Graduate Real Analysis A
Lecture Notes

By

Omer Tamuz

With

Lucas Abounader, Santiago Adams, MohammedSaid Alhalimi, Minakshi Ashok, Zhaojun Chen, Luis Soldevilla Estrada, Samuel Goodman, Holly Krynicki, Tal Hershko, Wei Hou, Joey Litvin, Eric Ma, Caden Mikkelsen, Eric Paul, Lara San Martin Suarez, Shreshth Srivastava, Bharathan Sundar, Elizabeth Xiao, Brian Yang and Jonah Yoshida

California Institute of Technology

2023

Acknowledgments

These lecture notes are Based on Folland's "Real Analysis". We are indebted to Kimberly Golubeva for her help, and thank Fan Wu for finding and correcting numerous mistakes.

Contents

Disclaimer

This a not a textbook. These are lecture notes.

1 Measures

1.1 Measuring and why it is hard

What is the area of the open unit disk $D = \{(x,y) : x^2 + y^2 < 1\}$? This is an old question, with a (maybe) simple answer. An interesting question that came much later is: What do we formally mean by "area"? For the unit disk there are a few different formal definitions, that all give the same answer. One of them is the following: Start by saying that a rectangle $[a,b] \times [c,d] \subset \mathbb{R}^2$ has area $(b-a)(d-c)$. Say that the disjoint union of a finite number of rectangles has area which is equal to the sum of the areas of the rectangles (here we already need to check that this is consistent, and it is). Finally, let the area of the disk equal the supremum of all areas of finite disjoint unions of rectangles that are contained in it.

While this gives a good answer for disks and other nice (i.e., open) shapes like disks. It has some nice properties:

1. The area of the unit square is one.
2. The area of the unit disk is π.
3. The area of a subset A is equal to the area of its translate $A + (x,y)$.
4. The area of a countable disjoint union $A_1 \cup A_2 \cup \cdots$ of open sets is equal to the sum of the areas of $A_1, A_2, \ldots$.

But this notion fails for sets that are not as nice as D. For example, what is the area of the unit disk minus its rational points: $D \setminus \mathbb{Q}^2$? According to this definition the answer is zero. The same holds for the complement of this set inside D, and so we do not have additivity for all sets. The same issue appears more generally in $\mathbb{R}^d$.

Ideally, what we want is a map $\mu\colon 2^{\mathbb{R}^d} \to [0,\infty]$ with the following properties:

1. $\mu([0,1)) = 1$.
2. If $A_1, A_2, \ldots$ are pairwise disjoint subsets of $\mathbb{R}^d$ then
$$\mu(\cup_n A_n) = \sum_n \mu(A_n).$$
3. $\mu(A + x) = \mu(A)$ for any $x \in \mathbb{R}^d$ and subset A of $\mathbb{R}^d$.

Unfortunately, such a map does not exist. To see this, consider already $n = 1$. For simplicity, we will in fact consider $S^1 = \{e^{2\pi i x} : x \in \mathbb{R}\} \subset \mathbb{C}$, and show that there does not exist a measure $\mu\colon 2^{S^1} \to [0,1]$ such that

1. $\mu(S^1) = 1$.
2. If $A_1, A_2, \ldots$ are pairwise disjoint subsets of S^1 then
$$\mu(\cup_n A_n) = \sum_n \mu(A_n).$$

equivalence class	V	A_1	A_2	$\cdots$
$[x]$	x	$T_{w_1}(x)$	$T_{w_2}(x)$	$\cdots$
$[y]$	y	$T_{w_1}(y)$	$T_{w_2}(y)$	$\cdots$
$[z]$	z	$T_{w_1}(z)$	$T_{w_2}(z)$	$\cdots$
$\cdots$				

Table 1: The construction of $(A_n)_n$. Each column corresponds to some A_n, and each row is an equivalence class. It is easy to see in this table that $A_n \cap A_m$ is empty when $n \neq m$.

3. $\mu(A \cdot z) = \mu(A)$ for any $z \in S^1$ and subset A of S^1.

For $w \in [0,1)$ define the map $T_w : S^1 \to S^1$ by $T_w(z) = z \cdot e^{2\pi i w}$. This is simply a rotation by $2\pi w$ radians. The third property is equivalent to $\mu(T_w(A)) = \mu(A)$ for all $A \subset S^1$.

Define an equivalence relation on S^1 by $e^{2\pi i x} \sim e^{2\pi i y}$ if $x - y \in \mathbb{Q}$. Equivalently, $z \sim z'$ if there is $w \in \mathbb{Q} \cap [0,1)$ such that $T_w(z) = z'$. Denote by $[x] = \{x \cdot e^{2\pi i q} : q \in \mathbb{Q} \cap [0,1)\}$ the equivalence class of x. Choose a representative for each equivalence class, and let V be the set of representatives.

Let $W = \{w_1, w_2, \ldots\} = \mathbb{Q} \cap [0,1)$ be an enumeration of $\mathbb{Q} \cap [0,1)$, and let $A_n = T_{w_n}(V)$. We claim that $A_n \cap A_m = \emptyset$ if $n \neq m$. To see this, suppose that $T_{w_n}(z) \in A_n$ and $T_{w_m}(z') \in A_m$ for $z, z' \in A$. If $[z] = [z']$ then $z = z'$, since both are representatives, and hence $T_{w_n}(z) \neq T_{w_m}(z')$. If $[z] \neq [z']$ then $[T_{w_n}(z)] = [z] \neq [z'] = [T_{w_m}(z')]$, and so $T_{w_m}(z) \neq T_{w_m}(z')$ (see Table 1). It follows that

$$\mu(\cup_n A_n) = \sum_n \mu(A_n).$$

Note the union $\cup_n A_n$ is equal to S^1, since $\{T_{w_1}(z) + T_{w_2}(z), \ldots\} = [z]$, and so the left-hand side is equal to 1. But $\mu(A_n) = \mu(A)$ by translation invariance, and so the right-hand side cannot equal 1.

Note that if we relax the countble additivty requirement and only require finite additivity, then such measures do exist. For $n = 2$ we can even have a measure that is also invariant to rotations. However, for $n \geq 3$ it is impossible, even with finite additivity, to be invariant to both translations and rotations. This is a consequence of the Banach-Tarski paradox.

1.2 Algebras, σ-algebras and measures

In the previous lecture we proved that we could not measure the length / area / volume of general sets in $\mathbb{R}^d$. Our solution will be to restrict ourselves to measures that are only defined on a subset $\Sigma \subset 2^{\mathbb{R}^d}$ of all sets.

For simplicity, consider the case of $\mathbb{R}$. Which set Σ should we choose? We want to make it as big as possible, so that we have many sets we can measure, but we cannot make it too big. We want to include open (say) intervals in Σ, since we know their measure. We want our measure to be additive, and so we want to include countable unions of open intervals. And likewise we want to include their complements. More generally, once we have included a set in Σ, we want to also include its complement, and we want to include all possible countable unions of all sets we have already included.

If we are more generally trying to measure a set Ω (rather than $\mathbb{R}^d$ specifically) then we want Σ to have the following properties:

1. $\Omega \in \Sigma$.
2. If $A \in \Sigma$ then $\Omega \setminus A \in \Sigma$.
3. If $A_1, A_2, \ldots \in \Sigma$ then $\cup_n A_n \in \Sigma$.

A collection of sets Σ that satisfies these properties is called a **σ-algebra**. We say that Σ is an **algebra** if it closed to finite unions, rather than countable unions. The pair (Ω, Σ) is called a **measurable space**.

Some (trivial) examples of σ-algebras:

1. $\Sigma = \{\emptyset, \Omega\}$. This is the trivial σ-algebra. It is contained in every σ-algebra on Ω.
2. $\Sigma = 2^\Omega$. This σ-algebra contains every σ-algebra on Ω.
3. Σ is finite if and only if there is a partition $\Omega = A_1 \cup A_2 \cup \ldots \cup A_n$ such that Σ consists of all unions of partitions elements.
4. If Ω is countable then every σ-algebra is of the form above, but where the partition is not necessarily finite.
5. For $\Omega = \Omega_1 \times \Omega_2$, $\Sigma_1 = \{A_1 \times \Omega_2 : A_1 \subseteq \Omega_1\}$.

For a non-trivial example (which is very important in probability), let $\Omega = \mathbb{R}^\mathbb{N}$, and let the tail σ-algebra be given by

$$T = \{A \subseteq \Omega : \forall \omega \in A \, \forall n \in \mathbb{N} \, \forall x \in \mathbb{R} \quad (\omega_1, \ldots, \omega_{n-1}, x, \omega_{n+1}, \ldots) \in A\}.$$

An example of $A \in T$ is

$$A = \{\omega : \lim_n \omega_n = 17\}.$$

Returning to our case of $\Omega = \mathbb{R}$, we want to include in Σ the open intervals. So, for Σ to be a σ-algebra, we have to include in it many more sets, because once we have added the open intervals we also have countable unions of open intervals, then we have the complements of these, then we have countable unions of these, etc.

To construct our σ-algebra, we make the following observation:

Lemma 1.1. *Let $(\Sigma_i)_{i \in I}$ be a collection of σ-algebras of a set Ω. Then $\cap_i \Sigma_i$ is a σ-algebra.*

It follows that for any collection $\Theta \subset 2^\Omega$ there is a minimal σ-algebra that contains Θ: the intersection of all σ-algebras that contain Θ. Note that this set is non-empty because it includes 2^Ω. We denote it by $\mathscr{M}(\Theta)$.

When Ω is a topological space, the minimal σ-algebra that includes all the open sets is called the **Borel σ-algebra** of Ω, and is denoted by $\mathscr{B}_\Omega$. This is the σ-algebra for which we will construct a measure on $\mathbb{R}$. For the case of $\mathbb{R}$, it is useful to note that $\mathscr{B}_\mathbb{R}$ is also equal to $\mathscr{M}(\Theta)$, where Θ is (for example) the collection of the half-open intervals: $\Theta = \{(a,b] : a < b \in \mathbb{R}\}$.

Given a σ-algebra Σ over a set Ω, a **measure** is a map $\mu \colon \Sigma \to [0,\infty]$ such that

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

2. If $A_1, A_2, \ldots \in \Sigma$ are pairwise disjoint, then

$$\mu(\cup_n A_n) = \sum_n \mu(A_n).$$

We say that μ is finite if $\mu(A) < \infty$ for all $A \in \Sigma$. We say that it is σ-finite if there exist $A_1, A_2, \ldots \in \Sigma$ such that $\cup_n A_n = \Omega$ and $\mu(A_n) < \infty$ for all n. In this course we will focus on finite and σ-finite measures. Maps that satisfy only finite additivity are called **finitely additive measures**. Given a measure μ on a measurable space (Ω, Σ), we will call (Ω, Σ, μ) a **measure space**.

Our goal in the next two lectures will be to prove the following theorem:

Theorem 1.2. *There exists a (unique) measure $\mu \colon \mathscr{B}_\mathbb{R} \to [0,\infty]$ such that $\mu([a,b]) = b - a$ for all $b \geq a$.*

Note that this theorem implies that this measure is translation invariant, i.e., that $\mu(A) = \mu(A + x)$ for all $A \in \mathscr{B}$ and $x \in \mathbb{R}$, since for a fixed x, the map $\mu_x(A) = \mu(A + x)$ is also a measure such that $\mu_x([a,b]) = b - a$.

The following are basic and useful observation about measures:

Claim 1.3. *Let $\mu \colon \Sigma \to [0,\infty]$ be a measure. Then*

1. *If $A \subseteq B \in \Sigma$ then $\mu(A) \leq \mu(B)$.*

2. *If $A_1, A_2, \ldots \in \Sigma$ then $\mu(\cup_n A_n) \leq \sum_n \mu(A_n)$.*

3. *If $A_1 \subseteq A_2 \subseteq \cdots \in \Sigma$ then $\mu(\cup_n A_n) = \lim_n \mu(A_n)$.*

4. *If $A_1 \supseteq A_2 \supseteq \cdots \in \Sigma$ then $\mu(\cap_n A_n) = \lim_n \mu(A_n)$, provided $\mu(A_n) < \infty$ for some n.*

The proof of the first claim follows from the fact that $B = A \cup B \setminus A$ and (finite) additivity. To see the third claim, define $A_0 = \emptyset$ and $B_n = A_n \setminus A_{n-1}$. Then $B_1, B_2, \ldots$ are pairwise disjoint and $\cup_n B_n = \cup_n A_n$. We can now apply additivity to conclude that $\mu(\cup_n B_n) = \sum_n \mu(B_n)$. Since $\sum_n \mu(B_n) = \lim_n \sum_{m=1}^n \mu(B_m) = \lim_n \mu(A_n)$ (by additivity) we are done.

1.3 Outer measures

Let Θ be the set of open rectangles in the plane:

$$\Theta = \{(a,b)\times(c,d) \subset \mathbb{R}^2\}.$$

We denote the area of a rectangle $R=(a,b)\times(c,d)$ by $\rho(R)=(b-a)\cdot(d-c)$. Given any subset $A\subseteq\mathbb{R}^2$, a natural upper bound to its area is what is known as its outer measure:

$$\mu^*(A) = \inf\left\{\sum_n \rho(R_n) : R_1,R_2,\ldots \in \Theta, A \subseteq \cup_n R_n\right\}.b \tag{1.1}$$

This defines a map $\mu^*: 2^{\mathbb{R}^2} \to [0,\infty]$. One can verify that—unlike the map proposed in Lecture 1.1—this one gives the desired answer for $D\setminus\mathbb{Q}^2$, as well to its complement in D, $D\cap\mathbb{Q}^2$. We will prove the former later in this lecture. To see the latter, enumerate $A = D\cap\mathbb{Q}^2 = \{(q_1,p_1),(q_2,p_2),\ldots\}$, fix any $\mathbb{Q}\ni\varepsilon>0$ and let $R_n = [q_n-\varepsilon 2^{-n/2}, q_n+\varepsilon 2^{-n/2}]\times[p_n-\varepsilon 2^{-n/2}, p_n+\varepsilon 2^{-n/2}]$, so that $\rho(R_n)=4\varepsilon 2^{-n}$. Then $A\subset\cup_n R_n$, and $\sum_n \rho(R_n)=4\varepsilon$.

More generally, let Ω be a nonempty set, and let Θ be a collection of subsets of Ω that includes $\emptyset$ and Ω. Let $\rho:\Theta\to[0,\infty]$ be such that $\rho(\emptyset)=0$, and define $\mu^*: 2^\Omega\to[0,\infty]$ as in (1.1). Then

Claim 1.4. *1. $\mu^*(\emptyset)=0$.*

2. For all $A\subseteq B\subseteq\Omega$ it holds that $\mu^*(A)\le\mu^*(B)$.

3. For all $A_1,A_2,\ldots\subseteq\Omega$, $\mu^*(\cup_n A_n)\le\sum_n\mu^*(A_n)$.

Proof. (1) and (2) are left to the reader. For (3), denote $A=\cup_n A_n$ and fix $\varepsilon>0$. Then for each A_n there is a sequence $E_1^n,E_2^n,E_3^n,\ldots\in\Theta$ such that $A\subseteq\cup_k E_k^n$ and $\sum_k\rho(E_k^n)\le\mu^*(A_n)+\varepsilon 2^{-n}$. Then $A\subseteq\cup_{n,k}E_k^n$ and $\sum_{n,k}\rho(E_k^n)\le\varepsilon$. ☐

Any map $\mu^*: 2^\Omega$ that satisfies the properties of this claim is called an **outer measure**. Returning to our example of $A=D\setminus\mathbb{Q}^2$, we note that $A\subset D$, and so $\mu^*(A)\le\mu^*(D)$. On the other hand, $D=(D\setminus\mathbb{Q}^2)\cup(D\cap\mathbb{Q}^2)$, and so

$$\mu^*(D)\le\mu^*(D\setminus\mathbb{Q}^2)+\mu^*(D\cap\mathbb{Q}^2)=\mu^*(D\setminus\mathbb{Q}^2).$$

It will be useful to define our outer measures μ^* using a ρ that has more structure. In the case of $\Omega=\mathbb{R}$, we can indeed extend ρ beyond intervals to the algebra generated by the intervals. In particular, let Θ_0 be any interval of the form $(-\infty,b]$. Let $\mathscr{A}$ be the algebra of sets generated by Θ_0. This includes the complements of the sets in $\mathscr{A}$, i.e., intervals of the form (a,∞), intersections of these sets, i.e., intervals of the form $(a,b]$, and finite unions of such intervals.

Let $\rho:\mathscr{A}\to[0,\infty]$ assign to each $A\in\mathscr{A}$ its length in the obvious way. Then

Claim 1.5. *1. $\rho(\emptyset)=0$.*

2. *If* $R_1, R_2, \ldots \in \mathscr{A}$ *are pairwise disjoint and* $R = \cup_n R_n$ *is in* $\mathscr{A}$ *then* $\rho(R) = \sum_n \rho(R_n)$.

The proof of (1) is immediate. The proof of (2) takes some work, but we will skip it.

A map ρ from an algebra to $[0,\infty]$ that satisfies these properties is called a **premeasure**. Consider now the outer measure μ^* defined by a premeasure ρ.

Claim 1.6. *For every* $R \in \mathscr{A}$, $\mu^*(R) = \rho(R)$.

Proof. Clearly $\mu^*(R) \le \rho(R)$, since $R \subseteq \cup_n R_n$ for $R = R_1 = R_2 = \cdots$. For the other direction, choose $R_1, R_2, \ldots \in \mathscr{A}$ such that $R \subseteq \cup_n R_n$ and $\sum_n \rho(R_n) \le \mu^*(R) + \varepsilon$. Define $S_n = R \cap R_n$. Then $R = \cup_n S_n$. Let $T_n = S_n \setminus \cup_{k=1}^{n-1} S_n$. Then again $R = \cup_n T_n$, and now $T_1, T_2, \ldots$ are pairwise disjoint. Since they are also in $\mathscr{A}$, we have that

$$\rho(R) = \sum_n \rho(T_n) \le \sum_n \rho(R_n) \le \mu^*(R) + \varepsilon,$$

and so $\rho(R) \le \mu^*(R)$. □

The next claim is an even more important property of μ^*.

Claim 1.7. *For every* $R \in \mathscr{A}$ *it holds for every* $A \subseteq \Omega$ *that*

$$\mu^*(A) = \mu^*(A \cap R) + \mu^*(A \cap R^c).$$

Proof. Since $A \cap R$ and $A \cap R^c$ are disjoint, it follows by subadditivity that

$$\mu^*(A) \le \mu^*(A \cap R) + \mu^*(A \cap R^c).$$

For the other direction, find $R_1, R_2, \ldots \in \mathscr{A}$ such that $A \subset \cup_n R_n$ and $\sum_n \rho(R_n) \le \mu^*(A) + \varepsilon$. Let $S_n = R_n \cap R$ and $T_n = R_n \cap R^c$. Then

$$\sum_n \rho(R_n) = \sum_n \rho(S_n) + \sum_n \rho(T_n),$$

by the countable additivity of ρ. Now, $A \cap R \subseteq \cup_n S_n$ and $A \cap R^c \subseteq \cup_n T_s$. Hence

$$\mu^*(A) + \varepsilon \ge \sum_n \rho(S_n) + \sum_n \rho(T_n) \ge \mu^*(A \cap R) + \mu^*(A \cap R^c)$$

□

This claim can be interpreted as follows: A set $R \in \mathscr{A}$ has the nice property that if we define the two maps $\mu_1^*(A) = \mu^*(A \cap R)$ and $\mu_2^*(A) = \mu^*(A \cap R^c)$ then $\mu^* = \mu_1^* + \mu_2^*$, as we expect from measures.

More generally, given any outer measure μ^* defined on Ω using some Θ and ρ, we will say that $S \subseteq \Omega$ is μ^*-measurable if

$$\mu^*(A) = \mu^*(A \cap S) + \mu^*(A \cap S^c)$$

for all $A \subseteq \Omega$. Note that for A that contains S, this translates to

$$\mu^*(S) = \mu^*(A) - \mu^*(A \cap S^c)$$

which is close to what we tried to do in the first lecture.

1.4 Carathéodory and extending measures (by Lucas Abounader and Santiago Adams)

In this lecture we will state (and prove!) Carathéodory's Extension Theorem, which allows us to take an arbitrary outer measure $\mu^* : 2^\Omega \to [0,\infty]$ and construct a measure $\mu = \mu^*|_\Sigma : \Sigma \subset 2^\Omega \to [0,\infty]$. In particular, we will find a good way to exclude the pathological sets (that confounded us in Lecture 1.1) in specifying the σ-algebra Σ, by looking only at the "nice" μ^*-measureable sets. Since premeasures induce outer measures (by 1.1), we will ultimately be able to construct measures from arbitrary premeasures as well.

For those that have previously seen a construction of the Lebesgue measure, recall that one starts with the more primitive concept of Lebesgue outer measure (and, preceding this, the specification of a desired premeasure on elementary sets); this procedure is now presented in a slightly more abstract setting. The punchline here (in a way, justifying abstraction) is some guarantee of uniqueness, given the premeasure. Thus, when we apply the results proven in this lecture to construct, e.g., Lebesgue-Stieltjes, Hausdorff, or product measures, we can pat ourselves on the back, satisfied with the knowledge that the fruits of our labor are the unique objects we're looking for.

- Recap of outer measure ($\mu^* : 2^\Omega \to [0,\infty]$ satisfying $\mu^*(\emptyset) = 0$, monotonicity for $A \subset B$, and countable *sub*additivity).
- Recap of μ^*-measurable ($\forall A,\ \mu^*(A) = \mu^*(A \cap R) + \mu^*(A \cap R^c) = \mu^*(A \cap R) + \mu^*(A \setminus R)$), highlighting that we want the set R to be able to split any set A in a way that recovers $\mu^*(A)$.

Theorem 1.8 (Carathéodory)**.** *If μ^* is an outer measure on Ω, then the collection Σ of μ^*-measurable subsets of Ω is a σ-algebra, and the restriction of μ^* to Σ is a measure.*

Proof. We start with the claim that Σ is more weakly an algebra:

- (nonempty). $\emptyset \in \Sigma$. In fact, for any *null set* N ($\mu^*(N) = 0$), we have $N \in \Sigma$:

  $$\mu^*(A) \le \mu^*(A \cap N) + \mu^*(A \setminus N) = \mu^*(A \setminus N) \le \mu^*(A),$$

  applying monotonicity ($A \cap N \subseteq N \implies \mu^*(A \cap N) \le 0$) and subadditivity in succession.
- (complements). Note the definition of Σ is symmetric with respect to R and R^c.
- (finite union). [Not going to go through this part in entirety; a bit tedious]. Suppose $A, B \in \Sigma$. For any $E \in 2^X$, we want to show

  $$\mu^*(E) = \mu^*(E \cap (A \cup B)) + \mu^*(E \setminus (A \cup B)). \tag{1.2}$$

  It helps to classify points in E based upon their inclusions in A and B. That is, with the partitions

  $$E_{00} := E \setminus (A \cup B); \quad E_{10} := (E \setminus B) \cap A; \quad E_{01} := (E \setminus A) \cap B; \quad E_{11} := E \cap A \cap B,$$

we can convert 1.2 into the equivalent

$$\mu^*(E_{00} \cup E_{10} \cup E_{01} \cup E_{11}) = \mu^*(E_{10} \cup E_{01} \cup E_{11}) + \mu^*(E_{00}). \tag{1.3}$$

Using A to split the sets E and $E \setminus E_{00}$ (applying μ^*-measurability):

$$\mu^*(E_{00} \cup E_{10} \cup E_{01} \cup E_{11}) = \mu^*(E_{10} \cup E_{11}) + \mu^*(E_{00} \cup E_{01}) \tag{1.4}$$

and

$$\mu^*(E_{10} \cup E_{01} \cup E_{11}) = \mu^*(E_{10} \cup E_{11}) + \mu^*(E_{01}), \tag{1.5}$$

while using B to split $E \setminus A$ we obtain

$$\mu^*(E_{00} \cup E_{01}) = \mu^*(E_{01}) + \mu^*(E_{00}). \tag{1.6}$$

Collecting 1.4, 1.5 and 1.6 we obtain 1.3, as desired.

Now we extend finite unions to countable ones:

- (σ-algebra). It suffices to consider only countable *disjoint* unions. (Why? Given $(A_j)_{j\in\mathbb{N}}$, replace with $(B_j)_{j\in\mathbb{N}}$ defined by $B_n = A_n \setminus \bigcup_{j=1}^{n-1} A_j$.) So we want

$$\mu^*(A) = \mu^*(A \cap \bigcup_{j=1}^{\infty} E_j) + \mu^*(A \setminus \bigcup_{j=1}^{\infty} E_j)$$

for $(E_j)_1^\infty \subset \Sigma$ a collection of pairwise disjoint sets. $\le$ is given by subadditivity, so we only need to prove $\ge$. Given N, define $B_N = \bigcup_{j=1}^N E_j$, and note that B_N is μ^*-measureable.

