

Sometimes we want to cover with *open* sets rather than h-intervals. Does it work if we replace $(a_i, b_i]$ by (a_i, b_i)?

On one hand, every open interval can be expressed as a countable union of h-intervals, so clearly

$$\inf\{\sum_{i=1}^{\infty} \mu_F((a_i, b_i]) : E \subset \bigcup_{i=1}^{\infty}(a_i, b_i]\} \leqslant \inf\{\sum_{i=1}^{\infty} \mu_F((a_i, b_i)) : E \subset \bigcup_{i=1}^{\infty}(a_i, b_i)\}.$$

The converse is also true. Given $E \subset \bigcup_{i=1}^{\infty}(a_i, b_i]$, since F is right-continuous, for each i there exists a δ_i such that $F(b + \delta_i) - F(b_i) < \epsilon 2^{-i}$. Then we can always find "ϵ-optimal" covers $\{(a_i, b_i + \delta_i)\}$, from which the other direction of inequality follows.

Remark. From the derivation above, given $E \subset \mathbb{R}$ and $\epsilon > 0$, we can approximate E by an open $G \supset E$ where $\mu(G - E) < \epsilon$. Hence we can approximate general Borel sets by open sets.

Beginning of Sept.17, 2021

1.4 Measure-Topology Connections

Lemma 1.4.1

In $\mathbb{R}^k$, let μ be a finite measure on $\mathcal{B}_{\mathbb{R}^k}$. Let E be closed and G open. Then

$$\mu(E) = \inf\{\mu(U) : U \supset E \text{ and } U \text{ open}\}, \tag{1}$$

and

$$\mu(G) = \sup\{\mu(K) : K \subset E \text{ and } K \text{ compact}\}. \tag{2}$$

If a Borel measure satisfies the equalities above, it is called **regular**.

Proof. Clearly $\leqslant$ in (1) and $\geqslant$ in (2) hold. To prove = in (1), we consider the " $1/n$ neighborhoods"

$$U_n := \{x : d(x,y) < 1/n \text{ for some } y \in E\}.$$

Taking the intersection of all U_n's we obtain equality. Since $U_1 \supset U_2 \supset \dots$ and μ is finite, $\bigcap_{n=1}^{\infty} U_n = E$ as E is closed, so $\lim \mu(U_n) = \mu(E)$.
To prove = in (2), let $x_0 \in G$ and $K_n := \{x : d(x,y) \geqslant 1/n \text{ for all } y \in G^c\} \cap \overline{B(x_0,n)}$ (i.e., a bonded set whose distance at least $1/n$ from the complement). Then the union of these K_n's is the original G: Since $K_1 \subset K_2 \subset \dots$, we have $\bigcup_{n=1}^{\infty} K_n = G$, so $\lim \mu(K_n) = \mu(G)$. □

Theorem 1.4.2

If μ is a finite measure on $\mathcal{B}_{\mathbb{R}^k}$ completed to $\mathfrak{M} : \{\text{all } \mu^* - \text{measurable sets}\}$, then *for all* $E \in \mathfrak{M}$,

(1) $\mu(E) = \inf\{\mu(U) : U \supset E, U \text{ open}\}$, and

(2) $\mu(E) = \sup\{\mu(K) : K \subset E, K \text{ compact}\}$.

Proof. Let $\mathcal{E}$ be the collection of Borel sets that satisfy (1) and (2). The lemma implies open sets satisfy (2), whereas for (1) we simply take $U = E$. Thus open sets meet both criteria. By the General Principle we want to show that $\mathcal{E}$ is a σ-algebra.
Let $E_1, E_2, \dots \in \mathcal{E}$. We want to show that their union is in $\mathcal{E}$. Indeed, let $\epsilon > 0$. Since $E_n \in \mathcal{E}$, there exists $U_n, K_n \subset E$ (open and compact, respectively) with $\mu(U_n) \leqslant \mu(E_n) + \epsilon/2^n$ and $\mu(K_n) \geqslant \mu(E_n) - \epsilon/2^n$.
(*Note that the union of finitely many compact sets is compact, but this does not hold for countable unions in general.*)
Note that

$$\mu(\bigcup_{n=1}^{N} K_n) \geqslant \mu(\bigcup_{n=1}^{N} E_n) - \sum_{n=1}^{N} \frac{\epsilon}{2^n}.$$

Using continuity from below, we can choose N sufficiently large such that $\mu(\bigcup_{n=1}^{\infty} E_n) - \mu(\bigcup_{n=1}^{N} E_n) < \epsilon$. Then the LHS above is $\leqslant 2\epsilon$ away from $\bigcup_{n=1}^{\infty} E_n$, so (2) holds for this set. For (1),

$$a\mu(\bigcup_{n=1}^{\infty} U_n) \leqslant \mu(\bigcup_{n=1}^{\infty} E_n) + \mu(\bigcup_{n=1}^{\infty} U_n - E_n) \leqslant \mu(\bigcup_{n=1}^{\infty} E_n) + \epsilon.$$

Therefore $\mathcal{E}$ is closed under countable unions. To check that it is closed under complementation, since μ is finite,

$$\begin{aligned}\mu(E^c) &= \mu(\mathbb{R}^k) - \mu(E) = \mu(\mathbb{R}^k) - \inf\{\mu(U) : U \supset E, U \text{ open}\}\\ &= \sup\{\mu(\mathbb{R}^k) - \mu(U) : U \subset E, U \text{ open}\}\\ &= \sup\{\mu(F) : F \subset E^c, F \text{ closed}\}.\end{aligned}$$

Finally, to upgrade from closed sets to compact sets, we need to intersect with "large closed balls", which would give us compact $K \subset E^c$ with measure close to $\mu(E^c)$. Showing E^c has (1) is much easier as the complement of a compact set is an open set – no nuisance. So $\mathcal{E}$ is indeed a σ-algebra.

Since $\mathcal{E}$ is a σ-algebra and it contains all open sets, it contains all Borels. Let

$$\mathfrak{M} = \{E \cup F : E \text{ Borel}, F \subset \text{ null Borel}\}.$$

Let $\epsilon > 0$ and $F \subset N$ null. We know that $E, N \in \mathcal{E}$, so there exists K compact and U open such that

$$K \subset E \subset E \cup F \subset E \cup N \subset U$$

with

$$\mu(K) > \mu(E) - \epsilon \qquad \mu(U) < \mu(E \cup N) + \epsilon.$$

But then

$$\mu(K) > \mu(E) - \epsilon = \mu(E \cup F) - \epsilon \qquad \mu(U) < \mu(E \cup N) + \epsilon = \mu(E \cup F) + \epsilon,$$

which shows that $\mathfrak{M} \subset \mathcal{E}$. □

Remark. If there exist $\{K_n\}$ (compact) and $\{U_n\}$ (open) that approximate E by $1/n$, then

$$\mu(\bigcup_{n=1}^{\infty} K_n) = \mu(E) = \mu(\bigcap_{n=1}^{\infty} U_n)$$

where the countable union is a F_σ set and the intersection a G_δ set. This shows that if μ, ν are regular and $\mu = \nu$ on open sets, then $\mu = \nu$ on Borels.

1.5 Lebesgue Measure

Definition 1.5.1: Lebesgue Measure

Lebesgue measure m on $\mathbb{R}$ is the complete measure associated to the distribution function $F(x) = x$. Its domain $\mathcal{L}$ is the completion of $\mathcal{B}_{\mathbb{R}}$ for μ_F.

Example 1.5.2. Recall that the Cantor set $\mathcal{C}$ is the set of all $x \in [0, 1]$ with ternary expansion consisting of only 0's and 2's. The *Cantor-Lebesgue function* is a bijection $F : \mathcal{C} \to [0, 1]$ by replacing 2's by 1's in binary.

We can extend the domain of F to $[0,1]$: for example

$$F'(.0220120021...) := F(02200222...).$$

This gives the *Devil's staircase*; F' is constant on removed intervals in $\mathcal{C}$.

Beginning of Sept.20, 2021

Note that F is bijective except for conutably many points, for example $F(1/3) = F(2/3) = 1/2$. We can also construct a "pseudo-inverse" $F : [0,1] \to \mathcal{C}$ by

$$F^{-1}(y) := \inf\{x : F(x) = y\},$$

so for example $F^{-1}(1/2) = 1/3$, not anything in $(1/3, 2/3]$. It follows that

$$F^{-1}(y) = x \iff y \leqslant F(x). \tag{*}$$

Since F is nondecreasing and has no jump discontinuity, F is continuous. Furthermore, F is differentiable (with derivative 0) except on a null set $\mathcal{C}$ (i.e., $F'(x) = 0$ **almost everywhere**). However, F is not continuous: $F(0) = 0$ and $F(1) = 1$. Then $\mu_F([0,1] - \mathcal{C}) = 0$, $\mu_F([0,1]) = 1$, and so $\mu_F(\mathcal{C}) = 1$. We say μ_F "lives on" $\mathcal{C}$.

Let $(X, \mathfrak{M}, \mu)$ be a measure space and $(Y, \mathfrak{G})$ be another measurable space (with or without a measure). Then $f : X \to Y$ gives a measure ν on $(Y, \mathfrak{G})$ by

$$\nu(E) = \nu(\{x \in X : f(x) \in E\}) = \nu(f^{-1}(E)).$$

(Note we require $f^{-1}(E) \in \mathfrak{M}$ for all $E \in \mathfrak{G}$.)

Example 1.5.3. Let $F : \mathcal{C} \to [0,1]$ and $F^{-1} : [0,1] \to \mathcal{C}$ be defined as above. Consider $([0,1], \mathcal{B}, m)$ where m is the Lebesgue number be the domain of F^{-1} and $([0,1], \mathcal{B})$ be the range. Now we analyze the push-forward of m by F^{-1}: for $x \in \mathcal{C}$, we define

$$m(\{y : F^{-1}(y) \in (0,x]\}) = m(\{y : y \leqslant F(x)\}) = m([0, F(x)]) = F(x).$$

Thus the push-forward for $x \in \mathcal{C}$ is just μ_F. Since F is constant on $[0,1] - \mathcal{C}$ so $\mu_F(\mathcal{C}^c) = 0$, this is also true for $x \in [0,1] - \mathcal{C}$.

In this sense, μ_F is the "uniform measure" on $\mathcal{C}$.

Chapter 2

Integration

2.1 Measurable Functions

Definition 2.1.1: Measurable Function

Let $(X, \mathfrak{M})$ and $(Y, \mathfrak{N})$ be measurable spaces. A function $f : X \to Y$ is said to be **$(\mathfrak{M}, \mathfrak{N})$-measurable** if

$$f^{-1}(E) \in \mathfrak{M} \qquad \text{for all } E \in \mathfrak{N}.$$

Recall that inverse images commute with set operations:

$$f^{-1}(E \cup F) = f^{-1}(E) \cup f^{-1}(F),$$

and

$$f^{-1}(E \cap F) = f^{-1}(E) \cap f^{-1}(F).$$

The unions and intersections can be finite, countable, or uncountable. Therefore,

$$\{E \in Y : f^{-1}(E) \in \mathfrak{M}\}$$

is a σ-algebra. In particular, if the set contains some collection $\mathcal{E}$ then it contains $\sigma(\mathcal{E})$.

Proposition 2.1.2

If $\mathcal{E}$ generates $\mathfrak{N}$ and $f^{-1}(E) \in \mathfrak{m}$ for every $E \in \mathcal{E}$, then f is $(\mathfrak{M}, \mathfrak{N})$-measurable.

Example 2.1.3. If $(Y, \mathfrak{N}) = (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, $\mathcal{E} := \{(-\infty, t] : t \in \mathbb{R}\}$, $(X, \mathfrak{M})$ is another space, then $f : X \to \mathbb{R}$ is $(\mathfrak{M}, \mathcal{B}_{\mathbb{R}})$-measurable if and only if

$$\{x : f(x) \leqslant t\} \in \mathfrak{M} \text{ for all } t \in \mathbb{R}.$$

(Since we can also take $\mathcal{E} := \{(-\infty, t) : t \in \mathbb{R}\}$, we can replace $\leqslant$ in the above inequality by $<$; $\geqslant$ and $>$ would also work.)

Example 2.1.4. Let $(Y, \mathcal{B}_Y)$ be generated by $\mathcal{E} :=$ open sets in $Y\}$. Then $f : X \to Y$ (where $X = (X, \mathcal{B}_X)$) is continuous if and only if $f^{-1}(U)$ is open for $U \subset \mathcal{E}$. If so, since $f^{-1}(U)$ is Borel for all $U \in \mathcal{E}$, we see that f is **Borel measurable.**

Beginning of Sept.22, 2021

We chose inverse images because forward images are not so "well-behaved". For example, consider $F : \mathcal{C} \to [0,1]$ (the Cantor function discussed earlier) and the Vitali (nonmeasurable) set A discussed in the beginning of this course. Consider $F^{-1}(A)$. The Lebesgue measure of C is 0, and since it is complete, we see that $F^{-1}(A)$ should be Lebesgue measurable and null. This shows that the image of a measurable set may become nonmeasurable.

If f, g are measurable, what about $f \circ g$, $f + g$, fg, etc.? What about $f_n \to f$?

Example 2.1.5: Composition of measurable function is measurable. Let $g : (X, \mathfrak{M}) \to (Y, \mathfrak{N})$ and $f : (Y, \mathfrak{N}) \to (Z, \mathfrak{Q})$ be measurable functions. Immediately we see by definition that if $E \in \mathfrak{Q}$ then

$$(f \circ q)^{-1}(E) = g^{-1}(\underbrace{f^{-1}(E)}_{\in \mathfrak{N}}) \in \mathfrak{M},$$

which means $f \circ g$ is measurable.

Example 2.1.6: Product of measurable functions. Consider $f : X \to \prod_{\alpha \in A} Y_\alpha$ where $Y_\alpha = (Y_\alpha, \mathfrak{N}_\alpha)$ are measurable spaces, and let $f_\alpha := \pi_\alpha \circ f$, the α^{th} coordinate function.
Claim: f is measurable into $\bigotimes_{\alpha \in A} \mathfrak{N}_\alpha$ if and only if f_α is measurable into $\mathfrak{N}_\alpha$ for each α.

Proof. Recall that $\bigotimes_{\alpha \in A} \mathfrak{N}_\alpha$ is generated by $\{\pi_\alpha^{-1}(E), E \in \mathfrak{N}_\alpha, \alpha \in A\}$, so each π_α inverse images of the generators and are hence measurable.
Therefore, if f is continuous, f_α, a composition of measurable functions, is measurable.
Conversely, suppose each f_α is measurable. To show f is measurable, it suffices to show that $f^{-1}(\pi_\alpha^{-1}(E))$ is measurable for all generator $\pi_\alpha^{-1}(E)$'s. However, this is nothing but $f_\alpha^{-1}(E)$ so it is indeed measurable. □

Proposition 2.1.7

If $f, g : X \to \mathbb{R}$ (or $\mathbb{C}$) are measurable, then so are $f + g$ and fg.

Proof. Write $x \mapsto f(x)g(x)$ as $x \mapsto (f(x), g(x)) \mapsto f(x) = g(x)$. Since $uv \mapsto u + v$ is continuous and $x \mapsto (f(x), g(x))$ is measurale, by definition we obtain measurability of $f + g$. Likewise for fg. □

2.2 Taking Limits

For $\mathbb{R}$, we allow finite values as well as $\pm\infty$, i.e., we use the extended real line system $\overline{\mathbb{R}}$. Recall that for a sequence $\{x_n\} \subset \mathbb{R}$, we have

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} (\sup_{k\geqslant n} x_k) = \inf_{n\to\infty} (\sup_{k\geqslant n} x_k)$$

and

$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} (\sup_{k\geqslant n} x_k) = \sup_{n\to\infty} (\sup_{k\geqslant n} x_k).$$

From this we have

$$\limsup_{n\to\infty} x_n < x \text{ if and only if } \sup_{k\geqslant n} x_k < t \text{ for some } n.$$

Also recall that

$$\limsup_{n\to\infty} x_n = \text{ largest subsequential limit}$$

(and parallel for $\liminf$).

For a sequence of functions $\{f_n\}$, we define $\limsup f_n$ to take the *pointwise* limit superior, i.e.,

$$f := \limsup_{n\to\infty} f_n \text{ by } f(x) := \limsup_{n\to\infty} f_n(x).$$

Finally, two equations that will be helpful later on:

$$\{x : \sup_n f_n(x) > t\} = \bigcup_{n=1}^{\infty} \{x : f_n(x) > t\} \tag{1}$$

(note the *strict* inequality), and

$$\{x : \inf_n f_n(x) < t\} = \bigcup_{n=1}^{\infty} \{x : f_n(x) < t\}. \tag{2}$$

Proposition 2.2.1

Suppose $\{f_n\}$ are measurable (extended) real-valued functions. Then

$$\sup_n f_n \quad \inf_n f_n \quad \limsup_n f_n \quad \liminf_n f_n$$

are all measurable.

Proof. Sup follows from (1), inf from (2). Using the definition $\limsup = \inf \sup$, the third claim follows from first using (1) and then using (2), and likewise the last one follows from using (2) and then (1). □

Corollary 2.2.2

If $\{f_n\}$ is a sequence of measurable $\mathbb{C}$-valued functions and $f_n \to f$ pointwise, then f is measurable. (Check real and complex parts separately.)

2.3 Integration

We begin by examining integration of nonnegative functions. To this end, we decompose f into nonnegative functions

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

It follows immediately that $f = f^+ = f^-$.

Definition 2.3.1: Simple Functions

A **simple function** is a (complex valued) function that takes only *finitely* any values. In particular, we can write a simple function as a linear combination of indicator functions:

$$f = \sum_{i=1}^{n} c_i \chi_{E_i}.$$

Beginning of Sept.24, 2021

Remark. A **standard representation** means that the E_i's are pairwise disjoint, so $E_i = f^{-1}(c_i)$.

Theorem 2.3.2: Approximation by Simple Functions

(1) If $f : X \to [0, \infty]$ is measurable, then there exist measurable functions $0 \leqslant \varphi_n \uparrow f$ uniformly on any set where f is bounded.

(2) If $f : X \to \mathbb{C}$, a similar claim holds with value replaced by magnitude.

Proof. For φ_n, we divide the y-axis into small intervals of length $1/2^n$, and we define a "2^{-n} floor function" by

$$\varphi_n(x) := \begin{cases} k/2^n & \text{if } k/2^n < f(x) \leqslant (k+1)/2^n \text{ and } f(x) \leqslant 2^n \\ 2^n & \text{if } f(x) > 2^n \\ 0 & \text{if } f(x) = 0. \end{cases}$$

Intuitively, as n gets large, the error in approximation is $< 1/2^n$ which tends to 0. It is also clear that $\{\varphi_n\}$ is an increasing sequence.

For the complex case, apply the real case to $\mathfrak{Re} f$ and $\mathfrak{Im} f$ separately. □

Example 2.3.3. If f_n is measurable and $f_n \to f$ pointwise, then f is measurable. However, almost everywhere convergence does not preserve this property: let $\mathcal{C}$ be the Cantor set and let $E \subset \mathcal{C}$ be non-Borel. Let $f_n \equiv 1$ for all n and let $f(x)\chi_{E^c}$ (so it is 1 almost everywhere). Then $f_n \to f$ a.e. but f is not Borel-measurable. (It is Lebesgue measurable, though, as the Lebesgue measure is complete.)

Proposition 2.3.4

Suppose μ is complete, all f_n'are measurable ($\mathbb{R}$- or $\mathbb{C}$-valued), and $f_n \to f$ a.e. (μ), then f is measurable.

Proof. Define $E := \{x : f_n(x) \to f(x)\}$. By assumption E^c is contained is a null set (if μ is incomplete, it is *contained* in a null set). We instead consider the functions $\tilde{f}_n, \tilde{f}$ where $\tilde{f}_n \equiv f_n$, $\tilde{f} \equiv f$ on E and 0 on E^c, that is,

$$\tilde{f}_n = \tilde{f}_n \chi_E \qquad \tilde{f} = f\chi_E.$$

Then $\tilde{f}_n \to \tilde{f}$ pointwise[!] so $\tilde{f}$ is measurable. Since $\tilde{f} = f$ a.e., we have f measurable too. ☐

Lebesgue Integeration

Consider $(X, \mathfrak{M}, \mu)$. We define $\mathcal{L}^+ := \{$all measurable $f : X \to [0, \infty]\}$. For a simple function φ with standard representation, we simply define

$$\int \varphi \, \mathrm{d}\mu := \sum_{i=1}^{n} a_i \mu(E_i).$$

(This is basically the same as Riemann integration: sum of value times size.)

Q: does this integral depend on the representation of φ?

A: if E_j's are disjoint, in each representation, we can group together disjoint pieces with the same φ value (hence the horizontal "slicing"). Hence the answer is no.