For the difference,

$$B_N \subseteq \bigcup_{j=1}^{\infty} E_j \implies \mu^*(A \setminus \bigcup_{j=1}^{\infty} E_j) \le \mu^*(A \setminus B_N).$$

For the intersection, we want

$$\mu^*(A \cap \bigcup_{j=1}^{\infty} E_j) \le \lim_{N\to\infty} \mu^*(A \cap B_N).$$

Since the B_N are μ^*-measureable, for any N

$$\mu^*(A \cap B_{N+1}) = \mu^*(A \cap B_N) + \mu^*(A \cap E_{N+1} \setminus B_N),$$

leading to the expressions

$$\mu^*(A \cap B_N) = \sum_{n=0}^{N-1} \mu^*(A \cap E_{n+1} \setminus B_n) \overset{(\lim)}{\implies} \lim_{N\to\infty} \mu^*(A \cap B_N) = \sum_{n=0}^{\infty} \mu^*(A \cap E_{n+1} \setminus B_n).$$

Finally we note that $\bigcup_{n=0}^{\infty}(A \cap E_{n+1} \setminus B_n)$ is just $A \cap \bigcup_{n=1}^{\infty} E_n$, so by countable subadditivity

$$\mu^*(A \cap \bigcup_{n=1}^{\infty} E_n) \le \sum_{n=0}^{\infty} \mu^*(A \cap E_{n+1} \setminus B_n),$$

as desired.

The second part of the theorem states that $\mu = \mu^*|_\Sigma$ is in fact a *measure*. The only thing μ^* is missing is (σ-)additivity.

- (additivity). We need to show

$$\mu^*(\bigcup_{j=1}^\infty E_j) \geq \sum_{j=1}^\infty \mu^*(E_j)$$

  for $E_1, E_2, \dots \in \Sigma$ disjoint. This turns out to be equivalent to *finite* additivity, as demonstrated:

$$(\forall N): \sum_{j=1}^N \mu^*(E_j) \leq \mu^*(\bigcup_{j=1}^N E_j) \leq \mu^*(\bigcup_{j=1}^\infty E_j) \implies \sum_{j=1}^\infty \mu^*(E_j) \leq \mu^*(\bigcup_{j=1}^\infty E_j),$$

  and the finite case follows directly from the key μ^*-measurability property of Σ: For $E, F \in \Sigma$ disjoint,

$$\mu^*(E \cup F) = \mu^*(E) + \mu^*(F).$$

$\square$

- Recap of premeasure; $\mu_0 : \mathscr{A} \to [0,\infty]$ on an algebra $\mathscr{A}$ with $\mu_0(\emptyset) = 0$, and $\mu_0(\bigcup_{j=1}^\infty R_j) = \sum_{j=1}^\infty \mu_0(R_j)$ whenever $R_1, R_2, \dots \in \mathscr{A}$ disjoint and the union lies in $\mathscr{A}$.

- Recall that

$$\mu_\rho^*(A) = \inf\left\{\sum_{j=1}^\infty \rho(R_j) : R_j \in \Theta,\ A \subseteq \bigcup_{j=1}^\infty R_j\right\}$$

  is an outer measure for any function $\rho : \Theta \to [0,\infty]$ with $\emptyset, X \in \Theta$ and $\rho(\emptyset) = 0$, and thus for any premeasure μ_0.

Theorem 1.9. *Let $\mathscr{A}$ be an algebra of subsets of Ω, let $\mu_0 : \mathscr{A} \to [0,\infty]$ be a premeasure, and let $\mathscr{B} = \mathscr{M}(\mathscr{A})$ be the σ-algebra generated by $\mathscr{A}$. Then $\mu = \mu^*|_{\mathscr{B}}$ is a measure on $\mathscr{B}$ that extends μ_0, and this measure is the unique extension if μ_0 is σ-finite. [If time] Even if μ_0 is not σ-finite, given another extension ν of μ_0, then we have $\nu(E) \leq \mu(E)$ for all E (so μ is maximal) and $\nu(E) = \mu(E)$ whenever $\mu(E) < \infty$.*

Proof.

- Equation 1.1 and Carathéodory's Theorem (Theorem 1.8) present the construction of $\mu : \Sigma \to [0,\infty]$ on a (potentially larger!) σ-algebra $\Sigma \supseteq \mathscr{A}$ of μ^*-measureable sets.

- Suppose that $E \in \mathscr{B}$, and take any cover $\bigcup_{j=1}^\infty A_j \supseteq E$ where $A_j \in \mathscr{A}$. Then if ν is another extension, $\nu(E) \leq \sum_{j=1}^\infty \nu(A_j) = \sum_{j=1}^\infty \mu_0(A_j)$, so $\nu(E) \leq \mu(E)$.

  Let $A = \bigcup_{j=1}^\infty A_j$; then

$$\nu(A) = \lim_{n\to\infty} \nu(\bigcup_{j=1}^n A_j) = \lim_{n\to\infty} \mu(\bigcup_{j=1}^n A_j) = \mu(A).$$

If $\mu(E) < \infty$ we can pick A_j such that $\mu(A) < \mu(E) + \varepsilon$, so $\mu(A \setminus E) < \varepsilon$ and

$$\mu(E) \leq \mu(A) = \nu(A) = \nu(E) + \nu(A \setminus E) \leq \nu(E) + \mu(A \setminus E) \leq \nu(E) + \varepsilon.$$

ε is arbitrary, hence $\nu(E) = \mu(E)$. This proves the last claim.

- Finally, suppose μ_0 is σ-finite, and take a countable partition $\bigcup_{j=1}^{\infty} A_j = \Omega$ with $\mu_0(A_j) < \infty$ (we can use the familiar trick and assume the A_j are disjoint). Then for any E:

$$\mu(E) = \sum_{j=1}^{\infty} \mu(E \cap A_j) = \sum_{j=1}^{\infty} \nu(E \cap A_j) = \nu(E).$$

$\square$

As a Corollary, we can prove Theorem 1.2. Take the premeasure ρ on the algebra $\mathscr{A}$ generated by $\{[a,b) : a, b \in \mathbb{R}\}$, use Theorem 1.9 to obtain a measure μ on $\Sigma = \mathscr{M}(\mathscr{A})$, and note that $\mathscr{M}(\mathscr{A}) = \mathscr{B}_{\mathbb{R}}$.

1.5 Complete measures and Borel measures on $\mathbb{R}$

Let (Ω, Σ) be a measurable space, and let $\mu \colon \Sigma \to [0, \infty]$ be a measure. We say that $A \in \Sigma$ is a null set if $\mu(A) = 0$. We say that something holds almost everywhere (or μ-almost everywhere) if the set of points at which this something holds has a null set as its complement.

Given (Ω, Σ, μ) we define $\Sigma_0 = \{N \in \Sigma : \mu(N) = 0\}$. Let $\bar{\Sigma}_0 = \{M \subseteq \Omega : M \subseteq N$ for some $N \in \Sigma_0\}$ be the collection of all sets that are contained in a null set. We would like to extend μ to a measure $\bar{\mu}$ that is also defined over $\bar{\Sigma}_0$, by assigning 0 to all $S \in \bar{\Sigma}_0$. It turns out that this is possible, and that furthermore this extension will have advantages beyond having a larger domain.

Since we want to include $\bar{\Sigma}_0$ and Σ in the domain of $\bar{\mu}$, we need to include $\bar{\Sigma} = \mathscr{M}(\bar{\Sigma}_0 \cup \Sigma)$. This σ-algebra is called the **completion** of Σ (with respect to μ. It happens to be simple to construct.

Claim 1.10. $\bar{\Sigma} = \{A \cup M : A \in \Sigma$ *and* $M \in \bar{\Sigma}_0\}$

Proof. Denote $T = \{A \cup M : A \in \Sigma$ and $M \in \bar{\Sigma}_0\}$. Clearly $T \subseteq \bar{\Sigma}$. Hence, to prove the claim it suffices to show that T is a σ-algebra.

While Σ_0 is not a σ-algebra (since it is not closed to taking complements), it is closed to countable unions, since by the subadditivity of μ it holds for $A_1, A_2, \ldots \in \Sigma_0$ that

$$\mu(\cup_n A_n) \leq \sum_n \mu(A_n) = 0.$$

Hence T is also closed to countable unions.

Since $\Omega \in T$, it remains to be shown that T is closed to complements. To this end, consider any $C = A \cup M \in T$ with $A \in \Sigma$ and $M \subseteq N$ for some $N \in \Sigma_0$. Let $M' = M \setminus A$ and $N' = N \setminus A$. Then $C = A \cup M'$ with $M' \subseteq N' \in \Sigma_0$, and furthermore $A \cup N' = A \bigcup N \in \Sigma$. Hence

$$C^c = (A \cup M')^c = (A \cup N')^c \cup (N' \setminus M')$$

which is also in T. $\square$

Theorem 1.11. *Given* (Ω, Σ, μ) *there is a unique measure* $\bar{\mu}\colon \bar{\Sigma} \to [0,\infty]$ *that extends* μ*.*

The measure $\bar{\mu}$ is defined as follows: Given $A \in \Sigma$ and $M \in \bar{\Sigma}_0$, let $\bar{\mu}(A \cup M) = \mu(A)$. It is easy to verify that $\bar{\mu}$ is well defined and is indeed a measure. Furthermore it is **complete**, in the sense that every subset of a null set is measurable.

When $\mu\colon \mathscr{B}_{\mathbb{R}} \to [0,\infty]$ is the unique measure such that $\mu([b-a]) = b-a$, we call $\bar{\mu}$ the Lebesgue measure, and $\bar{\mathscr{B}}_{\mathbb{R}}$ the set of Lebesgue-measurable sets.

A larger class of measures can be constructed in a similar way. Let $F\colon \mathbb{R} \to \mathbb{R}$ be an increasing, right continuous function, i.e., $\lim_{x \searrow x_0} F(x) = F(x_0)$. We can define a premeasure $\rho((b,a]) = F(b) - F(a)$, and from there a measure $\mu_F\colon \mathscr{B}_{\mathbb{R}} \to [0,\infty]$ that satisfies the same. Right continuity is essential, because the continuity of μ means that $\lim_{x \searrow x_0} \mu((x,b]) = \mu((x_0,b])$. Points of discontinuity of F correspond to atoms: $x \in \mathbb{R}$ such that $\mu(\{x\}) > 0$.

In the other direction, for every μ that assigns finite measure to bounded sets we can define $F(x) = \mu((0,x])$ for nonnegative x and $F(x) = -\mu((x,0])$ for negative x. Then $\mu = \mu_F$. When μ is a finite measure then we can take $\lim_{x \to -\infty} F(x) = 0$, and F is called the cumulative distribution function of μ.

The complete measure $\bar{\mu}_F\colon \bar{\mathscr{B}}_{\mathbb{R}} \to [0,\infty]$ is called the Lebesgue-Stieltjes measure associated with F. These measures enjoy some nice properties.

Claim 1.12. *For* $\mu = \mu_F$ *and* $A \in \mathscr{B}_{\mathbb{R}}$ *such that* $\mu(A) < \infty$ *then following holds*

1. *For each* $\varepsilon > 0$ *there exists a finite union* U *of open intervals such that* $\mu(A \triangle U) < \varepsilon$*.*

2. $\mu(A) = \inf\{\mu(U) : U$ *is open and contains* $A\}$*.*

3. *There is a* G_δ *set* B *containing* A *such that* $\mu(B \setminus A) = 0$*.*

4. $\mu(A) = \sup\{\mu(K) : K$ *is compact and is contained by* $A\}$*.*

5. *There is an* F_σ *set* C *contained in* A *such that* $\mu(A \setminus C) = 0$*.*

2 Integration

2.1 Integration of nonnegative functions

Let $(\Omega_1,\Sigma_1),(\Omega_2,\Sigma_2)$ be measurable spaces. A map $f:\Omega_1\to\Omega_2$ is **measurable** if $f^{-1}(A)\in\Sigma_1$ for all $A\in\Sigma_2$. In other words, if for any $A\in\Sigma_2$ it holds that $\{\omega_1\in\Omega : f(\omega_1)\in A\}$ is in Σ_1. When f is measurable, $f^{-1}(\Sigma_2)$ is a sub-σ-algebra of Σ_1. We can think of this sub-σ-algebra as what you can measure when you only observe $f(\omega_1)$ rather than ω_1. Given a measure $\mu_1\colon\Sigma_1\to[0,\infty]$, a measurable $f:\Sigma_1\to\Sigma_2$ induces what is known as the **pushforward measure** $\mu_2=f_*\mu_1\colon\Sigma_2\to[0,\infty]$ given by $\mu_2(A)=\mu_1(f^{-1}(A))$. Thus, when we can measure Ω_1, a measurable function $f:\Omega_1\to\Omega_2$ allows us to measure subsets of Ω_2.

Claim 2.1. *Let $(\Omega_1,\Sigma_1),(\Omega_2,\Sigma_2)$ be measurable spaces, and let $\Sigma_2=\mathscr{M}(\Theta)$. Then $f:\Omega_1\to\Omega_2$ is measurable iff $f^{-1}(R)\in\Sigma_1$ for all $R\in\Theta$.*

In the context of (Ω,Σ) we say that a map $f:\Omega\to\mathbb{R}$ is (Borel) measurable if it is measurable as a map to $(\mathbb{R},\mathscr{B}_{\mathbb{R}})$. By the claim above, to prove that $f:\Omega\to\mathbb{R}$ is measurable it suffices to show that $f^{-1}((-\infty,a])\in\Sigma$ for all $a\in\mathbb{R}$.

The set of measurable functions $f:\Omega\to\mathbb{R}$ is a vector space:

Claim 2.2. *If f,g are measurable then so are $f+g$ and $\lambda\cdot f$.*

Likewise, $\max\{f(x),g(x)\}$ is measurable. Given measurable functions $f_1,f_2,\ldots$, the functions $\sup_n f_n(x),\limsup_n f_n(x)$ are also measurable.

We say that a function $s\colon\Omega\to\mathbb{R}_{\geq 0}$ is an **indicator function** if there is an $A\subseteq\Omega$ such that $s(\omega)=1$ for $\omega\in A$ and $s(\omega)=0$ for $\omega\notin A$. We denote the indicator function of A by $\mathbb{1}_A$. Note that $\mathbb{1}_A$ is measurable if and only if $A\in\Sigma$; we will only consider measurable indicators henceforth.

Let (Ω,Σ) be a measurable space, and let $\mu\colon\Sigma\to[0,\infty]$ be a measure. We would like to define a notion of integral, or the "area" under a measurable function $f:\Omega\to\mathbb{R}_{\geq 0}$.

Formally, denote by L^+ the set of measurable functions $\Omega\to\mathbb{R}_{\geq 0}$. We would like a function $\Phi\colon L^+\to[0,\infty]$ that satisfies the following properties for any $f,g\in L^+$. We will call these properties the axioms of integration.

1. **Calibration.** If $f=\mathbb{1}_A$ then $\Phi(f)=\mu(A)$.

2. **Homogeneity.** $\Phi(\lambda\cdot f)=\lambda\Phi(f)$ for all $\lambda\geq 0$.

3. **Additivity.** $\Phi(f+g)=\Phi(f)+\Phi(g)$.

Note that additivity implies monotonicity: If $f\leq g$ then $\Phi(f)\leq\Phi(g)$. This is because $f\leq g$ means that $g-f\in L^+$, and so $\Phi(g)=\Phi(f+(g-f))=\Phi(f)+\Phi(g-f)\geq\Phi(f)$, because $\Phi(g-f)\in[0,\infty]$.

We will show that when μ is finite there is a unique Φ that satisfies these axioms for bounded functions, and that it furthermore admits a simple form.

The idea is the following. Given an indicator function $f = \mathbb{1}_A$, we define its integral by $\int f(\omega)\,\mathrm{d}\mu(\omega) = \mu(A)$. We say that a function is **simple** if it is a finite linear combination of indicator functions. Equivalently, a function $f\colon \Omega \to \mathbb{R}$ is simple if it has a finite image. Consider a simple function $s\colon \Omega \to \mathbb{R}$ given by $s(\omega) = \sum_{n=1}^{N} a_n \mathbb{1}_{A_n}(\omega)$. We define its integral by

$$\int s(\omega)\,\mathrm{d}\mu(\omega) = \sum_{n=1}^{N} a_n \mu(A_n).$$

Note that there may be multiple representations of s as finite linear combinations of indicator functions, so one needs to check that this is well-defined. The idea behind showing this is writing each simple function as a linear combination of the indicators of a finite partition of Ω. This partition is given by the preimages of the (finitely many) different elements of the image of s. That is, each simple s can be written in a canonical form given by

$$s(\omega) = \sum_{x \in \mathrm{Im}(s)} x \mathbb{1}_{s^{-1}(x)}.$$

Hence

$$\int s(\omega)\,\mathrm{d}\mu(\omega) = \sum_{x \in \mathrm{Im}(s)} x \cdot \mu\left(s^{-1}(x)\right).$$

The motivation for this definition is that if Φ satisfies calibration, homogeneity and additivity then it must be of this form for any simple function. Indeed, using the canonical form it can be shown that $\int (s_1 + s_2)\,\mathrm{d}\mu = \int s_1\,\mathrm{d}\mu + \int s_2\,\mathrm{d}\mu$.

We finally define the **integral** of any measurable $f\colon \Omega \to \mathbb{R}_{\geq 0}$ by

$$\int f(\omega)\,\mathrm{d}\mu(\omega) = \sup\left\{\int s(\omega)\,\mathrm{d}\mu(\omega) : s \text{ is simple and } 0 \leq s \leq f\right\}.$$

Theorem 2.3. *Suppose that μ is a finite measure and Φ satisfies the axioms of integration for every bounded measurable $f \in L^+$. Then for every such f*

$$\Phi(f) = \int f(\omega)\,\mathrm{d}\mu(\omega).$$

Proof. First, we need to show that the map $f \mapsto \int f\,\mathrm{d}\mu$ satisfies the axioms of integration. Indeed, this holds even without the assumption that μ is finite and f is bounded. Monotonicity and homogeneity are immediate from the definition. For calibration, suppose that $f = \mathbb{1}_A$. Then $\int f\,\mathrm{d}\mu \geq \mu(A)$ by the definition of the integral. For any simple s such that $0 \leq s \leq f$ it must hold that $s(\omega) = 0$ for $\omega \in A^c$ and $s(\omega) \leq 1$ for $s \in A$. Hence

$$\int s\,\mathrm{d}\mu = \sum_{x \in \mathrm{Im}(s)} x\mu\left(s^{-1}(x)\right) = \sum_{x \in \mathrm{Im}(s)} x\mu\left(s^{-1}(x) \cap A\right) + \sum_{x \in \mathrm{Im}(s)} x\mu\left(s^{-1}(x) \cap A^c\right) \leq \sum_{x \in \mathrm{Im}(s)} \mu\left(s^{-1}(x) \cap A\right) \leq \mu(A).$$

Subadditivity follows from the definition. We will show additivity later.

To show uniqueness, we observe that calibration, homogeneity and additivity imply that $\Phi(s) = \int s \, \mathrm{d}\mu$ for all simple s. It follows from the definition of the integral and monotonicity that $\Phi(f) \geq \int f \, \mathrm{d}\mu$ for all $f \geq 0$.

If $f \geq 0$ is bounded, then for every $\varepsilon > 0$ there exist simple functions s, t such that $s \leq f \leq t$ and $t - \varepsilon \leq f \leq s + \varepsilon$. To see this, for $n = 1, 2, \ldots$ let $A_n = \{\omega : f(\omega) \in [n\varepsilon, (n+1)\varepsilon)\}$, and let $s = \sum_n n\varepsilon \mathbb{1}_{A_n}$ and $t = \sum_n (n+1)\varepsilon \mathbb{1}_{A_n} = s + \varepsilon\mu(\Omega)$. Note that these sums have only finitely many non-zero summands since f is bounded. By monotonicity and additivity (for simple functions) we have that

$$\Phi(f) \leq \Phi(t) = \int t \, \mathrm{d}\mu = \int s \, \mathrm{d}\mu + \varepsilon\mu(\Omega) \leq \int f \, \mathrm{d}\mu + \varepsilon\mu(\Omega).$$

Since this holds for every $\varepsilon > 0$ we have shown both directions. □

2.2 Monotone convergence

Let (Ω, Σ, μ) be a measure space. We have claimed, but not shown, that $f \mapsto \int f \, \mathrm{d}\mu$ is additive: $\int (f+g) \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu$. In this lecture we will show this, and furthermore show that it is countably additive.

Recall that we denote by L^+ the set of measurable functions $\Omega \to \mathbb{R}_{\geq 0}$.

Claim 2.4. *For any $f \in L^+$, $\int f \, \mathrm{d}\mu = 0$ iff $\mu(\{\omega : f(\omega) > 0\}) = 0$.*

Proof. Let $B_n = f^{-1}([1/n, \infty))$ be the set of ω such that $f(\omega) \geq 1/n$, and denote $B = \cup_n B_n = \{\omega : f(\omega) > 0\}$. Note that $0 \leq \frac{1}{n}\mathbb{1}_{B_n} \leq f$, and so $\int f \, \mathrm{d}\mu \geq \frac{1}{n}\mu(B_n)$.

Suppose that $\int f \, \mathrm{d}\mu = 0$. Then $\mu(B_n) = 0$, and hence $\mu(B) = \mu(\cup_n B_n) = 0$.

Suppose that $\mu(B) = 0$. Then $\mu(B_n) = 0$. Towards a contradiction, suppose that $\int f \, \mathrm{d}\mu > 0$. Then there is some indicator $\mathbb{1}_A$ and $a > 0$ such that $a\mathbb{1}_A \leq f$ and $\mathbb{1}_A > 0$. Hence $f(\omega) \geq a$ for $\omega \in A$, and so $A \subseteq B_n$ for some n large enough. But then $\mu(A) = 0$, and we have reached a contradiction. □

Given $f \in L^+$ and $A \in \Sigma$ we denote

$$\int_A f \, \mathrm{d}\mu = \int f \cdot \mathbb{1}_A \, \mathrm{d}\mu,$$

Claim 2.5. *Given a simple $s \geq 0$, the map*

$$\begin{aligned} \mu_s \colon \Sigma &\to [0, \infty] \\ A &\mapsto \int_A s \, \mathrm{d}\mu \end{aligned}$$

is a measure on (Ω, Σ).

Proof. Let $A_1, A_2, \ldots \in \Sigma$ be disjoint, let $s = \sum_{k=1}^K b_k \mathbb{1}_{B_k}$, and denote $A = \cup_n A_n$. Then $s \cdot \mathbb{1}_A = \sum_{k=1}^K b_k \mathbb{1}_{B_k \cap A}$, and so

$$\mu_s(A) = \int_A s \, \mathrm{d}\mu = \int s(\omega)\mathbb{1}_A(\omega) \, \mathrm{d}\mu = \sum_{k=1}^K b_k \mu(B_k \cap A) = \sum_{k=1}^K b_k \sum_n \mu(B_k \cap A_n) = \sum_n \mu_s(A_n).$$

□

Theorem 2.6 (Monotone Convergence)**.** *Consider $f_1, f_2, \ldots \in L^+$ such that $f_{n+1} \geq f_n$ for all n and $f(\omega) = \lim_n f_n(\omega) < \infty$ for all ω. Then $\lim_n \int f_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$.*

Proof. By monotonicity $\int f \, \mathrm{d}\mu \geq \lim_n \int f_n \, \mathrm{d}\mu$. For the other direction, let s be a simple function such that $0 \leq s \leq f$, and fix any $\alpha \in (0,1)$. Let $A_n = \{\omega : f_n(\omega) \geq \alpha s(\omega)\}$. Since $\lim_n f_n = f$, and since $\alpha s(\omega) < f(\omega)$ for all ω such that $f(\omega) > 0$, we have that $A_n \subseteq A_{n+1}$ and $\cup A_n = \Omega$. It thus follows from Claim 2.5 that

$$\int \alpha s \, \mathrm{d}\mu = \int_\Omega \alpha s \, \mathrm{d}\mu = \mu_S(\Omega) = \lim_n \mu_S(A_n) = \lim_n \int_{A_n} \alpha s \, \mathrm{d}\mu \leq \lim_n \int_{A_n} f_n \, \mathrm{d}\mu \leq \lim_n \int f_n \, \mathrm{d}\mu.$$

Since this holds for every such s, we can conclude that

$$\alpha \int f \, \mathrm{d}\mu \leq \lim_n \int f_n \, \mathrm{d}\mu.$$

And since this holds for every $\alpha \in (0,1)$ this holds also for $\alpha = 1$.

□

Theorem 2.7. *For $f, g \in L^+$ it holds that $\int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu = \int (f+g) \, \mathrm{d}\mu$.*

Proof. Similarly to the proof of Theorem 2.3,we can find a sequence of simple functions $s_1, s_2, \ldots$ such that $0 \leq s_n \leq f$, $s_{n+1} \geq s_n$ and $\lim_n s_n(\omega) = f(\omega)$ for all ω: Let

$$A_n^k = \left\{\omega : f(\omega) \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)\right\}$$

and let $s_n = \sum_{k=1}^{2^{2n}} \frac{k}{2^n} \mathbb{1}_{A_n^k}$. Construct a similar sequence $t_1, t_2, \ldots$ for g.

Since $s_n + t_n \leq f + g$ and $\lim_n s_n + t_n = f + g$, it follows from the Monotone Convergence Theorem (Theorem 2.6) that

$$\int (f+g) \, \mathrm{d}\mu = \lim_n \int (s_n + t_n) \, \mathrm{d}\mu.$$

Since integration is additive for simple functions, this equals to $\int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu$. □

Theorem 2.8. *Let $f_1, f_2, \ldots \in L^+$. Then $\sum_n \int f_n \, \mathrm{d}\mu = \int \sum_n f_n \, \mathrm{d}\mu$.*

Proof. Let $g_n = \sum_{k=1}^n f_n$. Then $g_1, g_2, \ldots$ satisfy the conditions of the Monotone Convergence Theorem (Theorem 2.6) and so

$$\int \sum_n f_n \, \mathrm{d}\mu = \int \lim_n g_n \, \mathrm{d}\mu = \lim_n \int g_n \, \mathrm{d}\mu_n = \lim_n \int \sum_{k=1}^n f_n \, \mathrm{d}\mu_n.$$

Since we have finite additivity (Theorem 2.7) we can exchange the sum and integral, which concludes the proof. □

Given a measurable function $f : \Omega \to \mathbb{R}$, we denote

$$f^+(\omega) = \max\{f(\omega), 0\}$$
$$f^-(\omega) = \max\{-f(\omega), 0\}.$$

We say that f is integrable if $\int f^+ \, \mathrm{d}\mu$ is finite or $\int f^- \, \mathrm{d}\mu$ is finite. For integrable f we define

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

2.3 Fatou's Lemma and dominated convergence (by Joey Litvin and Bharathan Sundar)

We denote by $L = L(\Omega, \Sigma, \mu)$ the set of equivalence classes of measurable functions $\Omega \to \mathbb{R}$, under the equivalence relation in which f and g are equivalent if

$$\mu(\{\omega,\ f(\omega) \neq g(\omega)\}) = 0.$$

That is, if f and g agree almost everywhere.

Denote by $L^1 = L^1(\Omega, \Sigma, \mu) \subset L$ the set of equivalence classes of integrable functions $f : \Omega \to \mathbb{R}$ such that $\int f \, \mathrm{d}\mu \in \mathbb{R}$. As for L^0, we will sometimes abuse notation by writing $f \in L^1$, by which we mean that the equivalence class of f is in L^1.

Claim 2.9. *For $f, g \in L^1$, if $\mu(\{\omega : f \neq g\}) = 0$ then $\int_A f \, \mathrm{d}\mu = \int_A g \, \mathrm{d}\mu$ for all $A \in \Sigma$.*

In this lecture we will build upon the monotone convergence theorem and prove Fatou's Lemma. We will prove an important result about the convergence of a sequence of functions in L^1: the (Lebesgue) Dominated Convergence Theorem. Finally, we will prove an interesting corollary about a series of functions in L^1. Recall that we denote by L^+ the set of measurable functions $\Omega \to \mathbb{R}_{\geq 0}$.

Theorem 2.10 (Fatou's Lemma)**.** *Consider $f_1, f_2, \ldots \in L^+$. Then $\liminf_n \int f_n \, \mathrm{d}\mu \geq \int \liminf_n f_n \, \mathrm{d}\mu$.*

Proof. First notice that for $k \geq 1$ we have that $\inf_{n \geq k} f_n \leq f_j$ for any $j > k$. Thus we have

$$\int \inf_{n \geq k} f_n \, \mathrm{d}\mu \leq \int f_j \, \mathrm{d}\mu$$

when $j > k$. From this we can also deduce that

$$\int \inf_{n \geq k} f_n \, \mathrm{d}\mu \leq \inf_{j \geq k} \int f_j \, \mathrm{d}\mu$$

since our only condition on j was that $j \geq k$. Now taking $k \to \infty$ this inequality becomes:

$$\lim_{k \to \infty} \int (\inf_{n \geq k} f_n) \, \mathrm{d}\mu \leq \liminf \int f_n \, \mathrm{d}\mu$$

Finally note that the sequence $(\int \inf_{n \geq k} f_n \, \mathrm{d}\mu)_k$ satisfies $f_k \leq f_{k+1}$ for each $k \geq 1$. Hence we can apply the monotone convergence theorem to deduce that

$$\liminf \int f_n \, \mathrm{d}\mu \geq \lim_{k \to \infty} \int \inf_{n \geq k} f_n \, \mathrm{d}\mu = \int \liminf f_n \, \mathrm{d}\mu.$$

$\square$

Let us look at a concrete example where we have strict inequality. Consider the sequence of functions

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

. Here, we note that $\int \liminf_n f_n \,\mathrm{d}\mu = 0$, while $\liminf_n \int f_n \,\mathrm{d}\mu = 1$, so we have a case where equality is broken. Intuitively, Fatou's lemma shows us that that integrating after taking the lim inf can cause us to lose some "mass" in the process.

Now that we have the Monotone Convergence Theorem and Fatou's lemma under our belt, we have the machinery we need in order to present the main convergence result: the Dominated Convergence Theorem.

Theorem 2.11 (Dominated convergence)**.** *Consider $f_1, f_2, \ldots \in L^1$ such that $\lim_n f_n(\omega) = f(\omega)$ for some f and almost every $\omega \in \Omega$, and there exists a $g \in L^1$ such that $|f_n| \le g$ almost everywhere for all n. Then $f \in L^1$ and $\lim_n \int f_n \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu$.*

Proof. First we need to show that f is a measurable function. The rigorous proof of this is a bit tedious and not the most interesting, and so we will skip it. Also we have that $|f| \le g$ almost everywhere since $|f_n| \le g$ almost everywhere. From this we deduce that $f \in L^1$ since $g \in L^1$.
Next notice that

$$|f_n| \le g \implies 0 \le g - |f_n| \le g - f_n$$

and

$$|f_n| \le g \implies 0 \le g - |f_n| \le g + f_n$$

Now we want to apply Fatou's Lemma to get:

$$\int g \,\mathrm{d}\mu + \int f \,\mathrm{d}\mu = \int (g + \liminf f_n) \,\mathrm{d}\mu \le \liminf \int (g + f_n) \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu + \liminf \int f_n \,\mathrm{d}\mu$$

Again using Fatou's lemma with the other inequality we get:

$$\int g \,\mathrm{d}\mu - \int f \,\mathrm{d}\mu = \int (g - \limsup f_n) \,\mathrm{d}\mu \le \liminf \int (g - f_n) \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu - \limsup \int f_n \,\mathrm{d}\mu.$$

From these two inequalities we have that

$$\limsup_n \int f_n \,\mathrm{d}\mu \le \int f \,\mathrm{d}\mu \le \liminf \int f_n \,\mathrm{d}\mu$$

from which we are able to deduce that $\lim_{n\to\infty} \int f_n \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu$. $\square$

We now return to a version of the example we explore for Fatou's lemma, where $f_n = n\mathbb{1}_{(0,1/n]}$. Here, we have the functions f_n as "boxes" of length $\frac{1}{n}$, with height n. We note that no function dominates this sequence. As in the case of Fatou's lemma, the limit of the integral tends to 1, while the integral of the limit tends to 0.

Finally, we can use the Dominated Convergence Theorem to better understand a series of functions in L^1. In particular, we can prove a similar Dominated Convergence Theorem type result for series of functions.

Corollary 2.12 (Dominated Convergence Theorem for series of functions)**.** *Consider $f_1, f_2, \ldots \in L^1$ such that $\sum_n \int |f_n| \, \mathrm{d}\mu < \infty$. Then $\sum_n f_n$ converges almost everywhere to some $f \in L^1$, and $\sum_n \int f_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$.*

Proof. By Theorem 2.8, we have that $\int \sum_1^\infty |f_n| = \sum_1^\infty \int |f_n| < \infty$. Define $g = \sum_1^\infty |f_n|$, and note that $g \in L^1$ by assumption. Then, by the above claim, we note that $\sum_1^\infty |f_n(\omega)|$ is finite almost everywhere, and for almost every ω we have that $\sum_1^\infty f_n(\omega)$ converges. Now, we can define $h_k = \sum_{n=1}^k f_n$, and similarly $h = \lim_k h_k$. We note that $|h_k| \leq g$ almost everywhere by the triangle inequality $\forall k$. We now apply the Dominated Convergence Theorem to the sequence of partial sums. We write that

$$\sum_{n=1}^{\infty} \int f_n \, \mathrm{d}\mu = \lim_k \sum_{n=1}^{k} \int f_n \, \mathrm{d}\mu$$

. By finite additivity of integrals in L^1, we can interchange, so

$$= \lim_k \int \sum_{n=1}^{k} f_n \, \mathrm{d}\mu = \lim_k \int h_k \, \mathrm{d}\mu$$

Now applying the Dominated Convergence Theorem yields

$$= \int h \, \mathrm{d}\mu = \int \sum_{n=1}^{\infty} f_k \, \mathrm{d}\mu$$

as desired. □

2.4 Modes of convergence

Let (Ω, Σ, μ) be a measure space, and let $f, f_1, f_2, \ldots$ be measurable functions $\Omega \to \mathbb{R}$. There are a number of different interesting senses of convergence of the sequence $f_1, f_2, \ldots$ to f, which we will explore in this lecture.

A strong sense of convergence is **uniform convergence**: f_n converges uniformly to f if for all $\varepsilon > 0$ it holds for all n large enough that $|f_n(\omega) - f(\omega)| < \varepsilon$. This notion does not depend on the measure. As an example $f_n = n^{-1} \mathbb{1}_{[0,n]}$ converges uniformly to 0.

Another natural notion that does not depend on the measure is **pointwise convergence**, in which f_n converges to f if $\lim_n f_n(\omega) = f(\omega)$ for all $\omega \in \Omega$. Again, this does not depend on μ. Uniform convergence implies pointwise convergence, but the converse is not true unless Ω is finite. As an example, $f_n = n \mathbb{1}_{(0,n^{-1})}$ converges pointwise to 0, but not uniformly, as does $f_n = \mathbb{1}_{[n,n+1]}$.

Since we have a measure, a natural related notion is **almost everywhere pointwise convergence**, which holds when $\lim_n f_n(\omega) = f(\omega)$ for all ω in some co-null set. Clearly pointwise convergence implies almost everywhere pointwise convergence.

Let μ, η be two measures defined on (Ω, Σ). We say that they are in the same **measure class** (or just equivalent) if they have the same null sets: $\mu(A) = 0$ iff $\eta(A) = 0$. Note that, unlike pointwise convergence, a.e. convergences does depend on the measure, but only on the measure class.

Recall that L is the set of equivalence classes of measurable functions where f,g are equivalent if they agree almost everywhere, and that $L^1 \subseteq L$ is the subset of (equivalence classes) of f such that $\int |f| \mathrm{d}\mu < \infty$. We define a metric on L^1, given by $D_1 \colon L^1 \times L^1 \to \mathbb{R}_{\geq 0}$ given by $D_1(f,g) = \int |f-g| \mathrm{d}\mu$. Note that this is indeed well defined, satisfies the triangle inequality, and is equal to 0 iff f,g are equivalent. We say that f_n converges to f in L^1 if $\lim_n \int |f_n - f| \mathrm{d}\mu = 0$. Note that none of the above notions imply convergence in L^1. However, if μ is finite then a.e. uniform convergence implies convergence in L^1:

Claim 2.13. *If $\mu(\Omega) < \infty$ and $f_n \to f$ uniformly then $f_n \to f$ in L^1.*

Proof. Fix $\varepsilon > 0$. Since $f_n \to f$ uniformly, $|f_n - f| < \varepsilon$ almost everywhere for all n large enough. It follows that $\int |f_n - f| \mathrm{d}\mu < \varepsilon\mu(\Omega)$, and since this holds for all $\varepsilon > 0$ we have proved the claim. $\square$

In the other direction, convergence in L^1 does not even imply a.e. pointwise convergence, and not even in finite measure spaces. To see this, suppose that $\Omega = [0,1]$ and μ is the Lebesgue measure. For $n = 2^k + m$, $m < 2^k$, let $g_n = \mathbb{1}_{2^{-k}[m,m+1]}$. Then $g_n(\omega)$ does not converge for any ω, but $\int |g_n| \mathrm{d}\mu = 2^{-k}$, and so $f_n \to f$ in L^1.

We say that f_n converges to f **in measure** if for all $\varepsilon > 0$

$$\lim_n \mu(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) = 0.$$

As an example, g_n as defined above converges in measure to 0, as does $f_n = n\mathbb{1}_{(0,n^{-1})}$, but $f_n = \mathbb{1}_{(n,n+1)}$ does not. Hence convergence in measure does not imply pointwise convergence, or convergence in L^1, and is not implied by pointwise convergence.

Claim 2.14. *If $f_n \to f$ in L^1 then $f_n \to f$ in measure.*

Proof. Let $A_n = \{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}$. Then

$$0 = \lim_n \int |f_n - f| \mathrm{d}\mu \geq \lim_n \int_{A_n} |f_n - f| \mathrm{d}\mu \geq \lim_n \varepsilon\mu(A_n) \geq 0.$$

$\square$

Proposition 2.15. *If $f_n \to f$ in measure, then there exists a sequence $(n_k)_k$ such that $f_{n_k} \to f$ almost everywhere.*

Proof. For each k choose n_k large enough so that $\mu(A_k) \leq 2^{-k}$ where $A_k = \{\omega | f_{n_k}(\omega) - f(\omega)| > 2^{-k}\}$. Let $g_k = f_{n_k}$, and let

$$B_m = \cup_{k=m+1}^{\infty} A_k = \left\{\omega : |g_k(\omega) - f(\omega)| > 2^{-k} \text{ for some } k > m\right\}.$$

Then $\mu(B_m) \leq 2^{-m}$ and

$$B_m^c = \left\{\omega : |g_k(\omega) - f(\omega)| \leq 2^{-k} \text{ for all } k > m\right\}.$$

Hence on B_m^c it holds that $g_k \to f$ uniformly, and in particular pointwise.

Note that $(B_m)_m$ is a decreasing sequence and hence $(B_m^c)_m$ is an increasing sequence. Furthermore, $B = \cup_m B_m^c$ is co-null, and on B it holds that $g_k \to f$ pointwise. $\square$

2.5 Measures on $\mathbb{R}^d$ (by Jonah Yoshida and Wei Hou)

Recall the way to construct an outer measure from a premeasure, and the way to construct a measure using the Carathéodory extension theorem. We will construct a measure on the product σ-algebra $\Sigma_1 \bigotimes \Sigma_2$ from two σ-algebras Σ_1 and Σ_2. If we accomplish this, we can boostrap this step to get measures on $\mathbb{R}^n$. Let $\{X_\alpha\}_{\alpha\in A}$ be a collection of nonempty sets indexed by A. $\pi_\alpha : X \to X_\alpha$ the coordinate maps. If Σ_α is a σ-algebra on X_α for each α, **the product σ-algebra** on X is the σ-algebra generated by:

$$\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha \in \Sigma_\alpha, \alpha \in A\} \tag{2.1}$$

and denoted by $\bigotimes_{\alpha\in A} \Sigma_\alpha$

Proposition 2.16. *If A is countable, then $\bigotimes_{\alpha\in A} \Sigma_\alpha$ is the σ-algebra generated by $\{\Pi_{\alpha\in A} E_\alpha : E_\alpha \in \Sigma_\alpha\}$*

Now, in particular, if we are only considering the product σ-algebra generated by 2 σ-algebras, i.e. $\Sigma_1 \bigotimes \Sigma_2$, Prop. 2.16 still holds.

To define the product measure, we first construct a premeasure. Suppose C is a finite union of some disjoint sets $A_i \times B_i$ where $A_i \in \Sigma_2$ and $B_i \in \Sigma_2$. Then we define:

$$\pi(C) = \sum_i \mu(A_i)\nu(B_i) \tag{2.2}$$

this can be defined for all sets in the algebra generated by $\{A_i \times B_i\}_i$. Thus, π is a premeasure on this algebra. This premeasure can be used to define an outermeasure (Eq. 1.1). Using Theorem 1.8, we can define a measure on the product σ-algebra.

We now characterize integration under the product measure. Before doing this, we establish some notations and technical results that will help us construct the integrals under the product measure. Returning to the two measure spaces (X, Σ_1, μ) and (Y, Σ_2, ν). Define the x-section E_x and y-section E^y of E by:

$$E_x = \{y \in Y : (x, y) \in E\}, \quad E^y = \{x \in X : (x, y) \in E\} \tag{2.3}$$

and also if f is a function on $X \times Y$, we define the x-section f_x and y-section f^y of f by

$$f(x, y) = f^y(x) = f_x(y) \tag{2.4}$$

To integrate the functions slice by slice, we need the following Proposition:

Proposition 2.17. *We have the following:*

1. *If $E \in \Sigma_1 \bigotimes \Sigma_2$, then $E_x \in \Sigma_2\ \forall x \in X$ and $E^y \in \Sigma_1\ \forall y \in Y$*

2. *If f is $\Sigma_1 \bigotimes \Sigma_2$-measurable, then f_x is Σ_2-measurable $\forall x \in X$ and f^y is Σ_1-measurable $\forall y \in Y$*

Next, we define a **monotone class** on a space X to be a subset of 2^X that is closed under countable increasing unions and countable decreasing intersections. i.e.

$$E_i \in \mathscr{C} \text{ and } E_1 \subset E_2 \subset \ldots \Rightarrow \bigcup E_j \in \mathscr{C}$$
$$E_i \in \mathscr{C} \text{ and } E_1 \supset E_2 \supset \ldots \Rightarrow \bigcap E_j \in \mathscr{C}$$

Clearly, a σ-algebra is a monotone class, here we show that the monotone class generated by an algebra is the same as the σ-algebra generated by the same algebra.

Lemma 2.18 (The Monotone Class Lemma)**.** *If $\mathscr{A}$ is an algebra of subsets of X, then the monotone class $\mathscr{C}$ generated by $\mathscr{A}$ coincides with the σ-algebra Σ generated by $\mathscr{A}$*

Proof. Clearly, Σ is a monotone class (stable under countable union). Since $\mathscr{C}$ is the minimal monotone class containing $\mathscr{A}$, $\mathscr{C} \subset \Sigma$. Then, as long as we show that $\mathscr{C}$ is a σ-algebra, we can conclude that $\mathscr{C} = \Sigma$.

Since $\mathscr{A} \subset \mathscr{C}$, $\phi \in \mathscr{C}$ and $\Omega \in \mathscr{C}$. Then define the following:

$$\mathscr{C}(E) = \{F \in \mathscr{C} : E \backslash F,\, F \backslash E, \text{ and } E \cap F \text{ are in } \mathscr{C}\} \tag{2.5}$$

One can check that $\mathscr{C}(E)$ is a monotone class if $E \in \mathscr{C}$. Also, if $E \in \mathscr{A}$, then $\mathscr{A} \subset \mathscr{C}(E)$. Thus, $\mathscr{C} \subset \mathscr{C}(E)$. Therefore, $\mathscr{C} \subset \mathscr{C}(E)\, \forall E \in \mathscr{A}$.

In addition, it is easy to check $E \in \mathscr{C}(F)$ iff $F \in \mathscr{C}(E)$. Thus, $\forall E \in \mathscr{A}$ and $\forall F \in \mathscr{C}$, we have $E \in \mathscr{C}(F)$. This implied that $\forall F \in \mathscr{C}, \mathscr{A} \in \mathscr{C}(F)$. But $\mathscr{C}(E) \subset \mathscr{C}\, \forall E \in \mathscr{C}$, we have $\mathscr{C} = \mathscr{C}(E)\, \forall E \in \mathscr{C}$. Thus, if $E, F \in \mathscr{C}$, then $E \backslash F \in \mathscr{C}$ and $E \cap F \in \mathscr{C}$. Also, since $\mathscr{A} \in \mathscr{C}$, we have $\phi, \Omega \in \mathscr{C}$. We only need to check that $\mathscr{C}$ is closed under countable unions.

Let E_1, E_2, ... be a sequence of countable sets. Since $\mathscr{C}$ is closed under finite unions, define $K_n = \bigcup_{i=1}^n E_i$. We have $K_1 \subset K_2 \subset \ldots \in \mathscr{C}$. Since $\mathscr{C}$ is a monotone class, we have $\bigcup_{i=1}^\infty E_i = \bigcup_{i=1}^\infty K_i \in \mathscr{C}$. This shows closure under countable unions. $\square$

Now, with the Monotone Class Lemma, we are endowed with a tool to prove a theorem about how the product measure splits on the product σ-algebra. First, we create a collection of sets $\mathscr{C}$ in the product σ-algebra that satisfy the desired conclusions, then we prove that $\mathscr{C}$ is indeed a monotone class so that the Monotone Class Lemma guarantees $\mathscr{C}$ is in fact the entire product σ-algebra.

Theorem 2.19. *Suppose (X, Σ_1, μ) and (Y, Σ_2, ν) are σ-finite measure spaces. If $E \in \Sigma_1 \otimes \Sigma_2$, then the functions $x \mapsto \nu(E_x)$ and $y \mapsto \mu(E^y)$ are measurable on X and Y, respectively, and $[\mu \times \nu](E) = \int \nu(E_x) \mathrm{d}\mu(x) = \int \mu(E^y) \mathrm{d}\nu(y)$.*

Proof. Let $\mathscr{C} \subseteq \Sigma \otimes \Sigma_2$ such that $\forall E \in \mathscr{C}$, the theorem holds. We wish to show $\mathscr{C} = \Sigma \otimes \Sigma_2$. Assume μ and ν are finite on X and Y. Note that $E = A \times B \in \mathscr{C}$, as $\nu(E_x) = \mathbb{1}_A(x)\nu(B)$ and $\mu(E^y) = \mathbb{1}_B(y)\mu(A)$, so that

$$\int \nu(E_x)\, \mathrm{d}\mu(x) = \int \mathbb{1}_A(x)\nu(B)\, \mathrm{d}\mu(x) = \mu(A)\nu(B) = \int \mathbb{1}_B(y)\mu(A)\, \mathrm{d}\nu(y) = \int \mu(E^y)\, \mathrm{d}\nu(y).$$

Furthermore, by the additivity of both measures, all finite disjoint unions of rectangles lie in $\mathscr{C}$ and the rectangles form an algebra $\mathscr{A}$ as a subset of $X \times Y$ and $\mathscr{C}$. If we can now prove that $\mathscr{C}$ is a monotone class, the Monotone Class Lemma will yield the desired result. Let $\{E_n\}$ be an increasing sequence in $\mathscr{C}$. We wish to show that $E = \cup_n E_n \in \mathscr{C}$. Observe first that because $(E_n)^y \in \Sigma_2$, $f_n(y) = \mu((E_n)^y)$ is measurable $\forall n$ and $f_n \to f(y) = \mu(E^y)$. Now $f_n \in L^+$ with $f_{n+1} \geq f_n$ and $f(y) < \infty$ so that the Monotone Convergence Theorem applies:

$$\mu \times \nu(E) = \lim_n \mu \times \nu(E_n) = \lim_n \int f_n(y)\,\mathrm{d}\nu(y) = \int f(y)\,\mathrm{d}\nu(y) = \int \mu(E^y)\,\mathrm{d}\nu(y),$$

and the same holds for the $\mu \times \nu(E) = \int \nu(E_x)\,\mathrm{d}\mu(x)$ property. Therefore, $E \in \mathscr{C}$. Now, if $E = \cap_n E_n$ for a decreasing sequence $\{E_n\}$, we have that $g(y) := y \mapsto \mu((E_1)^y)$ satisfies $g \in L^1, |f_n| \leq g$. Therefore, Dominated Convergence implies the same sufficient condition for $E \in \mathscr{C}$: $\lim_n \int f_n(y)\,\mathrm{d}\nu(y) = \int f(y)\,\mathrm{d}\nu(y)$.