If $A \subset X$ and $\varphi = \sum_{i=1}^{n} a_i \chi_{E_i}$ simple, then we define

$$\int_A \varphi \, \mathrm{d}\mu := \int_X \varphi \chi_A \, \mathrm{d}\mu = \sum_{i=1}^{n} a_i \mu(E_i \cap A).$$

(so we "force" φ to be 0 outside A).

Proposition 2.3.5: Properties of the Lebesgue Integral

(1) Integrals can take infinite value (as we are only dealing with addition of integral of positive simple functions).

(2) (Linearity) $\int c\varphi \, \mathrm{d}\mu = c \int \varphi \, \mathrm{d}\mu$ for $c \geqslant 0$,

(3) (Linearity) $\int (\varphi + \psi) \, \mathrm{d}\mu = \int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\mu$.

(4) (Monotonicity) If $\varphi \leqslant \psi$ then $\int \varphi \, \mathrm{d}\mu \leqslant \int \psi \, \mathrm{d}\mu$.

(5) The mapping $A \mapsto \int_A \varphi \, \mathrm{d}\mu$ defines a measure on A.

For (2), (3), (4), simply "divide" the corresponding $\{E_i\}, \{F_i\}$ into a collection of smaller sets.

Beginning of Sept.27, 2021

Proof. For (5), the only nontrivial claim is countable additivity. If $A_1, A_2, \ldots$ are disjoint then

$$\begin{aligned}
\int_{\bigcup A_n} \sum_{j=1}^{k} a_j \chi_{E_j} \, \mathrm{d}\mu &= \sum_{j=1}^{k} a_j \mu(E_j \cap (\bigcup_{n=1}^{\infty} A_n)) \\
&= \sum_{j=1}^{k} a_j \sum_{n=1}^{\infty} \mu(E_j \cap A_n) \\
&= \sum_{n=1}^{\infty} \sum_{j=1}^{k} a_j \mu(E_j \cap A_n) = \sum_{n=1}^{\infty} \int_{A_n} \varphi \, \mathrm{d}\mu.
\end{aligned}$$

□

Remark. Since $A \mapsto \int_A \varphi \, d\mu$ defines a measure, continuity from below states that, if $A_1 \subset A_2 \subset \ldots$ and $A = \bigcup_{n=1}^{\infty} A_n$, then

$$\int_{A_n} \varphi \, d\mu \to \int_A \varphi \, d\mu.$$

Defining Lebesgue Integral on General (Nonnegative) Functions

For $f \geqslant 0$, we are tempted to define $\int f \, d\mu$ by choosing a sequence $\varphi_n \uparrow f$ and let $\int f \, d\mu = \lim_{n\to\infty} \int \varphi_n \, d\mu$. However, this might depend on a particular sequence of φ_n. A better definition uses

$$\int f \, d\mu = \sup\{\int \varphi \, d\mu : 0 \leqslant \varphi \leqslant f, \varphi \text{ simple}\}.$$

(Note that this is well-defined for simple functions too.)

Basic question. If $f_n \to f$ a.e., does the integral converge?
Answer. No. Let $f_n := n^2 \chi_{[0,1/n]}$. They converge (almost) everywhere on $(0, 1]$ whereas the integral $\to \infty$. Alternatively, consider $f_n := 1/n$ on $\mathbb{R}$: pointwise converging to 0 but integral is always ∞. We've given an example of "*large values on small sets*" and an example of "*small values on large sets*". Without balance, issues like this arise. This question, however, has a positive answer for *monoetone* sequences of functions.

We define $\mathcal{L}^+ := \{\text{all measurable functions } f \geqslant 0\}$.

Theorem 2.3.6: Monotone Convergence Theorem (MCT)

Suppose $f_1 \leqslant f_2 \leqslant \ldots$ with $f_n \in \mathcal{L}^+$ and f is the pointwise limit of f_n. Then

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

Proof. The $\geqslant$ direction is clear as $\int f \geqslant \int f_n$ for all n. For $\leqslant$, let $\epsilon > 0$ be given and we want to show that $\lim_{n\to\infty} \int f_n \geqslant (1-\epsilon) \int f$. For this, we show that $\lim_{n\to\infty} \int f_n \geqslant \int (1-\epsilon)\varphi$ for every simple $\varphi \leqslant f$. (*Idea: for large n we have $f_n \geqslant (1-\epsilon)\varphi$, but we can't just say this because the convergence is not uniform.*)
Let $\varphi \leqslant f$ be given. Let $E_n := \{x \in X : f_n(x) \geqslant (1-\epsilon)\varphi(x)\}$. By pointwise convergence, for each n, there exist n_x such that

$$n \geqslant n_x \implies f_n(x) \geqslant (1-\epsilon)\varphi(x) \implies x \in E_n.$$

Hence $E_1 \subset E_2 \subset \ldots$ and $\bigcup_{n=1}^{\infty} E_n = X$. Also,

$$\int f_n \geqslant \int_{E_n} f_n \geqslant \int_{E_n} (1-\epsilon)\varphi,$$

so taking the limit of E_n (recall that $E_n \mapsto \int_{E_n} \varphi$ defines a measure!) gives

$$\lim_{n\to\infty} \int f_n \geqslant \lim_{n\to\infty} \int_{E_n} (1-\epsilon)\varphi = \int (1-\epsilon)\varphi \qquad \text{for all simple } \varphi \leqslant f.$$

Letting $\epsilon \to 0$, we obtain our claim. □

Definition 2.3.7

For a general real-valued f (not $f \geqslant 0$), recall that $f = f^+ - f^-$. We define

$$\int f \, \mathrm{d}\mu := \int f^+ \, \mathrm{d}\mu - \int f^- \, \mathrm{d}\mu$$

(the integral is not defined if the above subtraction takes form $\infty - \infty$). If both are finite (so $\int f$ is finite), we say f is **integrable**. Note that since $|f| = f^+ + f^-$, being integrable is equivalent to $\int |f| \, \mathrm{d}\mu < \infty$.
If f is $\mathbb{C}$-valued, when both exist, we define

$$\int f \, \mathrm{d}\mu := \int \mathfrak{Re} f \, \mathrm{d}\mu + i \int \mathfrak{Im} f \, \mathrm{d}\mu.$$

Also note that saying this is finite is equivalent to saying $\int |f| \, \mathrm{d}\mu < \infty$. To this end, we define

$$L^1 := \{\text{all } \mathbb{C} - \text{valued integral functions } f \text{ on } (X, \mathfrak{M}, \mu)\}.$$

Proposition 2.3.8

For $f \in L^1$, $\left| \int f \mathrm{d}\mu \right| \leqslant \int |f| \mathrm{d}\mu$.

Proof. For $\mathbb{R}$-valued f,

$$\left| \int f \right| = \left| \int f^+ - \int f^+ \right| \leqslant \int f^+ + \int f^- = \int |f|.$$

For a $\mathbb{C}$-valued function, we "rotate" f to make it $\mathbb{R}$-valued: choose θ such that

$$\int f = e^{-i\theta} \left| \int f \right|$$

so $\int e^{i\theta} f$ is real. Therefore,

$$\left| \int f \right| = \left| \int e^{i\theta} f \right| = \left| \int (e^{i\theta} f) \right| \overset{*}{\leqslant} \int |\mathfrak{Re}(e^{i\theta} f)| \leqslant \int |e^{i\theta} f| = \int |f|,$$

where $(*)$ is by the first part of this proof. ☐

Beginning of Sept.27, 2021

Remark: MCT on downward convergence. What if $f_n \downarrow f$? If $g \geqslant f$ then we write $g - f_n \uparrow g - f$. If g *is integrable* (i.e., the following does not take the form $\infty - \infty$) then the MCT gives

$$\int (g - f_n) \to \int (g - f) \implies \int g - \int f_n \to \int g - \int f \implies \int f_n \to \int f.$$

However, if g is not integrable, this claim fails: consider $f_n(x) := 1/n$ on $\mathbb{R}$.

Corollary 2.3.9

If $f_n \geqslant 0$ then $\sum_{i=1}^{k} f_n \uparrow \sum_{n=1}^{\infty} f_n$. If each $f_n \in \mathcal{L}^+$ and $f := \sum_{n=1}^{\infty} f_n$, then $\int f = \sum_{n=1}^{\infty} \int f_n$.

Proof. First we verify finite additivity: for f_1, f_2 we have sequences of nonnegative simple functions $\varphi_n \uparrow f_1$ and $\psi_n \uparrow f_2$. By linearity of integral for simple functions,

$$\int (\varphi_n + \psi n) = \int \varphi_n + \int \psi_n.$$

Letting $n \to \infty$ we see $\int (f_1 + f_2) = \int f_1 + \int f_2$. The rest of the claim follows from MC on $\sum_{n=1}^{k} f_n$. □

Proposition 2.3.10

For $f \in L^+$, the integral $\int f = 0$ if and only if $f = 0$ almost everywhere.

Proof. For $\Leftarrow$, notice that for simple functions $\varphi = \sum_{i=1}^{n} a_i \chi_{E_i} = 0$ a.e. We have $m(E_i) = 0$ for all i so $\int \varphi = 0$. For general f with $f = 0$ a.e., since any simple function below it has integral 0, $\int f = 0$ as well.

For $\Rightarrow$, notice that ($f = 0$ almost everywhere) $\Leftrightarrow$ ($\mu(\{x : f(x) > 1/n\}) = 0$ for all n). Let E_n be the set corresponding to $1/n$. Since $f \geqslant 1/n \cdot \chi_{E_n}$, $\int f \geqslant \mu(E_n)/n$. Therefore $\mu(E_n) = 0$ for all n, which shows that $f = 0$ almost everywhere. □

Remark. By the same token, in general if $f \geqslant 0$ and $\int f < \infty$, then $\mu(E_n) < \infty$ for all n.

Corollary 2.3.11

If $\int_E f = \int_E g$ for all E (and f, g are integrable), then $f = g$ a.e.

Proof. Consider $\{x : f(x) > g(x)\} = (f - g)^{-1}((0, \infty))$, a measurable set. Let it be E. If $f = g$ a.e. then $f - g = 0$ a.e., so $\int_E (f - g) = 0$ for all E, and so $\int_E f = \int_E g$ a.e. □

Theorem 2.3.12: Fatou's Lemma

For $\{f_n\} \subset \mathcal{L}^+$,

$$\int \liminf_{n\to\infty} f_n \leqslant \liminf_{n\to\infty} \int f_n.$$

Proof. Notice that $\inf_{n \geqslant k} f_n$ monotonically increases as k increases and its limit is $\liminf_{n\to\infty} f_n(x)$. By MCT

$$\int \liminf_{n\to\infty} f_n = \lim_{k\to\infty} \int \inf_{n\geqslant k} f_n = \liminf_{k\to\infty} \int \inf_{n\to k} f_n \leqslant \liminf_{k\to\infty} \int f_k.$$

□

Proposition 2.3.13

(a) For $f \in L^1$, $\{x : f(x) \neq 0\}$ is σ-finite.

(b) For $f, g \in L^1$,

$$\int_E f = \int_E g \text{ for all } E \in \mathfrak{M} \Leftrightarrow \int |f - g| = 0 \Leftrightarrow f = g \text{ a.e.}$$

Proof.

(a) We showed that $\{x : f(x) \neq 0\} = \bigcup_{n=1}^{\infty} \{x : |f(x)| > 1/n\}$ where each "$1/n$" set (call it E_n) satisfies

$$\frac{1}{n}\mu(E_n) \leqslant \int |f| < \infty.$$

(b) We've shown $\int |f - g| = 0$ if and only if $|f - g| = 0$ a.e., i.e., $f = g$ a.e. For the first $\Leftrightarrow$, the $\Leftarrow$ is given by

$$\left| \int_E f - \int_E g \right| = \left| \int (f - g)\chi_E \right| \leqslant \int |f - g|\chi_E \leqslant \int |f - g|.$$

Conversely, suppose $\int_E f = \int_E g$ for all $E \in \mathfrak{M}$. For real-valued functions take $E := \{x : f(x) > g(x)\}$. Then

$$0 = \int_E f - \int_E g = \int \underbrace{(f - g)\chi_E}_{\geqslant 0},$$

so $(f - g)\chi_E = 0$ almost everywhere. A similar argument follows for $\{x : f(x) < g(x)\}$. The claim follows.

For $\mathbb{C}$-valued functions, we apply the same argument to $\mathfrak{Re}(f - g)$ and $\mathfrak{Im}(f - g)$.

□

Definition 2.3.14: L^1 Norm

Note that $\int f$ makes sense for f even if f is undefined on a null set. Define

$$\rho(f, g) := \int |f - g| \qquad \text{for } f, g \in L^1$$

the "L^1 distance". (*Note that this is a* ***pseudometric*** *since* $\rho(f, g) = 0$ *only implies* $f = g$ *a.e.*) To this end, we define an equivalence relation $f \sim g$ if $f = g$ a.e. Then ρ is a metric on these equivalence classes.

Note that for indicators, $|\chi_E - \chi_F| = \chi_{E\Delta F}$, so

$$\mu(E\Delta F) = \int |\chi_E - \chi_F| = \rho(\chi_E, \chi_F)$$

also gives a (pseudo)metric on sets. This verifies one of the HW2 problems.

Beginning of Oct.1, 2021

Example 2.3.15: Fatou for limsup. Fatou's lemma states that $\liminf \int f_n \geqslant \int \liminf f_n$. For the limsup argument, *if there exists* g *integrable with* $g \geqslant f_n$ *for all* n,

$$g(x) - \limsup_{n\to\infty} f_n(x) = \liminf_{n\to\infty}(g(x) - f(x))$$

so in this case

$$\int \liminf(g(x) - f(x)) \leqslant \liminf \int (f_n - g_n)$$

and thus $\int \limsup f_n \geqslant \limsup \int f_n$.

Example 2.3.16. If $f_n := \chi_{E_n}$ then

$$\limsup_{n\to\infty} \chi_{E_n}(x) = \begin{cases} 1 & \text{if } x \in E_n \text{ for infinitely many } n \\ 0 & \text{if not.} \end{cases}$$

Hence $\limsup_{n\to\infty} E_n = \{x : x \in E_n \text{ for infinitely many } E_n\text{'s}\}$. The limsup variation of Fatou implies

$$\mu(\limsup_{n\to\infty} E_n) \geqslant \limsup_{n\to\infty} \mu(E_n).$$

(Integrating χ gives measure of set.) For example if finitely many E_n's have measure $\geqslant 1/3$, then

$$\mu(\{x : x \in E_n \text{ for infinitely many } n\}) \geqslant \frac{1}{3}.$$

Example 2.3.17: Counting Measure. Let $\mu(E) := |E|$ the cardinality (either finite or infinite). On $\mathbb{N}$, we can write $\mu := \sum_{n=1}^{\infty} \delta_n$. We can take $\mathfrak{M} := \mathbb{P}(\mathbb{N})$. We can think of a function f on $\mathbb{N}$ as a sequence $\{a_n\}_{n\geqslant 1}$.
We claim that $\int f \mathrm{d}\mu = \int f \,\mathrm{d}(\sum_{i=1}^{\infty} \delta_n) = \sum_{n=1}^{\infty} \int f \,\mathrm{d}\delta_n = \sum_{n=1}^{\infty} a_n$.
To see this, we first show that the claim holds for φ (simple functions): if $\varphi := \sum_{j=1}^{k} c_j \chi_{E_j}$, then

$$\int \varphi \,\mathrm{d}\mu = \sum_{j=1}^{k} c_j |E_j|$$

where $|E_j|$ is the number of times c_j appears in the sequence.
Now for general $g \geqslant 0$: we know

$$\int f \,\mathrm{d}\mu = \sup\{\int \varphi \,\mathrm{d}\mu : 0 \leqslant \varphi \leqslant f, \varphi \text{ simple}\}.$$

In particular consider the truncated sequence $\{a_n\}_{n=1}^{k}$; the supremum evaluates to $\sum_{n=1}^{\infty} f(n)$.
For general f, taking $f = f^+ - f^-$, the claim holds provided $\sum_{n=1}^{\infty} f(n)$ converges absolutely.

Example 2.3.18: Nonnegative double arrays. How to compare $\sum_{k=1}^{\infty}\sum_{n=1}^{\infty} a_{k,n}$ and $\sum_{n=1}^{\infty}\sum_{k=1}^{\infty} a_{k,n}$?

Solution. We can denote $a_{k,n}$ by $f_k(n)$ and let μ be the counting measure. Then the LHS is $\sum_{k=1}^{\infty} \int f_k \, \mathrm{d}\mu$ and the RHS is $\int \sum_{k=1}^{\infty} f_k \, \mathrm{d}\mu$. If $a_{k,n} \geqslant 0$ then the MCT gives =. The general case is shown later. □

Example 2.3.19: Do integrals define measures?. Suppose μ, ν are finite Borel measures on $[0,1]$ and $\int f \, \mathrm{d}\mu = \int f \, \mathrm{d}\nu$ for all *continuous* f. Does $\mu = \nu$?

Solution. It is enough to examine indicators $\chi_{[0,t]}$ for $t \in [0,1]$ (as μ, ν are both generated by distribution functions).

For a given t, there exist continuous functions $f_n \downarrow f := \chi_{[0,t]}$. Consider MCT on $1 - f_n$:

$$1 - \int f \, \mathrm{d}\mu = \int (1-f) \, \mathrm{d}\mu = \lim_{n\to\infty} \int (1-f_n) \, \mathrm{d}\mu = \lim_{n\to\infty} \int (1-f_n) \, \mathrm{d}\nu.$$

Therefore $F_\mu = F_\nu$ and so $\mu = \nu$.

Theorem 2.3.20: Lebesgue's Dominated Convergence Theorem (DCT)

Let $f_n \in L^1$ with $f_n \to f$ a.e. and suppose that there exists $g \in L^1$ with $|f_n| \leqslant g$ for all n. Then $f \in L^1$ and

$$\int f_n \to \int f.$$

Example 2.3.21. Let μ be a finite Borel measure on $\mathbb{R}$. Define the **Fourier transform**

$$F(t) := \int e^{itx} \, \mu(\mathrm{d}x) \qquad t \in \mathbb{R}.$$

Question: does $F'(t) = \int ixe^{itx} \, \mu(\mathrm{d}x)$?

Solution. Notice that

$$\frac{F(t+h) - F(t)}{h} = \int \underbrace{\frac{e^{ix(t+h)} - e^{itx}}{h}}_{=:\psi_n(x)} \, \mu(\mathrm{d}x).$$

Take $h = h_n$ and let $h_n \to 0$. Then $\lim_{n\to\infty} \psi_n(x) = ixe^{itx} =: \psi(x)$. Thus we have $\psi_n \to \psi$ and we wonder if $\int \psi_n \to \int \psi$. Indeed:

$$|\psi_n(x)| = |e^{itx}| \left| \frac{e^{ixh_n} - 1}{h_n} \right| = \left| \frac{e^{ixh_n} - 1}{h_n} \right|,$$

and the numerator is

$$e^{ixh_n} - 1 = \int_0^{h_n x} i^{-1} \left(\frac{\mathrm{d}}{\mathrm{d}u} e^{iu} \right) \mathrm{d}u = \int_0^{h_n x} e^{iu} \, \mathrm{d}u$$

where $|ie^{iu}| \leqslant 1$. Hence

$$|e^{ixh_n} - 1| \leqslant \int_0^{|h_n x|} 1 \, \mathrm{d}\mu = |h_n x|.$$

This means $|\psi_n(x)| \leqslant |x|$. Thus if $\int |x| \, \mathrm{d}\mu < \infty$, the claim holds according to DCT.

Beginning of Oct.4, 2021

More generally, given a function $G(t,x)$, does $\frac{\mathrm{d}}{\mathrm{d}t}\int G(t,x)\mu(\mathrm{d}x) = \int \frac{\partial}{\partial t}G(t,x)\mu(\mathrm{d}x)$?

By the same token, we take $h_n \to 0$ and $\psi_n(x) := \frac{G(t+h_n,x) - G(t,x)}{h_n}$ and $\psi(x) = \frac{\partial}{\partial t}G(t,x)$.

It is sufficient that all $|\psi_n|$ is bounded by some integrable g. MVT states

$$|\psi_n(x)| \leqslant \sup_t \left|\frac{\partial G}{\partial t}(t,x)\right|$$

so it is sufficient if we can find $g \in L^1$ with $\sup_t \left|\frac{\partial}{\partial t}G(t,x)\right| \leqslant g(x)$.

Proof of the Lebesgue DCT. First change f_n, f to 0 on a null set where pointwise convergence fails. Then $f_n \to f$ everywhere. Then $|f_n| \leqslant g$ means $|f| \leqslant g$ and so $f \in L^1$.

WLOG assume f is real-valued (complex-valued proof is analogous by considering $\mathfrak{Re} f_n, \mathfrak{Im} f_n$ separately). It suffices to show that $\liminf \int f_n \geqslant \int f \geqslant \limsup \int f_n$.