Now, if μ and ν are σ-finite but not necessarily both finite, $X \times Y = \cup_j \{X_j \times Y_j\}$ with $\{X_j \times Y_j\}$ increasing and each $\{X_j \times Y_j\}$ of finite measure. Now, $\forall E \in \Sigma \otimes \Sigma_2$, $E \cap (X_j \times Y_j)$ is finite $\forall j$ so that the above argument applies and we have

$$\mu \times \nu(E \cap (X_j \times Y_j)) = \int \mathbb{1}_{X_j}(x)\nu(E_x \cap Y_j)\,\mathrm{d}\mu(x) = \int \mathbb{1}_{Y_j}(y)\mu(E^y \cap X_j)\,\mathrm{d}\nu(y).$$

Applying the Monotone Convergence Theorem one last time, we have

$$\begin{aligned}
\mu \times \nu(E \cap (X \times Y)) &= \lim_j \mu \times \nu(E \cap (X_j \times Y_j)) \\
&= \lim_j \int \mathbb{1}_{X_j}(x)\nu(E_x \cap Y_j)\,\mathrm{d}\mu(x) \\
&= \int \mathbb{1}_X(x)\nu(E_x \cap Y)\,\mathrm{d}\mu(x) \\
&= \int \nu(E_x \cap (X \times Y))\,\mathrm{d}\mu,
\end{aligned}$$

and the same equality for $\mu(E^y \cap (X \times Y))$. $\square$

We now know how the product measure can be computed: taking E^y or E_x slices and integrating the measure of these slices over points x or y in the other space. The theorem above thus gives us necessary tools to prove the infamous Fubini-Tonelli Theorem.

Theorem 2.20 (Fubini-Tonelli)**.** *Suppose (X, Σ, μ) and (Y, Σ_2, ν) are σ-finite measure spaces.*

1. *(Tonelli) If $f \in L^+(X \times Y)$, then the functions $g(x) = \int f_x\,\mathrm{d}\nu$ and $h(y) = \int f^y\,\mathrm{d}\mu$ are in $L^+(X)$ and $L^+(Y)$, respectively, and $\int f\,\mathrm{d}(\mu\times\nu) = \int[\int f(x,y)\,\mathrm{d}\nu(y)]\,\mathrm{d}\mu(x) = \int[\int f(x,y)\,\mathrm{d}\mu(y)]\,\mathrm{d}\nu(x)$.*
2. *(Fubini) If $f \in L^1(\mu \times \nu)$, then $f_x \in L^1(\nu)$ for a.e. $x \in X, f^y \in L^1(\mu)$ for a.e. $y \in Y$, the a.e.-defined functions $g(x) = \int f_x\,\mathrm{d}\nu$ and $h(y) = \int f^y\,\mathrm{d}\mu$ are in $L^1(\mu)$ and $L^1(\nu)$, respectively, and Tonelli holds.*

Proof. Observe that Tonelli's theorem follows immediately from Theorem 2.19 when f is an indicator function. By the linearity of the integrals in Theorem 2.19, the first part holds for non-negative simple functions. Let $f \in L^+(X \times Y)$. Then, $\exists s_n \to f$ with $s_n \leq f_n$ simple. Furthermore, $\exists g_n$ and h_n that satisfy the desired integrals. By the Monotone Convergence Theorem, $\lim_n g_n = \lim_n \int (s_n)_x \, \mathrm{d}\nu = \int f_x \mathrm{d}\nu = g$ and $\lim_n h_n = \lim_n \int (s_n)^y \, \mathrm{d}\mu = \int h^y \, \mathrm{d}\mu = h \implies g, h$ measurable. Furthermore, $\int g \, \mathrm{d}\mu = \lim \int g_n \, \mathrm{d}\mu = \lim \int s_n \, \mathrm{d}(\mu \times \nu)$, by Tonelli's on g_n, $= \int f \, \mathrm{d}(\mu \times \nu)$ and $\int h \, \mathrm{d}\mu = \lim \int h_n \, \mathrm{d}\mu = \lim \int s_n \, \mathrm{d}(\mu \times \nu)$, by Tonelli's on h_n, $= \int f \, \mathrm{d}(\mu \times \nu)$, as desired.

Now, if $f \in L^1(\mu \times \nu)$, $f \in L^+(X \times Y)$ so that we may apply Tonelli's theorem. As a result, $\int g \, \mathrm{d}\mu = \int f \, \mathrm{d}(\mu \times \nu) < \infty$ by $f \in L^1(\mu \times \nu) \implies g \in L^1(\mu)$ and $\int h \, \mathrm{d}\mu = \int f \, \mathrm{d}(\mu \times \nu) < \infty$ by $f \in L^1(\mu \times \nu) \implies h \in L^1(\mu)$. Lastly, expanding g and h by their definitions in Tonelli yields $f^y \in L^1(\mu)$ for a.e. $y \in Y$ and $f_x \in L_1(\nu)$ for a.e. $x \in X$. $\square$

3 Differentiation

3.1 Signed measures

Let (Ω,Σ,μ) be a measure space, and fix $f \in L^+$. Then it can be shown that

$$\nu(A) = \int_A f \,\mathrm{d}\mu$$

is also a measure on (Ω,Σ). If μ is the Lebesgue measure on $\mathbb{R}$ then we can think of f as capturing the density of mass along $\mathbb{R}$, and so ν measures the total mass of a subset of $\mathbb{R}$.

We can do something similar for f that is not necessarily non-negative, and likewise define

$$\eta(A) = \int_A f \,\mathrm{d}\mu.$$

This will not be well defined for $C = A \cup B$ if A and B are disjoint, $\eta(A) = \infty$ and $\eta(B) = -\infty$. So we would need some additional assumption about f, namely that it is integrable. Note that we can write $\eta = \mu_1 - \mu_2$, where μ_1,μ_2 are measures given by

$$\mu_1(A) = \int f^+ \,\mathrm{d}\mu$$
$$\mu_2(A) = \int f^- \,\mathrm{d}\mu.$$

Note that the two measures μ_1 and μ_2 “live in different places”. We now do this more formally.

Let (Ω,Σ) be a measurable space, and let μ_1,μ_2 be σ-finite measures on it. We say that μ_1,μ_2 are **mutually singular** if there exists a $P,N \in \Sigma$ such that P is co-null for μ_1 and $N = P^c$ is co-null for μ_2. Informally, this means that μ_1 and μ_2 exist on disjoint sets.

In this lecture we will think of P as having positive measure, and $N = P^c$ as having negative measure. The total measure of some $A \in \Sigma$ will be the measure of its intersection with P minus the measure of its intersection with N. To avoid taking differences of infinities, we will need to assume that at least one of μ_1,μ_2 is a finite measure.

Accordingly, we will say that $\eta\colon \Sigma \to [-\infty,\infty]$ is a **signed measure** if $\eta = \mu_1 - \mu_2$, for mutually singular μ_1,μ_2, where at least one of μ_1,μ_2 is finite, and both are σ-finite. Note that η cannot attain both $+\infty$ and $-\infty$. To simplify the exposition we will assume in this lecture that η never attains $+\infty$. Equivalently, we assume that μ_1 is a finite measure. Note that for P witnessing the mutual singularity of μ_1 and μ_2 it holds that

$$\eta(A) = \mu_1(A \cap P) - \mu_2(A \cap P^c).$$

Proposition 3.1. *Suppose that $\eta = \mu_1 - \mu_2 = \nu_1 - \nu_2$ for some mutually singular μ_1,μ_2 and likewise mutually singular ν_1,ν_2. Then $\mu_1 = \nu_1$ and $\mu_2 = \nu_2$.*

To prove this xproposition we will need the notion of a positive set. We say that $A \in \Sigma$ is **positive** if $\eta(B) \geq 0$ for all $B \subseteq A$. If $A_1 \cap A_2 = \emptyset$ and$\mu_1(P^c) = \mu_2(P) = 0$, then P is a positive set. Analogously, a set A is **negative** if $\eta(B) \leq 0$ for all $B \subseteq A$.

Claim 3.2. *The collection of positive sets is closed under countable unions.*

This follows from the continuity of measures.

Note that if $P \subseteq \Omega$ is a positive set then $\mu_1(A) = \eta(P \cap A)$ is a measure.

Proof of Proposition 3.1. Since μ_1, μ_2 are mutually singular there exists a $P \in \Sigma$, such that $\mu_1(P^c) = \mu_2(P) = 0$. Likewise, there is a Q such that $\nu_1(Q^c) = \nu_2(Q) = 0$. Hence

$$\eta(A) = \mu_1(A \cap P) - \mu_2(A \cap P^c) = \nu_1(A \cap Q) - \nu_2(A \cap Q^c).$$

Hence $P \cup Q$ is a positive set and $P^c \cup Q^c$ is a negative set. Since a set that is both negative and positive is null (i.e., η assigns zero to all of its subsets), it follows that $P' = P \cap Q$ is positive, $N' = P^c \cap Q^c$ is negative, and the remainder $\Omega \setminus (P' \cup N')$ is null. Hence

$$\eta(A) = \eta(A \cap (P' \cup N')).$$

Hence, for any $A \in \Sigma$,

$$\mu_1(A) = \eta(A \cap P) = \eta(A \cap P') = \eta(A \cap Q) = \nu_1(A).$$

A similar calculation shows that $\mu_2(A) = \nu_2(A)$. $\square$

Given $\eta = \mu_1 - \mu_2$ (written as the difference of mutually singular measures), we denote by $|\eta| = \mu_1 + \mu_2$ the **total variation** of η, which is a measure. By Proposition 3.1 this is well defined. We say that ν is finite if its total variation is finite.

The next theorem gives an intrinsic characterization of signed measures.

Theorem 3.3 (Jordan Decomposition Theorem)**.** *For a map $\eta \colon \Sigma \to [-\infty, +\infty)$ the following are equivalent.*

1. *η is a signed measure.*
2. *η has the following properties:*
   (a) *$\eta(\emptyset) = 0$.*
   (b) *If $A_1, A_2, \ldots \in \Sigma$ are disjoint with $A = \cup_n A_n$ then $\eta(A) = \sum_n \eta(A_n)$.*

Note that (2b) implies that the sum in (2a) converges absolutely if $\eta(A) < \infty$.

A consequence of this result is that if μ_1 is finite and μ_2 is σ-finite, then $\eta = \mu_1 - \mu_2$ is a signed measure even if μ_1, μ_2 are not mutually singular. To see this, verify that η satisfies (2).

Proof of Theorem 3.3. That (1) implies (2a) and (2b) follows from the definition of measures and signed measures.

Suppose η satisfies (2a) and (2b). We claim that if $A_1, A_2 \ldots \in \Sigma$ is an increasing sequence and $A = \cup_n A_n$ then $\eta(A) = \lim_n \eta(A_n)$. The same holds if $A_1, A_2 \ldots \in \Sigma$ is a decreasing sequence and $A = \cap_n A_n$. The proof is the same as for measures. Note that this implies that η is bounded from above.

Denote by $\mathscr{P} \subset \Sigma$ the collection of positive sets, and let $m = \sup_{A \in \mathscr{P}} \eta(A)$. Let $Q_1, Q_2, \ldots \in \mathscr{P}$ be a sequence such that $\lim_n \eta(Q_n) = m$, let $P_n = \cup_{k=1}^n Q_k$, and let $P = \cup_n P_n$. Then by Claim 3.2 P is also a positive set, and since $P_1, P_2, \ldots$ is an increasing sequence, $\eta(P) = m < \infty$.

Denote $N = P^c$. Then by (2b) we have that for any $A \in \Sigma$

$$\eta(A) = \eta(A \cap P) + \eta(A \cap N).$$

Since P is positive, we can define $\mu_1 \colon \Sigma \to [0, \infty]$ by

$$\mu_1(A) = \eta(A \cap P),$$

and μ_1 is a measure, by (2ab).

Let $N = P^c$. We will show that N is negative. Note that N cannot have any positive subsets with positive measure, since if $Q \subseteq N$ is positive and $\eta(Q) > 0$, then $Q \cup P$ is positive and $\eta(Q \cup P) > m$. It follows that if $\eta(A) > 0$ for some $A \subseteq N$ then, because A cannot be positive, $\eta(C) < 0$ for some $C \subset A$, and so if we set $B = A \setminus C$ then we have found a $B \subset A$ such that $\eta(B) > \eta(A)$.

Towards a contradiction, suppose that N is not negative, so that $\eta(A) > 0$ for some $A \subseteq N$. Let $A_1 = A$. Given A_n, let $\varepsilon_n = \sup_{B \subset A_n} \eta(B) - \eta(A_n) \in (0, \infty)$, and choose A_{n+1} to be some subset of A_n such that $\eta(A_{n+1}) > \eta(A_n) + \frac{1}{2}\varepsilon_n$. Let $A' = \cap_n A_n$. Then $\eta(A') \geq \frac{1}{2}\sum_n \varepsilon_n + \eta(A)$. Since $\eta(A') < \infty$, we have that $\lim_n \varepsilon_n = 0$. Find $B \subseteq A'$ and $\varepsilon > 0$ such that $\eta(B) = \eta(A') + \varepsilon$. Then $B \subseteq A_n$ for all n, and for some (indeed, all) n large enough $\varepsilon > \varepsilon_n$, in contradiction to the definition of ε_n. We have thus shown that N is a negative set.

Since P is positive and $P^c = N$ is negative we can define two measures

$$\mu_1(A) = \eta(A \cap P)$$
$$\mu_2(A) = -\eta(A \cap N).$$

These are mutually singular, and by the additivity of η we have that $\eta = \mu_1 - \mu_2$.

□

3.2 Lebesgue-Radon-Nikodym (by MohammedSaid Alhalimi & Eric Paul)

We prove the Lebesgue-Radon-Nikodym Theorem. Before stating and proving the theorem, we review the definitions from last lecture.

Two signed measures ν and μ on (Ω, Σ) are mutually singular if there exists $A \in \Sigma$ such that A is null for μ and A^c is null for ν. Furthermore, we know that we can write any signed measure ν on (Ω, Σ), as $\nu^+ - \nu^-$ where ν^+ and ν^- are unique positive measures. We then define the total variation of ν to be $\nu^+ + \nu^-$ and we denote this measure as $|\nu|$.

As was explained in the previous lecture, given a signed measure ν and positive measure μ on (Ω, Σ), we say that ν is absolutely continuous with respect to μ if for any $E \in \Sigma$, $\mu(E) = 0$ implies that $\nu(E) = 0$. We denote this as $\nu \ll \mu$.

Since this definition is crucial to the Lebesgue-Radon-Nikodym Theorem, we first motivate the name and then provide examples. The name "absoutely continuous" is actually a reasonable choice as seen by the following theorem.

Theorem 3.4. *Let ν be a finite signed measure and μ a positive measure on (Ω, Σ). Then $\nu \ll \mu$ if and only if for all $\epsilon > 0$ there exists $\delta > 0$ such that if $\mu(E) < \delta$, then $|\nu(E)| < \epsilon$.*

Proof. We begin by showing that it is sufficient to prove this for ν a positive measure. The overall claim is that we can just prove this for $|\nu|$ as $\nu \ll \mu$ if and only if $|\nu| \ll \mu$ and $|\nu(A)| \leq |\nu|(A)$. Prove this using the fact that $\nu = \nu^+ - \nu^-$ where $\nu^+ \perp \nu^-$.

So for the following we assume that $\nu = |\nu|$.

($\Leftarrow$) Let $A \in \Sigma$ such that $\mu(E) = 0$. Then for all $\epsilon > 0$, we have a $\delta > 0$ and since $\mu(A) < \delta$, $\nu(A) < \epsilon$. Thus, $\nu(A) = 0$.

($\Rightarrow$) We prove the contrapositive. So we assume that there exists $\epsilon > 0$ such that for all $\delta > 0$, there exists $A \in \Sigma$ such that $\mu(A) < \delta$ but $\nu(A) \geq \epsilon$. Thus, for all $n \in \mathbb{N}$, there exists $A_n \in \Sigma$ such that $\mu(A_n) < 2^{-n}$ and $\nu(A) \geq \epsilon$. Then we let $B_k = \bigcup_{i=k}^{\infty} A_i$ and $B = \bigcap_{k=1}^{\infty} B_k$. By continuity from above,

$$\mu(B) = \lim_{k\to\infty} \mu(B_k) < \lim_{k\to\infty} \sum_{n=k}^{\infty} 2^{-n} = \lim_{k\to\infty} 2^{1-k} = 0.$$

However, $\nu(B_k) \geq \epsilon$ for all k and ν is finite so $\nu(B) = \lim_{k\to\infty} \nu(B_k) \geq \epsilon$. So it is not the case that $\nu \ll \mu$. ☐

We now look at an example of absolute continuity. Like in the previous lecture, let μ be any measure on (Ω, Σ) and f be an extended μ-integrable function (measurable and $\int f^+ \mathrm{d}\mu$ or $\int f^- \mathrm{d}\mu$ is finite). Then we define a new measure ν as

$$\nu(A) = \int_A f \,\mathrm{d}\mu$$

Density example: Let l measure the length and m measure the mass. Then if f is the mass density, $m(A) = \int_A f \,\mathrm{d}l$. We often notate mass density as $\frac{\mathrm{d}m}{\mathrm{d}l}$. We can likewise define a derivative of a measure ν with respect to μ, $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$, as being equal to the function f such that $\nu(A) = \int_A f \,\mathrm{d}\mu$. This is called the Radon-Nikodym derivative of ν with respect to μ.

We see thus that defining $\nu(A) = \int_A f \,\mathrm{d}\mu$ means that $\nu \ll \mu$ and that we get to talk about a derivative. Now, one might wonder whether the other direction holds: if $\nu \ll \mu$, can we find an f such that $\nu(A) = \int_A f \, d\mu$? This is exactly what (half of) the Lebesgue-Radon-Nikodym Theorem answers! It says that if ν and μ are σ-finite, then such an f does exist.

The other half of the Lebesgue-Radon-Nikodym Theorem tells us that if ν is not absolutely continuous with respect to μ, then we can decompose the ν into two mutually singular measures such that one of then is absolutely continuous with respect to μ.

Before we introduce the main result of this section, we prove some technical lemmas. Assume (Ω, Σ) is a measurable space.

Lemma 3.5. *Assume $\mu \colon \Sigma \to [0, \infty]$ is a positive measure and $\nu \colon \Sigma \to [-\infty, \infty]$ is a signed measure. Suppose that $\nu \ll \mu$ and $\nu \perp \mu$. Then $\nu = 0$.*

Proof. Since $\nu \perp \mu$, there exists some A such that A is ν-null and A^c is μ-null. Fix any $Q \in \Sigma$. Since $Q \cap A \subseteq A$ so $\nu(Q \cap A) = 0$. Since $Q \cap A^c \subseteq A^c$ so $\mu(Q \cap A^c) = 0$ which implies (as $\nu \ll \mu$) that $\nu(Q \cap A^c) = 0$. Thus, we have $\nu(Q \cap A) + \nu(Q \cap A^c) = \nu(Q) = 0$, as claimed. ☐

Lemma 3.6. *Fix some measure $\mu \colon \Sigma \to [-\infty, \infty]$. Suppose that $\lambda_i \colon \Sigma \to [-\infty, \infty]$ is a sequence of measures such that $\lambda_i \perp \mu$. Then $\sum_i \lambda_i \perp \mu$. Furthermore, if $\lambda_i \ll \mu$ for all i then $\sum_i \lambda_i \ll \mu$.*

Proof. By assumption, we can furnish a sequence of sets $\{A_i\}_i$ such that A_i is λ_i-null and A_i^c is μ-null. Clearly, $\bigcap_i A_i$ is $\sum_i \lambda_i$-null and $(\bigcap_i A_i)^c = \bigcup_i A_i^c$ is μ-null. Thus, $\sum_i \lambda_i \perp \mu$. For the second part, if $A \in \Sigma$ then clearly $\lambda_i(A) = 0$ for all i then $\sum_i \lambda_i(A) = 0$, i.e. $\sum_i \lambda_i \ll \mu$. ☐

Lemma 3.7. *Fix two positive, finite measures $\mu, \nu \colon \Sigma \to [0, \infty]$. Then one of the following holds:*

1. *$\mu \perp \nu$*
2. *There exists $\epsilon > 0$ and $A \in \Sigma$ such that $\mu(A) > 0$ and $\nu - \epsilon\mu \geq 0$ on A.*

Proof. For each $n \in \mathbb{N}$, we can use Hahn decomposition to furnish a pair of sets (P_n, N_n) with $P_n \cup N_n = X$ and $P_n \cap N_n = \emptyset$ such that P_n and N_n are positive and negative (respectively) for $\nu - \frac{1}{n}\mu$. Put $P = \cup_n P_n$ and $N = \cap_n N_n$. We claim that $\nu(N) = 0$. This is easy to see as N is negative for $\nu - 1/n \cdot \mu$ so $0 \leq \nu(N) \leq 1/n \cdot \mu(N)$ for all n implies $\nu(N) = 0$. If $\mu(P) = 0$ then (1) holds. Otherwise, there exists n_0 such that $\mu(P_{n_0}) > 0$. But then $\nu - \frac{1}{n}\mu \geq 0$ on P_n by construction. Hence, (2) holds with $A = P_n$ and $\epsilon = 1/n$. ☐

Now, we have the tools to prove theorem of this section.

Theorem 3.8. *Suppose that $\mu \colon \Sigma \to [0, \infty]$ is positive, sigma-finite measure and $\nu \colon \Sigma \to [-\infty, \infty]$ is signed, sigma-finite measure. Then there exists unique signed measures $\rho, \lambda \colon \Sigma \to [-\infty, \infty]$ such that the following holds:*

1. *$\rho \ll \mu$*
2. *$\lambda \perp \mu$*
3. *$\nu = \lambda + \rho$*

4. *There exists a μ-integrable $f : X \to \mathbb{R}$ such that $\rho(A) = \int_A f \, d\mu$ for any $A \in \Sigma$.*

Proof. First, we assume that ν, μ are finite and positive. Let

$$\tilde{F} = \left\{ f : \Omega \to \mathbb{R} : \int_A f \, \mathrm{d}\mu < \nu(A) \text{ for all } A \in \Sigma \right\}.$$

Observe that $\tilde{F}$ is nonempty because it contains the zero function $0 \in \tilde{F}$. Further, $\tilde{F}$ is closed under max. To see why, fix any $f_1, f_2 \in \tilde{F}$ and set $A = \{\omega : f_1(\omega) \geq f_2(\omega)\}$. Then for any $B \in \Sigma$, we have

$$\int_B \max\{f_1, f_2\} \, \mathrm{d}\mu = \int_{B \cap A} f_1 \, \mathrm{d}\mu + \int_{B \cap A^c} f_2 \, \mathrm{d}\mu \leq \nu(B \cap A) + \nu(B \cap A^c) = \nu(B)$$

so $\max\{f_1, f_2\} \in \tilde{F}$.

Let $\alpha = \sup\{\int f \, d\mu : f \in \tilde{F}\}$. By assumption, $0 \leq \alpha < \infty$. Let $\{f_n\}_{n \geq 1}$ be a sequence of functions $f_n \in \tilde{F}$ such that $\int f_n d\mu \to \alpha$. Let $g_n = \max\{f_1, \ldots, f_n\}$. Then $g_n \in \tilde{F}$ and $\int g_n d\mu \to \alpha$. Let $f = \sup_n f_n$. Observe that $g_n \to f$ pointwise. Since $\int g_n d\mu < \infty$ and $g_{n+1} \geq g_n$, by the Montone Convergence Theorem, we get that

$$\alpha = \lim_n \int g_n \, \mathrm{d}\mu = \int \lim_n g_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu.$$

Furthermore, for any $A \in \Sigma$, we have $\int_A g_n \, \mathrm{d}\mu \leq \nu(A)$. The inequality also holds for the limit $\int_A f \, \mathrm{d}\mu \leq \nu(A)$ and hence $f \in \tilde{F}$.

We claim that f, $\rho = \int f \, \mathrm{d}\mu$ and $\lambda = \nu - \rho$ satisfy the theorem. It is clear that $\rho + \lambda = \nu$. If $A \in \Sigma$ then $\mu(A) = 0$ implies $\int_A f \, \mathrm{d}\mu = 0$ so $\rho \ll \mu$. To see why $\lambda \perp \mu$ holds, assume otherwise. By Lemma 3.7, there exists some $\epsilon > 0$ and $A \in \Sigma$ with $\mu(A) > 0$ such that $\lambda - \epsilon\mu \geq 0$ on A. Define $f' = f + \epsilon \mathbb{1}_A$. It follows that $f' \in \tilde{F}$ because for any $Q \in \Sigma$, we have

$$\begin{aligned}
\int_Q f' \, \mathrm{d}\mu &= \int_{Q \cap A} f' \, \mathrm{d}\mu + \int_{Q \cap E^c} f' \, \mathrm{d}\mu \\
&\leq \int_{Q \cap A} f \, \mathrm{d}\mu + \epsilon\mu(Q \cap A) + \nu(Q \cap A^c) \\
&\leq \int_{Q \cap A} f \, \mathrm{d}\mu + \left(\nu(A \cap Q) - \int_{A \cap Q} f \, \mathrm{d}\mu \right) + \nu(Q \cap A^c) \\
&= \nu(Q).
\end{aligned}$$

But clearly $\int f' \, \mathrm{d}\mu = \alpha + \epsilon\mu(A) > 0$ which contradicts the maximality of α. Hence, we conclude that $\lambda \perp \mu$. This shows existence. To show uniqueness, suppose that $\lambda' = \nu - \int f' \, d d\mu$. Then $\lambda - \lambda' = \int (f - f') \mathrm{d}\mu$. Since $\lambda \perp \mu$ and $\lambda' \perp \mu$ then $\lambda - \lambda' \perp \mu$ by Lemma 3.6. Similarly, $\int (f - f') \mathrm{d}\mu \ll \mu$. As $\lambda - \lambda' \perp \mu$ and $\lambda - \lambda' \ll \mu$ then $\lambda = \lambda'$ by Lemma 3.5 and $f = f'$ μ-almost everywhere.