Since $f_n \geqslant -g$, $f_n + g \geqslant 0$, so Fatou's lemma gives

$$\liminf \int f_n + \int g = \liminf \int (f_n + g) \geqslant \int \liminf(f_n + g) = \int (f+g).$$

Since g is integrable, subtracting gives $\liminf \int f_n \geqslant \inf f$. Similarly, for $-f_n$ we have

$$-\limsup \int f_n = \liminf \int -f_n \geqslant \int (-f)$$

so taking "$-$" gives the claim. □

Example 2.3.22: DCT on series. If ($f_n \in L^1$ and) $\sum_{n=1}^{\infty} \int |f_n| < \infty$, then $\sum_{n=1}^{\infty} f_n \in L^1$ with $\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n$.

Sometimes showing $\int_X f_n \to \inf_X f$ requires different methods on different parts of X, e.g., MCT on A and DCT on A^c. Consider the fowling example.

Example 2.3.23. Let $f \in L^1$, $f > 0$. Note that $\frac{\mathrm{d}}{\mathrm{d}t}u^t\big|_{t=0} = \log u$. Consider the equality

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_X (f^\epsilon - 1) = \int_X \log f.$$

To prove this, let ϵ_n be a sequence converging to 0.

To use DCT, we need

$$\left|\frac{f(x)^\epsilon - 1}{\epsilon}\right| \leqslant g(x) \in L^1 \text{ for small } \epsilon.$$

In other words, do we have $|(t^\epsilon - 1)/\epsilon| \leqslant Ct$ for some C, all t, and small ϵ? The answer is *not for small* t, because this tends to $|\log t|$ as $\epsilon \to 0$ and $|\log t| \to -\infty$ as $t \to 0$.

However, if we only consider $t \geqslant 1$, this is equivalent to considering $t^\epsilon \leqslant 1 + \epsilon Ct$ for small ϵ. Define $\varphi(t) := t^\epsilon$ and $\psi(t) := 1 + \epsilon Ct$. Then

$$\varphi(0) < \psi(0) \text{ and } \frac{\varphi'(t)}{\psi'(t)} = \frac{t^{\epsilon-1}}{C} \leqslant \frac{1}{C}.$$

Taking $C = 1$ then $\varphi' \leqslant \psi'$ so $\varphi \leq \psi$ for $t \geqslant 1$; that is, for ϵ small and $t \in [1, \infty)$, DCT works on $A := \{x : f(x) \geqslant 1\}$.

For A^c, $f(x) = t < 1$, so $0 > (f^\epsilon - 1)/\epsilon$; as $\epsilon \downarrow 0$, this quotient monotonically decreases to $\log f$. We can then multiply by -1 and apply MCT on A^c. Thus,

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{\epsilon_n} \int (f^{\epsilon_n} - 1) &= \lim_{n\to\infty} \frac{1}{\epsilon_n} \int_A (f^{\epsilon_n} - 1) + \lim_{n\to\infty} \frac{1}{\epsilon_n} \int_{A^c} (f^{\epsilon_n} - 1) \\ &= \underbrace{\int_A \log f}_{\text{DCT}} + \underbrace{\int_{A^c} \log f}_{\text{MCT}} = \int \log f.\end{aligned}$$

Example 2.3.24: Other limits besides limits. Consider $F(t) := \int f(t,x)\mu(\mathrm{d}x)$ as $t \to t_0$. If $(\cdot, x)$ is continuous at t_0 for all x, i.e., $f(t,x) \to f(t_0,x)$ as $t \to t_0$ for all x and $|f(t,x)| \leqslant g(x)$ for all x and t near t_0, then by DCT $F(t) \to F(t_0)$.

Example 2.3.25. Find $\lim_{t\to 0} \int_0^a \frac{\sin(tx)}{tx}\,\mathrm{d}x$. (We define the quotient to be 1 if the division is of form 1/1.)

Indeed $(t,x) \to f(0,x)$ as $t \to 0$. Also, the function is dominated by 1, and 1 is integrable on $[0,a]$. Thus

$$\lim_{t\to 0} \int_0^a \frac{\sin(tx)}{tx}\,\mathrm{d}x = \int_0^a 1\,\mathrm{d}x = a.$$

Beginning of Oct.6, 2021

Approximation in L^1

Theorem 2.3.26

Integrable simple functions are dense in L^1. (The distance is w.r.t. to $d(f,g) = \int |f - g|$.)

Proof. Let $f \in L^1$. Recall that there exists a sequence $\{\varphi_n\}$ of simple functions with $|\varphi_n| \uparrow |f|$ pointwise, i.e., $|\varphi_n - f| \to 0$ pointwise. Since $|\varphi_n - f|$ is dominated by $2\|f\| \in L^1$, by DCT $\int |\varphi_n - f| \to 0$. □

Definition 2.3.27: Step Function

A **step function** is a simple function that is constant on intervals.

Theorem 2.3.28

For μ a Lebesgue-Stiltjes measure on $\mathbb{R}$, step functions are dense in L^1.

Proof. Let $\epsilon > 0$ and $f \in L^1(\mu)$. From the previous theorem there exists a simple $\varphi = \sum_{j=1}^{m} a_i \chi_{E_j}$ with $\int |f - \varphi| < \epsilon$. We may assume $a_j \neq 0$ and all E_j's are bounded (if not, intersect it with $[-M, M]$ for sufficiently large M, because $\mu(E_j \cap [-n, n]) \to \mu(E_j)$ as $n \to \infty$).
Fix j. We can approximate E_j by a bounded open $U \supset E$ with $\mu(U - E_j) < \epsilon/(m|a_j|)$. Write U as $\bigcup_{k \geqslant 1} I_k$ a union of open intervals. Then there exists n such that

$$\mu(U - \bigcup_{k=1}^{n} I_k) < \frac{\epsilon}{m|a_j|}.$$

Defining $F_j := \bigcup_{k=1}^{n} I_k$, we have $\int |\chi_{E_j} - \chi_{F_j}| < \frac{2\epsilon}{m|a_j|}$.
If we define $\psi := \sum_{j=1}^{m} a_j \chi_{F_j}$ a step function, we obtain the desired approximation:

$$\int |f - \psi| \leqslant \int |f - \varphi| + \int |\varphi - \psi| < \epsilon + \sum_{j=1}^{m} |a_j| \cdot \frac{2\epsilon}{m|a_j|} \leqslant 3\epsilon.$$

□

Theorem 2.3.29

For μ a Lebesgue-Stieltjes measure on $\mathbb{R}$, continuous functions are dense in L^1.

Proof. Let $\epsilon > 0$ and $f \in L^1$. There exists a step function $\psi = \sum_{j=1}^{m} a_j \chi_{I_j}$ with $\int |f - \psi| < \epsilon$. We take I_j's open with $I_j = (a_j, b_j)$. Main idea: approximate each χ_{I_j} by a piecewise linear function (from $(a_j, 0)$ to some $(a_j + 1/n, 1)$ to $(b_j - 1/n, 1)$ to $(b, 0)$) so that the L^1 between χ_{I_j} and this function is $< \epsilon/(m|a_j|)$.

□

Lebesgue and Riemann Integrals

Beginning of Oct.8, 2021

Let $f : [a, b] \to \mathbb{R}$ and let $P_k = \{t_0, \dots, t_k\}$ with $a = t_0 < \dots < t_n = b$ be a partition. Let M_j, m_j be the supremum and infimum in $[t_{j-1}, t_j]$. In Riemann integration, we defined the upper and lower sums to be G_{P_k} and g_{P_k} such that

$$G_{P_k} \equiv M_j \qquad g_{P_k} \equiv m_j \text{ on } [t_{j-1}, t_j].$$

Define the integrals of G_{P_k}, g_{P_k} to $U_{P_k} f, L_{P_k} f$, respectively. Suppose $G_{P_k} \downarrow G$ and $g_{P_k} \uparrow g$. Also define

$$H(x) := \limsup_{y \to x} f(y) \qquad \text{and} \qquad h(x) := \liminf_{y \to x} f(y).$$

It follows that $G(x) = H(x)$ if $x \notin \bigcup_{k \geqslant 1} P_k$. To make everything rigorous:

Theorem 2.3.30

Let f be a bounded $\mathbb{R}$-valued function on $[a,b]$. Then

(a) f is Riemann integrable $\Rightarrow$ f is Lebesgue integrable and the integrals agree;

(b) f is Riemann integrable $\Leftrightarrow$ $\{x : f \text{ is discontinuous at } x\}$ is a (Lebesgue) null set.

Proof.

(a) Note that $\int G_{P_k}\,\mathrm{d}m = U_{P_k} f \downarrow \overline{\int_a^b} f(x)\,\mathrm{d}x$ and $\int g_{P_k}\,\mathrm{d}m = L_{P_k} f \uparrow \underline{\int_a^b} f(x)\,\mathrm{d}x$. Since f is bounded,

$$\int G_{P_k}\,\mathrm{d}m \to \int G\,\mathrm{d}m = \int H\,\mathrm{d}m$$

and

$$\int g_{P_k}\,\mathrm{d}m \to \int g\,\mathrm{d}m = \int h\,\mathrm{d}m.$$

Thus the upper and lower Riemann sums are the same if and only if $\int (H-h)\,\mathrm{d}m = 0$. Since $H \geqslant h$, this is equivalent to $H = h$ a.e., i.e., f continuous almost everywhere. Hence

$$f \text{ Riemann integrable} \Rightarrow H = f = g \text{ a.e.} \Rightarrow f \text{ Lebesgue measurable.}$$

□

When does $\int f_n \not\to \int f$ even if $f_n \to f$ a.e. and $\mu(X) < \infty$?

Fix a large K. We truncate f_n, f at $\pm K$ such that

$$f_n^{(k)}(x) = \begin{cases} -K & f_n < -K \\ f_n(x) & f_n(x) \in [-K, K] \\ K & f_n > K. \end{cases}$$

By assumption $f_n \to f$ a.e., we see $f_n^{(k)}(x) \to f^{(k)}(x)$, a.e. Since $\mu(X) < \infty$, all functions are bounded by the integrable function K, so by Lebesgue DCT

$$\int f_n^{(k)} \to \int f^{(k)} \qquad \text{for all } K.$$

Therefore, failure of $\int f_n \to \inf$ must be caused by *large* values (positive or negative) of f on small sets, for example $x_n := n^2 \chi_{(0,1/n]}$.

Modes of Convergence

Previously we talked about a.e. convergence, L^1 convergence, and uniform convergence. Here we introduce two more modes of convergence:

Definition 2.3.31: Convergence in Measure

We say $f_n \to f$ **in measure** if for every $\epsilon > 0$,

$$\mu(\{x : |f_n(x) - f(x)| > \epsilon\}) \to 0 \qquad \text{as } n \to \infty.$$

Definition 2.3.32: Cauchy-ness in Measure

We say $\{f_n\}$ is **Cauchy in measure** if for every $\epsilon > 0$,

$$\mu(\{x : |f_n(x) - f_m(x)| > \epsilon\}) \implies 0 \qquad \text{as } \min(m,n) \to \infty.$$

Questions:

(1) Do these two new modes imply each other?

(2) How do they relate to the previous modes?

(3) Does there exist a metric d of functions such hat $f_n \to f$ in measure $\Leftrightarrow d(f_n, f) \to 0$?

Example 2.3.33. Some examples with Lebesgue measure on $\mathbb{R}$:

f_n	f	pointwise	a.e.	uniform	in L^1	in measure
$n^{-1}\chi_{(0,n]}$	$f \equiv 0$	yes	yes	yes	no	yes
$n\chi_{(0,1/n]}$	$f \equiv 0$	yes	yes	yes	no	yes
*	$f \equiv 0$	no	no	no	yes	yes

The starred one refers to the indicator functions of

$$[0,1), [0,1/2), [1/2,1), [0,1/3), [1/3,2/3), [2/3,1), [0,1/4), \ldots$$

We refers to this as the "scanning interval".

Proposition 2.3.34: L^1 Convergence $\Rightarrow$ Convergence in Measure

If $f_n \to f$ in L^1 then $f_n \to f$ in measure.
Heuristically, "violation eventually stops" implies "violation eventually gets small".

Proof. Fix ϵ. Define

$$E_{n,\epsilon} = \{x : |f_n(x) - f(x)| > \epsilon\}.$$

Then $|f_n - f| \geqslant \epsilon\chi_{E_{n,\epsilon}}$, so $\int |f_n - f| \geqslant \epsilon\mu(E_{n,\epsilon})$. Since the LHS $\to 0$ as $n \to \infty$, we must have $\mu(E_{n,\epsilon}) \to 0$ as $n \to \infty$. Since ϵ is arbitrary, this completes the proof. □

Beginning of Oct.18, 2021

Theorem 2.3.35

Let $\{f_n\}$ be measurable functions.

(1) $\{f_n\}$ is Cauchy in measure if and only if they converge in measure.

(2) If $\{f_n\}$ is Cauchy (or converges in measure), then some subsequence $\{f_{n_k}\}$ converges a.e.

(3) Unique limit in measure: if $f_n \to f$ and $f_n \to g$ in measure, then $f = g$ a.e.

Proof of (2). Choose $\{n_j\}$ so that $\mu(\{x : |f_{n_j}(x) - f_m(z)| \geqslant 3^{-j}\} \leqslant 2^{-j}$ for all $m \geqslant n_j$ (the bad sets). Define correspondingly $E_j := \{x : |f_{n_j}(x) - f_{n_{j+1}(x)}| \geqslant 3^{-j}\}$. Let $F_k := \bigcup_{j \geqslant k} E_j$ so $\mu(F_j) \leqslant 2^{-k+1}$. Let $F = \bigcap_{k \geqslant 1} F_k$ (the collection of x for which violations never stop). It follows that for $x \notin F$, the violation stops, so

$$f(x) = \lim_{j \to \infty} f_{n_j}(x)$$

exists. Thus for $x \notin F$,

$$|f_{n_j}(x) - f(x)| \leqslant \sum_{m=j}^{\infty} |f_{n_m}(x) - f_{n_{m+1}}(x)| \leqslant \sum_{m=j}^{\infty} 3^{-m} = \frac{3}{2} e^{-j}$$

so $\mu(\{x : |f_{n_j}(x) - f(x)| > 3/2 \cdot e^{-j}\}) \leqslant \mu(F_j) \leqslant 2^{-j+1}$. Hence given $\epsilon > 0$, for large j we have $3/2 \cdot 3^{-j} < \epsilon$ and

$$\mu(\{|f_{n_j}(x) - f(x)| > \epsilon\}) \leqslant 2^{-j+1} \to 0.$$

This says $f_{n_j} \to f$ in measure. □

Proof of (1). We now want to show that the full sequence $f_n \to f$ in measure. We compare the subsequence to the full sequence:

$$\{x : |f_n(x) - f(x)| > \epsilon\} = \{x : |f_n(x) - f_{n_j}(x)| > \epsilon/2\} \cup \{x : |f_{n_j}(x) - f(x)| > \epsilon/2\}.$$

Take $n_j > n$. Then $n \to \infty$ implies $n_j \to \infty$. Cauchy in measure says μ of the first set $\to 0$. On the other hand, the convergence in measure of subsequence implies that μ of the second set also $\to 0$.

The converse is trivial. □

Proof of (3). Suppose $f_n \to f$ and $f_n \to g$ in measure. Then

$$\mu(\{x : |f(x) - g(x)| > \epsilon\}) \leqslant \mu(\{x : |f(x) - f_n(x)| > \epsilon/2\}) + \mu(\{x : |g(x) - f_n(x)| > \epsilon/2\}) \to 0$$

for all ϵ. Note that the original expression does not contain f_n's. Thus $f = g$ a.e. □

Remark. There is no metric for pointwise convergence of functions on uncountable X. None for a.e. convergence either.

Proof of Lebesgue case. We consider $[0, 1]$ with Lebesgue. Suppose there were a metric d such that $f_n \to f$ a.e. implies $d(f_n, f) = 0$. In a metric space, if $y_n \nrightarrow y$, then there exists $\epsilon > 0$ and $\{y_{n_k}\}$ such that $d(y_{n_k}, y) > \epsilon$. Hence no further subsequence $\{y_{n_{k(\ell)}}\}$ can possibly converge to y.

We recall that the scanning interval example $f_n \to 0$ in measure but not a.e. But then every subsequence $f_{n_k} \to 0$ in measure, so does any further subsequence $f_{n_{k(\ell)}}$. But this is not possible in a metric space, so there is no metric for which $f_n \to f$ a.e. if and only if $d(f_n, f) \to 0$. □

Lemma: Riemann-Lebesgue Lemma

Let $f : \mathbb{R} \to \mathbb{R}$, $f \in L^1(m)$. Then

$$\int_{\mathbb{R}} e^{-itx} f(x)\, m(\mathrm{d}x) \to 0 \text{ as } t \to \infty.$$

Beginning of Oct.20, 2021

Example 2.3.36. Define h_n on $[0,1]$ by

$$h_n(x) = \sum_{j=1}^{n} (-1)^j \chi_{((j-n_1)/n, j/n]}(x).$$

Suppose f is integrable on $[0,1]$. Then $\lim_{n\to\infty} \int_{[0,1]} f h_n \,\mathrm{d}m = 0$.

Proof. We can use a density argument and examine step functions. If $\varphi = \chi_{(a,h]}$, most $1/n$-intervals cancel each other out except up to three $1/n$-intervals. Hence

$$\left| \int_{[0,1]} \chi_{(a,h]} h_n \,\mathrm{d}m \right| \leqslant \frac{3}{n} \to 0.$$

For more general integrable f: let $\epsilon > 0$ be given. There exists a step function φ with $\int |f - \varphi| < \epsilon$. Then

$$\left| \int h_n f \right| \leqslant \left| \int h_n (f - \varphi) \right| + \left| \int h_n \varphi \right| \leqslant \int |h_n||f-\varphi| + \left| \int h_n \varphi \right|.$$

We've shown that the second term $\to 0$. For the first one, since $|h_n| \leqslant 1$, $\int |h_n||f - \varphi| \leqslant \int |f - \varphi|$, so

$$\limsup_{n\to\infty} \left| \int h_n f \right| \leqslant \epsilon.$$

□

Theorem 2.3.37: Riemann-Lebesgue (Lemma)

Let $f : \mathbb{R} \to \mathbb{R}$ be integrable. Then

$$\int_{\mathbb{R}} e^{-itx} f(x)\, m(\mathrm{d}x) \to 0 \qquad \text{as } t \to \infty.$$

Definition 2.3.38: Almost Uniform Convergence

We say $f_n \to f$ **almost uniformly** if for all $\epsilon > 0$, there exists $E \subset X$ with $\mu(E) < \epsilon$ such that $f_n \to f$ uniformly on E^c.

Example 2.3.39. Consider $f_n(x) := 1/(nx)$ for $x \in [0, 1]$. The convergence is obviously not uniform, but if we take out $[0, \epsilon]$ then the convergence on $(\epsilon, 1]$ is uniform.

Theorem 2.3.40: Egoroff's Theorem

If $\mu(X) < \infty$, then a.e. convergence implies almost uniform convergence.

Proof. We chance f_n, f on a null set so that $f_n \to f$ pointwise. For each *fixed* x, the violations of bound $/k$ eventually stops: for big $n \geqslant N_k(x)$,

$$|f_n(x) - f(x)| \leqslant 1/k.$$

On the other hand, "uniform on E^c" means

$$n \geqslant N_k(x) \implies |f_n(x) - f(x)| \leqslant \frac{1}{k} \qquad \text{for all } x \in E^c. \tag{*}$$

We want to combine (over k) the "bad sets" where (*) fails, with the combined set still small.

We write the bad sets

$$\{x : |f_n(x) - f(x)| > 1/k \text{ for some } n \geqslant m\}.$$

By pointwise convergence, these sets eventually shrink to $\varnothing$ as $m \to \infty$. Hence for each k, there exists n_k where

$$\mu(\{x : |f_n(x) - f(x)| > 1/k \text{ for some } n \geqslant n_k\} < \epsilon 2^{-k}.$$

Then the union E of these sets has measure $< \epsilon$. For $x \in E^c$, the violation eventuallly stops: $n \geqslant n_k$ implies $|f_n(x) - f(x)| \leqslant 1/k$ for all $x \in E^c$. This shows almost uniform convergence. □

Corollary 2.3.41

If $\mu(X) < \infty, \epsilon > 0$, $f_n \to f$ a.e., and f_n continuous, there exists a E such that $\mu(E) < \epsilon$ and $f_n \to f$ uniformly on E, so $f|_{E^c}$ is continuous.

Note that "f is continuous at x for all $x \in E^c$" is stronger than "$f|_{E^c}$ is continuous!" For example consider $[0, 1]$, $F :=$ irrationals, and $f = \chi_F$. Then $f|_F \equiv 1$ but clearly f is not continuous at any point.

Summary of Convergence.