Now, suppose ν is σ-finite. We can furnish a disjoint sequence $\{A_j\}$ in Σ such that $\mu(A_j) < \infty$ and $\nu(A_j) < \infty$ with $\Omega = \bigcup_1^\infty A_j$. Applying the previous case to $\mu_j(E) = \mu(E \cap A_j)$ and $\nu_j(E) = \nu(E \cap A_j)$, we find sequences $\{\lambda_j\}$ and $\{f_j\}$ satisfying the conditions. Letting $\lambda = \sum_j \lambda_j$ and $f = \sum_j f_j$ gives the result. Similarly, if ν is signed, we can split into ν^+ and ν^- and apply the previous cases to both measures to produce $\lambda_+ - \lambda_-$ and $f_+ - f_-$. This concludes the theorem. $\square$

3.3 Differentiation on $\mathbb{R}^d$

In this lecture we will consider $(\Omega, \Sigma) = (\mathbb{R}^d, \mathscr{B}_{\mathbb{R}^d})$ and λ the Lebesgue measure. We will prove a version of the Fundamental Theorem of Calculus, showing that Radon-Nikodym derivatives (of measures that are absolutely continuous with respect to λ) indeed correspond to the usual notion of differentiation. We will omit λ in integrals, so that $\int f \,\mathrm{d}\lambda(x)$ will be written as $\int f \,\mathrm{d}x$.

We have not proved this, but the measure of the open ball

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

is $\lambda(B_r(x)) = C_d r^d$ for some constant $C_d > 0$.

Let L^1_{loc} denote the **locally integrable** functions: measurable $f : \mathbb{R}^d \to \mathbb{R}$ such that

$$\int_K |f(x)| \,\mathrm{d}x < \infty$$

for all bounded sets K.

Theorem 3.9. *For any $f \in L^1_{\text{loc}}$ it holds for almost every $x \in \mathbb{R}^d$ that*

$$\lim_{r \to 0} \frac{1}{\lambda(B_r(x))} \int_{B_r(x)} f \,\mathrm{d}x = f(x).$$

A corollary of our main theorem will be the following. Given a function $F : \mathbb{R} \to \mathbb{R}$, say that F is differentiable at $x_0 \in \mathbb{R}$ if

$$\lim_{r \to 0} \frac{F(x+r) - F(x-r)}{2r}$$

exists, in which case we say that $F'(x_0)$ is equal to this limit.

Corollary 3.10. *For any $f \in L^1$ it holds for almost every $x_0 \in \mathbb{R}^d$ that $F(a) = \int_{(-\infty,a]} f(x) \,\mathrm{d}x$ is differentiable at x_0 with $F'(x_0) = f(x_0)$.*

Given $f \in L^1_{\text{loc}}$ and $r > 0$, denote by $A^f_r : \mathbb{R}^d \to \mathbb{R}$ the function

$$A^f_r(x_0) = \frac{1}{\lambda(B_r(x_0))} \int_{B_r(x)} f \,\mathrm{d}x = C_d r^{-d} \int_{B_r(x_0)} f \,\mathrm{d}x$$

Geometrically, $A^f_r(x)$ is the average of f on the ball of radius r around x. Note that $A^{f+g}_r = A^f_r + A^g_r$, by the additivity of the integral. Theorem 3.9 states that $\lim_{r \to 0} A^f_r(x) = f(x)$ a.e.

Claim 3.11. *For any $f \in L^1$ and $\varepsilon > 0$ there exists a continuous $g \in L^1$ such that $\int |f - g| \,\mathrm{d}x < \varepsilon$.*

The idea of the proof is to approximate f by a simple function (which we know we can do by the definition of integration). Then approximate this function by a simple function whose preimages are each a finite union of open balls, and then approximate this simple function by a continuous function.

Proof of Theorem 3.9. We first note that for continuous f, $\lim_{r\to 0} A_r^f(x) = f(x)$ for all $x \in \mathbb{R}^d$. This is because for every $x_0 \in \mathbb{R}^d$ and $\delta > 0$ it holds for all r small enough that $|f(x)-f(x_0)| < \delta$ for $x \in B_r(x_0)$, and so $|A_r^f(x) - f(x)| < \delta$.

We prove for $f \in L^1$; the extension to L^1_{loc} is simple. Fix $\varepsilon > 0$. By Claim 3.11 there is a continuous $g \in L^1$ such that $\int |f-g| \, \mathrm{d}x < \varepsilon$. Since g is continuous $\lim_{r\to 0} A_r^g(x) = g(x)$ for all $x \in \mathbb{R}^d$. Hence

$$\begin{aligned}
\limsup_{r\to 0} |A_r^f(x) - f(x)| &= \limsup_{r\to 0} |A_r^{f-g}(x) + A_r^g(x) - g(x) - (f-g)(x)| \\
&\le \limsup_{r\to 0} |A_r^{f-g}(x)| + |A_r^g(x) - g(x)| + |(f-g)(x)| \\
&= \limsup_{r\to 0} |A_r^h(x)| + |h(x)|
\end{aligned}$$

Where $h = f - g$. We would like to show that the set on which this takes large values has measure 0. For $a > 0$, let

$$A_a = \left\{ x : \limsup_{r\to 0} |A_r^f(x) - f(x)| > a \right\}.$$

To complete the proof, it suffices to show that $\lambda(A_a) = 0$, since $\lim_{r\to 0} A_r^f(x) = f(x)$ on the complement of $\cup_n A_{1/n}$.

By the above,

$$A_{2a} \subseteq B_a \cup C_a,$$

where

$$\begin{aligned}
C_a &= \left\{ x : \limsup_{r\to 0} |A_r^h(x)| > a \right\} \\
B_a &= \{ x : |h(x)| > a \}
\end{aligned}$$

The set B_a is small, because

$$\varepsilon > \int |h(x)| \, \mathrm{d}x \ge \int_{B_a} |h(x)| \, \mathrm{d}x > a\lambda(B_a),$$

so $\lambda(B_a) < \varepsilon/a$.

To control C_a, we note that

$$\limsup_{r\to 0} |A_r^h(x)| \le \sup_{r>0} A_r^{|h|}(x)$$

and accordingly define the **Hardy-Littlewood maximal function** H^h by

$$H^h(x) = \sup_{r>0} A_r^{|h|}(x).$$

Let

$$D_a = \{ x : H^h(x) > a \},$$

so that $C_a \subseteq D_a$. By Theorem 3.12 below, $\lambda(D_a) \le 3^d \varepsilon/a$, and hence $\lambda(A_a) \le (3^d + 1)\varepsilon/a$. Since this holds for all $\varepsilon > 0$ we are done. $\square$

3.4 The Maximal Theorem and a.e. differentiability of increasing functions

In the previous lecture we used the following theorem, called the **Maximal Theorem**:

Theorem 3.12. *Suppose that $f \in L^1$. For $a > 0$ let $D_a = \{x : H^f(x) > a\}$. Then*

$$\lambda(D_a) \leq \frac{3^d}{a} \int |f| \, \mathrm{d}x.$$

Note: It actually takes some work to show that D_a is measurable, but we will skip this.
To prove this we will need the following lemma.

Lemma 3.13. *Let R be a collection of open balls in $\mathbb{R}^d$, and let U be their union. Suppose that $\lambda(U) < \infty$. Then for every $\varepsilon > 0$ there exist disjoint $B_1, \ldots, B_k \in R$ such that $\sum_{n=1}^k \lambda(B_n) \geq (1-\varepsilon)3^{-d}\lambda(U)$.*

Proof. By Claim 1.12 (or, more precisely, by its analogue for $\mathbb{R}^d$) there is a compact $K \subseteq U$ such that $\lambda(K) > (1-\varepsilon)\lambda(U)$. Since K is compact and R is an open cover of K there is a finite $R' \subseteq R$ such that $K \subseteq V = \cup_{A \in R'} A \subset U$. In particular $\lambda(V) \geq (1-\varepsilon)\lambda(U)$.

Let B_1 be the largest ball in R'. Remove from R' all balls that intersect B_1, and let B_2 be the largest remaining ball. Continue until there are no balls left. Let $R'' = \{B_{r_1}(x_1), \ldots, B_{r_k}(x_k)\} \subseteq R'$.

Now, for each $A \in R'$ that is not in R'' there is some largest $B_r(x) \in R''$ such that $A \cap B_r(x) \neq \emptyset$, and $B_r(x)$ is larger than A. Hence A is contained in $B_{3r}(x)$. Hence A is contained in $W = \cup_n B_{3r}(x_n)$, and so V is contained in W. Thus

$$\sum_n \lambda(B_{r_n}(x_n)) = \sum_n 3^{-d}\lambda(B_{3r_n}(x_n)) \geq 3^{-d}\lambda(W) \geq 3^{-d}\lambda(V) > (1-\varepsilon)3^{-d}\lambda(U).$$

□

Proof of Theorem 3.12. Fix some $M > 0$ and let $D'_a = D_a \cap B_M(0)$. For each $x \in D'_a$ choose an r_x such that $A^{|f|}_{r_x}(x) > a$, and let $R = \{B_{r_x}(x) : x \in D'_a\}$. By Lemma 3.13 above for every $\varepsilon > 0$ there is a finite $E_a \subseteq D'_a$ such that $B_{r_x}(x)$ and $B_{r_y}(y)$ are disjoint for all $x \neq y \in E_a$ and such that $(1-\varepsilon)3^{-d}\lambda(D'_a) < \lambda(U)$, where $U = \cup_{x \in E_a} B_{r_x}(x)$. Note that by definition the average of f on U is more than a:

$$\frac{1}{\lambda(U)} \int_U |f| \, \mathrm{d}x > a.$$

Hence

$$\lambda(U) < \frac{1}{a} \int_U |f| \, \mathrm{d}x \leq \frac{1}{a} \int |f| \, \mathrm{d}x.$$

Since $\lambda(D'_a) \leq \frac{3^d}{1-\varepsilon}\lambda(U)$ we have that

$$\lambda(D'_a) \leq \frac{3^d}{(1-\varepsilon)a} \int |f| \, \mathrm{d}x.$$

Since this holds for every $\varepsilon > 0$, and since $D'_a = D_a \cap B_M(0)$ we have shown that

$$\lambda(D_a \cap B_M(0)) \leq \frac{3^d}{a} \int |f| \, \mathrm{d}x.$$

Since this holds for every M we are done. $\square$

Theorem 3.14. *Let $F : \mathbb{R} \to \mathbb{R}$ be increasing. Then F is continuous except on a countable set, and F is differentiable almost everywhere.*

Proof. First, we can assume without loss of generality that F is constant outside of some interval $[-n, n]$, since if we prove it for every

$$F_n(x) = \begin{cases} F(-n) & \text{for } x \leq -n \\ F(x) & \text{for } x \in [-n, n] \\ F(n) & \text{for } x \geq n \end{cases}$$

then the claim follows. We can also assume that $\lim_{x \to -\infty} F(x) = 0$, by adding a constant to F.

For the first part, note that since F is increasing, the intervals of the form

$$I_x = \left(\lim_{x \nearrow x_0} F(x), \lim_{x \searrow x_0} F(x) \right)$$

are disjoint, and so

$$\lambda(\cup_x I_x) = \sum_x \lambda(I_x) \leq F(n) - F(-n) < \infty.$$

Hence $\lambda(I_x)$ is positive for at most countably many $x \in (-n, n)$. Since $\lambda(I_x) > 0$ iff F is discontinuous at x, we have shown the first part.

Let $G(x_0) = \lim_{x \searrow x_0} F(x)$, so that G is increasing and right continuous. Hence we can, as in §1.5, define the measure μ_G by

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

By our assumption on F we have $\mu((-\infty, b]) = G(b)$.

Using Theorem 1.5, write $\mu_G = \mu_1 + \mu_2$, where $\mu_1 = \int f \, \mathrm{d}\lambda$ for some $f \in L^+$ and μ_2 is mutually singular with λ. Since μ_G is finite so are μ_1 and μ_2, and in particular $f \in L^1(\lambda)$.

Let $G_1(b) = \mu_1((-\infty, b])$ and $G_2(b) = \mu_2((-\infty, b])$, so that $G = G_1 + G_2$. By Theorem 3.9 we know that $G'_1(x) = f(x)$ for λ-a.e. x. To complete the proof, we show that $G'_2(x) = 0$ for a.e. x.

Since μ_1, μ_2 are mutually singular there is a set $B \in \mathscr{B}$ that is co-null for μ_1 and null for μ_2. Let

$$A_n = B \cap \left\{ x : \limsup_{r \to 0} \frac{\mu_2((x - r, x + r))}{2r} > \frac{1}{n} \right\}.$$

By Claim 1.12, for every $\varepsilon > 0$ find an open $U \supseteq A$ such that $\mu_2(U) < \varepsilon$. For every $x \in A_n$ there is some $r_x > 0$ such that $(x - r_x, x + r_x) \subseteq U$ and that $\mu_2((x - r_x, x + r_x)) > 2r_x/n$. Let

$$R = \{(x - r_x, x + r_x) : x \in A_n\}$$

be the collection of these open intervals, denote their union by V, and note that $A_n \subseteq V \subseteq U$. Hence, by Lemma 3.13, if $\lambda(A_n) > c$ then there exists a finite sub-collection $R' = \{(x_1 - r_1, x_1 + r_1), \ldots, (x_k - r_k, x_k + r_k)\} \subseteq R$ of disjoint intervals such that

$$c < 3 \sum_{m=1}^{k} 2r_m < 3n\lambda(V) \leq 3n\varepsilon.$$

Hence $\lambda(A_n) \leq \frac{1}{3n}\varepsilon$, and since this holds for every $\varepsilon > 0$ we are done. □

3.5 Bounded variation (by Tal Hershko and Elizabeth Xiao)

In previous lectures, we have seen the following correspondence between Borel measures on $\mathbb{R}$ and monotone functions $F : \mathbb{R} \to \mathbb{R}$.

Theorem 3.15. *Let $\mathscr{M}$ be the set of Borel measures $\mu : \mathscr{B}(\mathbb{R}) \to [0,\infty]$ which are finite on bounded sets. Let $\mathscr{F}$ be the set of (weakly) increasing functions $F : \mathbb{R} \to \mathbb{R}$ with $F(0) = 0$ which are right-continuous. Then there is a one-to-one correspondence between $\mathscr{M}$ and $\mathscr{F}$ defined as follows:*

1. *For every $\mu \in \mathscr{M}$, we define $F_\mu : \mathbb{R} \to \mathbb{R}$ as*

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

2. *For every $F \in \mathscr{F}$, we define μ_F as the unique Borel measure satisfying $\mu_F((a,b]) = F(b) - F(a)$ for every $a < b$.*

Example 3.16. *Let μ be the measure defined by $\mu(\{3\}) = 1$ and $\mu(\mathbb{R} \setminus \{3\}) = 0$. Then it corresponds to the function*

$$F(x) = \begin{cases} 0 & x < 3 \\ 1 & x \geq 3 \end{cases}.$$

Remark 3.17. *The requirement $F(0) = 0$ serves for "normalization" purposes. That is, it is simply a standard way to choose one function representing μ, among all other functions which differ by a constant.*

This correspondence naturally translates concepts from the world of measures to the world of functions, and vice versa. For example:

1. μ is finite $\iff$ F is bounded.

2. F is continuous $\iff$ μ is continuous.[1]

3. $F'(x) = \frac{d\mu}{d\lambda}(x)$ for almost every $x \in \mathbb{R}$, where λ is the Borel-Lebesgue measure and $\frac{d\mu}{d\lambda}$ is the Radon-Nikodym derivative.

We would like to extend this correspondence to signed measures; more specifically, let us focus on finite signed measures. Which functions correspond to them? This question leads us to the notion of *total variation* of a function.

Let $F : \mathbb{R} \to \mathbb{R}$ be any function. We define the *total variation* of F on an interval I as

$$T_F I = \sup \left(\sum_{i=1}^{n} |F(x_i) - F(x_{i-1})| \right)$$

where the supremum is taken over all $x_0 < x_1 < \cdots < x_n$ with $x_0, x_1, \ldots, x_n \in I$.

If $T_F I < \infty$, we say that F is of *bounded variation* on I and denote $F \in \mathrm{BV}(I)$. We also write $\mathrm{BV} = \mathrm{BV}(\mathbb{R})$ for short.

Example 3.18.

- *Every bounded monotone function is of bounded variation.*
- $\sin(x)$ *is of bounded variation on any finite interval, but not on* $\mathbb{R}$*.*
- *The function*

$$F(x) = \begin{cases} x \sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

  is not of bounded variation on $[0,1]$*, although it is continuous.*

Let $F : \mathbb{R} \to \mathbb{R}$ be any function. We define its total variation function $T_F : \mathbb{R} \to [0, \infty]$ as $T_F(x) = T_F(-\infty, x]$.

Lemma 3.19. *Let* $F : \mathbb{R} \to \mathbb{R}$*. Then the functions* $T_F + F$ *and* $T_F - F$ *are increasing.*

Proof. Take $x < y$. Then

$$T_F(x) + |F(y) - F(x)| \leq T_F(x) + T_F[x, y] = T_F(y).$$

Therefore

$$\begin{aligned} T_F(x) + F(y) - F(x) &\leq T_F(y), \\ T_F(x) - F(y) + F(x) &\leq T_F(y). \end{aligned}$$

Rearranging, we get

$$\begin{aligned} T_F(x) - F(x) &\leq T_F(y) - F(y), \\ T_F(x) + F(x) &\leq T_F(y) + F(y). \end{aligned}$$

□

[1] A Borel measure $\mu : \mathscr{B}(\mathbb{R}) \to [0, \infty]$ is said to be continuous if $\mu(\{x\}) = 0$ for all $x \in \mathbb{R}$.

Corollary 3.20. $F \in \mathrm{BV} \iff$ *it is the difference of two bounded increasing functions on* $\mathbb{R}$.

Proof. $\implies$: If $F \in \mathrm{BV}$ then F and T_F are both bounded, and then $F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F)$.

$\impliedby$: A bounded increasing function is in BV, and the difference of two functions in BV is in BV. ☐

Remark 3.21. *In particular, if* $F \in \mathrm{BV}$ *then the limit* $\lim_{x\to-\infty} F(x)$ *exists. We denote it* $F(-\infty)$ *for short.*

One more preparation before we present the generalized correspondence. Instead of using the condition $F(0) = 0$ for normalization, it will now be more convenient to use the condition $F(-\infty) = 0$. We also need to take right continuity into account. We therefore define

$$\mathrm{NBV} = \{F \in \mathrm{BV} : F(-\infty) = 0 \text{ and } F \text{ is right continuous}\}.$$

Corollary 3.20 can now be easily extended as follows.

Proposition 3.22. $F \in \mathrm{NBV} \iff$ *it is the difference of two increasing functions in* NBV.

Proof. This follows from the fact that if F is right continuous then T_F is right continuous. This is not hard to prove, but we skip it here. ☐

Theorem 3.23. *There is a one-to-one correspondence between the set* $\mathcal{N}$ *of finite signed Borel measures* $\mu : \mathscr{B}(\mathbb{R}) \to \mathbb{R}$ *and the set* NBV *of (normalized) functions of bounded variation, defined as follows:*

1. *For every* $\mu \in \mathcal{N}$, *we define* $F_\mu : \mathbb{R} \to \mathbb{R}$ *as* $F_\mu(x) = \mu((-\infty, x])$.
2. *For every* $F \in \mathrm{NBV}$, *we define* μ_F *as the unique signed Borel measure satisfying* $\mu_F((a,b]) = F(b) - F(a)$ *for every* $a < b$.

Proof. First let $\mu \in \mathcal{N}$. Use the Jordan decomposition to write $\mu = \mu_+ - \mu_-$ where μ_+, μ_- are two finite Borel measures. Let

$$F_\pm(x) = \mu_\pm((-\infty, x]).$$

These are the corresponding functions from Theorem 3.15 (except the normalization is different). Then $F_\pm \in \mathrm{NBV}$. Therefore $F_\mu = F_+ - F_- \in \mathrm{NBV}$.

Now let $F \in \mathrm{NBV}$. From Proposition 3.22, we can write $F = F_+ - F_-$ where $F_\pm$ increasing and in NBV. Let μ_+, μ_- be the corresponding functions from Theorem 3.15. Then they are finite Borel measures. Then $\mu_F = \mu_+ - \mu_- \in \mathcal{N}$. ☐

Remark 3.24. *This correspondence also naturally translates concepts from the world of measures to the world of functions. For example, if* μ *corresponds to* F, *then* $|\mu|$ *corresponds to* T_F.

Recall the notion of absolutely continuity of measures, which had two equivalent formulations. If η is a signed measure and μ a measure, then η is **absolutely continuous with respect to** μ, written $\eta \ll \mu$, if, for any A in the σ-algebra,

1. $\mu(A) = 0 \implies \eta(A) = 0$.

2. For every $\epsilon > 0$ there exists some $\delta > 0$, such that $\mu(A) < \delta \implies |\eta(A)| < \epsilon$.

The term absolute continuity originates from real analysis; it is a stronger form of continuity than uniform continuity, which itself is stronger than continuity.

The second formulation, which was proved in Theorem 3.4 to be equivalent to the first, justifies the use of the same term in reference to measures.

A function $F : \mathbb{R} \to \mathbb{R}$ is **absolutely continuous** if for every $\epsilon > 0$, there exists a $\delta > 0$ such that for any finite collection of disjoint intervals $(a_1, b_1), \ldots, (a_N, b_N)$,

$$\sum_{j=1}^{N} (b_j - a_j) < \delta \implies \sum_{j=1}^{N} |F(b_j) - F(a_j)| < \epsilon.$$

In particular, F is **absolutely continuous on** $[a, b]$ if this holds whenever all the intervals (a_j, b_j) are in $[a, b]$.

Remark 3.25. *The left-hand sum is the measure of the union of the intervals under the Lebesgue measure λ. Moreover, if $F(x) = \mu_F((-\infty, x])$ for the unique measure μ_F, then the right-hand sum is at most the total variation of μ_F.*

There is a correspondence between signed measures and normalized BV functions. It seems natural that there should be a relationship between $F \in \mathrm{NBV}$ being absolutely continuous as a function, and μ_F being absolutely continuous with respect to λ as a measure.

The relationship is not immediate because the definition for functions involves only finite disjoint intervals, whereas arbitrary Borel subsets could look very different. But it turns out that for NBV functions, having this condition for intervals is enough to show that it holds for any Borel subset.

Before making this connection explicit, let's state some facts about absolutely continuous functions.

- If F is absolutely continuous, then it is uniformly continuous: set N, the number of terms in the sum, to 1.

- If F is uniformly continuous, it is not necessarily absolutely continuous. An example is

$$F(x) = \begin{cases} x \sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

  which is uniformly continuous on $[0, 1]$ (in fact, on any bounded interval), but not absolutely continuous. Note that the same function is also not in $\mathrm{BV}([0, 1])$.

- If F is absolutely continuous on a bounded interval $[a, b]$, then it is in $\mathrm{BV}([a, b])$.

Theorem 3.26. *Suppose $F \in \mathrm{NBV}$. Then F is absolutely continuous $\iff \mu_F \ll \lambda$.*

Proof.
- $\Longleftarrow$: Suppose $\mu_F \ll \lambda$. For any $\epsilon > 0$ there is a $\delta > 0$ such that $\lambda(A) < \delta$ implies $|\mu_F|(A) < \epsilon$. In particular, if $A = \bigcup_{j=1}^N (a_j, b_j]$ is a union of disjoint intervals, then

$$\lambda(A) = \sum_{j=1}^N (b_j - a_j) < \delta \implies \epsilon > |\mu_F|(A) \geq \sum_{j=1}^N |F(b_j) - F(a_j)|$$

so F is absolutely continuous.

- $\Longrightarrow$: Conversely, suppose F is absolutely continuous. Let $\epsilon > 0$ and choose $\delta > 0$ to satisfy absolute continuity for F with respect to $\epsilon/2$, i.e.,

$$\sum_{i=1}^N (b_i - a_i) < \delta \implies \sum_{i=1}^N |F(b_i) - F(a_i)| < \frac{\varepsilon}{2}.$$

Let A be a Lebesgue-measurable set satisfying $\lambda(A) < \delta$. By Claim 1.12, A is contained in open Borel subsets of arbitrarily small measure. In particular, there is a decreasing sequence of open sets $U_1 \supset U_2 \supset \cdots \supset A$ such that $\lambda(U_1) < \delta$, hence $\lambda(U_k) < \delta$ for all k. Furthermore, $\lim_{k\to\infty} \mu_F(U_k) = \mu_F(A)$.

For now, fix an index k. Since U_k is open, it can be expressed as a countable disjoint union of intervals $(a_i^{(k)}, b_i^{(k)})$. For each N, we then have

$$\sum_{i=1}^N (b_i^{(k)} - a_i^{(k)}) = \lambda\Big(\bigcup_{i=1}^N (a_i^{(k)}, b_i^{(k)})\Big) \leq \lambda(U_k) < \delta,$$

so by absolute continuity of F,

$$\sum_{i=1}^N |F(b_i^{(k)}) - F(a_i^{(k)})| < \frac{\varepsilon}{2}.$$

Because this holds for every N,

$$\sum_{i=1}^\infty |F(b_i^{(k)}) - F(a_i^{(k)})| \leq \frac{\varepsilon}{2}.$$

But $F(b_i^{(k)}) - F(a_i^{(k)}) = \mu_F((a_i^{(k)}, b_i^{(k)})]$ by definition, so

$$\sum_{i=1}^\infty |F(b_i^{(k)}) - F(a_i^{(k)})| = \sum_{i=1}^\infty |\mu_F((a_i^{(k)}, b_i^{(k)}))| \geq \Big|\sum_{i=1}^\infty \mu_F((a_i^{(k)}, b_i^{(k)}])\Big| = |\mu_F(U_k)|$$

which follows from the triangle inequality and countable additivity of signed measures. Hence $|\mu_F(U_k)| \leq \varepsilon/2$ for every k. But $\lim_{k\to\infty} \mu_F(U_k) = \mu_F(A)$, so finally

$$|\mu_F(A)| \leq \frac{\varepsilon}{2} < \varepsilon.$$

We conclude that $\mu_F \ll \lambda$, by the second formulation of absolute continuity for measures.

$\square$

It is also natural to relate normalized BV functions with their derivative. This leads to a generalization of the Fundamental Theorem of Calculus.

Proposition 3.27. *If $F \in \mathrm{NBV}$, then $F' \in L^1$. Furthermore,*

- $\mu_F \perp m \iff F' = 0$ *a.e.*
- $\mu_F \ll m \iff F(x) = \int_{-\infty}^{x} F'(t) dt$.

Theorem 3.28 (The Fundamental Theorem of Calculus for Lebesgue Integrals)**.** *Given a bounded interval $[a,b]$ and a function $F : [a,b] \to \mathbb{R}$, the following are equivalent:*

1. *F is absolutely continuous on $[a,b]$.*
2. *$F(x) - F(a) = \int_a^x f(t)\,\mathrm{d}t$ for some $f \in L^1([a,b])$.*
3. *F is differentiable a.e. on $[a,b]$, $F' \in L^1([a,b])$ and $F(x) - F(a) = \int_x^a F'(t)\,\mathrm{d}t$.*

4 Topological vector spaces

4.1 Normed vector spaces

We say that the real numbers **act on a set** X if there is a map $\mathbb{R}\times X \to X$, denoted $(\lambda, x) \mapsto \lambda x$ such that $(\lambda_1 \cdot \lambda_2)x = \lambda_1(\lambda_2 x)$.

An **abelian group** is a set A endowed with an associative and commutative binary operation + having an element $0 \in A$ such that (i) $a + 0 = a$ for all $a \in A$, and (ii) each $a \in A$ has an inverse $-a$, i.e, $a + (-a) = 0$.

A **real vector space** V is an abelian group on which the reals act and such that $\lambda(v + w) = \lambda v + \lambda w$.

Examples:

- $\mathbb{R}^d$.
- If V, W are vector spaces then $V \times W$ is a vector space.
- $\mathbb{R}^S$ for some set S.
- Functions $f : \mathbb{N} \to \mathbb{R}$ with finite support.
- Measurable functions on (Ω, Σ).
- Bounded measurable functions on (Ω, Σ, μ).
- $L^1(\Omega, \Sigma, \mu)$.
- Finite signed measures on $(\mathbb{R}, \mathscr{B})$.
- NBV.
- $C([0,1])$, the continuous functions on $[0,1]$.

A **norm** $\|\|$ on a real vector space V is a function $V \to \mathbb{R}_{\geq 0}$, $v \mapsto \|x\|$ such that

1. **Homogeneity.** $\lambda\|v\| = |\lambda|\|v\|$.
2. **Triangle inequalty.** $\|u + w\| \leq \|u\| + \|w\|$.
3. **Positive definitiveness.** $\|v\| = 0$ only if $v = 0$.

Examples:

- $\mathbb{R}^d$:
  - $\|(x_1, \ldots, x_n)\|_1 = \sum_i |x_i|$.
  - $\|(x_1, \ldots, x_n)\|_2 = \sqrt{\sum_i x_i^2}$.
  - $\|(x_1, \ldots, x_n)\|_\infty = \max_i |x_i|$.

- V, W are normed vector spaces:
  - $\|(v,w)\| = \|v\| + \|w\|$.
  - $\|(v,w)\| = \sqrt{\|v\|^2 + \|w\|^2}$.
  - $\|(v,w)\| = \max\{\|v\|, \|w\|\}$.
- $\mathbb{R}^S$ for some set S: There is a norm for any S, but in general we would need the axiom of choice to construct it.
- Functions $f : \mathbb{N} \to \mathbb{R}$ with finite support: $\|f\| = \max_n |f(n)|$.
- Measurable functions on (Ω, Σ): Likewise no useful norm.
- Bounded measurable functions on (Ω, Σ, μ).
  - $\|f\| = \sup_{\omega \in \Omega} |f(\omega)| = \inf\{a \in \mathbb{R} : |f| \le a\}$.
  - $\|[f]\| = \inf\{a \in \mathbb{R} : |f| \le a \text{ a.e.}\}$.
- $L^1(\Omega, \Sigma, \mu)$: $\|f\| = \int |f| \, \mathrm{d}\mu$.
- Finite signed measures on $(\mathbb{R}, \mathscr{B})$: $\|\eta\| = |\eta|(\mathbb{R})$.
- NBV: $\|F\| = |\eta_F|(\mathbb{R})$.
- $C([0,1])$, the continuous functions on $[0,1]$:
  - $\|f\|_\infty = \max_x |f(x)|$.
  - $\|f\|_1 = \int |f| \, \mathrm{d}\lambda$.

  Note that $\|f\|_1 \le \|f\|_\infty$, but it is possible that $\lim_n \|f_n\|_1 = 0$ but $\|f_n\|_\infty = 1$.

Given a norm $\|\cdot\|$, the map $\rho : V \times V \to \mathbb{R}_{\ge 0}$ given by $\rho(v,w) = \|v - w\|$ is a metric on V. It is translation invariant:

$$\rho(u + v, w + v) = \rho(u, w).$$

The topology that this metric defines is the norm topology. It is generated by the open balls $B_r(u) = \{w \in V : \|w - v\| < r\}$. Equivalently, this is the topology in which $\lim_n v_n = v$ if $\lim_n \|v_n - v\| = 0$. Because ρ is translation invariant, this topology is translation invariant: if $U \subseteq V$ is open and $v \in V$ then $U + v$ is open.

Claim 4.1. *Let V be a normed vector space. Then the map $\mathbb{R} \times V \times V \to V$, $(\lambda, u, w) \mapsto \lambda u + w$ is continuous.*

Proof. Suppose $\lim_n(\lambda_n, u_n, w_n) = (\lambda, u, w)$, i.e.,

$$\lim_n |\lambda_n - \lambda = \lim_n \|u_n - u\| = \lim_n \|w_n - w\| = 0.$$

Then by the triangle inequality

$$\lim_n \|\lambda_n u_n + w_n - (\lambda u + w)\| \le \lim_n \|\lambda_n u_n - \lambda u\| + \lim_n \|w_n - w\|.$$

The last term vanishes. We write $\lambda_n u_n = \lambda u_n + (\lambda_n - \lambda)u_n$, and so by another application of the triangle inequality

$$\le \lim_n \|\lambda u_n - \lambda u\| + \|(\lambda_n - \lambda)u_n\| = \lim_n |\lambda_n - \lambda| \|u_n\| = 0.$$

□

A **Cauchy sequence** in a normed real vector space V is a sequence $v_1, v_2, \ldots \in V$ such that for every $r > 0$ there is an n such that the suffix $\{v_n, v_{n+1}, \ldots\}$ is contained in $B_r(v_n)$. If every Cauchy sequence in V converges (i.e., V is a complete normed space) then we say that V is a **Banach space**.

We say that a series $v_1, v_2, \ldots$ is **absolutely convergent** if $\sum_n \|x_n\| < \infty$.

Claim 4.2. *Let V be a normed real vector space. The following are equivalent:*

1. *If $u_1, u_2, \ldots$ is absolutely convergent then $w_n = \sum_{k=1}^{n} u_k$ converges.*
2. *V is a Banach space.*

A map $T\colon V \to W$ between vector spaces is **linear** if $T(\lambda v + u) = \lambda T v + T u$. If the spaces are normed it is called an **isometry** if $\|Tv\|_W = \|v\|_V$. We will usually be interested in maps that do not necessarily preserve the norm, but only respects the topologies they induce, i.e., is continuous.

Claim 4.3. *Let V, W be normed vector spaces, let $T\colon V \to W$ be linear, and fix $v \in V$. The following are equivalent:*

1. *T is continuous.*
2. *T is continuous at v.*

Proof. Clearly (1) implies (2). Assume (2), so that whenever $\lim_n v_n = v$ then $\lim_n T(v_n) = T(v)$. Let $\lim_n w_n = w$, i.e., $\lim_n \|w_n - w\| = 0$. Let $v_n = w_n - w + v$ Then

$$\lim_n \|v_n - v\| = \lim_n \|w_n - w\| = 0,$$

so that $\lim_n v_n = v$. It then follows from Claim 4.1 that

$$\lim_n T(w_n) = \lim_n T(v_n + w - v) = T(w) - T(v) + \lim_n T(v_n) = T(w).$$

□

We say that two norms on V are equivalent if there exists a constant $C > 0$ such that for all $v \in V$ it holds that $\|v\|_1 \le C\|v\|_2$ and likewise $\|v\|_2 \le C\|v\|_1$. It follows from Claim 4.3 that two norms are equivalent iff they induce the same topology. As an example of equivalent norms we can take all the norms on $\mathbb{R}^d$. As an example of unequivalent norms we can take the two norms on $C([0,1])$.

Linear maps $\mathbb{R}^n \to \mathbb{R}^m$ are always continuous. This is not true more generally. To see this, consider the normed vector space V of finitely supported functions $f: \mathbb{N} \to \mathbb{R}$, with the norm $\|f\| = \max_i |f(i)|$. Let $T: V \to V$ be the linear map given by $[Tf](i) = if(i)$. Consider the sequence $f_1, f_2, \dots$ given by

$$f_n(i) = \begin{cases} 1/n & i = n \\ 0 & i \neq n. \end{cases}$$

Then $\|f_n\| = 1/n$, and so $\lim_n f_n = 0$. But $\lim_n Tf_n \neq 0$ since $\|f_n\| = 1$. Hence T is not continuous. Note also that T is not **bounded**, in the following sense: for every $C > 0$ there exists $f \in V$ such that $\|Tf\| > C\|f\|$. That is, a linear map $T: V \to W$ between normed linear spaces is bounded if there exists a $C > 0$ such that $\|Tv\|_W \le C\|v\|_W$ for all $v \in V$. Another equivalent definition is the following. Let

$$\|T\| = \inf\{C : \|Tv\| \le C\|v\|\} = \sup\{\|Tv\| : \|v\| = 1\} = \sup\left\{\frac{\|Tv\|}{\|v\|} : v \neq 0\right\}$$

be the **operator norm** of T. Then T is bounded if it has finite norm.

Theorem 4.4. *Let V, W be normed vector spaces. Then a linear map $T: V \to W$ is continuous iff it is bounded.*

Proof. Suppose T is bounded. Let $\lim_n v_n = 0$, so that $\lim_n \|v_n\| = 0$. Then $\lim_n \|Tv_n\| \le \lim_n \|T\|\|v_n\| = 0$, and T is continuous at 0. It follows from Claim 4.3 that T is continuous.

Suppose T is not bounded. Let $v_1, v_2, \dots$ be unit vectors such that $\lim_n \|Tv_n\| = \infty$. Let $w_n = \frac{v_n}{\|Tv_n\|}$. Then $\|w_n\| = 1/\|Tv\|$, and so $\lim_n w_n = 0$. But $\|Tw_n\| = 1$, and so $\lim_n Tw_n \neq 0 = T(\lim_n w_n)$. ∎

Let V, W be normed vector spaces. We denote by $L(V, W)$ the set of bounded linear operators. We can equip it with a normed vector space structure using the obvious linear operators and the operator norm. It turns out that if W is a Banach space then so is $L(V, W)$.

4.2 Linear functionals

Let V be a real vector space. A **linear functional** is a linear map from V to $\mathbb{R}$. Note that by Claim 4.3, when V is normed, a linear functional $\varphi: V \to \mathbb{R}$ is continuous iff $\lim_n \varphi(v_n) = 0$ whenever $\lim_n v_n = 0$. By Theorem 4.4, it is continuous iff there exists some $C > 0$ such that $|\phi(v)| \le C$ whenever $\|v\| = 1$.

When V is normed we call $L(V, \mathbb{R})$ the **dual space** of V and denote it by V^*. Since $\mathbb{R}$ is a Banach space, so is V^*.

Every linear functional $F: \mathbb{R}^d \to \mathbb{R}$ is of the form $F(x_1,\ldots,x_d) = \sum_k \lambda_k x_k$. If V are the bounded signed measures on $(\mathbb{R},\mathscr{B})$ then, given a bounded measurable $f: \mathbb{R} \to \mathbb{R}$, the functional $F(\eta) = \int f\,\mathrm{d}\eta$ is a linear functional which is moreover bounded. Given a Borel probability measure μ on $[0,1]$, a linear functional on $C([0,1])$ is $\varphi(f) = \int f\,\mathrm{d}\mu$. This functional is bounded under the norm $\|f\|_\infty = \max_x |f(x)|$, since $\varphi(f) \le \|f\|_\infty$. However, under the norm $\|f\|_1 = \int f\,\mathrm{d}\lambda$ it is not necessarily bounded. For example, if $\mu = \delta_0$, then $\varphi(f) = |f(0)|$, and clearly we can find $f_n \in C([0,1])$ such that $\|f_n\|_1 \to 0$ but $f(0) = 1$.

In the remainder of this lecture we will prove the Hahn-Banach Separating Hyperplane Theorem. A **hyperplane** is a linear subspace $H \subseteq V$ such that $H \ne V$ and there is some $v \in V \setminus H$ such that $V = H + \mathbb{R}v$.

An equivalent definition is the following. Given a subspace $W \subseteq V$, define an equivalence relation on V by $u \sim v$ if $u - v \in W$. It is easy to check that V/W inherits a vector space structure (furthermore, if V is normed and W is closed, then $\|u + W\| = \inf_{w\in W} \|u + w\|$ is a norm on V/W). A Hyperplane is a subspace H such that V/H has dimension 1.

One can show that if φ is a non-zero linear functional then for every $x \in \mathbb{R}$ the set $\varphi^{-1}(x)$ is a hyperplane.

Recall that $C \subseteq V$ is convex if $u,w \in C$ implies that $\alpha u + (1-\alpha)w \in C$ for all $\alpha \in (0,1)$.

Theorem 4.5 (Separating Hyperplane Theorem)**.** *Let V be a normed vector space, let C,D be open, convex, disjoint subsets of V. Then there exists a linear functional φ and $\alpha \in \mathbb{R}$ such that $\varphi(u) < \alpha < \varphi(v)$ for all $u \in C$ and $v \in D$.*

To prove this we will prove the analytic Hahn-Banach Theorem. It allows us to construct linear functionals by extending functionals from subspaces to the entire space. Instead of considering normed spaces we will do this for a larger class.

Let V be a linear space. A **sublinear functional** is a map $p: V \to \mathbb{R}$ such that

1. **Subadditivity.** $p(u+w) \le p(u) + p(w)$.

2. **Positive homogeneity.** $p(\lambda v) = \lambda p(v)$ for all $\lambda \ge 0$.

Every norm is a sublinear functional, as is every linear functional. As an example of a sublinear functional that is neither, let V be the finitely supported functions $f: \mathbb{N} \to \mathbb{R}$, and let $p(f) = |f(1)|$. We say that a linear functional $\varphi: V \to \mathbb{R}$ is dominated by p if $\varphi(v) \le p(v)$ for all $v \in V$. Note that if V is normed then a linear functional φ is bounded iff it is dominated by some equivalent norm.

Theorem 4.6 (Hahn-Banach)**.** *Let V be a real vector space, and let p be a sublinear functional on V. Suppose that $W \subset V$ is a linear subspace and $\varphi: W \to \mathbb{R}$ is a linear functional that is dominated by p. Then φ extends to a linear functional $\psi: V \to \mathbb{R}$ that is dominated by p.*

Before proving this theorem we will use it to prove the separating hyperplane theorem. Given a convex, open $C \subset V$ containing 0, define its **gauge** $p: V \to \mathbb{R}$ by $p(v) = \inf\{a > 0 : v \in aC\}$. Positive homogeneity of p is immediate, and subadditivity follows from the convexity of C. Hence p is a sublinear functional. It can be checked that that $p(v) < 1$ if $v \in C$.

Proof of Theorem 4.5. We prove a simpler claim, by showing that if $0 \notin C$ then there is a linear functional φ such that $\varphi(v) < 0$ for all $v \in C$. The more general statement can be shown from this one by considering the convex set $C - D = \{u - w : u \in C, w \in D\}$, which does not contain 0.

Choose any $v_0 \in C$, and let $C_0 = C - v_0$. Let p be the gauge of C_0. Note that $-v_0 \notin C_0$, because $0 \notin C$. Hence $p(-v_0) \geq 1$. Define a linear functional on $\mathbb{R}v_0$ by $\varphi(\lambda v_0) = -\lambda$. Then φ is dominated by p.

Using Theorem 4.6, extend ϕ to a linear functional on V that is dominated by p. Then for every $v \in C$

$$\psi(v) = \psi(v - v_0) + \psi(v_0) \leq p(v - v_0) - 1 < 0,$$

because $v - v_0 \in C_0$, and so $p(v - v_0) < 1$.

□

Proof of Theorem 4.6. If $W = V$ then clearly we are done. Otherwise, choose $v \in W \setminus V$. We show that we can extend φ to a linear functional $\psi \colon W + \mathbb{R}v$ that is dominated by p. From there, the proof follows by a Zorn's Lemma argument.

Note that since $\varphi(v) \leq p(v)$, we have that for all $w_1, w_2 \in W$

$$\varphi(w_1) + \varphi(w_2) = \varphi(w_1 + w_2) \leq p(w_1 + w_2) \leq p(w_1 - v) + p(w_2 + v).$$

Rearranging we get

$$\varphi(w_1) - p(w_1 - v) \leq p(w_2 + v) - \varphi(w_2).$$

Since this holds for all w_1, w_2 there is some $\alpha \in \mathbb{R}$ such that

$$\varphi(w_1) - p(w_1 - v) \leq \alpha \leq p(w_2 + v) - \varphi(w_2)$$

for all $w_1, w_2 \in W$.

Define $\psi \colon W + \mathbb{R}v \to \mathbb{R}$ by $\psi(w + \lambda v) = \varphi(w) + \lambda\alpha$. This is well defined, because if $w_1 + \lambda_1 v = w_2 + \lambda_2 v$ then $(\lambda_1 - \lambda_2)v = w_1 - w_2$, which can only happen if $\lambda_1 = \lambda_2$. It follows that for all $\lambda > 0$

$$\varphi(w + \lambda v) = \lambda[\varphi(w/\lambda) + \alpha] \leq \lambda[\varphi(w/\lambda) + p(w/\lambda + v) - \varphi(w/\lambda)] = p(w + \lambda v).$$

A similar calculation shows that the same holds when $\lambda < 0$. □

4.3 Baire Category Theorem (by Holly Krynicki and Lara San Martin Suarez)

Let (Ω, Σ, μ) be a measure space. We think of co-null sets as being "almost everything". As such, they have the following useful property: if $A_1, A_2, \ldots$ are co-null then so is their intersection. This is useful for proving the existence of certain objects: if we can write a property A as the countable intersection of co-null properties $A_1, A_2, \ldots$, then we know that there exists an object with property A, even if we cannot construct it directly.

Let X be a topological space. Today we will introduce a purely topological notion that corresponds to being "almost everything". The idea is to start by thinking of dense open sets as being "almost everything", but this does not exactly work, since a countable intersection of open sets need not be open. However, if we think of countable intersections of dense open sets (i.e., dense G_δ sets) as "almost everything," then the Baire Category Theorem shows that when X is a complete metric space this works.

Definition 4.7. A space X is *separable* if there is a countable subset $A \subset X$ such that $\overline{A} = X$, or, equivalently, A is *dense* in X.

Definition 4.8. A set B is called a G_δ *set* if $B = \bigcap_{i=1}^{\infty} U_i$ where each U_i is open.

Example 4.9. *The set of rationals in* $[0,1]$ *is not a* G_δ *set.*

Definition 4.10. A subset A of a space X is *nowhere dense* if $\overline{A}$ contains no non-empty open set. Equivalently, A is nowhere dense if $X - \overline{A}$ is dense.

Definition 4.11. A set that is a countable union of nowhere dense sets is *meager* or a *set of the first category* (Cat I). A set that is not Cat I is a *set of the second category* (Cat II). A *residual* set is a Cat II set whose complement is Cat I.

Theorem 4.12 (Baire Category Theorem)**.** *Let X be a complete metric space. Then:*

1. *A countable intersection of dense G_δ sets is a dense G_δ set.*

2. *X is Cat II.*

Proof. If $\{A_m\}$ is a sequence of dense G_δ sets, then we know each A_m is a countable intersection of open dense sets. The intersection of all A_m will then still be a countable intersection of open dense sets, thus we need only consider the open dense sets that contain each A_m. Let $\{U_n\}_{n \in \mathbb{N}}$ be a sequence of open dense sets (G_δ sets) and let $V = \bigcap_{n \in \mathbb{N}} U_n$.

For (1), it suffices to show that if $W \subset X$ is open and nonempty, then W intersects V. If W did not intersect V, then V would be closed and since W is nonempty, so $\overline{V} = V \subsetneq X$.

We know $U_1 \cap W$ is open and nonempty since U_1 is an open dense set. Thus, we can choose $x_0 \in X$ and $r_0 \in \mathbb{R}^+$ such that $B(x_0, r_0) \subset U_1 \cap W$ and $0 < r_0 < 1$. Since each ball is open and nonempty, we can inductively choose $x_n \in X$ and $r_n \in \mathbb{R}^+$ for all $n > 0$ such that:

$$\overline{B(x_n, r_n)} \subset U_n \cap B(x_{n-1}, r_{n-1}), \text{ and}$$
$$0 < r_n < 2^{-n}$$

Since $r_n \to 0$ and for all $N \in \mathbb{N}$, $x_m \in B(x_N, r_N)$, $\{x_n\}$ is a Cauchy sequence.
X is complete, so $x = \lim x_n$ exists, and we have for all N:

$$x \in \overline{B(x_N, r_N)} \subset U_N \cap B(x_1, r_1) \subset U_N \cap W$$

That is, $W \cap V$ is open, nonempty, and since W was arbitrary, we can conclude that V is dense, and thus V is a G_δ set.

For (2), we need to show that X is not Cat I. Let $\{A_n\}$ be a sequence of nowhere dense sets. Then, $\{X - \overline{A}_n\}$ is a sequence of open dense sets. Then $\bigcap (X - \overline{A}_n) \neq \emptyset$, so

$$\bigcup A_n \subset \bigcup \overline{A}_n \neq X$$

□

Remark 4.13. *If X is a complete metric space and no point is isolated, then X contains no countable dense G_δ set. Note a point x of subset $A \subset X$ is an isolated point if there exists some open $U \subset X$ such that $U \cap A = \{x\}$.*

Example 4.14. *Consider $\mathbb{R}$ union sequence $\{r_i\}$ for each $r \in \mathbb{Q}$. Define $d(r_i, r) = \frac{1}{i}$ and for $x \in \mathbb{R}$,*

$$d(r_i, x) = \frac{1}{i} + d(r, x)$$
$$d(r_i, s_j) = \frac{1}{i} + \frac{1}{j} + d(r, s)$$

This is a complete metric such that each r_i is isolated. The set $\bigcup_r (\bigcup_i r_i)$ is open, countable, and dense.

Let us see how we can apply Baire's Category Theorem in the theory of linear maps.