(1) We showed (convergence in L^1) $\Rightarrow$ (convergence in measure).

(2) For $\mu(X) < \infty$:

(i) (convergence a.e.) $\Rightarrow$ (almost uniform) (Egoroff)

(ii) (convergence a.e.) $\Rightarrow$ (convergence in measure)

(3) For countable X, say $X = \mathbb{N}$: we can assume $\mu(\{x\}) > 0$ for all x by throwing out all points at which the measure is 0. Suppose $\mu(X) < \infty$ so convergence a.e. $\Rightarrow$ convergence in measure. Suppose $f_n \to f$ in measure. Fix $x_0 \in \mathbb{N}$ and $\epsilon > 0$. If $|f_n(x_0) - f(x_0)| > \epsilon$ for infinitely many n, then for these n,

$$\mu(\{x : |f_n(x) - f(x)| > \epsilon\}) \geqslant \mu(\{x_0\}) > 0,$$

a contradiction. Hence if $f_n \to f$ in measure, $f_n \to f$ must happen on all these points with positive measure.

Beginning of Oct.22, 2021

2.4 Product Algebra and the Fubini-Tonelli Theorem

Recall in calculus we have integrals of form $\iint f(x,y)\,\mathrm{d}x\mathrm{d}y$ or $\iint_D f(x,y)\,\mathrm{d}A$.

General picture: we want product measures. More specifically, let $(X.\mathfrak{M},\mu)$ and $(Y,\mathfrak{N},\nu)$ be two measure spaces, and let $\mathfrak{M}\otimes\mathfrak{N}$ be the product σ-algebra generated by $\{A\times B : A\in\mathfrak{M}, B\in\mathfrak{N}\}$, and we want a measure $\mu\times\nu$ such such that $(\mu\times\nu)(A\times b) = \mu(A)\nu(B)$.

Example 2.4.1. In $\mathbb{R}\times\mathbb{R}$, the finite unions of geometric rectangles form an elementary family. We can use this to get $m\times m$ (where m is the Lebesgue measure).

Example 2.4.2. In probability, let S, T be random quantities defined on $(\Omega,\mathfrak{F},\mathbb{P})$. We have the "push-forward distribution"

$$\mu(A) = \mathbb{P}(S\in A) \qquad \nu(B) = \mathbb{P}(T\in B)$$

where $A, B\in\mathcal{B}_{\mathbb{R}}$. If S, T are independent then

$$\mathbb{P}(S\in A, T\in B) = \mathbb{P}(S\in A)\mathbb{P}(T\in B),$$

i.e.,

$$(\mu\times\nu)(A\times B) := \mathbb{P}((S,T)\in A\times B) = \mu(A)\nu(B).$$

In general, let

$$\mathcal{A} := \{\text{all finite disjoint unions of abstract rectangles}\}.$$

Then $\mathcal{A}$ is an algebra which generates $\mathfrak{M}\otimes\mathfrak{N}$. To construct $\mu\times\nu$ we need a *premeasure* on $\mathcal{A}$; for

$$C = \bigcup_{j=1}^{n} A_j\times B_j,$$

we define

$$(\mu\times\nu)(C) := \sum_{j=1}^{n}\mu(A_i)\nu(B_i).$$

We need to first check if this notion is well-defined; that is, if the representation is not unique, for example $C = \bigcup_{i=1}^{m} E_i \times F_i$, we need the definition to be consistent. *Heuristically we can decompose the coordinates in X into smaller pieces from which we can both assemble A_i and A_j and likewise we can break Y into smaller pieces to assemble B_i and B_j. The idea is that both original sums will become the same thing under this representation. This shows that $\mu \times \nu$ is well-defined.*

We can again verify $\mu \times \nu$ is a premeasure but the proof is omitted. Therefore $\mu \times \nu$ on $\mathcal{A}$ has an extension to a measure on $\sigma(A) = \mathfrak{M} \otimes \mathfrak{N}$. This extended $\mu \times \nu$ is called the **product measure**.

A similar definition holds for $\mu_1 \times \ldots \times \mu_n$ on $\mathfrak{M}_1 \otimes \ldots \otimes \mathfrak{M}_n$.

Definition 2.4.3: Sections of a set

Let $E \subset X \times Y$. Then

$$E^y := \{x \in X : (x, y) \in E\}$$

and

$$E_x := \{y \in Y : (x, y) \in E\}$$

are called the **sections** of E in the product space $X \times Y$. (The first one is a horizontal slice and the second a vertical slice.)

Sections commute with a set of operations: for a fixed x,

$$(E \cup F)_x = E_x \cup F_x \qquad (E \cap F)_x = E_x \cap F_x \qquad (E_x)^c = (E^c)_x.$$

The same holds for a fixed y. Unions and intersections also hold for arbitrary number of sets.

Similarly, we can apply the definitions to functions and define

$$f_x(\cdot) := f(x, \cdot) \qquad f^y(x) := f(\cdot, y).$$

Then

$$f_x^{-1}(B) = (f^{-1}(B))_x \qquad (f^y)^{-1}(A) = (f^{-1}(A))^y.$$

Proposition 2.4.4

(1) If $E \subset \mathfrak{M} \times \mathfrak{N}$, then all sections E_x, E^y are measurable.

(2) If f is a function on $X \times Y$. If f is measurable with respect to $\mathfrak{M} \otimes \mathfrak{N}$, then so are any f_x, f^y.

Proof.

(1) The claim is true for $A \times B \in \mathfrak{M} \otimes \mathfrak{N}$. Let $\mathcal{E} := \{E \in \mathfrak{M} \otimes \mathfrak{N} :$ all sections are measurable$\}$. We want to show $\mathcal{E}$ is a σ-algebra. Since

$$(\bigcup_j E^{(j)})_x = \bigcup_j E_x^{(j)}$$

we get $\mathcal{E}$ is closed. Similarly $\mathcal{E}$ is closed under complementation. Therefore $E \supset \mathfrak{M} \otimes \mathfrak{N}$. Clearly $\mathcal{E} \subset \mathfrak{M} \otimes \mathfrak{N}$, so they are equal.

(2) Recall that $f_x^{-1}(B) = (f^{-1}(B))_x$. Since f is measurable, $f^{-1}(B)$ is measurable. Hence $f_x^{-1}(B)$ is measur-

able, and this says f_x is measurable. Likewise for f_y.

□

Beginning of Oct.25, 2021

Question. Does it hold in general that

$$(\mu \times \nu)(E) = \int \mu(E^y)\nu(\mathrm{d}y) = \int \nu(E_x)\mu(\mathrm{d}x)? \tag{Δ}$$

Last time we show that each slice is indeed measurable, so $\nu(E_x), \mu(E^y)$ are well-defined. However a new question arises — is $\nu(E_x)$ a measurable function of x, and is $\mu(E^y)$ a measurable function of y?

Recall the general principle. We define a property P by saying E has the property if $\nu(E_x), \mu(E^y)$ are measurable functions of x, y and (Δ) holds for E.

Notice that originally the general principle requires that $\{$sets with property $P\}$ to be a σ-algebra, but there's an easier alternative:

Definition 2.4.5: Monotone class

A **monotone class** is a collection $\mathcal{C}$ of subsets of X such that $\mathcal{C}$ is closed under *countable increasing* unions and *countable decreasing* intersections.

Note that σ-algebras are always monotone classes but not the converse.

Just like σ-algebra, a set E generates a monotone class, which is defined to be the intersection of all monotone classes containing E. Hence this monotone class is a (possibly non-strict) subset of $\sigma(E)$.

Lemma

Let $\mathcal{A}$ be an *algebra* and $\mathfrak{M}, \mathcal{C}$ be the σ-algebra and monotone class generated by $\mathcal{A}$, respectively. Then $\mathfrak{M} = \mathcal{C}$.

Upshot. For an algebra $\mathcal{A}$, the general principle works exactly the same if we show $\{$sets with property $P\}$ is a monotone class.

Proof. By definition $\mathcal{C} \subset \mathfrak{M}$ so it suffices to show $\mathcal{C}$ is a σ-algebra so $\mathcal{C} \supset \mathfrak{M}$.
We say (E, F) forms a *good pair* if $E \cap F, E - F, F - E$ are all in $\mathcal{C}$. We will show that all $E, F \in \mathcal{C}$ make good pairs.
Clearly, if $E, F \in \mathfrak{A}$ then this works by the definition of an algebra.
More generally, fix $E \in \mathcal{C}$. Define $\mathcal{C}(E) := \{F \in \mathcal{C} : (E, F) \text{ is a good pair}\}$.
Claim. $\mathcal{C}(E)$ is a monotone class.
For countable increasing unions, suppose $F_1 \subset F_2 \subset \ldots$, all in $\mathcal{C}(E)$. We show that $(E, \bigcup_{n\geqslant 1} F_n)$ is a good pair so the union is in $\mathcal{C}(E)$:

$$E \cap (\bigcup_{n\geqslant 1} F_n) = \bigcup_{n\geqslant 1}(E \cap F_n)$$

where $E \cap F_n$ forms an increasing sequence. Since $\mathcal{C}$ is a monotone class, the union is in $\mathcal{C}$.

$$E - (\bigcup_{n\geqslant 1} F_n) = \bigcap_{n\geqslant 1}(E - F_n)$$

where $E - F_n$ forms a decreasing sequence, so the intersection is again in $\mathcal{C}$. Finally,

$$\Big(\bigcup_{n\geqslant 1} F_n\Big) - E = \bigcup_{n\geqslant 1}(F_n - E),$$

again an increasing union, which is therefore in $\mathcal{C}$. Similarly we can show $\mathcal{C}(E)$ is closed under decreasing intersections. Hence $\mathcal{C}(E)$ is a monotone class.

Now notice that if $E \in \mathcal{A}$ then $\mathcal{C}(E) \supset \mathcal{A}$, so $\mathcal{C}(E) \supset$ all of $\mathcal{C}$. Hence (E, F) is a good pair for all $E \in \mathcal{A}$ and $F \in \mathcal{C}$. Equivalently $A \subset \mathcal{C}(F)$ for all $F \in \mathcal{C}$. But we showed $\mathcal{C}(F)$ is a monotone class, so $\mathcal{A} \subset \mathcal{C}(F)$ ensures $\mathcal{C} \subset \mathcal{C}(F)$. This says (E, F) is a good pair for all $E, F \in \mathcal{C}$. Hence $\mathcal{C}$ is closed under the three "good pair" operations.

Finally, since $\varnothing, X$ are in $\mathcal{A}$, they are also in $\mathcal{C}$. Also, the "good pair" operations also ensure that $\mathcal{C}$ is closed under complements. And it's closed under finite intersections and countable increasing unions. Therefore $\mathcal{C}$ is also closed under countable unions, i.e., it is a σ-algebra. □

Now we show that $\nu(E_x)$ and $\mu(E^y)$ are measurable functions of x and y.

Theorem 2.4.6

Let $(X, \mathfrak{M}, \nu)$ and $(Y, \mathfrak{N}, \nu)$ be σ-finite [!!] Then for all $E \in \mathfrak{M} \otimes \mathfrak{N}$,

(1) The functions $x \mapsto \nu(E_x)$ and $y \mapsto \mu(E^y)$ are measurable, and

(2) $(\mu \times v)(E) = \int \nu(E_x)\mu(\mathrm{d}x) = \int \mu(E^y)\nu(\mathrm{d}y)$.

Proof. We first consider *finite* measures μ and ν. Recall that it is enough to show that

$$\mathcal{C} := \{E : (1),(2) \text{ holds}\}$$

contains $\mathcal{A} := \{$all finite unions of abstract rectangles$\}$ and is either a σ-algebra or a monotone class. For the integration part, i.e., (2), the monotone class argument is easier because we can use MCT.

For an abstract rectangle,

$$\nu(E_x) = \begin{cases} \nu(B) & x \in A \\ 0 & \text{otherwise} \end{cases}$$

which is a function taking two values and therefore measurable. Similarly $\mu(E^y)$ is measurable in this case. Then

$$\int \nu(E_x)\mu(\mathrm{d}x) = \int \nu(B)\chi_A\mu(\mathrm{d}x) = \nu(B)\mu(A) = (\mu \times \nu)(E)$$

and the other one follows analogously. Thus $E \in \mathcal{C}$. It is clear that *finite* unions of abstract rectangles (disjoint) are also in $\mathcal{C}$.

Now we show that $\mathcal{C}$ is closed under increasing unions. Suppose $E_1 \subset E_2 \subset \ldots$, all in $\mathcal{C}$, and let E be their union. Define $f_n(y) := \mu((E_n)^y)$. By continuity from below, $f_n(y) \uparrow \mu(E^y)$. Call this limit $f(y)$. Then f as a limit of measurable f_n's is measurable, and

$$\begin{aligned} \int \mu(E^y)\nu(\mathrm{d}y) &= \lim_{n\to\infty} \int \mu((E_n)^y)\nu(\mathrm{d}y) && \text{(MCT)} \\ &= \lim_{n\to\infty} (\mu \times \nu)(E_n) = (\mu \times \nu)(E). \end{aligned}$$

By the assumption that μ, ν are finite, we can do the same for decreasing intersections: if $F_1 \supset F_2 \supset \ldots$ and $F :=$ the intersection then $F \in \mathcal{C}$ (again, requires finite measure to use MCT). Therefore $\mathcal{C}$ is a monotone class generated by $\mathcal{A}$ so it contains $\sigma(A) = \mathfrak{M} \otimes \mathfrak{N}$.

Now we move to *σ-finite* measures μ and ν. This means there are sets of finite measure $A_n \uparrow X$ and $B_n \uparrow Y$. Therefore for all E, $E \cap (A_n \times B_n) \uparrow E$, and (1) (2) are true for each $E \cap (A_n \times B_n)$. Using MCT we can take limits as $n \to \infty$ so that E satisfies (1) and (2) too. □

This theorem says that under the same assumptions

$$\nu(E_x) = \int \chi_{E_n} \mathrm{d}\nu = \int \chi_E(x, \cdot)\, \mathrm{d}\nu$$

for all fixed x. (2) says

$$\int \chi_E \,\mathrm{d}(\mu \times \nu) = \int \left[\int \chi_E(x,y)\, \nu(\mathrm{d}y) \right] \mu(\mathrm{d}x) = \int \left[\int \chi_E(x,y)\, \nu(\mathrm{d}x) \right] \nu(\mathrm{d}y).$$

Example 2.4.7. Calculus is not good enough! Define

$$f(x,y) = 0 \text{ at origin and } \frac{x^2 - y^2}{(x^2+y^2)^2} \text{ at everywhere else.}$$

For fixed x, its antiderivative in y is $y/(x^2+y^2)$, so

$$\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\, \mathrm{d}y = \frac{y}{x^2+y^2}\Big|_0^1 = \frac{1}{1+x^2}.$$

Then

$$\int_0^1 \left[\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\, \mathrm{d}y \right] \mathrm{d}x = \int_0^1 \frac{1}{1+x^2}\, \mathrm{d}x = \frac{\pi}{4}.$$

However, since x, y appear symmetrically up to a minus sign, if we integrate x first we obtain

$$\int_0^1 \left[\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\, \mathrm{d}x \right] \mathrm{d}y = -\frac{\pi}{4}.$$

Things break down. This is because f has a singularity at $(0,0)$, and the integral involves $\infty - \infty$, resulting in a non-measurable function.

Beginning of Oct.27, 2021

Example 2.4.8. We provide a discrete analogue which might be easier to understand. Consider the array

$$(a_{i,j}) := \begin{matrix} 1 & 0 & 0 & 0 & 0 & \ldots \\ -1 & 1 & 0 & 0 & 0 & \ldots \\ 0 & -1 & 1 & 0 & 0 & \ldots \\ 0 & 0 & -1 & 1 & 0 & \ldots \\ 0 & 0 & 0 & -1 & 1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{matrix}$$

We claim that $\sum_{i=1}^{\infty} \sum_{j}^{\infty} a_{i,j} \neq \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{i,j}$. Indeed, if we sum over columns first and compute the sum over columns, we get $0 + 0 + \ldots = 0$, whereas the row sums are $1, 0, 0, \ldots$ so the sum is 0.

However, if we have absolute convergence, i.e., $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}|a_{i,j}| < \infty$ then the claim would have held. (We can find m, n such that the partial sum $\sum_{i=1}^{m}\sum_{j=1}^{n}$ is $< \epsilon$ away from the sum.)

Remark. Problems arise when the positive and negative contributions to a sum or integral are both infinite (so we have $\infty - \infty$). In this case, they cancel in different ways depending on the order of summation, so order matters. The weird counterexample f above has

$$\iint_{\{(x,y):f\geqslant 0\}} f(x,y)\,\mathrm{d}y\mathrm{d}x = \infty$$

and

$$\iint_{\{(x,y):f<0\}} f(x,y)\,\mathrm{d}x\mathrm{d}y = -\infty.$$

Theorem 2.4.9: Fubini-Tonelli

This theorem generalizes the previous theorem, in which we only focused on indicator functions.
Suppose $(X, \mathfrak{M}, \mu)$ and $(Y, \mathfrak{N}, \nu)$ are σ-finite (like before!).

(1) (Tonelli) If $f \in L^+(\mu \times \nu)$, the functions defined by integral along slices

$$x \mapsto \int f_x\,\mathrm{d}\nu \qquad y \mapsto \int f^y\,\mathrm{d}\mu$$

are measurable, i.e., the first one $\in L^+(X)$ and the second $L^+(Y)$, and

$$\int f\,\mathrm{d}(\mu\times\nu) = \int\left[\int f(x,y)\nu(\mathrm{d}y)\right]\mu(\mathrm{d}x) = \int\left[\int f(x,y)\mu(\mathrm{d}x)\right]\nu(\mathrm{d}y). \qquad (*)$$

(They can all be ∞.)

(2) (Fubini) If $f \in L^1(\mu \times \nu)$, we have

$$\int_Y f(x,y)\nu(\mathrm{d}y) \qquad \int_X f(x,y)\mu(\mathrm{d}x)$$

finite a.e. for x and y, respectively. Also, the a.e. defined functions

$$x \mapsto \int_Y f(x,y)\nu(\mathrm{d}y) \qquad y \mapsto \int_X f(x,y)\mu(\mathrm{d}x)$$

are in $L^1(\mu)$ and $L^1(\nu)$, respectively. Furthermore, (*) holds (with finite value too!).

Remark. Tonelli requires $f \geqslant 0$ and Fubini requires f to be integrable. Neither allows positive and negative contributions to both have infinite magnitude.

Proof. For simple functions, this follows from the previous theorem. Since integration is linear, the theorem also holds for simple functions too. For general $f \in L^+(X \times Y)$, let f_n be simple functions converging $\uparrow$ to f. By MCT,

$$g_n(x) := \int_Y f_n(x,y)\nu(\mathrm{d}y) \uparrow \int_Y f(x,y)\nu(\mathrm{d}y) =: g(x)$$

and similarly

$$h_n(y) := \int_X f_n(x,y)\mu(\mathrm{d}x) \uparrow \int_X f(x,y)\mu(\mathrm{d}x) =: h(y).$$

Since each g_n, h_n is measurable, g, h are measurable too. Then, using MCT once more,

$$\begin{aligned}\int_X \left[\int_Y f(x,y)\nu(\mathrm{d}y)\right]\mu(\mathrm{d}x) &= \int_X g(x)\mu(\mathrm{d}x) \\ &= \lim_{n\to\infty}\int_X g_n(x)\mu(\mathrm{d}x) = \lim_{n\to\infty}\int_{X\times Y} f_n \,\mathrm{d}(\mu\times\nu) \\ &= \int_{X\times Y} f \,\mathrm{d}(\mu\times\nu)\end{aligned}$$

where the second last = is because g_n is simple. This proves Torelli.

Now for Fubini, we first assume $f \geqslant 0$ (for general functions, decompose into positive and negative parts; for complex functions, decompose into real and imaginary parts). If f is integrable, then by above, $g < \infty$ and $h < \infty$ a.e., so $f_x \in L^1(\nu)$ for a.e. x and $f^y \in L^1(\mu)$ for a.e. y. □

Example 2.4.10. Suppose μ is a finite measure and g measurable. Suppose $f(x,y) := g(x) - g(y)$ is measurable. We claim that g is integrable.

Proof. Fubini says that since f is integrable, then $f(x) = g(x) - g(y)$ is integrable over x for a.e. y. That is, for a constant $g(y)$, $g(x) - g(y)$ is integrable. Since $\mu(X) < \infty$, subtracting makes sense, and we see that g is integrable. □

Example 2.4.11. (A non-example) Consider the product space of $([0,1], \mathcal{B}, m)$ and $([0,1], \mathcal{B}, \nu)$ where ν is the counting measure. Define

$$E := \{(x,x) : x \in [0,1]\}$$

a closed set and therefore a measurable set. Does

$$\int_{[0,1]}\left[\int_{[0,1]} \chi_E \,\mathrm{d}\nu\right] \mathrm{d}m = \int_{[0,1]}\left[\int_{[0,1]} \chi_E \,\mathrm{d}m\right] \mathrm{d}\nu?$$

Solution. For a fixed x,

$$\int_{[0,1]} \chi_E(x,y)\nu(\mathrm{d}y) = 1$$

because $\chi_E(x,y) = 1$ for just one point $y = x$, and counting measure gives $\nu(\{x\}) = 1$. Hence the LHS in the example is 1. However, for fixed y, $\chi_E(x,y) = 0$ for a.e. x, so the inner integral is always 0 and so is the entire one.

This does *not* violate Fubini-Tonelli, as ν is not σ-finite!

Example 2.4.12. Here we provide another example in which we prove some property for simple functions, then nonnegative functions, and finally general functions.

Let $(X, \mathfrak{M}, \mu)$ be a measure space and $g : X \to \mathbb{R}$. Define ν_g to be the "push-forward" of μ to Borels in $\mathbb{R}$:

$$\nu_g(A) := \mu(g^{-1}(A)) \text{ for } A \in \mathcal{B}_{\mathbb{R}}.$$

For $f : \mathbb{R} \to \mathbb{R}$, what is $\int f \mathrm{d}\nu_g$?

Solution. First, for an indicator function χ_E:

$$\int_{\mathbb{R}} \chi_E \,\mathrm{d}\nu_g = \nu_g(E) = \mu(g^{-1}(E)) \int_X \chi_{g^{-1}(E)} \,\mathrm{d}\mu = \int_X \chi_E \circ g \,\mathrm{d}\mu.$$

Linearity shows that if $\varphi \geqslant 0$ is simple then

$$\int_{\mathbb{R}} \varphi \,\mathrm{d}\nu_g = \int_X \varphi \circ g \,\mathrm{d}\mu.$$

For general measurable $f \geqslant 0$, let $\varphi_n \uparrow f$ be a sequence of simple functions. MCT implies

$$\int_{\mathbb{R}} f \,\mathrm{d}\nu_g = \lim_{n\to\infty} \int_{\mathbb{R}} \varphi_n \,\mathrm{d}\nu_g = \lim_{n\to\infty} \int_X \varphi_n \circ g \,\mathrm{d}\mu = \lim_{n\to\infty} \int_X f \circ g \,\mathrm{d}\mu.$$

Finally, for general f, we just need to decompose it into f^+, f^-, or real and complex parts, if necessary.

2.5 Some Remarks on Lebesgue Measure m on $\mathbb{R}^n$

We know how m works on $\mathbb{R}$. However, we *cannot* define m on $\mathbb{R}^2$ simply by taking the product of two copies of m on $\mathbb{R}$. In particular, **the product of two complete measures is not necessarily complete**.

For example, consider $D = \{(x,x) : x \in \mathbb{R}\}$, the diagonal line of $\mathbb{R}^2$, a measurable (null) set for $m \times m$. Let A be a nonmeasurable subset of $\mathbb{R}$ and let $A_D := \{(x,x) : x \in \forall\}$. so that $A_D \subset D$.
We claim that A_D is not $\mathcal{L} \times \mathcal{L}$ measurable, which means that $m \times m$ is not complete.

Proof. Let $f : \mathbb{R} \to \mathbb{R}^2$ defined by $f(x) := (x,x)$. We show f is measurable from $\mathcal{L}$ to $\mathcal{L} \times \mathcal{L}$. Indeed, generators of $\mathcal{L} \times \mathcal{L}$ are the product of generators of $\mathcal{L}$, i.e.,

$$\{(E \cup F) \times (G \cup H) : E, G \in \mathcal{B}_{\mathbb{R}}, F, H \text{ contained in null Borel}\}.$$

Since

$$f^{-1}((E \cup F) \times (G \cup H)) = (E \cup F) \cap (G \cup H) \in \mathcal{L}$$

we see f is indeed measurable. If A_D were $\mathcal{L} \times \mathcal{L}$ measurable then $A = f^{-1}(A_D) \in \mathcal{L}$, contradiction. Hence $m \times m$ is not complete!! □

To this end, we modify the definition of m^n on $\mathbb{R}^n$ by completion:

Definition 2.5.1: Lebesgue measure m^n on $\mathbb{R}^n$

m^n on $\mathbb{R}^n$ is defined to be the *completion* of $m \times \ldots \times m$.

Theorem 2.5.2

m^n is translation-invariant.
The proof begins by checking rectangles and finite unions of them. Finite unions form an algebra $\mathcal{A}$. $\mu(E); = m^n(E + x)$ defines a premeasure. Since μ, m^n are σ-finite, μ can be uniquely txtended to $\mathcal{B}_{\mathbb{R}^n}$ so $\mu = m^n$ on $\mathcal{B}_{\mathbb{R}^n}$.

So do the completion.

Chapter 3

Signed Measures and Differentiation

3.1 Signed Measures

In physics, for example, when integrating over charges, we sometimes need $\mu(A) < 0$.
For $f \in L^+$ we know $E \mapsto \int_E f \, d\mu$ defines a measure. This should also give us "something like a measure" for functions that become negative sometimes. We therefore consider general *countably additive set functions* (functions taking sets as input that are countably additive for disjoint sets) that are not necessarily $\geqslant 0$. However, this poses some difficulties.

Example 3.1.1. Suppose μ is *finitely* additive on an algebra $\mathcal{A}$ (and it might take negative values). Suppose for some $A, B \in \mathcal{A}$ we have $\mu(A) = \infty, \mu(B) - \infty$. What can

$$\mu(A - B) \qquad \mu(A \cap B) \qquad \mu(B - A)$$

be? The first two add to ∞ whereas the last two add to $-\infty$. This is possible if and only if

$$\mu(A - B) = \infty \qquad \mu(A \cap B) \text{ finite} \qquad \mu(B - A) = -\infty.$$

However,

$$\mu(A \Delta B) = \mu(A - B) + \mu(B - A) = \infty - \infty$$

which contradicts finite additivity of disjoint sets.

Example 3.1.2. Suppose μ is a countably additive set function and $A_1, A_2, \ldots$ are disjoint. Split them into "positive" and "negative" sets:

$$\mathcal{J} := \{j : \mu(A_j) \geqslant 0\}.$$

Suppose

$$\sum_{j \in \mathcal{J}} \mu(A_j) = \infty \qquad \text{and} \qquad \sum_{j \notin \mathcal{J}} \mu(A_j) = -\infty.$$

Then

$$\mu(\bigcup_{j \in \mathcal{J}} A_j) = \infty \qquad \mu(\bigcup_{j \notin \mathcal{J}} A_j) = -\infty.$$

Again we have the problem encountered in the previous example when we try to take μ of the union of all A_j's.

We see that we run into issue unless one between

$$\sum_{j\in\mathcal{J}} \mu(A_j) \qquad \sum_{j\notin\mathcal{J}} \mu(A_j)$$

is finite. In particular, if $\sum_{j\geqslant 1} \mu(A_j) < \infty$, then both above must be finite. Furthermore, in this case the terms in this series must converge absolutely.

Definition 3.1.3: Signed measure

A **signed measure** on $(X, \mathfrak{M})$ is a function $\mu : \mathfrak{M} \to [-\infty, \infty]$ satisfying

(i) $\mu(\varnothing) = 0$,

(ii) μ assumes at most at most one of the values $\pm\infty$ (in other words we cannot have $\mu(A) = \infty$ and $\mu(B) = -\infty$ for some A, B), and

(iii) for $E_1, E_2, \ldots$ disjoint, $\mu(\bigcup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} \mu(E_j)$, with absolute convergence if the sum is finite.

Note that our previous (positive) measures all satisfy these criteria.

Example 3.1.4.

(1) A function $f : X \to [-\infty, \infty]$ is called **extended μ-integrable** if $\int f^+ \, \mathrm{d}\mu, \int f^- \, \mathrm{d}\nu$ are not both ∞. Then

$$\nu(E) := \int_E f \, \mathrm{d}\mu$$

defines a signed measure. f is called the **density** of ν with respect to μ.

(2) $\mu = \mu_1 - \mu_2$ where μ_1, μ_2 are positive measures and at least one being finite, is a signed measure.

Some important continuity properties of signed measures are inherited from positive measures:

(1) Signed measures are continuous from below.

(2) If $\mu(X) < \infty$, then μ is continuous from above.

Example 3.1.5. Let μ have density $\sin x$ with respect to the Lebesgue measure on $[0, 2\pi]$.
If we strict the domain to $[0, \pi]$, the map $E \mapsto \int_{[0,\pi]} \sin(x) \, \mathrm{d}m$ is a positive measure. If we restrict it to $[\pi, 2\pi]$ then μ is purely negative.
Question: can we always split the space into two parts, one on which the measure is positive, the other one on which the measure is negative?
Also, another complication: $\mu(E) = 0$ no longer implies E is "small": consider $[\pi/2, 3\pi/2]$ for example. What we get is instead a cancellation. This will give rise to our revised definition of null sets.

Definition 3.1.6: Positive, negative, & null set

A measurable set E is a **positive set** for μ if $\mu(F) \geqslant 0$ for all $F \subset E$. (Similar for negative set.)
It is called a **null set** if $\mu(F) = 0$ for all $F \subset E$.

Beginning of Oct.29, 2021

Example 3.1.7. Let A, B, μ be such that $\mu(A) = \mu(B) = 3$, $\mu(A \cap B) = 8$, and $\mu(A - B) = \mu(B - A) = -5$. Thus we can have $\mu(A), \mu(B) > 0$ but $\mu(A \cup B) < 0$!

Lemma

A countable union of positive sets is positive.

Proof. Suppose $E_1, E_2, \ldots$ are positive and $B \subset \bigcup_{j \geqslant 1} E_j$. We make the sets disjoint by defining $F_j := E_j - \bigcup_{i<j} E_i$. Then

$$\nu(B) = \sum_{j \geqslant 1} \nu(B \cap F_j) \geqslant 0$$

since $\nu(B \cap F_j) \geqslant 0$ for each j (it's a subset of E_j). □

Note for the upcoming proof: suppose ν is a signed measure and some set F contains *no* positive set but does contain a set of positive measure.
Let $E \subset F$ with $\nu(E) > 0$. By assumption E cannot be positive so it contains some H with $\nu(H) < 0$. This means if we *discard* it from E (resulting in $E - H$), the set gets *bigger*!

$$\nu(E - H) = \nu(E) - \nu(H) > \nu(E).$$

Thus, for all $E \subset F$ with $\nu(E) > 0$, the measure can be *increased* by discarding some set. This never happens for a positive measure.

Theorem 3.1.8: Hahn Decomposition Theorem

Let ν be a signed measure on $(X, \mathfrak{M})$. Then there exists a positive set P and a negative set N with $X = P \cup N$ and $P \cap N = \varnothing$. The decomposition is unique up to null sets. That is, if $X = P' \cup N'$ then $P \Delta P', N \Delta N'$ are both null.

Proof. We may assume that ν is never $+\infty$ (similar if it's never $-\infty$). Define m by

$$m := \sup\{\nu(E) : E \text{ is positive}\}.$$

(We don't know if it's achieved yet.) There exists a sequence of positive sets P_i such that $\nu(P_j) \to m$. Also, ν never attains $+\infty$ implies m is finite. By the previous lemma, $P := \bigcup_{j \geqslant 1} P_j$ is also positive. Since $P \supset P_j$ we know $\nu(P) \geqslant \nu(P_j)$. Therefore we must have $\nu(P) = m$, i.e. the supremum is achieved!
We are half way done with the decomposition. Now it remains to show that $N := P^c$ is a negative set.
First, notice that if N has a subset E with $\nu(E) > 0$ then E cannot be positive. Otherwise, $P \cup E$ would also be positive, contradicting the maximality of P.

In addition, this implies any $E \subset N$ with $\nu(E)$ contains a negative measure set C. That is, for $B = E - C$,

$$\nu(B) = \nu(E) - \nu(C) > \nu(E).$$

(Discarding C from E increases the measure!)

Now assume for contradiction that N is *not* a negative set. We can iterate this discarding. Let n_1 be the least integer such that there exists $E_1 \subset N$ with $\nu(E_1) > 1/n_1$. (This happens for some n_1 because N is not a negative set so it contains subsets with positive measure.)

Since E_1 cannot be positive, we can discard something from E_1 to make the measure bigger by our first observation: we pick n_2 to be the least integer such that there exists $E_2 \subset E_1$ with $\nu(E_2) > \nu(E_1) + 1/n_2$. (Note $n_1 < n_2$ by construction.)

Done iteratively, the sets E_i are getting smaller but $\nu(E_i)$ is increasing. We define

$$E := \bigcap_{j \geqslant 1} E_j$$

the undiscarded part. Continuity from above (recall we said ν cannot take $+\infty$) says $\nu(E_j) \uparrow \nu(E)$, so

$$\infty > \nu(E) = \lim_{j \to \infty} \nu(E_j) > \sum_{j \geqslant 1} \frac{1}{n_j}.$$

The series converges so we must have $n_j \to \infty$. Also, from the definition of n_j (the smallest integer that...), we have

$$\frac{1}{n_j - 1} \geqslant \nu(E_j - E_{j-1}) = \nu(E_j) - \nu(E_{j-1}) > \frac{1}{n_j}.$$

So it is impossible to discard any larger set from E_j.

However, we can discard *something* from E. Since E has positive measure, by our first observation, there exists $\epsilon > 0$ and $D \subset E$ where $\nu(D) > \nu(E) + \epsilon$ (discard the negative set $E - D$). Since $D \subset E_j$, this means

$$\epsilon < \frac{1}{n_j - 1} \qquad \text{for all } j.$$

This contradicts $n_j \to \infty$. Hence N must be a negative set. This proves the existence claim.

For uniqueness, suppose P, N, P', N' satisfy the assumptions. Then $P - P' \subset P$ so it's a positive set and $P - P' = N' - N \in N'$ so it's a negative set. Hence every subset of $P - P'$ has measure both $\geqslant 0$ and $\leqslant 0$, i.e., they must have measure 0. This means $P - P'$ is null. Similar for N' and N. ☐

We say $X = P \cup N$ is the **Hahn decomposition** of X with respect to the measure μ.

Definition 3.1.9: Variations

Let $X = P \cup U$ according to the Hahn decomposiion. We define the **positive variation** and **negative variation** of ν by

$$\nu^+(E) := \nu(E \cap P) \qquad \nu^-(E) = -\nu(E \cap N).$$

This implies ν^+, ν^- are positive measures and $\nu(E) = \nu^+(E) - \nu^-(E)$.

We define

$$|\nu|(E) := \nu^+(E) + \nu^-(E)$$

to be the **total measure**.

Remark. $|\nu|(E)$ is not the same as $|\nu(E)|$! For example consider

$$\nu(E) := \int_E \sin x \, dx \qquad \text{on } [0, 2\pi].$$

Let $E = [\pi/2, 3\pi/2]$. Clearly $P = [0, \pi]$ and $N = (\pi, 2\pi]$. Then

$$\nu^+(E) = \nu([\pi/2, \pi]) = 1 = -\nu((\pi, 3\pi/2])\nu^-(E)$$

so $|\nu|(E) = 2$. However, $|\nu(E)| = 0$.

Definition 3.1.10: Mutual singularity

Two measures μ, ν are **mutually singular**, written $\mu \perp \nu$, if there exist disjoint A, B with $A \cup B = X$ such that A is null for μ and B is null for ν. That is, μ "lives on" B and ν "lives on" A.

Example 3.1.11. Let ν be a signed measure. Then ν^+ lives on P and ν^- lives on N. $X = P \cup N$ is disjoint, so $\nu^+ \perp \nu^-$.
This implies Hahn decomposition gives $\nu = \nu^+ - \nu^-$, so we can express ν as a difference of two mutually singular positive measures. Question: is this unique?
Suppose also that $\nu = \mu^+ - \nu^-$ with μ^+, μ^- veubg mutually singular positive measures. Suppose μ^+ lives on E and μ^- on $F = E^c$. But then E is a positive set for ν and F a negative set, with $X = E \cup F$. By Hahn decomposition, the decomposition is unique up to null sets, so $\mu^+ = \nu\big|_E = \nu\big|_P = \nu^+$ (since $E \Delta P$ is null) and similarly $\nu^- = \nu\big|_N = \nu^-$. Therefore the decomposition is unique!

Theorem 3.1.12: Jordan Decomposition Theorem

Every signed measure ν can be *uniquely* decomposed as $\nu = \nu^+ - \nu^-$ with ν^+, ν^- positive and mutually singular.

Example 3.1.13. Let m be the Lebesgue measure on $[0, 1]$. Let μ be the "uniform measure on Canter set $\mathcal{C}$". Then μ is supported on $\mathcal{C}$ (null for m) and m is supposed on $\mathcal{C}^c$ (null for μ), so $\mu \perp m$.

Example 3.1.14. Consider the sum of point masses. Recall

$$\delta_x(A) := \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

Suppose $\mu := \sum_{i=1}^{\infty} a_i \delta_{x_i}$. Let $B = \{x_1, x_2, \ldots\}$. Then $\mu(A)$ is the amount of mass in A where there is mass a_i at each x_i. Then μ lives on B ($\mu = 0$ on B^c) and m lives on B^c (it is null on a countable set), so $\mu \perp m$.

Beginning of Nov.1, 2021

Example 3.1.15. Let μ, ν be finite signed measures. Then $|\mu + \nu| \leqslant |\mu| + |\nu|$.

Proof. By decomposition, $\rho = \rho^+ - \rho^- \leqslant \rho^+ + \rho^- = |\rho|$ and similarly $-\rho \leqslant |\rho|$. Therefore $|\rho(A)| \leqslant |\rho|(A)$. (The first one is absolute value; second one is total variation.) Thus, for subsets for P,

$$|\mu + \nu| = \mu + \nu \leqslant |\mu| + |\nu|$$

and for subsets of N,

$$|\mu + \nu| = -\mu - \nu \leqslant |\mu| + |\nu|.$$

Together we recover the original claim since any E splits into $E \cap P$ and $E \cap N$. ☐

3.2 Absolute Continuity

This is heuristically the opposite of being mutually singular.

Definition 3.2.1: Absolute Continuity

Let μ be a fixed *positive* measure and let ν be a signed measure on $(X, \mathfrak{M})$. We say ν is **absolutely continuous** with respect to μ, written $\nu \ll \mu$, if $\mu(E) = 0$ implies $\nu(E) = 0$.

Remark. $\nu \ll \mu$ means that if μ lives on E then ν lives on E too.

Remark. This definition is related to having the notion of density: recall that "ν has a density with respect to μ" if there exists f such that

$$\nu(E) = \int_e f \, \mathrm{d}\mu \qquad \text{for all } E \in \mathfrak{M}, \tag{1}$$

that is, $\int_E f \, \mathrm{d}\mu = \int_E f \, \mathrm{d}\mu$ or $\mathrm{d}\nu = f \mathrm{d}\mu$ for shorthand.

Under (1), if $\mu(E) = 0$ then $\nu(E) = 0$, so we obtain absolutely continuous measures from the notion of density.

Question. Are there other ways to get $\nu \ll \mu$? (The answer is no, at least for the σ-finite case. See Lebesgue-Radon-Nikodym later.)

Definition 3.2.2: Equivalent measures

We say μ and ν (both positive) are **equivalent** if $\mu \ll \nu$ and $\nu \ll \mu$. In particular, they have the same sets: μ lives on E if and only if ν lives on E.

Example 3.2.3. Let $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x & x \in [0, \infty) \\ 0 & x \in (-\infty, 0). \end{cases}$$

Let $\mathrm{d}\nu = f\mathrm{d}m$. Then $m(E) = 0 \Rightarrow \nu(E) = 0$, so $\nu \ll m$. However,

$$\nu((-\infty, 0)) = 0 \text{ but } m((-\infty, 0)) = \infty \neq 0,$$

so $m \not\ll \nu$. m and ν are not equivalent.

Example 3.2.4. Following the previous example, if $\mathrm{d}\nu = f\mathrm{d}\mu$ and $f > 0$ μ-a.e., then μ,ν are equivalent.

Example 3.2.5. Consider two random processes evolving in time: $Y(t), Z(t), t \in [0,1]$. That is, Y, Z may be viewed as mappings from some $(X, \mathfrak{M}, \mathbb{P})$ to $C([0,1])$. Consider the induced push-forward measures

$$\mu(A) := \mathbb{P}(Y \in A) \qquad \nu(A) = \mathbb{P}(Z \in A).$$

Suppose we observe a process over $t \in [0,1]$ and we don't know if it's produced by Y or Z.
If $\mu \perp \nu$, then we can (in principle) tell *in one observation* whether we say Y or Z, as there exists a partition of $C([0,1])$ into $A \cup B$ such that μ lives on A and ν lives on B, except for null sets. Then if our observed process is in A if and only if the process is Y and likewise for B and Z.
A typical result: if $Y, Z \sim$ Gaussians, then either $\mu \perp \nu$ or μ, ν are equivalent.