Theorem 4.15 (Open Mapping Theorem)**.** *Let $T: V \to W$ be a bounded linear operator between Banach spaces. If T is surjective then T is open.*

Proof. Denote by $B_r := B_r(0)$ the ball of radius $r > 0$ centered at 0. It will be clear from context whether this ball is taken to be in V or W. Note that it suffices to show that $\exists r > 0$ such that $B_r \subset T(B_1)$ in W by the linearity of T.

Since $V = \bigcup_{i=1}^{\infty} B_i$, we can express $W = \bigcup_{i=1}^{\infty} T(B_i)$ by the surjectivity of T. Here is where we apply Baire's Category Theorem: W is complete and so $T(B_n)$ cannot be nowhere dense $\forall n \geq 1$. Taking the homeomorphism $w \mapsto nw$ mapping $T(B_1)$ to $T(B_n)$, this is equivalent to saying $T(B_1)$ cannot be nowhere dense. That is, $\exists w_0 \in W, r > 0$ such that $B_{4r}(w_0) \subset \overline{T(B_1)}$.

Let us pick $v_1 \in V$ (or equivalently, $w_1 := Tv_1 \in W$) such that $||w_1 - w_0|| < 2r$. Then, we have that $B_{2r}(w_1) \subsetneq B_{4r}(w_0) \subseteq T(B_1)$ since $B_{4r}(w_0) \subseteq \overline{T(B_1)}$. So choosing a w with $||w|| < 2r$, we can write $w = Tv_1 + (w - w_1)$. Since $w - w_1 \in \overline{T(B_1)}$ and T is linear, we deduce $w \in \overline{T(v_1 + B_1)} \subsetneq \overline{T(B_2)}$ since $v_1 \in B_1$.

Overall, dividing the above expression by 2, we have found that there exists $r > 0$ such that, for $w \in W$,

$$||w|| < r \Rightarrow w \in \overline{T(B_1)}.$$

To show that T is open, we just have left to check that, by shrinking our radius, we can replace $\overline{T(B_1)}$ by $T(B_1)$. It is in this part of the proof where we will use that V is also a Banach space.

Suppose $||w|| < r/2$. Then, there is a $v_1 \in B_{1/2}$ with $||w - Tv_1|| < r/4$. Inductively, there is $v_n \in B_{2^{-n}}$ with $||w - \sum_{j=1}^{n} Tv_j|| < r2^{-n-1}$. Notice that we are using that T is linear and thus $||w|| < r2^{-n} \Rightarrow w \in \overline{T(B_{2^{-n}})}$.

Since V is complete, we have that $\sum_{n=1}^{\infty} v_n \to v \in V$ such that $||v|| \leq \sum_{n=1}^{\infty} ||v_n|| = \sum_{n=1}^{\infty} 2^{-n} = 1$ and $w = Tv$. In order words, $T(B_1)$ contains all w with $||w|| < r/2$ and we are done. $\square$

Corollary 4.16. *If V and W are Banach spaces and $T\colon V \to W$ is a bijective linear map, then T^{-1} is also a bounded linear operator.*

Definition 4.17. We say a linear map $T\colon V \to W$ is *closed* if its graph $\Gamma(T) := \{(v, w) \in V \times W : w = Tv\}$ is closed as a subspace of $V \times W$.

Note that this definition is equivalent to saying that for any $\{x_n\}_n$ sequence of elements in V such that $x_n \to x$, if $Tx_n \to y$ then $y = Tx$.

Remark 4.18. *If $T\colon V \to W$ is continuous, then it is closed.*

Note, however, that the converse is not true. Take for example the linear map

$$\frac{d}{dx}\colon \left(\mathscr{C}^1([0,1]), ||\cdot||_\infty\right) \to (\mathscr{C}([0,1]), ||\cdot||_\infty).$$

This map is closed, but not bounded.

To show that indeed it is not bounded, consider the sequence of functions $\{f_n\}_n$ in $\left(\mathscr{C}^1([0,1]), ||\cdot||_\infty\right)$ defined as

$$f_n(x) = e^{-n^4x^2}$$

Then, $\frac{d}{dx} f_n = -2n^4 x e^{-n^4x^2}$ and we have that $\left|\left|\frac{d}{dx} f_n\right|\right|_\infty \geq 2n^2/e$, which means

$$\left|\left|\frac{d}{dx}\right|\right| = \sup \frac{\left|\left|\frac{d}{dx} f\right|\right|_\infty}{||f||_\infty} : f \in \mathscr{C}^1([0,1]) \geq \frac{\left|\left|\frac{d}{dx} f_n(x)\right|\right|_\infty}{||f_n||_\infty} \geq \frac{2n^2}{e}$$

The next application of Baire's Category Theorem tells us that, if we consider an extra assumption in our spaces V and W are complete, the converse will also hold.

Theorem 4.19 (Closed Graph Theorem)**.** *Every closed linear map $T\colon V \to W$ between Banach spaces is bounded.*

In the example above, what fails is that $\left(\mathscr{C}^1([0,1]), ||\cdot||_\infty\right)$ is not complete. Indeed, take $\{f_n\}_n$ a sequence of functions in $\mathscr{C}^1([0,1])$ defined as $f_n(t) = \frac{1}{n}\sin nt$. We have that $||fn||_\infty \leq \frac{1}{n}$, meaning that $\lim_{n\to\infty} ||f_n - 0||_\infty = 0$. However, $\frac{df_n}{dt} = \cos nt$ and $\lim_{n\to\infty} ||\cos nt||_\infty \neq \frac{d0}{dt} = 0$.

Proof. Consider π_V and π_W the projections of $\Gamma(T)$ onto V and W respectively. Notice that V and W being complete implies that $V \times W$ is complete, and thus $\Gamma(T)$ is too since T is closed. π_V is a bijection from $\Gamma(T)$ to V and by Corollary 4.16, π_V^{-1} is bounded. Then, $T = \pi_W \circ \pi_V^{-1}$ is a composition of bounded operators and we are done. $\square$

As a last application, the next theorem provides us with uniform estimates from pointwise estimates in certain situations.

Theorem 4.20 (Uniform Boundedness Principle)**.** *Suppose that V and W are normed vector spaces and $\mathscr{A}$ a subset of $\mathscr{L}(V,W)$. Then,*

1. *If* $\sup_{T\in\mathscr{A}} ||Tv|| < \infty$ *for all v in some non-meager subset of V, then* $\sup_{T\in\mathscr{A}} ||T|| < \infty$.
2. *If V is a Banach space and* $\sup_{T\in\mathscr{A}} ||Tv|| < \infty$ *for all $v \in V$, then* $\sup_{T\in\mathscr{A}} ||T|| < \infty$.

4.4 Topological vector spaces (by Minakshi Ashok and Luis Soldevilla Estrada)

In this section, we generalize some results of normed topological vector spaces to *semi-normed* topological vector spaces.

Considered a semi-norm $p : X \to \mathbb{R}$. Seminorms have the following properties:

1. **Triangle Inequality**: $p(\boldsymbol{x}+\boldsymbol{y}) \leq p(\boldsymbol{x}) + p(\boldsymbol{y})$
2. **Absolute Homogeneity**: $p(\lambda\boldsymbol{x}) = |\lambda| p(\boldsymbol{x})$

Notice that semi-norms have all the properties of norms *except* that a non-zero vector $\boldsymbol{x}$ can have a zero seminorm. Since norms are also seminorms, the results that follow always apply to normed vector spaces.

Why should we study seminormed spaces in the first place? It turns out that in some cases, the topological space arising from seminorms cannot be recreated using norms.

Let's consider a concrete example: take the space of all real-valued sequences $\mathbb{R}^{\mathbb{N}}$. Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{\mathbb{N}}$. Define a family semi-norms $\{p_\alpha\}_{\alpha\in\mathbb{N}}$ as the absolute difference between the α^{th} elements of $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$p_\alpha = |\boldsymbol{x}_\alpha - \boldsymbol{y}_\alpha|. \tag{4.1}$$

One can check that p_α satisfies the conditions for a seminorm. Additionally, the topology generated by this seminorm is that of pointwise convergence. It is not possible to obtain the same topology using seminorms.

To see this, consider the sequence $f_n(i) = 1$ if $i = n$ and zero otherwise. The family of seminorms induces a topology of pointwise-convergence, and the sequence converges to $f = (0,0,0,0......)$ in the point-wise sense.

Now, if there were some norm $\|\cdot\|$ such that $f_n \to f$, then using the definition of convergence of sequences in normed spaces, $\|f_n - f\| \to 0$. However $\|f_n\| = 1 \forall n$, and $\|f\| = 0$, which implies that $f_n \to f$ cannot be satisfied. This is a consequence of the fact that there is no norm inducing the topology of pointwise convergence for countably infinite spaces.

The above example demonstrates why topologies induced by families of seminorms are interesting in their own right.

Recall that a norm naturally induces a notion of a metric, which in turn induces a topology. Metric topologies are T4, and any two points in the topology can be separated by open sets (Hausdorff). This is often a nice property to have.[2] When can we construct Hausdorff spaces using seminorms?

Proposition 4.21 (Hausdorff Seminormed TVS). *Let $\{p_\alpha\}_{\alpha\in A}$ be a family of seminorms defined on a vector space X. Consider the topological vector space (X,τ) generated by open balls $U^\alpha_\epsilon(\boldsymbol{x}) = \{\boldsymbol{y} \in X | p_\alpha(y-x) < \epsilon\}$*[3].

X is Hausdorff iff for each $\boldsymbol{x} \neq 0, \exists \alpha \in A$ such that $p_\alpha(\boldsymbol{x}) \neq 0$.

Proof.

($\implies$) Given X is Hausdorff. Then, for any distinct points $\boldsymbol{x},\boldsymbol{y} \in X$, $\exists$ open sets $U,V \in \tau$ such that $\boldsymbol{x} \in U, \boldsymbol{y} \in V$ and $U \cap V = \phi$. U,V can be written as a union of open balls as $U = \cup_i U^{\alpha_i}_{\epsilon_i}(\boldsymbol{x}_i), V = \cup_j V^{\alpha_j}_{\epsilon_j}(\boldsymbol{x}_j)$, $U^{\alpha_i}_{\epsilon_i}(\boldsymbol{x}_i), V^{\alpha_j}_{\epsilon_j}(\boldsymbol{x}_j) \in \tau$, since τ is generated from the set of open balls. This implies that $\exists \alpha \in A, \epsilon > 0$ such that $p_\alpha(\boldsymbol{x}-\boldsymbol{y}) > \epsilon$. Since $\boldsymbol{x},\boldsymbol{y}$ are distinct points, $\boldsymbol{z} = \boldsymbol{x}-\boldsymbol{y}$ is non-zero. Therefore, $\forall$ non-zero vectors $\boldsymbol{z}$, $\exists \alpha$ such that $p_\alpha(\boldsymbol{z}) \neq 0$.

($\impliedby$) Given for each $\boldsymbol{x} \neq 0, \exists \alpha \in A$ such that $p_\alpha(\boldsymbol{x}) \neq 0$. Pick two distinct $\boldsymbol{x},\boldsymbol{y} \in X$. Then, $\boldsymbol{x}-\boldsymbol{y}$ is non-zero and $\exists \alpha \in A$ such that $p_\alpha(\boldsymbol{x}-\boldsymbol{y}) \neq 0$. Define $\delta = \frac{p_\alpha(\boldsymbol{x}-\boldsymbol{y})}{2}$. Then, $U^\alpha_\delta(\boldsymbol{x})$ and $U^\alpha_\delta(\boldsymbol{y})$ are two disjoint open balls around $\boldsymbol{x}$ and $\boldsymbol{y}$. Such a construction can be made for any choice of distinct $\boldsymbol{x},\boldsymbol{y} \in X \implies (X,\tau)$ is Hausdorff.

∎

Proposition 4.22 (Metrizability of Hausdorff Seminormed TVS). *Let $\{p_\alpha\}_{\alpha\in A}$ be a family of seminorms defined on a vector space X. Consider the topological vector space (X,τ) generated by open balls $U^\alpha_\epsilon(\boldsymbol{x}) = \{\boldsymbol{y} \in X | p_\alpha(y-x) < \epsilon\}$.*

If X is Hausdorff and A is countable, then X is metrizable with a translation invariant metric (i.e., $\rho(\boldsymbol{x},\boldsymbol{y}) = \rho(\boldsymbol{x}-\boldsymbol{z},\boldsymbol{y}-\boldsymbol{z}), \forall \boldsymbol{x},\boldsymbol{y},\boldsymbol{z} \in X$).

Proof. Construct $\rho(\boldsymbol{x},\boldsymbol{y}) = \Sigma_\alpha \frac{1}{2^\alpha} min\{1, p_\alpha(\boldsymbol{x}-\boldsymbol{y})\}$. We can verify that ρ satisfies the conditions for a metric:

(1) **Distance from a point to itself is zero.** $\rho(\boldsymbol{x},\boldsymbol{x}) = \Sigma_\alpha \frac{1}{2^\alpha} min\{1, p_\alpha(\boldsymbol{x}-\boldsymbol{x})\} = \Sigma_\alpha \frac{1}{2^\alpha} min\{1, p_\alpha(0)\} = 0$.

(2) **Symmetric.** Since seminorms are symmetric (why?), $\rho(\boldsymbol{x},\boldsymbol{y}) = \Sigma_\alpha \frac{1}{2^\alpha} min\{1, p_\alpha(\boldsymbol{x}-\boldsymbol{y})\} = \Sigma_\alpha \frac{1}{2^\alpha} min\{1, p_\alpha(\boldsymbol{y}-\boldsymbol{x})\} = \rho(\boldsymbol{y},\boldsymbol{x})$.

(3) **Triangle Inequality.**

Consider $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z} \in X$. Then,

$\rho(\boldsymbol{x},\boldsymbol{y}) = \Sigma_\alpha \frac{1}{2^\alpha} min\{1, p_\alpha(\boldsymbol{x}-\boldsymbol{y})\} \leq \Sigma_\alpha \frac{1}{2^\alpha} min\{1, p_\alpha(\boldsymbol{x}-\boldsymbol{z}) + p_\alpha(\boldsymbol{z}-\boldsymbol{y})\} \leq \Sigma_\alpha \frac{1}{2^\alpha} min\{1, p_\alpha(\boldsymbol{x}-\boldsymbol{z})\} + \Sigma_\alpha \frac{1}{2^\alpha} min\{1, p_\alpha(\boldsymbol{z}-\boldsymbol{y})\} \implies \rho(\boldsymbol{x},\boldsymbol{y}) \leq \rho(\boldsymbol{x},\boldsymbol{z}) + \rho(\boldsymbol{z},\boldsymbol{y})$.

[2] Most spaces we usually encounter are Hausdorff. Without it, we begin to encounter some weird properties.
[3] Note that the textbook uses the notation $U_{x\alpha\epsilon}$ instead.

Additionally, we can verify ρ is translational invariant. Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in X$. Then: $\rho(\boldsymbol{x}-\boldsymbol{z}, \boldsymbol{y}-\boldsymbol{z}) = \Sigma_\alpha \frac{1}{2^\alpha} min\{1, p_\alpha(\boldsymbol{x}-\boldsymbol{z}+\boldsymbol{z}-\boldsymbol{y})\} = \Sigma_\alpha \frac{1}{2^\alpha} min\{1, p_\alpha(\boldsymbol{x}-\boldsymbol{y})\} = \rho(\boldsymbol{x}, \boldsymbol{y})$ ☐

In general, a topological space (X, τ) is metrizable if it is first-countable and Hausdorff (Birkhoff-Kakutani Theorem).

Theorem 4.23 (Convergence of Nets)**.** *If $< x_i >_{i \in I}$ is a net in X, then $x_i \to x$ iff $p_\alpha(x_i - x) \to 0\, \forall \alpha \in A$.*

Proof.

($\implies$) Given $x_i \to x$. Recall that $x_i \to x \implies \forall$ open neighbourhoods $U of x$, the net x_i is eventually in U.

Therefore, $\forall \alpha \in A, \forall \delta > 0, \exists N(\delta, \alpha) \in \mathbb{N}$ such that $p_\alpha(x - x_i) \le \delta$, where $i > N(\delta, \alpha) \implies p_\alpha(x - x_i) \to 0 \forall \alpha \in A$.

($\impliedby$) Given $p_\alpha(x_i - x) \to 0\, \forall \alpha \in A \implies \forall$ open neighbourhoods $U of x$, the net x_i is eventually in U. Therefore, $x_i \to x$.

☐

It's worth mentioning that the notion of nets can be used to define continuity of a function. Given spaces X, Y, $f : X \to Y$ is continuous at $x \in X$ iff $\forall$ net $< x_\alpha > \to x$, $< f(x_\alpha) > \to f(x)$. See Proposition 4.19 in Folland for more details.

(if time):

Definition 4.24. A set A is a **convex set** if $\forall x, y \in A$, $tx + (1-t)y \in A$, where $t \in (0,1)$.

Definition 4.25. A topological vector space is **locally convex** is there is a base for the topology consisting of convex sets.

Theorem 4.26 (Locally Convexity)**.** *Let $\{p_\alpha\}_{\alpha \in A}$ be a family of seminorms defined on a vector space X. Consider the topological vector space (X, τ) generated by the basis $\mathbb{B} = \{U^\alpha_\epsilon(\boldsymbol{x}) = \{\boldsymbol{y} \in X | p_\alpha(y - x) < \epsilon\}\}$ (that is, set of open balls defined by the seminorms).*

Then, (X, τ) is a locally convex topological vector space. .

Proof. Pick any $U^\alpha_\epsilon(\boldsymbol{x}) \in \mathbb{B}$. Then, $\forall y, z \in U$, $p_\alpha(x - ty - (1-t)z) = p_\alpha(tx + (1-t)x - ty - (1-t)z) = p_\alpha(t(x-y) + (1-t)(x-z)) \le p_\alpha(t(x-y)) + p_\alpha((1-t)(x-z)) = tp_\alpha(x-y) + (1-t)p_\alpha(x-z) \le t\epsilon + (1-t)\epsilon = \epsilon$

$\implies p_\alpha(x - ty - (1-t)z) \le \epsilon$

$\implies ty + (1-t)z \in U^\alpha_\epsilon(\boldsymbol{x})$

$\implies \mathbb{B}$ is a set of convex sets. Therefore, X, τ is locally convex ☐

A few lectures ago, we showed that continuous linear maps between normed vector spaces are bounded. This is useful when studying bounded linear operators. We can prove a similar result for seminormed vector spaces:

Proposition 4.27 (Continuity of Linear Maps)**.** *Suppose X, Y are vector spaces with topologies defined, respectively, by the family of seminorms $\{p_\alpha\}_{\alpha\in A}$ and $\{p_\beta\}_{\beta\in B}$. Let $T : X \to Y$ be a linear map.*

T is continuous iff $\forall \beta, \exists \alpha_1, \dots\dots, \alpha_k \in A$ such that $q_\beta(Tx) \le C\sum_1^k p_{\alpha_k}(x)$.

Proof. Let $(X, \tau_X), (Y, \tau_Y)$ be the topologies generated from the families of seminorms $\{p_\alpha\}_{\alpha\in A}$, $\{p_\beta\}_{\beta\in B}$ respectively.

$\implies$ Assume that T is continuous, i.e. $x_i \to x$ implies that $Tx_i \to Tx$. More specifcally, for any $\beta \in B$ there exists a neighborhood U of $0 \in X$ such that $q_\beta(Tx) < 1$ for all $x \in U$. By construction of U_ϵ^α we know that finite intersection of these form a basis neighborhood around x. Therefore, let $U = \bigcap_{j=1}^k U(x)_{\alpha_j}^{\epsilon_j}$ and take $\epsilon = min(\epsilon_1, \dots, \epsilon_k)$. This gives that $q_\beta(Tx) < 1$ whenever $p_{\alpha_j}(x) < \epsilon$ for all j. Consider two cases

- If $p_{\alpha_j}(x) > 0$ for some j, then we define

$$y = \frac{\epsilon x}{\sum_{j=1}^k p_{\alpha_j}(x)}$$

Therefore, we have $p_{\alpha_j}(y) < \epsilon$ and

$$q_\beta(Tx) = \sum_{j=1}^k \epsilon^{-1} p_{\alpha_j}(x) q_\beta(Ty) \le \epsilon^{-1} \sum_{j=1}^k p_{\alpha_j}(x)$$

- Otherwise, we must have that $p_{\alpha_j}(x) = 0$ and $p_{\alpha_j}(rx) = 0$ for all $r > 0$ and all j. Thus, $rq_\beta(Tx) = q_\beta(T(rx)) < 1$ for all r. This immediately implies that $q_\beta(Tx) \le \epsilon^{-1}\sum_{j=1}^k p_{\alpha_j}(x)$, as desired.

$\impliedby$ Given $q_\beta(Tx) \le C\sum_1^k p_{\alpha_k}(x)$. Therefore, for every converging net $x_\alpha \to x$, $p_\alpha(x - x_i) \to 0\ \forall \alpha \in A$ (from Theorem 4.23). Therefore, $q_\beta(Tx - Tx_i) \to 0\ \forall \beta \in B \implies Tx_i \to Tx$. If $Tx_i \to Tx$ for every $x - x_i \to 0$, then T is continuous[4].

□

We wrap up this first half of the lecture by giving another example showing how seminorms give more structure to certain spaces and operators. For this, consider the space of infinitely differentiable functions $C^\infty([0,1])$ and the linear map $d/dx \colon C^\infty([0,1]) \to C^\infty([0,1])$ given by differentiation. For $f_\lambda = e^{\lambda x}$, we have that

$$\frac{d}{dx}(f_\lambda) = \lambda e^{\lambda x} = \lambda f_\lambda$$

which implies that $\frac{||(d/dx)(f_\lambda)||}{||f_\lambda||} = |\lambda|$ and, thus, in the operator norm $||d/dx||$ is $\ge \lambda$ for an arbitrary λ. Therefore, d/dx is an unbounded linear map. We have proved that there is no

[4]See Proposition 4.19 in Folland.

possible norm in $C^\infty([0,1])$ so that d/dx is bounded (more specifically, continuous). To fix this issue we use seminorms and the topology generated by these. Consider the seminorms

$$p_k(f) = \sup_{0 \le x \le 1} |f^{(k)}(x)|$$

for each $k \in \mathbb{N}$.

This makes $C^\infty([0,1])$ into a Frechet space (a complete topological vector space with countable seminorms) and makes $\frac{d}{dx}$ continuous, as desired. Indeed, for any $k \in \mathbb{N}$

$$\left|\left|p_k(\frac{d}{dx}f)\right|\right| = \left|\left|p_{k+1}(f)\right|\right|$$

By Proposition 4.27, $\frac{d}{dx}$ is a continuous linear map.

Now, we explore other useful ways to topologize a vector space X. A natural way to do this is to consider a vector space X, a normed space Y and a family of linear maps $\{T_\alpha : X \to Y_\alpha\}_{\alpha \in A}$. We set $\mathscr{T}$ to be the weakest topology in X that makes each T_α continuous in the sense of Proposition 4.27. More specifically, this topology is generated by the seminorm

$$p_\alpha(x) = ||T_\alpha(x)||$$

By proposition 4.21, this topology is generated by

$$U_{x\alpha\epsilon} = \{x_0 \in X : ||T_\alpha(x) - T_\alpha(x_0)|| < \epsilon\}$$

Following the example of $C^\infty([0,1])$, such norm is generated by $T_k(f) = f^{(k)}$. The notion of inducing the weakest topology in X is very useful because it will give different notions of compactness on the spaces of interest. This is reflected in the following theorem

Theorem 4.28. *If $\mathscr{X}$ is a normed vector space, then the closed unit ball $B^* = \{f \in \mathscr{X}^* : ||f|| \le 1\}$ in $\mathscr{X}^*$ is compact in the weak* topology.*

The importance of this theorem comes from its analogous in finite-dimensional vector spaces. If V is a normed vector space, then the unit ball is compact iff V is finite-dimensional. Therefore, by considering weak topologies, we are "uncovering" more compact sets that we would have missed if we were only considering normed vector spaces.

Before proving the above theorem, we introduce the apropppiate terminology. For any topological vector space X, we consider its dual space X^* consisting of continuous linear functionals.

- The **weak topology** on X is generated by X^*, i.e. it is generated by the seminorms

$$p_\lambda(x) = |\lambda(x)| \; for \; \lambda \in X^*$$

- The **weak * topology** on X^* is generated by X, i.e. it is generated by the seminorms

$$p_x(\lambda) = |\lambda(x)| \; for \; x \in X$$

The former implies that in X , $x_n \to x$ weakly iff $\lambda(x_n) \to \lambda(x)$ for all $\lambda \in X^*$. For the latter, $f_n \to f$ weakly iff $f_n(x) \to f(x)$ for all $x \in X$. We are in position to prove the theorem

Proof. **Proof of Theorem 4.28**. For each $x \in \mathscr{X}$ consider

$$D_x = \{z \in \mathbb{R} : |z| \le ||x||\}$$

By Tychonoff's theorem, we know that $D = \prod_{x \in \mathscr{X}} D_x$ is compact in the product topology. Each element in D can be identified with the real-valued functions f from $\mathscr{X}$ such that $|f(x)| \le ||x||$ for all $x \in \mathscr{X}$ and the elements of B^* are the linear functions.