Question. Why the word "continuity" in absolute continuity?

Theorem 3.2.6: $\ll$ is related to continuity

Let μ be positive and ν a finite signed measure. Then

$$\nu \ll \mu \iff \text{for all } \epsilon > 0, \text{ there exists } \delta > 0 : \mu(E) < \delta \Rightarrow |\nu(E)| < \epsilon.$$

Proof. It suffices to prove this for $|\nu|$ because we showed previously that $|\nu(E)| \leqslant |\nu|(E)$. Hence, we may assume that ν is a positive measure too.
If the $\epsilon - \delta$ holds, and if $\mu(E) = 0$, then this trivially implies $\mu(E) < \delta$ for all δ, and $\mu(E) < \epsilon$ for all ϵ, so $\nu(E) = 0$, i.e., $\nu \ll \mu$, as claimed.
For the other direction, suppose the $\epsilon - \delta$ fails. We want to find E with $\mu(E) = 0$ but $\nu(E) > 0$ (so $\nu \not\ll \mu$). Failure means that for some $\epsilon > 0$, no δ works. In particular, for all n, there exists E_n with $\mu(E_n) < 2^{-n}$ but $\nu(E_n) \geqslant \epsilon$. Define

$$E := \limsup_{n\to\infty} E_n \bigcap_{k\geqslant 1} \overbrace{\bigcup_{n\geqslant k} E_n}^{:=F_k} = \{x : x \in E_n \text{ infinitely many times}\}.$$

It is clear that for all k,

$$\mu(E) \leqslant \mu(F_k) \leqslant \sum_{n\geqslant k} \mu(E_n) \leqslant 2^{-k+1},$$

so $\mu(E) = 0$. However, continuity from above says (recall ν is finite!)

$$\nu(E) = \nu(\bigcap_{k\geqslant 1} F_k) = \lim_{k\to\infty} \nu(F_k) \geqslant \limsup_{k\to\infty} \nu(E_k) \geqslant \epsilon.$$

This provides a contradiction. □

3.3 The Lebesgue-Radon-Nikodym Theorem

Theorem 3.3.1: Lebesgue-Radon-Nikodym

Let μ be a positive measure and let ν be signed, *both σ-finite*. Then there exists a *unique* decomposition $\nu = \lambda + \rho$ with $\lambda \perp \mu$ and $\rho \ll \mu$, and ρ has a density: that is, there exists an extended μ-integrable f (that is, f^+ and f^- are both both ∞ upon integration) such that $\mathrm{d}\rho = f\mathrm{d}\mu$. Also, f is unique μ-a.e.

Remark. μ should be viewed as an "underlying measure". λ is singular and ρ is absolute continuous, both *relative to* μ.

Remark. If $\nu \ll \mu$, then one decomposition is $\nu = 0 + \nu$. Uniqueness says this is the only one, and the theorem says ν has a density, $\mathrm{d}\nu = f\mathrm{d}\mu$. This is the **Radon-Nikodym Theorem**.

Remark. λ is called the **singular part** of ν and ρ the **absolutely continuous part**.

Remark. An identity we'll use in the proof later:

$$\int_F \chi_E \,\mathrm{d}\mu = \int \chi_{E\cap F} \,\mathrm{d}\mu = \mu(E \cap F).$$

Remark. Let μ be positive and σ-finite on $\mathbb{R}$. The **atoms** of μ are $\{x : \nu(\{x\}) > 0\}$. Then σ-finiteness implies there are at most countably many atoms, say mass a_i at each x_i. Then $\sum\limits_{i\geqslant 1} \alpha_i \delta_{x_i}$ is the **discrete part** of ν.
If we decompose ν relative to m: $\mathrm{d}\nu = \lambda + f\mathrm{d}m$ with $\lambda \perp m$, i.e., λ lives on a m-null set. Subtracting the discrete part from λ,

$$\gamma = \lambda - \sum_{i\geqslant 1} \alpha_i \delta_{x_i}.$$

This remaining γ lives on a m-null set (since λ does) but it has no atoms. Such a measure is called **continuous singular**. Thus we have

$$\mathrm{d}u = \underbrace{\sum_{i\geqslant 1} \alpha_i \delta_{x_i}(\text{discrete}) + \gamma(\text{cont. singular})}_{\text{singular}} + \underbrace{f\mathrm{d}m}_{\text{abs. cont.}}.$$

Remark. Notation: we write $f \vee g := \max\{f, g\}$, the pointwise maximum.

The main difficulty in proving the L.-R.-N. Theorem: we have to come up with a function f with *certain* properties, when there is no function explicitly mentioned in the hypothesis. Where to get it?

Lemma

Let μ, ν be finite positive measures on $(X, \mathfrak{M})$. Then either $\mu \perp \nu$ or there exists $\epsilon > 0$ and $E \in \mathfrak{M}$ with $\mu(E) > 0$ such that $\nu \geqslant \epsilon\mu$ on E. That is, $\nu(F) \geqslant \epsilon\mu(F)$ for all $F \subset E$, or equivalently $\nu - \epsilon\mu$ is a positive measure when restricted on E.

Alternatively, we can say $\mathrm{d}\nu \geqslant \epsilon\chi_E \mathrm{d}\mu$ on all of X, since

$$\nu(A) \geqslant \nu(A \cap E) \geqslant \epsilon\mu(A \cap E) = \int_A \epsilon\chi_E \, \mathrm{d}\mu.$$

To put informally, there is "*a little bit of density*" inside ν: if $\nu \pm \mu$, then

$$\mathrm{d}\nu = \underbrace{\epsilon\chi_E}_{\text{density}} \mathrm{d}\mu + (\text{some positive measure}).$$

Beginning of Nov.3, 2021

Idea of proof of L-R-N. Consider the case $\nu \ll \mu$. If a density f exists, how do we identify it?

If some function g is "too big" ($g > f$ on some set E of positive measure), then

$$\int_E g \, \mathrm{d}\mu > \nu(E).$$

Equivalently, if for all E, $\int_E g \, \mathrm{d}\mu \leqslant \nu(E)$ then $g \leqslant f$ while $\int_E f \, \mathrm{d}\mu = \nu(E)$. That is, f is the *largest* function with $\int_E f \, \mathrm{d}\mu \leqslant \nu(E)$ for all E. However, we cannot use this as a definition because we don't know whether such f exists or not a priori. (This description still holds even if $\nu \not\ll \mu$.)

Proof of Lemma. For all n, consider $\nu - \mu/n$. It's a signed measure so it admits a Hahn decomposition $P_n \cup N_n$. Let $P = \bigcup_{n\geqslant 1} P_n$ and $N = \bigcap_{n\geqslant 1} N_n$. Clearly N is a negaitve set for all $\nu - \mu/n$. If $E \subset N$ then $\nu(E) \leqslant \mu(E)/n$ for all n. Hence $\nu(E) = 0$. Since E is arbitrary, $\nu(N) = 0$.

If $\mu(P) = 0$ then $\mu \perp \nu$.

Otherwise, if $\mu(P) > 0$, then $\mu(P_n) > 0$ for some n. Then P_n is a positive set for $\nu - \mu/n$, so in particular $(\nu - \mu/n)(P_n) \geqslant 0$, i.e., $\nu \geqslant \mu/n$ on P_n. □

Proof of L-R-N. We first assume that the measures are positive and finite. For this case, we first try to find a density f. Define

$$\mathcal{F} := \{f : X \to [0, \infty] : \int_E f \, \mathrm{d}\mu \leqslant \nu(E) \text{ for all } E \in \mathfrak{M}\}.$$

Claim: if $f, g \in \mathcal{F}$ then the pointwise maximum $f \vee g \in \mathcal{F}$. (This shows that there cannot be two different maximal functions.) Indeed, if so, for all E,

$$\begin{aligned}\int_E f \vee g \, \mathrm{d}\mu &= \int_{E\cap\{f\geqslant g\}} f \, \mathrm{d}\mu + \int_{E\cap\{f<g\}} g \, \mathrm{d}\mu \\ &\leqslant \nu(E \cap \{f \geqslant g\}) + \nu(E \cap \{f < g\}) = \nu(E).\end{aligned}$$

Now we look for the largest f by looking for the f maximizing $\int f \, \mathrm{d}\mu$. Define

$$a := \sup\{\int f \, \mathrm{d}\mu : f \in \mathcal{F}\}.$$

By definition there exists a sequence $\{f_n\}$ in $\mathcal{F}$ whose integrals converge to a. Define $g_n := f_1 \vee \ldots \vee f_n$ and

$f := \sup g_n$. (This is well-defined because the pointwise limit either exists or is infinite.) Hence $g_n \uparrow f$ and $g_n, f \in \mathcal{F}$.
Now, for all $h \in \mathcal{F}$, $\int_X h \, \mathrm{d}\mu \leqslant \nu(X)$, so $\alpha \leqslant \nu(X) < \infty$ (we assumed ν to be finite for now). Thus, we have

$$\alpha \geqslant \int g_n \, \mathrm{d}\mu \geqslant \int f_n \, \mathrm{d}\mu \to a,$$

which, by MCT,

$$\int f \, \mathrm{d}\mu = \lim_{n\to\infty} \int g_n \, \mathrm{d}\mu = a.$$

Since $a < \infty$, we have $f < \infty$ a.e.-μ. This proves the existence of a maximal f.

Now we show that f is the "right density function". That is, λ given by $\mathrm{d}\lambda = \mathrm{d}\nu - f\mathrm{d}\mu$ satisfies $\lambda \perp \mu$. Suppose λ and μ are not singular. By the previous lemma, there exists $\epsilon > 0$ and a set E such that $\mathrm{d}\lambda = \epsilon\chi_E \mathrm{d}\mu + \mathrm{d}\tau$ for some positive measure τ. That is,

$$\mathrm{d}\nu = f\mathrm{d}\mu + \epsilon\chi_E \mathrm{d}\mu + \mathrm{d}\tau$$

where the term $(f + \epsilon\chi_E)\mathrm{d}\mu$ implies that f is not the maximal element in $\mathcal{F}$. Contradiction, so $\lambda \perp \mu$ as desired.

For uniqueness: suppose $\mathrm{d}\lambda + f\mathrm{d}\mu = \lambda' + f'\mathrm{d}\mu$ with $\lambda, \lambda' \perp \mu$, then

$$(\lambda - \lambda') = (f' - f)\mathrm{d}\mu.$$

The LHS is something singular and the RHS something absolutely continuous, so both must be zero. Uniqueness a.e. follows.

For the positive and σ-finite case, we have, by definition, $A_1 \subset A_3 \subset \ldots$ with $\mu(A_n) < \infty$ and $\bigcup_{n\geqslant 1} A_n = X$ and similarly $B_1 \subset B_2 \subset$ with $\nu(B_n) < \infty$ and $\bigcup_{n\geqslant 1} B_n = X$.
We define

$$C_n := A_n - \bigcup_{m<n} A_m \qquad D_n = B_n - \bigcup_{m<n} B_m$$

so C_n's are dis disjoint and D_n's are disjoint, and they both fill up the whole space. Hence

$$X = \bigcup_{m,n} (C_n \cap D_m).$$

Name these sets as $E_1, \ldots, E_n$ so that $\mu(E_n), \nu(E_n) < \infty$. For each E_n, we can apply the previous case, giving $\nu_j = \lambda_j + f_j \mathrm{d}\mu$ with $f_j = 0$ on E_j^c. By definition each λ_j is supported on E_j so $\lambda_j \perp \mu$.
Now we combine all of them! Define

$$\lambda = \sum_{j\geqslant 1} \lambda_j \qquad f := \sum_{j\geqslant 1} f_j.$$

(In particular $f \equiv f_j$ on E_j.) Then $\lambda \perp \mu$ and $\nu = \sum_{j\geqslant 1} \nu_j = \lambda + f\mathrm{d}\mu$.

Finally, for the more general case, we apply above to ν^+ and ν^- separately. □

Remark. When $\nu \ll \mu$, L-R-N guarantees the existence of a function f with $\mathrm{d}\nu = f\mathrm{d}\mu$. We can write f as $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$. Then the very definition becomes

$$\nu(E) = \int_E \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\, \mathrm{d}\mu \qquad \text{for all } E \in \mathfrak{M}.$$

Remark. Making sense of $\mathrm{d}\nu = f\mathrm{d}\mu$ — does it mean that $\int g\, \mathrm{d}\mu = \int gf\, \mathrm{d}\mu$? At least the definition implies that this holds for indicator functions $g = \chi_E$.

Lemma

Suppose μ is a positive measure and $\mathrm{d}\nu = f\mathrm{d}\mu$ is a signed measure. Then for all $g \in L^1(\nu)$, we have $fg \in L^1(\mu)$ and

$$\int g\, \mathrm{d}\nu = \int gf\, \mathrm{d}\mu.$$

Note that for signed ν, $\int g\, \mathrm{d}\nu$ means $\int g\, \mathrm{d}\nu^+ - \int g\, \mathrm{d}\nu^-$ when both are finite.

Proof. Since $f = f^+ - f^-$, we may assume μ is positive and $f \geqslant 0$.
The claim holds for indicator functions. If $g_n \uparrow g$ then $g_n f \uparrow gf$, so by MCT, the claim holds for measurable functions $g \geqslant 0$. The general case follows by decomposing g into g^+ and g^-. □

Remark. **Chain rule?** Does $\frac{\mathrm{d}\nu}{\mathrm{d}\lambda} = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$, where μ, λ (denominators) are positive measures?

Proposition 3.3.2

Let μ, λ be σ-finite positive measures and ν a σ-finite signed measure. If $\nu \ll \mu$ and $\mu \ll \lambda$ then $\mu \ll \lambda$ and the chain rule holds. (We need σ-finite for everything because of LRN.)

Proof. We may assume $\nu \geqslant 0$ (otherwise decompose it). Applying the previous lemma to $g = \chi_E \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ gives

$$\nu(E) = \int \chi_E\, \mathrm{d}\nu = \int_E \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\, \mathrm{d}\mu = \int_E \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\, \mathrm{d}\lambda.$$

□

Remark. If $|\nu|(E) = 0$ then $\nu^+(E) = \nu^-(E) = 0$ so $\nu(E) = 0$. That is, $\nu \ll |\nu|$. LRN says $\frac{\mathrm{d}\nu}{\mathrm{d}|\nu|}$ is well-defined. What does it look like?
Again we decompose $X = P \cup N$. On P, $|\nu| = \nu$ so for all $E \subset P$, $\nu(E) = |\nu|(E)$ so $\frac{\mathrm{d}\nu}{\mathrm{d}|\nu|} = 1$ on P. Similarly we can show that $\frac{\mathrm{d}\nu}{\mathrm{d}|\nu|} = -1$ on N.

Beginning of Nov.5, 2021

Application of R-N in Probability

Let $(X, \mathfrak{M}, \mu)$ be a measure space and let $\mathfrak{N} \subset \mathfrak{M}$ be another σ-algebra. Suppose $f \geqslant 0$ is $\mathfrak{M}$-measurable but not $\mathfrak{N}$-measurable. Define

$$\nu(E) ;= \int_e f \, \mathrm{d}\mu$$

a measure on $\mathfrak{M}$ with $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} = f$.

Key point: when ν is a measure on some σ-algebra $\mathcal{F}$, then $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ must be $\mathcal{F}$-measurable. Hence if we restrict μ, ν to $\mathfrak{N}$ and call then μ_n and ν_n, then $\frac{\mathrm{d}\nu_n}{\mathrm{d}\mu_n}$ is *not* f. it must be some other $\mathfrak{N}$-measurable function g satisfying

$$\nu(E) = \int_E f \, \mathrm{d}\mu = \int_E g \, \mathrm{d}\mu \qquad \text{for all } E \in \mathfrak{N}.$$

Example 3.3.3. Let $X = [0,1]$, m the Lebesgue measure, and let A_1, A_2, A_3 be a partition of X. Let $\mathfrak{M}$ be the Borel sets and $\mathfrak{N} := \sigma(A_1, A_2, A_3)$.
First observation: $\mathfrak{N}$-measurable functions need to be constant on each A_i.
Now we ask — what $\mathfrak{N}$-measurable g has the same integral as f over each A_i? That is,

$$g(x) = \text{ average of } f \text{ on } A_i = \frac{1}{m(A_i)} \int_{A_i} f \, \mathrm{d}m \qquad \text{for all } x \in A_i, i \in \{1,2,3\}.$$

Then such a piecewise constant function g is $\mathfrak{N}$-measurable and has the same integral as f over any union of A_i's.
Connection to probability: if we only have partial information about a random $x \in X$ (we know whether $x \in E$ only for sets in $E \in \mathfrak{N}$, not for all $E \in \mathfrak{M}$; in this case we only know if x lies in A_1, A_2, or A_3, but not exactly where on X), then g is **expected value** of f given this information. Hence the Radon-Nikodym derivative $g = \frac{\mathrm{d}\nu_n}{\mathrm{d}\mu_n}$ defines the **conditional expectation** of f given $\mathfrak{N}$.

3.4 Complex Measure

For a complex measure, we require $\nu(E) \in \mathbb{C}$ for all E (so it cannot attain ∞) and we require absolute continuity for disjoint sets, i.e.,

$$\nu(\bigcup_{j \geqslant 1} E_j) = \sum_{j \geqslant 1} \nu(E_j).$$

The measure should also be able to be expressed as $\nu = \nu_r + i\nu_i$ with ν_r, ν_i being (real-valued) signed measures.

Lemma

For ν a complex measure, there exists a function θ and a *unique* positive measure $|\nu|$ such that

$$\nu(\mathrm{d}x) = e^{i\theta(x)} |\nu|(\mathrm{d}x).$$

Proof. If ν has a density $f(x) = r(x)e^{i\theta(x)}$ then this claim is trivial.
More generally, we choose μ positive with $\nu \ll \mu$, say $\mu = |\nu_r| + |\nu_i|$ (this indeed works) and let $f = \frac{d\nu}{d\mu}$. We write $f(x)$ as $r(x)e^{i\theta(x)}$ so $r(x) = |f(x)|$ and define $|\nu|$ by $d|\nu| = r d\mu$. Then

$$e^{i\theta} d|\nu| = re^{i\theta}\mu = d\nu.$$

Uniqueness omitted. □

3.5 Differentiation (not the usual kind) on $\mathbb{R}^n$

For a continuous function on $\mathbb{R}$ and $x \in \mathbb{R}$, we know that

$$\lim_{r\to 0} \frac{1}{2r} \int_{x-r}^{x+r} f(u)\, du = f(x).$$

In measure terms, letting $I_r := (x-r, x+r)$ and $d\nu = f dm$, we have

$$\frac{\nu(I_r)}{m(I_r)} = \int_{I_r} f\, dm/m(I_r) \to f(x) \qquad \text{as } r \to 0.$$

What happens in general (in $\mathbb{R}^n$) if f is measurable but not necessarily continuous? What if ν has a singular point?

Consider a measurable f and $d\nu = f dm$. Let $B(r,x)$ be the ball of radius r centered at x and $S(r,x)$ the corresponding sphere. We say f is **locally integrable** if

$$\int_K f\, dm < \infty \qquad \text{for all compact } K$$

and we write $L^1_{\text{loc}} = \{$ all locally integrable functions$\}$. For example, $f(x) := x^2 \in L^1_{\text{loc}}(m)$. Define the average

$$A_r f(x) := \text{ average of } f \text{ on } B(r,x) = \frac{1}{m(B(r,x))} \int_{B(r,x)} f\, dm.$$

We know that the above converges to 0 for continuous f. What about general $f \in L^1(m)$? First, by density of continuous functions in L^1, there exists a continuous g with

$$\int |f-g|\, dm < \epsilon.$$

Then,

$$|A_r f(x) - f(x)| \leqslant |A_r f(x) - A_r g(x)| + |A_r g(x) - g(x)| + |g(x) - f(x)|.$$

The second term $\to 0$ because g is continuous. By construction,

$$\int |f-g|\, dm$$

is small, so $|g(x) - f(x)|$ is small "for most x". Does this imply $|A_r(f-g)(x)|$ is small "for most x"?
We define the **Hardy-Littlewood Maximal Function**

$$(\text{HF})(x) := \sup_{r>0} A_r|f|(x) = \text{ largest average of } |f| \text{ on a ball.}$$

If f is continuous then clearly $|f|$ is bounded above by HF.

Theorem 3.5.1: Maximal Theorem

For all n, there exists $c > 0$ such that for all $f \in L^1(m)$ on $\mathbb{R}^n$ and for all $\alpha > 0$, we have

$$m(\{x : \mathrm{HF}(x) > \alpha\}) \leqslant \frac{c}{\alpha} \int |f| \, \mathrm{d}m.$$

(To be proven later.)

Applying this to $\int |f - g| \, \mathrm{d}m$ which can be made arbitrarily small, then

$$\{x : \mathrm{HF}(f-g)(x) > \alpha\}$$

can be made arbitrarily small, so

$$\{x : A_r(f-g)(x) > \alpha\}$$

can also be made arbitrarily small.

The following lemma will be used to prove the maximal theorem:

Lemma

Let $\mathcal{C}$ be a collection of open balls in $\mathbb{R}^n$ and $U := \bigcup_{B \in \mathcal{C}} B$. Then for all $c < m(U)$, there exist disjoint balls $B_1, \dots, B_k \in \mathcal{C}$ with

$$\sum_{j=1}^{k} m(B_j) \geqslant 3^{-n} c.$$

Example 3.5.2. If $d = 2$, we can make disjoint balls that cover almost $1/9 = 3^{-2}$ of U. In $d = 3$ this becomes $1/27$.

Beginning of Nov.8, 2021

Proof.

Idea: if we triple the radius of each B_k then we can cover almost all of U.

Since m is regular, there exists a compact K with $c < m(K) \leqslant m(U)$. By compactness there exists a finite subcover of K, say $K \subset \bigcup_{j \leqslant J} B_j$ with $B_j \in \mathcal{C}$.

We remove balls from the subcover one at a time. First we remove the largest one, and then we remove the largest *whole* ball remaining, and so on, until there is no *whole* ball remaining. Relabel the removed ones to be $B_1, \dots, B_k$. Let the centers be $z_1, \dots, z_k$. We claim that these removed B_i's cover K.

Suppose for contradiction that $x \in K$ is not covered by $\bigcup_{i=1}^{k} B_i$. We know $x \in B$ for some $B \in \mathcal{C}$, and B must intersect some B_j (take B_j largest if B intersects multiple B_j's). Then by construction B must be smaller than B_j, and

$$d(x, z_j) \leqslant \mathrm{radius}(B_j) + \mathrm{diam}(B_j) \leqslant 3 \cdot \mathrm{radius}(B_j).$$

Thus the balls centered at z_j with thrice the radii cover K. That is,

$$3^n m(\bigcup_{j \leqslant k} B_j) \geqslant m(K) > c.$$

□

Proof of Maximal Theorem. Define $E_\alpha := \{x : \mathrm{Hf}(x) > \alpha\}$. If $x \in E_\alpha$, then there exists a ball $B(r_x, x)$ where the average of $|f|$ exceeds α. From the lemma above, there exist disjoint balls $B_j := B(r_{x_j}, x_j)$ covering almost a 3^{-n} fraction of E_α. That is, given $\epsilon > 0$, there exist finitely many B_j's such that

$$\frac{1}{(1+\epsilon)3^n} m(E_\alpha) \leqslant m(\bigcup_j B_j).$$

Hence, the average of $|f| \geqslant \alpha$ on B_j implies

$$\alpha m(\bigcup_j B_j) \leqslant \int_{\bigcup B_j} |f| \,\mathrm{d}m.$$

Therefore,

$$m(E_\alpha) \leqslant \frac{(1+\epsilon)3^n}{\alpha} \int |f| \,\mathrm{d}m.$$

Setting $c := (1+\epsilon)3^n$ gives our claim. □

Application of Maximal Theorem

Theorem 3.5.3

If $f \in L^1_{\text{loc}}(m)$ on $\mathbb{R}^n$, then $A_r(f)x \to f(x)$ as $r \to 0$ for a.e. x.

Proof. It is sufficient to show that the claim holds for a.e. x with $|x| \leqslant N$ for all N. We may assume $f \in L^1$. Let $\epsilon > 0$ and $\alpha > 0$. By density there exists a continuous g with

$$\int |f - g| \,\mathrm{d}m < \epsilon.$$

Using triangle inequality, it suffices to show that

$$m(\{x : \limsup_{r\to 0} |A_r(f-g)| > \alpha/2\})$$

$$m(\{x : \limsup_{r\to 0} |A_r g(x) - g(x)| > 0\})$$

$$m(\{x : |g(x) - f(x)| > \alpha/2\})$$

are all small. The first one, by the maximal theorem, is bounded by $2C\epsilon/\alpha$ for some C. The second term = 0 by continuity of g. The third one, by Markov's inequality, is bounded by

$$\frac{2}{\alpha} \int |g - f| \,\mathrm{d}m < \frac{2\epsilon}{\alpha}.$$

Hence

$$m(\{x : \limsup_{r\to 0} |A_r f(x) - f(x)| > \alpha\}) \leqslant \frac{2C\epsilon}{\alpha} + 0 + \frac{2\epsilon}{\alpha}.$$

Since ϵ is arbitrary, the claim holds for any $\alpha > 0$. Hence the claim holds. □

Remark. We can replace the ball $B(r, x)$ (for average) by any set E_r occupying some minimal fraction of the ball. For example, as $r \to 0$, as long as all E_r occupy $1/10$ of $B(r, x)$, i.e.,

$$\frac{m(E_r)}{m(B(r,x))} > \beta > 0 \qquad \text{for all } r.$$

If so, we say E_r **shrinks nicely** to x, and the original claim still holds.

Theorem 3.5.4: Lebesgue Differentiation Theorem

For $f \in L^1_{\text{loc}}$, then for a.e. x, for all $\{E_r\}$ which shrinks nicely to x, the averages

$$\frac{1}{m(E_r)} \int_{E_r} |f(y) - f(x)| \, m(\mathrm{d}y) \to 0.$$

We replaced (average of f converging to f) by (average of $|f - f(x)|$ converging to 0).

Beginning of Nov.10, 2021

Recall that a regular measure satisfies

$$\mu(E) = \sup\{\mu(K) : K \subset E \text{ compact }\} = \inf\{\mu(U) : U \supset E \text{ open}\}.$$

In $\mathbb{R}^n$, for Borel sets, the above is equivalent to

$$\mu(K) < \infty \text{ for all compact } K \qquad \mu(E) = \inf\{\mu(U) : U \supset K \text{ open}\}.$$

Theorem 3.5.5

Let ν be a regular signed or complex measure on $\mathbb{R}^n$ with Lebesgue-Radon-Nikodym representation $\mathrm{d}\nu = \mathrm{d}\lambda + f\mathrm{d}m$ where the first part $\perp m$ and the second $\ll m$. Then for m-a.e. $x \in \mathbb{R}^n$, for all $\{E_r\}$ that shrinks nicely to x,

$$\lim_{r \to 0} \frac{v(E_r)}{m(E_r)} = f(x)$$

(as if $\mathrm{d}\lambda$ weren't there).

Proof. Since ν is finite on compact sets, so is $f\mathrm{d}m$, so $f\mathrm{d}m$ is regular and $f \in L^1_{\text{loc}}$. Thus the $f\mathrm{d}m$ part gives the limit $f(x)$, and it remians to show

$$\frac{\lambda(E_r)}{m(E_r)} \to 0 \qquad \text{for a.e.} x, \{E_r\}.$$

By definition of "shrinking nicely", there fists $\beta > 0$ such that $m(E_r) \geqslant \beta m(B(r,x))$, so

$$\left|\frac{\lambda(E_r)}{m(E_r)}\right| \leqslant \frac{|\lambda|(E_r)}{\beta m(B(r,x))} \leqslant \frac{|\lambda|(B(r,x))}{\beta m(B(r,x))}.$$

Hence it suffices to prove the limit claim for balls only. Since $\lambda \perp m$, λ lives on a m-null set A^c (so m lives on A). We will show that

$$F_k := \{x \in A : \limsup_{r \to 0} \frac{|\lambda|(B(r,x))}{m(B(r,x))} \geqslant 1/k\}$$

has measure $m(F_k) = 0$ for all k. We enlarge A to an open set $U \subset A$ with $\lambda(U) < \lambda(A) + \epsilon = \epsilon$.
By definition, for all $x \in F_k$, there exists a ball B_x centered at x and contained in U such that $\lambda(B_x)/m(B_x) > 1/k$, i.e., $m(B_x) < k\lambda(B_x)$. Fix k and let $V = \bigcup_{x \in F_k} B_x$. Then $V \subset U$ covers F_k. By the previous covering lemma, there exist disjoint balls $B_{x_1}, \ldots, B_{x_j}$ covering almost $1/3^n$ fraction of V. Since

$$m(F_k) \leqslant m(V) \leqslant 2 \cdot 3^n m(\bigcup_{i \leqslant j} B_{x_i}) < 2 \cdot 3^n k \lambda(\bigcup_{i \leqslant j} B_{x_i}) \leqslant \epsilon(U) = \epsilon,$$

and since ϵ is arbitrary, we must have $m(F_k) = 0$. ☐

Example 3.5.6. Let $H(x)$ be the distribution function of the "uniform measure" on the Cantor set $\mathcal{C}$. Define $F(x) = H(x)+x$, so $\mu_F = \mu_H+m$. By L-R-N this is the unique decomposition of μ_F, as $m \ll m$ and $\mu_H \perp m$.The previous theorem says

$$\lim_{r\to 0} \frac{\mu_F((x,x+r])}{m((x,x+r])} = \lim_{r\to 0} \frac{F(x+r)-F(x)}{r} = 1$$

for m-a.e. x. This says the F is differential m-a.e. with derivative 1 whereas $F(1)-F(0) = 2 \neq \int_0^1 F'(x)\,\mathrm{d}x$.

So far we have the following decompositions and properties:

$$\begin{cases} v \ll m, \nu \perp m \\ \mathrm{d}\nu = \mathrm{d}\lambda + f\mathrm{d}m \\ \nu = \nu^+ - \nu - . \end{cases}$$

For Lebesgue-Stieltjes measures μ_F on $\mathbb{R}$, how do we see these in F?

Theorem 3.5.7

Let $F : \mathbb{R} \to \mathbb{R}$ be nondecreasing and define $G(x) :=$ the right limit of $F(x)$, written $F(x^+) = \lim_{y\downarrow x} F(y)$. Then

(1) $\{x : F \text{ is discontinuous}\}$ is at most countable, so F, G differ at most at countable points,

(2) F, G are differentiable a.e. with $F' = G'$ a.e., and

(3) G is non-decreasing and right-continuous.

Beginning of Nov.12, 2021

Proof.

(1) Trivial since each jump must contain some rational.

(3) For right continuity, fix x and let $x_n \downarrow x$. Let D be the discontinuity set and assume $x_n \notin D$. Then $F(x_n) = F(x_n^+) = G(x_n)$, whereas the first term converges to $F(x^+)$ and the third to $G(x^+)$. Hence $G(x) = F(x^+) = G(x^+)$ and G is right-continuous.

(2) Let g be the density of the absolute continuous part of μ_G, so

$$\frac{G(x+h)-G(x)}{h} = \begin{cases} \dfrac{\mu_G((x,x+h])}{m((x,x+h])} & h > 0 \\ \dfrac{\mu_G((x+h,x])}{m((x+h,x])} & h < 0 \end{cases}$$

which converges to $g(x)$ for m-a.e. x, as $h \to 0$, by a previous theorem. Hence $G'(x) = g(x)$ a.e.

For differentiability of F, define $H(x) = G(x) - F(x)$. We know $H(x) \equiv 0$ except on a countable set, and it's nonzero at jumps, taking the size of the jump, i.e., $F(x^+) - F(x)$. Let $\{x_i\}$ be the enumeration of discontinuity points and define $I_{x_j} := (F(x_j), F(x_j^+))$. Clearly I_{x_j}'s are disjoint. Then, for each N,

$$\sum_{x_j \in [-N,N]} H(x_j) \leqslant F(N) - F(-N) < \infty.$$

We can define a discrete measure: let $\mu := \sum_j H(x_j)\delta_{x_j}$, which by above is finite on bounded intervals, and $\mu \perp m$. Therefore,

$$m(E^c) := m(\{x : \frac{\mu((x-r, x+r))}{2r} \nrightarrow 0\}) = 0.$$

Notice that

$$\left|\frac{H(x+h) - H(x)}{h}\right| \leqslant \frac{H(x+h) + H(x)}{|h|} \leqslant 4\frac{\mu((x-2|h|, x+2|h|))}{4|h|}$$

so this converges to 0 for a.e. x.

□

Example 3.5.8. Consider discrete measure $\mu = \sum_j a_j \delta_{x_j}$ on $\mathbb{R}$ with distribution F.

(1) At each x_j, F is discontinuous so not differentiable,

(2) If no $x_j \in (a, b)$ then F is constant there, i.e., $F' \equiv 0$.

We can in fact have $F'(x) = c \neq 0$ at some x. For example consider two functions $h \geqslant g$ with h, g tangent at x with nonzero derivative. Consider a "staircase" function that starts off from one point on the graph of g, bounces vertically and horizontally, all the way till it converges to $(x, g(x)$. Then, for $\Delta x > 0$,

$$\frac{g(x+\Delta x) - g(x)}{\Delta x} \leqslant \frac{F(x+\Delta x) - F(x)}{\Delta x} \leqslant \frac{h(x+\Delta x) - h(x)}{\Delta x}$$

Since the first and the last term both converge to $g'(x) = h'(x)$, the limit for the middle must also converge to that quantity. Nevertheless, $F'(x) = 0$ for a.e. x since the absolutely continuous part of μ is 0.

3.6 Total Variation Functions

Consider a function F. We want to define a function $T_F(x)$ to be the "total up-down movement" by $F(x)$. In the case of a "nice" differentiable function, this can be easily computed by first finding the critical points and add up the differences. For a more complicated differentiable function, e.g., $f(x) := x^2 \sin(1/x)$, use the following way.

$$T_F(x) := \sup\{\sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| : n \geqslant 0, \{x_i\} \text{ partitions } \mathbb{R}\}.$$

We define this to be the **total variation** of F. (Notice that as n increases, the supremum only gets bigger, not smaller.)

Alternatively, we can fix a only consider partitions containing a. Hence, for $a < b$,

$$T_F(b) + T_F(a) + \sup\{\sum_{i=1}^{n} |F(x_j) - F(x_{j-1})| : \{x_i\}_{i=0}^{n} \text{ partitions } [a, b]\}.$$

We define this to be the **total variation** of F on $[a, b]$.

Since T_F is an increasing, we define $T_F(\infty) := \lim_{x\to\infty} T_F(x)$ (possibly ∞). If finite, we say F is a function of **bounded variation** on $\mathbb{R}$. The definition of bounded variation on $[a,b]$ is defined analogously.

Beginning of Nov.15, 2021

Example 3.6.1. Suppose F is differentiable in $[a,b]$ and F' is bounded by K ($|F'| \leqslant K$). Then F has BV on $[a,b]$ by MVT: for any partition $\{x_i\}$, there exist $\xi_i \in (x_{i-1}, x_i)$ satisfying the MVT property:

$$\sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| = \sum_{j=1}^{n} F(\xi_j)(x_j - x_{j-1}) \leqslant \sum_{i=1}^{n} K(x_j - x_{j-1}) = K(b-a).$$

Example 3.6.2. If F is monotone on $[a,b]$ then F has BV.

Example 3.6.3. If $F = \chi_{\mathbb{Q}}$ on $[a,b]$ then picking partitions consisting of alternating rationals and irrationals implies F does not have BV.

Decomposition of Variation of F

If $F(-\infty) = 0$ then

$$F(x) = (\text{total upward in } (-\infty, x]) - (\text{total downward in } (-\infty, x]),$$

the difference of two increasing functions (similar to how a signed measure is decomposed).

Clearly for a "nice" function,

$$\frac{F(x) + T_F(x)}{2} = \text{upward} \qquad \frac{T_F(x) - F(x)}{2} = \text{downward}.$$

Lemma

For a real valued F with BV, $T_F + F$, $T_F - F$ are nondecreasing.

Proof. Let $x < y$. We will show that $F_F(x) + F(x) \leqslant T_F(y) + F(y)$. Consider a partition $x_0, x_1, \dots, x_n = x, x_{n+1} = y$. Then

$$T_F(y) + F(y) \geqslant \sum_{i=1}^{n} |F(x_i) - F(x_{i-1})| + |F(y) - F(x)| + F(y)$$

so taking supreum gives

$$\begin{aligned} T_F(y) + F(y) &\geqslant T_F(x) + \underbrace{|F(y) - F(x)| + F(y) - F(x)}_{\geqslant 0} + F(x) \\ &\geqslant T_F(x) + F(x). \end{aligned}$$

Similar for $T_F - F$, as $T_F(x) = T_{(-F)}(x)$ (so it's equivalent to above with $-F$). □

Theorem 3.6.4

For $F : \mathbb{R} \to \mathbb{R}$, F has BV if and only if F is the difference of two bounded nondecreasing functions.
With this theorem, we can define the **Jordan decomposition** of F by

$$F = \frac{T_F + F}{2} - \frac{T_F - F}{2}.$$

Proof. For $\Rightarrow$, F has BV then T_F is bounded by definition. Also, we claim that F is bounded: consider the partition $x = x_0 < x_1 = y$, which gives

$$T_F(y) - T_F(x) = \sup \ldots \geqslant |F(y) - F(x)|.$$

Since x, y are arbitrary and T_F bounded, F must be bounded too. Therefore $\Rightarrow$ holds by the previous lemma.
Conversely, if $F = F_1 - F_2$ where F_1, F_2 are bounded and nondecreasing, then F_1, F_2 have BV, so $F_1 - F_2$ has BV as well.

□

3.7 Absolutely Continuous Functions

Definition: Absolutely Continuous

We say $f : \mathbb{R} \to \mathbb{C}$ is **absolutely continuous** if for all $\epsilon > 0$, there exists $\delta > 0$ such that

$$\{(a_j, b_j)\}_{i=j}^{n} \text{ disjoint and } \sum_{j=1}^{n} |b_j - a_j| < \delta \implies \sum_{j=1}^{n} |F(b_j) - F(a_j)| < \epsilon. \qquad (*)$$

In particular this implies uniform continuity.
Note that if (*) holds then it also holds for $\left| \sum_{i=1}^{n} (F(b_j) - F(a_j)) \right|$, so if F is a distribution function, then

$$E = \bigcup_{i=1}^{n} (a_j, b_j) \text{ disjoint union with } m(E) < \delta \implies \mu_F(E) < \epsilon.$$

(Cf. Theorem 3.5 on notes p.169.)

Using MVT, if F is differentiable and F' is bounded, then F is absolutely continuous.

Definition: Normalized BV (NBV)

We define NBV to be the collection of right-continuous BV functions with $F(-\infty) = 0$.

Proposition 3.7.1

For $F \in$ NBV, F is absolutely continuous if and only if $\mu_F \ll m$, if and only if μ_F has a density F', i.e.,

$$F(x) = \int_{(-\infty, x]} F' \, \mathrm{d}m \qquad \text{for all } x.$$

Corollary 3.7.2

For $f \in L^1(m)$ on $\mathbb{R}$, $F(x) := \int_{-\infty}^{x} f(t)\,\mathrm{d}t$ is in NBV and absolutely continuous, with $F' = f$ a.e. Conversely, if $F \in$ NBV is absolutely continuous then $F' \in L^1(m)$ (exists a.e.) and

$$F(x) = \int_{-\infty}^{x} F'(t)\,\mathrm{d}t \qquad \text{for all } x.$$

Beginning of Nov.17, 2021

Proof. If $\mu_F \ll m$, then for $\epsilon > 0$, there exists $\delta > 0$ such that (by a previous theorem on absolute continuity) such that

$$\sum_{j=1}^{n}(b_j - a_j) < \delta \implies \left|\sum_{j=1}^{n} F(b_j) - F(a_j)\right| < \epsilon.$$

For absolute continuity, a similar argument needs to hold but the magnitude needs to be inside, i.e., $\sum|\cdot|$. We split the j's into

$$\{j : F(b_j) - F(a_j) < 0\} \qquad \{j : F(b_j) - F(a_j) < 0\}.$$

Then

$$\sum_{j=1}^{n}(b_j - a_j) < \delta \implies \begin{cases} \sum_{j=1}^{n}[F(b_j) - F(a_j)]^+ < \epsilon \\ \sum_{j=1}^{n}[F(b_j) - F(a_j)]^- < \epsilon. \end{cases}$$

Then $\sum_{j=1}^{n}|F(b_j) - F(a-j)| < 2\epsilon$. Thus F is absolutely continuous.

Conversely, suppose F is absolutely continuous and $m(E) = 0$. We want to show that $\mu_F(E) = 0$. Let $\epsilon > 0$ so that there exists a corresponding δ satisfying the definition. Then there exists $U_1 \supset E$ with $m(U_1) < \delta$. Also, there exist $U_2 \supset U_3 \supset \ldots \supset E$ with $\mu_F(U_j) \to \mu_F(E)$. We may assume $U_1 \supset U_2 \supset \ldots$ (otherwise let U_j be $U_j \cap U_1$). These are open sets with $m(U_j) < \delta$. Suppose U_j consists of intervals $(a_j^k, b_j^k)_{k \geq 1}$. Applying absolute continuity of F to the first n of these gives

$$\left|\mu_F(\bigcup_{k=1}^{n}(a_j^k, b_j^k))\right| \leq \sum_{k=1}^{n}|F(b_j^k) - F(a_j^k)| < \epsilon.$$

Letting $n \to \infty$, $|\mu_F(U_j)| \leq \epsilon$ (note that the definition of absolutely continuous function requires *finite* number of intervals that add up to length $< \delta$, not countable; this is why we start with the finite case first). Therefore $|\mu_F(E)| < \epsilon$. Since ϵ is arbitrary, $\mu_F \ll m$. □

Normalizing a BV Function

Recall the Jordan decomposition

$$F = \frac{T_F + F}{2} - \frac{T_F - F}{2}.$$

Claim. If $F(-\infty) = 0$ then

$$\frac{T_F + F}{2}(x) = \sup\{\sum_{i=1}^{n}(F(x_i) - F(x_{i-1}))^+, n \geq 0, -\infty < x_0 < \ldots < x_n = x\}.$$

Proof. Note that

$$2\sum_{i=1}^{n}(F(x_i)-F(x_{i-1}))^+ = \sum_{i=1}^{n}|F(x_i)-F(x_{i-1})| + \sum_{i=1}^{n}(F(x_i)-F(x_{i-1})).$$

Taking the supremum, given M, we can assume $x_0 < -M$. Thus we may let $x_0 \to \infty$ as the sum approaches the supremum. Then the second telescoping sum approaches $F(x) - F(-\infty) = F(x)$. Rearranging gives the original claim. □

Corollary 3.7.3

The conclusions about (the $G(x) = F(x^+)$ theorem) hold for $F \in \text{BV}$.
If F_1, F_2 are right-continuous nondecreasing functions, then $F = F_1 - F_2$ is the distribution function of a signed measure.

Beginning of Nov.19, 2021

Lemma

If $F \in \text{BV}$ then $\lim\limits_{x\to-\infty} T_F(x) = 0$. If $F \in \text{BV}$ is right-continuous, then T_F is also right-continuous.

Proof. We first show that $T_F(x+h) - T_F(x) \to 0$ as $h \downarrow 0$. Let $x = x_0 < x_1 < \ldots = x_n = x+h$. The sum of increments between x_1 and x_n satisfies $\sum\limits_{i=2}^{n}|F(x_i)-F(x_{i-1)}| \leqslant T_F(x+h) - T_F(x_1)$, and intuitively $|F(x_1) - F(x_0)|$ is small if $|x_1 - x|$ is small.
Given $\epsilon > 0$, if h is sufficiently small,

$$\sum_{i=1}^{n}|F(x_i)-F(x_{i-1})| \leqslant T_F(x+h) - T_F(x_1) + \epsilon.$$

Taking sup of the LHS, we can assume $x_1 \downarrow x$. Then,

$$T_F(x+h) - T_F(x) \leqslant \epsilon + T_F(x+h) - T_F(x+).$$

In other words,

$$T_F(x) \geqslant T_F(x+h) - \epsilon.$$

Since ϵ is arbitrary, we are done.
Now we show that $\lim\limits_{x\to-\infty} T_F(x) = 0$. Let $-\infty < x_0 < \ldots < x_n = x$. By definition,

$$\sum_{i=1}^{n}|F(x_i)-F(x_{i-1})| \leqslant T_F(x) - T_F(x_0).$$

Taking sup, we can assume $x_0 \to -\infty$. Then we have $T_F(x) \leqslant T_F(x) - T_F(x_0)$ so $T_F(x_0) \to 0$ □

Connection to Jordan Decomposition

Consider finite measures (maybe complex). Recall that $F \in \text{NBV}$ if and only if μ_F is a measure.
For $\mathbb{R}$-valued measures, we have

(1) Jordan decomposition: $\mu_F = \mu_F^+ - \mu_F^-$

(2) Total variation: $|\mu_F| = \mu_F^+ + \mu_F^-$,

(3) Jordan decomposition of F: $F = (T_F + F)/2 - (T_F - F)/2$

(4) Total variation function for F: $T_F = (T_F + F)/2 + (T_F - F)/2$.

Question. How do these correspond? How are the properties $\mu_F \ll m, \mu_F \perp m$ reflected in F? We know F' is the density of the continuous part of μ_F by L-R-N.

Theorem 3.7.4

There is a 1-1 correspondence between the NBV functions and signed measures μ given by $F(x) = \mu((-\infty, x])$. The total variation has an analogous statement: $|\mu_F| = \mu_{T_F}$.

Proof. For $F \in$ NBV, by definition F is right continuous. Hence F is the distribution function of some finite μ_F. Conversely, given a finite signed measure μ, its distribution function is right-continuous; continuity from above with finite total measure implies that $F(-\infty) = \lim_{x\to-\infty} \mu((-\infty, x]) = 0$. The other part is omitted. □

We know μ_F has density F' for the absolutely continuous part. What about $|\mu_F|$? Guess? $|F'|$?

Lemma

Yes. For convenience denote the absolutely continuous and dsingular singular parts of ν by ν_a, ν_s, respectively. Claim: for $F \in$ NBV, $|\nu_F|_a$ has density $|F'|$.

Proof. L-R-N gives $\mu_F = (\mu_F)_a + (\mu_F)_s$. Hahn decomposition of $(\mu_F)_a, (\mu_F)_s$ separately gives us four disjoint (up to null) sets that partition the space: P_a, N_a, P_s, N_s such that $(\mu_F)_a^+ - (\mu_F)_a^- + (\mu_F)_s^+ - (\mu_F)_s^- = \mu_F$. Then $|\mu_F|$ is the sum of these parts. Since $(\mu_F)_a^+, (\mu_F)_a^-$ are both absolutely continuous to m whereas the other two $\perp m$, and the sum of the first two lives on a set disjoint from that on which the sum of the last two lives. By uniqueness of decomposition, we must have $|\mu_F|_a = (\mu_F)_a^+ + (\mu_F)_a^-$. Then,

$$(\mu_F)_a = (\mu_F)_a^+ - (\mu_F)_a^- = F'\mathrm{d}m = (F')^+\mathrm{d}m - (F')^-\mathrm{d}m$$

where $(F')^+$ and $(F')^-$ are disjointly supported. Hence this must also be the Jordan decomposition of F'. Therefore,

$$|\mu_F|_a^+ = (F')^+\mathrm{d}m \qquad \text{and} \qquad |\mu_F|_a^- = (F')^-\mathrm{d}m,$$

and

$$|\mu_F|_a = (F')^+\mathrm{d}m + (F')^-\mathrm{d}m = |F'|\mathrm{d}m.$$

□

Beginning of Nov.22, 2021

Previously we've shown that T_F is the distribution function of $|\mu_F|$, so $(T_F + F)/2$ is the distribution function of $(|\mu_F| + \mu_F)/2 = \mu_F^+$ and similarly $(T_F - F)/2$ is the distribution function of μ_F^-.
We also know that

$$\mathrm{d}\mu_F = \mathrm{d}(\mu_F)_s + F'\mathrm{d}m. \tag{*}$$

Proposition: (3.30)

For $F \in \text{NBV}$, we have $F' \in L^1(m)$. Also, $\mu_F \perp m$ if and only if $F' = 0$ a.e., and $\mu_F \ll m$ if and only if $F(x) = \int_{(-\infty,x]} F' \, \mathrm{d}m$.

Proof. The "iff" statements follow from (*). To show the first claim,

$$\int |F'| \, \mathrm{d}m = |\mu_F|_a(\mathbb{R}) \leqslant |\mu_F|(\mathbb{R}) = T_F(\infty) < \infty.$$

□

Proposition: (3.32)

For $F \in \text{NBV}$, F is absolutely continuous if and only if $\mu_F \ll m$.

Corollary: (3.33)

For $f \in L^1(m)$ on $\mathbb{R}$, the function

$$F(x) := \int_{(-\infty,x]} f \, \mathrm{d}m$$

is in NBV, absolutely continuous, and $F' = f$ a.e.

Conversely, if $F \in \text{NBV}$ is absolutely continuous, then $F' \in L^1(m)$ and $F(x) = \int_{(-\infty,x]} F' \, \mathrm{d}m$.

In other words, for $F \in \text{NBV}$:

$$\begin{aligned} F \text{ is absolutely continuous} &\iff F \text{ is the integral of } \textit{some } f \\ &\iff F \text{ is the integral of } F'. \end{aligned}$$

Proof. If $f \in L^1$, then $F(x) := \int_{(-\infty,x]} f \, \mathrm{d}m$ is the distribution function of a *finite* signed measure. By (theorem 3.29), F is in NBV. Then F is absolutely continuous by (proposition 3.32), and by (proposition 3.30), F' is also a density of μ_F so $\mu_F = f$ a.e.

Conversely, if $F \in \text{NBV}$, then $\mu_F \ll m$ by (proposition 3.32). Then by (proposition 3.30), $F(x) = \int_{(-\infty,x]} F' \, \mathrm{d}m$.

□

Lemma

For BV on an bounded interval, the converse in (corollary 3.33), $F \in \text{BV}$ is a consequence of F being absolutely continuous. In other words, if F is absolutely continuous on $[a,b]$, then $F \in \text{BV}([a,b])$.

Proof. Use the (ϵ, δ) definition and take $\epsilon = 1$. Then for some δ,

$$\sum_{j=1}^{n} (b_j - a_j) < \delta \implies \sum_{j=1}^{n} |F(b_j) - F(a_j)| < 1.$$

Fix k large so that $(b-a)/k < \delta$. Fix the points $a = x_0, x_1 = a + 1/k, \ldots, x_k = b$. Grouping any partition in into these subintervals shows that the total variation $\leqslant k$. □

Example 3.7.5. (p.192 notes). Let $H(x)$ be the Cantor-lebesgue function and let $F(x) := H(x) + x$. Then $F(1) - F(0) = 2$ whereas

$$\int_{[0,1]} F' \, \mathrm{d}m = \int_{[0,1]} H' \, \mathrm{d}m + \int_{[0,1]} 1 \, \mathrm{d}m = 1.$$

Therefore F is continuous but *not* absolutely continuous.

Example 3.7.6: Integration by parts?. When does the following hold?

$$f(b)g(b) - f(a)g(a) = \int_{[a,b]} fg' \, \mathrm{d}m + \int_{[a,b]} f'g \, \mathrm{d}m?$$

Clearly, letting $f = g$ both as the Cantor function, the claim fails. However, if f, g are both absolutely continuous then the claim holds. (A HW8b problem showed that ratio of absolutely continuous functions is absolutely continuous, and analogously the product also is.) If so, fg is also absolutely continuous, so

$$f(b)g(b) - f(a)g(a) = \int_{[a,b]} (fg)' \, \mathrm{d}m = \int_{[a,b]} fg' + f'g \, \mathrm{d}m$$

as $(fg)'(x) = \lim_{h\to 0}\left[\frac{f(x+h) - f(x)}{h} g(x) + f(x+h)\frac{g(x+h) - g(x)}{h}\right] = f'(x)g(x) + f(x)g'(x)$ for a.e. x.

Example 3.7.7: Term-by-term differentiation?. Suppose $F_n \geqslant 0$ are increasing right-continuous functions on $[0,1]$ that are summable a.e. Let $F := \sum_{n=1}^{\infty} F_n$. Does $F' = \sum_{n=1}^{\infty} F_n'$ a.e.?

The answer is yes. We change it to an equivalent question about measures μ_{F_n} and μ_F so $\mu_F = \sum_{n=1}^{\infty} \mu_{F_n}$. The definition of F implies that F' is the density of the absolutely continuous part of μ_F and we wonder if summation "messes up" the absolute continuous parts of each element in the summation. We know

$$\begin{aligned} \mu_F &= \sum_{n=1}^{\infty} (\mu_{F_n}^s + F_n' \mathrm{d}m) \\ &= \sum_{n=1}^{\infty} \mu_{F_n}^s + \sum_{n=1}^{\infty} F_n' \mathrm{d}m =: \lambda + \rho. \end{aligned}$$

It suffices to check that $\lambda \perp \mu$ and $\rho \ll m$. That is, does $\rho = (\sum_{n=1}^{\infty} F_n')\mathrm{d}m$? Yes; see the notes.

Chapter 5

A Bit of Functional Analysis

Beginning of Nov.29, 2021

In finite dimensional spaces, linear maps, i.e., multiplication by matrices, is always continuous.

As we shall see, bonudedness and continuity for linear operator are equivalent, so there exists a $C > 0$ such that $\|Tx\| \leqslant C\|x\|$ for all x.

In general, linear operators need not to be bounded. For example consider the derivative operator $D : C^1([0,1]) \to C([0,1])$ defined by $T : f \mapsto f'$ where both spaces are equipped with the sup norm. Consider the functions $\sin(nx)$ which are uniformly bounded by 1 but derivatives $\to \infty$.

Subspace also need not to be closed in infinite-dimension. For exmaple let $X = C([-1,1])$ equipped with the sup norm and let Y be the subspace of differentiable functions. For example consider $f_n(x) := \sqrt{x^2 + 1/n}$ which converge uniformly to $|x|$, a non-differentiable function.

Definition 5.0.1: Seminorm

Let X be a vector space over $\mathbb{R}$ and $\mathbb{C}$. A **seminorm** is a function $\|\cdot\| : X \to [0,\infty)$ such that $\|\lambda x\| = |\lambda|\|x\|$ and $\|x + y\| \leqslant \|x\| + \|y\|$ for all $\lambda \in \mathbb{R}$ or $\mathbb{C}$ and $x, y \in X$. A norm has the additional property with $\|x\| = 0 \Rightarrow x = 0$.

Topological Spaces

In a metric space, we say $x_n \to x$ if for all open set G containing x, the tail of the sequence eventually is enclosed by G.

Heuristically, in a general topological space, we specify a collection $\mathfrak{J}$ of subsets of X; we say $x_n \to x$ means for all $G \in \mathfrak{J}$, $x_n \in G$ eventually.

A **topological space** is a pair $(X, \mathfrak{J})$ with $\mathfrak{J} \subset \mathbb{P}(X)$ with $\varnothing, X \in \mathfrak{J}$, $\mathfrak{J}$ closed under *all* union and *finite* intersections. We say sets in $\mathfrak{J}$ are open sets, and we call complements of open sets closed sets.

The **interior** of $A \subset X$ is A°, the union of all open $G \subset A$. A **neighborhood** of $x \in X$ is a set A with $x \in A^\circ$.

We say x_n **converges** to x if for every open neighborhood U of x, $\{x_n\}$ is eventually in U.

Given a norm $\|\cdot\|$, we obtain an **induced metric** $d(x,y) := \|x - y\|$, which gives open sets, forming a topology called the **norm topology**.

Two norms $\|\cdot\|_1, \|\cdot\|_2$ are called **equivalent** (or **compatible**) if there exist $c_1, c_2 \in (0,\infty)$ such that

$$c_1\|x\|_1 \leqslant \|x\|_2 \leqslant c_2\|x\|_1 \qquad \text{for all } x.$$

Definition 5.0.2: Banach Space

A **Banach space** is a normed linear space complete in its norm topology.

Beginning of Dec.1, 2021

Proposition 5.0.3

$L^1(\mu)$ is complete.

Proof. Recall that if a Cauchy sequence has a convergent subsequence then the entire sequence converges to the same limit.

Suppose $\{f_n\}$ is Cauchy in L^1. Cauchy in L^1 implies Cauchy in measure, so we have a.e. convergence of a subsequence. Now we show that the convergence is in L^1 too.

Let $\epsilon > 0$. For a fixed k, for onward terms,

$$|f_{n_k} - f_{n_j}| \to |f_{n_k} - f| \text{ a.e. as } j \to \infty.$$

By Fatou,

$$\int \liminf_{j\to\infty} |f_{n_k} - f_{n_j}| = \int |f_{n_k} - f| \leqslant \liminf_{j\to\infty} \int |f_{n_k} - f_{n_j}| < \epsilon$$

for k large. Hence $f_{n_k} \to f$ in L^1. □

Beginning of Dec.3, 2021

Note that for $x \in (0,2)$,

$$\frac{1}{x} = \frac{1}{1-(1-x)} = \sum_{n=0}^{\infty}(1-x)^n.$$

Claim. Similarly, for $\|I-T\| < 1$,

$$T^{-1} = \sum_{n=0}^{\infty}(I-T)^n,$$

i.e., for all x, $T^{-1}x = \sum_{n=0}^{\infty}(I-T)^n x = \lim_{n\to\infty}\sum_{k=0}^{n}(I-T)^k x$.

In general, a series of operators $\sum_{n=0}^{\infty} T_n$ with Banach codomain converges if the norm series $\sum_{n=0}^{\infty}\|T_n\|$ converges: for large $j < k$,

$$\left\|\sum_{n=j}^{k} T_n x\right\| \leqslant \sum_{n=j}^{k}\|T_n x\| \leqslant \sum_{n=j}^{k}\|T_n\|\|x\| \to 0.$$

Proof. Note that $\|(I-T)^n\| \leqslant \|I-T\|^n \to 0$ so we do have a convergent series, so the claim is well-defined. Hence

$$\begin{aligned} T\sum_{n=0}^{\infty}(I-T)^n &= (I-(I-T))\sum_{n=0}^{\infty}(I-T)^n \\ &= \sum_{n=o}^{\infty}(I-T)^n - \sum_{n=1}^{\infty}(I-T)^n = I \end{aligned}$$

and likewise for the other direction. □

Baire Category Theorem

When is a set "topologically small"?
Recall we say $B \subset X$ is dense in X if $\overline{B} = X$. Equivalently,

(1) every open neighborhood of every $x \in X$ intersects B nontrivially, or

(2) every open set intersects B.

On the opposite, a set $A \subset X$ is **nowhere dense** if $\overline{A}$ (closure!) contains no nonempty open set of X:

$$\begin{aligned} A \text{ is nowhere dense} &\iff \text{no open subset of } \overline{A} \\ &\iff \text{every open set intersects } \overline{A}^c \\ &\iff \overline{A}^c \text{ is dense and open.} \end{aligned}$$

If $G \subset \overline{E}$ then $E \cap G$ is dense in G.
Example. Let $\{q_n\}$ be an enumeration of $\mathbb{Q}$ and let $G := \bigcup_{n=1}^{\infty} (q_n - 2^{-n}, q_n + 2^{-n})$. Then G is open and dense in $\mathbb{R}$ with measure 2 whereas G^c cannot contain *any* interval. Hence G^c is nowhere dense.

Theorem 5.0.4: Baire Category Theorem

If X is complete, then

(1) if U_n is open and dense in X for all n (so U_n^c is nowhere dense), then $\bigcap_{n=1}^{\infty} U_n$ is dense; and

(2) X is *not* a union of countably many nowhere dense sets (follows from (1) to the complements).

We say $E \subset X$ is **meager** if it is the union of countably many nowhere dense sets. In particular X itself is not meager.

Examples:

(1) If $E_1, E_2, \ldots \subset \mathbb{R}$ and their union is $\mathbb{R}$, then at least one of the E_n's must contain an interval.

(2) Let X be complete. To show that "there exists $x \in X$ with property P", it suffices to show that $\{x \in X : x \text{ lacks } P\}$ is a countable union of nowhere dense sets (so the union cannot be the whole space).

(3) Consider $C([0,1])$ and let P be "nowhere differentiable". Define

$$A_n := \{f \in C([0,1]) : \text{ there exists } x_0 \text{ where } |(f(x) - f(x_0))/(x - x_0)| \leqslant n \text{ for all } x \neq x_0\}.$$

If f' exists at x_0 then f is contained in some A_n. With the BCT, it suffices to show that A_n is (closed and) nowhere dense.

Roughly, we want to show that for all $f \in C([0,1])$ and all $\epsilon > 0$, there exists g with $\|g - f\| < \epsilon$ and $g \notin A_n$. Heuristically given the ϵ-tube we construct a sawtooth fucntion similar to the Weierstraß monster disobeying the bound for A_n. This will show that A_n is nowhere dense so "at least one point differential" functions are meager.

(4) (**Open Mapping Theorem**.) Let X, Y be Banach and $T \in L(X, Y)$ a bijection. Then T^{-1} exists but is not necessarily bounded. (We know T^{-1} is bounded if and only if T^{-1} is continuous at 0, so $T(U)$ needs to contain a ball centered at 0 for every neighborhood U.)

We can assume that $U = B(0, \epsilon)$ is itself a ball. By a scaling argument, the choice of ϵ does not matter. Hence T^{-1} is bounded if and only if $T(B(0, n))$ contains a neighborhood of 0 for some (and by scaling, all) n. Then

$$Y = \bigcup_{n=1}^{\infty} T(B(0, n))$$

since T is onto. This means some $T(B(0, n))$ is *somewhere* dense. Eventually this implies the inverse of a bounded bijection is necessarily bounded.