Note that the product and weak* topology are the pointwise convergence topology in $\mathscr{X}$. Thus, it suffices to show that B^* is closed, i.e. the pointwise limit of linear functions is a linear function. Let $f_n \to f$ be a pointwise limit and note that

$$f(ax+by) = \lim_n f_n(ax+by) = \lim_n af_n(x) + bf_n(y) = \lim_n af_n(x) + \lim_n bf_n(y) = af(x) + bf(y)$$

□

To conclude the lecture, one may also ask if we can endow the same kind of topology on the space of operators between Banach spaces X, Y. Indeed, we can and we say

- We say that $L(\mathscr{X}, \mathscr{Y})$ has the strong operator topology when $T_n \to T$ iff $T_n(x) \to T(x)$ for all $x \in X$ in the norm topology of Y
- We say that $L(\mathscr{X}, \mathscr{Y})$ has the weak operator topology when $T_n \to T$ iff $T_n(x) \to T(x)$ for all $x \in X$ in the weak topology of Y

4.5 Hilbert spaces I

Let ℓ^2 be the set of functions $f : \mathbb{N} \to \mathbb{R}$ such that $\sum_i f(i)^2 < \infty$. Then it can be shown that $\mathscr{H}$ is a Banach space with the norm $\|f\| = \sum_i f(i)^2$. The subspace of finitely supported f is also a normed vector space, but is not complete.

The space ℓ^2 admits more structure than a general Banach space, and is in some sense the infinite dimensional space that is closest to $\mathbb{R}^d$. The extra structure is given by the **inner product**:

$$\langle f, g\rangle = \sum_i f(i)g(i).$$

A Hilbert space is a generalization of ℓ^2. Moreover, as we will see, every reasonable (i.e., separable) Hilbert space is in fact isomorphic to ℓ^2. To define Hilbert spaces we will first define inner products. Let V be a vector space. An inner product on V is a map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ with the following properties, which are satisfied by the usual inner product in $\mathbb{R}^d$.

1. **Symmetry.** $\langle v, w\rangle = \langle w, v\rangle$.

2. **Bilinearity.** $\langle \lambda v + u, w\rangle = \lambda\langle v, w\rangle + \langle u, w\rangle$.

3. **Positive definiteness.** $\langle v, v\rangle > 0$ for all $v \neq 0$.

Theorem 4.29 (Cauchy-Schwarz Inequality)**.** *For every $u, w \in V$ it holds that $\langle u, w\rangle^2 \leq \langle u, u\rangle\langle w, w\rangle$, with equality iff* $\operatorname{span}\{u\} = \operatorname{span}\{w\}$*.*

Proof. Fix $u, w \in V$. If $w = 0$ then the result is immediate. Otherwise, define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \langle u - \lambda w, u - \lambda w\rangle.$$

By positive definiteness $f \geq 0$, and by bilinearity and symmetry we have that

$$f(x) = \langle u, u\rangle - 2\lambda\langle u, w\rangle + \lambda^2\langle w, w\rangle.$$

Hence f is a non-negative quadratic, and achieves its minimum at $x_0 = \langle u, w\rangle / \langle w, w\rangle$. Now,

$$f(x_0) = \langle x, x\rangle^2 - \langle u, w\rangle^2 / \langle w, w\rangle^2,$$

and since $f \geq 0$ we have that this expression is also greater than 0, yielding the inequality, with equality iff $u - x_0 w = 0$. □

To an inner product we can associate a norm given by $\|v\| = \sqrt{\langle v, v\rangle}$. To see that this satisfies the triangle inequality, apply the Cauchy-Schwarz inequality:

$$\begin{aligned}
\|u + w\|^2 &= \langle u + w, u + w\rangle \\
&= \langle u, u\rangle + 2\langle u, w\rangle + \langle w, w\rangle \\
&\leq \langle u, u\rangle + 2\sqrt{\langle u, u\rangle\langle w, w\rangle} + \langle w, w\rangle \\
&= \|u\|^2 + 2\|u\|\|w\| + \|w\|^2 \\
&= (\|u\| + \|w\|)^2.
\end{aligned}$$

A **Hilbert space** is a vector space equipped with an inner product that is complete under the topology induced by the associated norm.

Given a measure space (Ω, Σ, μ), we define $L^2(\Omega, \Sigma, \mu)$ as the space of measurable functions (up to a.e. equivalence) $f : \Omega \to \mathbb{R}$ such that $\int f^2 \,\mathrm{d}\mu < \infty$. We equip it with the inner product

$$\langle f, g\rangle = \int f(\omega)g(\omega)\,\mathrm{d}\mu(\omega).$$

We will not show this now, but $L^2(\Omega, \Sigma, \mu)$ is a vector space, this is an inner product, and $L^2(\Omega, \Sigma, \mu)$ is moreover a Hilbert space. The space ℓ^2 is a particular case for $\Omega = \mathbb{N}$ and μ the counting measure.

The space $C([0,1])$ can be equipped with an inner product given by $\langle f, g\rangle = \int f g \,\mathrm{d}\lambda$, where λ is the Lebesgue measure. This is not a Hilbert space, however, as it can be shown that it is not complete.

The natural notion of an isomorphism between Hilbert spaces $\mathscr{H}_1, \mathscr{H}_2$ is that of an **orthogonal map** (for complex Hilbert spaces these operators are called unitary) $T\colon \mathscr{H}_1 \to \mathscr{H}_2$, which preserves the inner product:

$$\langle Tv, Tw\rangle_2 = \langle v, w\rangle_1.$$

This implies that T is also an isometry. An important example is the following. Consider $\mathscr{H} = L^2 = L^2(\mathbb{R}, \mathscr{B}, \lambda)$, and let $\sigma\colon \mathbb{R} \to \mathbb{R}$ be given by $\sigma(x) = x+1$. The operator $U\colon L^2 \to L^2$ given by

$$Uf = f \circ \sigma$$

can be seen to be orthogonal. More generally, if $\mathscr{H} = L^2(\Omega, \Sigma, \mu)$ and $\sigma\colon \Omega \to \Omega$ is **measure preserving**—i.e., $\sigma_* \mu = \mu$—then $Uf = f \circ \sigma$ is an orthogonal operator.

Given a Hilbert space $\mathscr{H}$, we say that $u, w \in \mathscr{H}$ are **orthogonal** if $\langle u, w\rangle = 0$. We say that u is orthogonal to a subset $S \subseteq \mathscr{H}$ if it is orthogonal to every $w \in S$.

It can be shown that any finite set $\{v_1, \ldots, v_n\} \subset \mathscr{H}$ spans a space that is isomorphic to some $\mathbb{R}^d$ as a Hilbert space. The idea is to construct a basis of d orthogonal unit vectors to $\mathrm{span}\{v_1, \ldots, v_n\}$ using Gram-Schmidt, and identify this basis with the usual unit vectors of $\mathbb{R}^d$.

This allows us to reduce any question involving only finitely many vectors in $\mathscr{H}$ to a question about $\mathbb{R}^d$.

For example, the next three results only involve finitely many vectors in $\mathscr{H}$. These not dot require completeness.

Theorem 4.30 (Parallelogram Law)**.** *For all $u, w \in \mathscr{H}$ it holds that*

$$\|u+w\|^2 + \|u-w\|^2 = 2(\|u\|^2 + \|w\|^2).$$

Theorem 4.31 (Pythagoras)**.** *If $v_1, \ldots, v_n \in \mathscr{H}$ are orthogonal then*

$$\left\| \sum_{i=1}^n v_i \right\|^2 = \sum_{i=1}^n \|v_i\|^2.$$

Theorem 4.32 (Convexity of the norm)**.** *Fix $v_0 \in \mathscr{H}$, and suppose that $c = \|v_0 - w\| = \|v_0 - u\| > 0$. If $u \neq w$ then $\|v_0 - (u+w)/2\| < c$.*

Note that this does not hold in general for normed vector spaces; for example, take the ℓ^1-norm on $\mathbb{R}^2$.

Another nice operation that Hilbert spaces share with $\mathbb{R}^d$ is that of **orthogonal projection**. The only caveat is that here things will only be nice if we project to a closed subspace. Note that a closed subspace of a Hilbert space is also a Hilbert space, and in particular is complete.

As an example of a non-closed subspace we can take $C([0,1]) \subset L^2([0,1], \mathscr{B}, \lambda)$. The following are examples of closed subspaces:

1. Let $\mathscr{H} = L^2(\Omega, \Sigma, \mu)$, and let $\Sigma_1 \subset \Sigma$ be a sub-sigma algebra. Then $L^2(\Omega, \Sigma_1, \mu)$ is a closed subspace.
2. Let $T : \mathscr{H}_1 \to \mathscr{H}_2$ be a continuous linear operator. Then the kernel of T, $\{v \in \mathscr{H} : Tv = 0\}$ is a closed subspace of $\mathscr{H}_1$.
3. Let $U : \mathscr{H} \to \mathscr{H}$ be a continuous linear operator. Then $\{v : Uv = v\}$ is a closed subspace.
4. Let S be a subset of $\mathscr{H}$. Then the set of vectors $S^\perp$ that are orthogonal to S is a closed vector space.

Let $W \subset \mathscr{H}$ be a closed subspace. We define the projection map $P_W : \mathscr{H} \to \mathscr{H}$ by

$$P_W(v) = \operatorname*{argmin}_{w \in W} \|v - w\|.$$

Proposition 4.33. *The map P_W is well defined, and $v - P_w(v) \in W^\perp$.*

Proof. Let $c = \inf_{w \in W} \|v - w\|$ and let $\|v - w_n\|^2 \le c + 1/n$ for some $w_1, w_2, \ldots \in W$. By Theorem 4.30, and because $(w_n - w_m)/2 \in W$,

$$\begin{aligned} \|w_n - w_m\|^2 &= 2\|w_m - v\|^2 + 2\|w_n - v\|^2 - 4\|(w_n + w_m)/2 - v\|^2 \\ &\le 2\|w_m - v\|^2 + 2\|w_n - v\|^2 - 4c^2 \\ &\le 4/n. \end{aligned}$$

Hence $w_1, w_2, \ldots$ is a Cauchy sequence, and thus by completeness converges to some $u \in W$, because W is closed. It follows that $\|v - u\|^2 = c$, and so $\|v - u\| = \min_{w \in W} \|v - w\|$. By Theorem 4.32, w is the unique minimizer, and so P_W is well defined.

To see that $v - P_W v \in W^\perp$ choose any $w \in W$. Then

$$f(x) = \|v - P_W v - xw\|^2 = \|v - P_W v\|^2 + 2x\langle v - P_W v, w\rangle + x^2\|w\|^2$$

is a quadratic function of x whose minimum must be achieved at $x = 0$. Hence $0 = f'(0) = 2\langle v - P_W v, w\rangle$. $\square$

Claim 4.34. *If W is a closed subspace of $\mathscr{H}$ then $\mathscr{H} = W \oplus W^\perp$.*

Proof. By Proposition 4.33, we can write any $v \in \mathscr{H}$ as $v = P_W v + (v - P_W v) \in W \oplus W^\perp$. Suppose that $v = w + w^\perp$ for some $w \in W, w^\perp \in W^\perp$. Then $P_W v - w = w^\perp - (v - P_W v)$. The left hand side is in W, the right hand is in $W^\perp$, and so both are in $W \cap W^\perp$ and are thus orthogonal to themselves and equal to 0. $\square$

Claim 4.35. *$P_W : \mathscr{H} \to \mathscr{H}$ is a continuous linear operator.*

Proof. Let $v = \lambda v_1 + v_2$. By Claim 4.34 above, $v_i = P_W v_i + (v_i - P_W v_i)$ is the unique representation of v_i as an element of $W \otimes W^\perp$. Hence $v = \lambda P_W v_1 + P_W v_2 + \lambda(v_1 - P_W v_1) + (v_2 - P_W v_2)$ is the unique such represenation of v, and so it must be that $P_W v = \lambda P_W v_1 + P_W v_2$.

To see that P_W is continuous, note that from its definition it follows that $\|P_W v\| \le \|v\|$, since $0 \in W$. Hence P_W is bounded and thus continuous. $\square$

4.6 Hilbert spaces II (by Samuel Goodman and Brian Yang)

In the previous lecture, we gave the definition of a **Hilbert space**, and established some of its basic properties, such as the Cauchy-Schwarz inequality. For the following lecture, $\mathscr{H}$ denotes a Hilbert space.

Now consider the following setup. If $y \in \mathscr{H}$, the Schwarz inequality shows that the formula $f_y(x) = \langle x, y\rangle$ defined a bounded linear functional on $\mathscr{H}$ such that $||f_y|| = ||y||$ (recall definition of operator norm). Thus, the map $\mathscr{H} \to \mathscr{H}^*,\ y \mapsto f_y$ is a conjugate-linear isometry. The next theorem claims it is surjective:

Theorem 4.36. *If $f \in \mathscr{H}^*$, there is a unique $y \in \mathscr{H}$ such that $f(x) = \langle x, y\rangle$ for all $x \in \mathscr{H}$.*

Proof. First we show existence. If $f = 0$, then we may take $y = 0$. Otherwise, $\ker(f)$ is a proper closed subspace of $\mathscr{H}$, so in the decomposition $\mathscr{H} = \ker(f) \oplus \ker(f)^\perp$, we have $\ker(f)^\perp \neq \{0\}$. Thus, there is $z \in \ker(f)^\perp$ with $||z|| = 1$. Given $x \in \mathscr{H}$, notice $u = f(x)z - f(z)x \in \ker(f)$, so that

$$0 = \langle u, z\rangle = f(x)||z|^2 - f(z)\langle x, z\rangle = f(x) - \langle x, \overline{f(z)}z\rangle,$$

i.e., $f(x) = \langle x, \overline{f(z)}z\rangle$ for all $x \in \mathscr{H}$.

Now we verify uniqueness. If $y, y' \in \mathscr{H}$ such that $\langle x, y\rangle = \langle x, y'\rangle$ for all $x \in \mathscr{H}$, then by taking $x = y - y'$, deduce $||y - y'||^2 = 0$, hence $y - y' = 0$. □

Thus, $\mathscr{H}^*$ is *naturally isomorphic* to the conjugate of $\mathscr{H}$. Consequently, the natural map $\mathscr{H} \to \mathscr{H}^* \to \mathscr{H}^{**}$ is a linear isomorphism. This coincides with the usual natural map $\mathscr{H} \to \mathscr{H}^{**}$ i.e. $\mathscr{H}$ is **reflexive**.

Now let us recall some linear algebra. A subset $\{u_\alpha\}_{\alpha\in A}$ of $\mathscr{H}$ is called **orthonormal** if $||u_\alpha|| = 1$ for all α and $\langle u_\alpha, u_\beta\rangle$ when $\alpha \neq \beta$. The **Gram Schmidt process** turns any linearly independent sequence $\{x_n\}_{n=1}^\infty$ in $\mathscr{H}$ into an orthonormal sequence $\{u_n\}$ such that the span of $x_1, x_2, \ldots, x_N$ is the same as that of $u_1, u_2, \ldots, u_N$ for all $N \in \mathbb{N}$. The first step is to set $u_1 = \frac{x_1}{||x_1||}$. Having defined $u_1, \ldots, u_{N-1}$, set

$$v_N = x_N - \sum_{n=1}^{N-1} \langle x_N, u_n\rangle u_n, \quad u_N = \frac{v_N}{||v_N||}.$$

Proposition 4.37 (Bessel's Inequality)**.** *If $\{u_\alpha\}_{\alpha\in A}$ is an orthonormal set in $\mathscr{H}$, then for any $x \in \mathscr{H}$, and any enumeration of a countable subset $u_1, u_2, \ldots$ of $\{u_\alpha\}_{\alpha\in A}$:*

$$\sum_{j=1}^{\infty} |\langle x, u_j\rangle|^2 \leq ||x||^2$$

In particular, $\{\alpha : \langle x, u_\alpha\rangle \neq 0\}$ is countable.

Proof. We show $\sum_{\alpha\in F}|\langle x,u_\alpha\rangle|^2 \le \|x\|^2$ for any finite $F\subseteq A$:

$$\begin{aligned}
0 &\le \left\| x - \sum_{\alpha\in F}\langle x,u_\alpha\rangle u_\alpha \right\|^2 \\
&= \|x\|^2 - 2\mathrm{Re}\left\langle x, \sum_{\alpha\in F}\langle x,u_\alpha\rangle u_\alpha \right\rangle + \left\| \sum_{\alpha\in F}\langle x,u_\alpha\rangle u_\alpha \right\|^2 \\
&= \|x\|^2 - 2\sum_{\alpha\in F}|\langle x,u_\alpha\rangle|^2 + \sum_{\alpha\in F}|\langle x,u_\alpha\rangle|^2 \qquad \text{(by Pythagorean theorem)} \\
&= \|x\|^2 - \sum_{\alpha\in F}|\langle x,u_\alpha\rangle|^2,
\end{aligned}$$

as needed. □

Now let us give the definition of an **orthonormal basis**. Note that this is not quite the traditional definition of a basis of a vector space!

Theorem 4.38. *If $\{u_\alpha\}_{\alpha\in A}$ is an orthonormal set in $\mathscr{H}$, the following are equivalent:*

*(a) (**Completeness**) For $x\in\mathscr{H}$, if $\langle x,u_\alpha\rangle = 0$ for all $\alpha\in A$, then $x=0$.*

*(b) (**Parseval's Identity**) $\|x\|^2 = \sum_{\alpha\in A}|\langle x,u_\alpha\rangle|^2$ for all $x\in\mathscr{H}$.*

(c) For any $x\in\mathscr{H}$, $x=\sum_{\alpha\in A}\langle x,u_\alpha\rangle u_\alpha$, where this sum has only countably many nonzero terms and converges in the norm topology for any enumeration $\alpha_1,\alpha_2\dots$ of the α's for which $\langle x,u_\alpha\rangle\neq 0$.

Proof. Fix the following notation: $x\in\mathscr{H}$, and $\alpha_1,\alpha_2,\dots$ is any enumeration of the α's for which $\langle x,u_\alpha\rangle\neq 0$.

(b) $\implies$ (a): immediate.

(c) $\implies$ (b): We have $\|x-\sum_{j=1}^n\langle x,u_{\alpha_j}\rangle u_{\alpha_j}\|^2 \to 0$ as $n\to\infty$ by (c). However, for any $n\in\mathbb{N}$, notice $\|x\|^2-\sum_{j=1}^n|\langle x,u_{\alpha_j}\rangle|^2 = \|x-\sum_{j=1}^n\langle x,u_{\alpha_j}\rangle u_{\alpha_j}\|^2$ by the computation in the proof of Bessel's inequality, so (b) follows.

(a) $\implies$ (c): Note that by the Pythagorean theorem, we have for any $n\le m$:

$$\left\| \sum_{j=n}^m \langle x,u_{\alpha_j}\rangle u_{\alpha_j} \right\|^2 = \sum_{j=n}^m |\langle x,u_{\alpha_j}\rangle|^2.$$

However, the sequence $\sum_{j=1}^\infty|\langle x,u_{\alpha_j}\rangle|^2$ converges by Bessel's inequality. In conjunction with the above identity, we deduce that the sequence $\sum_{j=1}^n\langle x,u_{\alpha_j}\rangle u_{\alpha_j}$ is Cauchy in the norm topology. Hence, $\sum_{j=1}^\infty\langle x,u_{\alpha_j}\rangle u_{\alpha_j}$ converges by $\mathscr{H}$ complete.

Now, set $y = x-\sum_{j=1}^\infty\langle x,u_{\alpha_j}\rangle u_{\alpha_j}$. Then, for any $\alpha\in A$, we see that

$$\langle y,u_\alpha\rangle = \langle x,u_\alpha\rangle - \sum_{j=1}^\infty \langle x,u_{\alpha_j}\rangle\langle u_{\alpha_j},u_\alpha\rangle$$

(we have already seen by Cauchy-Schwarz that the inner product commutes with convergent sequences of vectors in $\mathscr{H}$). Hence, $\langle y,u_\alpha\rangle = 0$, meaning $y=0$ by (a), so that (c) follows. □

An orthonormal set $\{u_\alpha\}_{\alpha\in A}$ of $\mathscr{H}$ is called a **orthonormal basis** if $\{u_\alpha\}$ satisfies any of the equivalent conditions of Theorem 4.38. Again, this is not the standard notion of an orthonormal basis, but it is the natural structure lying behind spaces like $\ell_2(\mathbb{N})$, where we can still have convergence at the level of partial sums. Nonetheless, we can show that it has many of the desirable properties that standard bases have.

Proposition 4.39. *Every Hilbert space has an orthonormal basis.*

Proof. Consider the set S of orthonormal subsets of $\mathscr{H}$. Any single nonzero vector may be normalized to have norm 1, and so S is nonempty. Furthermore, any chain of sets in S has an upper bound (namely the union of all sets in the chain). Thus Zorn's Lemma implies that S has a maximal element $T = \{u_\alpha\}_{\alpha\in A}$, which is a maximal orthornomal subset. We claim that T has the completeness property of Theorem 4.32. Suppose that for some nonzero x, we have $\langle x, u_\alpha\rangle = 0$ for all $u_\alpha \in T$. Since x is nonzero, we may normalize x to some x' so that $||x'|| = 1$. But then appending x' to the u_αs gives an orthonormal subset of $\mathscr{H}$ strictly containing T, contradicting the maximality of T. Thus we conclude that if $\langle x, u_\alpha\rangle = 0$ for all $u_\alpha \in T$, we must have $x = 0$, verifying the completeness property. But then by the Theorem 4.32, it follows that T is a basis for $\mathscr{H}$, completing the proof. ☐

Proposition 4.40. *A Hilbert space $\mathscr{H}$ is separable iff it has a countable orthonormal basis, in which case every orthonormal basis is countable.*

Proof. If $\mathscr{H}$ is separable, then we can find a countable dense subset $\{x_n, n \in \mathbb{N}\}$. Iteratively removing the x_n that are in the span of $\{x_1, \cdots, x_{n-1}\}$, we obtain a subset of $\mathscr{H}$ with span dense in $\mathscr{H}$. Applying Gram-Schmidt to what remains gives an orthonormal subset of $\mathscr{H}$. Let $\{y_n, n \in \mathbb{N}\}$ be the resulting sequence. As the span is dense in $\mathscr{H}$, completeness implies that it is actually a basis for $\mathscr{H}$, and so $\{y_n, n \in \mathbb{N}\}$ is a countable orthonormal basis for $\mathscr{H}$. Conversely, given a countable orthonormal basis $\{u_n, n \in \mathbb{N}\}$, we can choose a countable dense subset S of $\mathbb{C}$ and then consider all finite linear combinations of the u_i s with coefficients in S, which yields a countable dense subset of $\mathscr{H}$, implying that $\mathscr{H}$ is in fact separable. Then given another orthonormal basis $\{a_\alpha, \alpha \in A\}$, the sets $A_n = \{\alpha \in A, \langle a_\alpha, u_n\rangle \neq 0\}$ is countable for each n by definition, and thus $B = \bigcup_{n\in\mathbb{N}} A_n$ is countable. Now any element of $\mathscr{H}$ not in the span of the $\{a_\alpha, \alpha \in B\}$ is orthogonal to each u_n, so by completeness, is 0, which also cannot occur as 0 is in the span. Thus $\{a_\alpha, \alpha \in B\}$ spans $\mathscr{H}$ and so $A = B$, implying that A is countable. ☐

Hilbert spaces encode quite a lot of structure, and so a natural question to ask is which maps between Hilbert spaces accurately capture all that structure. Given Hilbert spaces $\mathscr{H}_1$ and $\mathscr{H}_2$ equipped with inner products $\langle\cdot,\cdot\rangle_1$ and $\langle\cdot,\cdot\rangle_2$, a unitary map $\mathscr{H}_1 \to \mathscr{H}_2$ is an invertible linear map U that preserves the inner products, namely for any $x_1, x_2 \in \mathscr{H}_1$, we have $\langle x, y\rangle_1 = \langle Ux, Uy\rangle_2$. Note that upon letting $x = y$, this implies that $\langle x, x\rangle_1 = \langle Ux, Ux\rangle_2$, which means that $||x||_1 = ||Ux||_2$ and thus that any such U is an isometry. Since unitary maps preserve norms, they preserve the topologies on $\mathscr{H}_1$ and $\mathscr{H}_2$. Thus the condition of preserving inner products essentially means that all information (inner product, norms, and topologies) on $\mathscr{H}_1$ are preserved by a unitary map, and so unitary maps are said to be the true "isomorphisms" in the category of Hilbert spaces.

Theorem 4.41. *Let $\{e_\alpha, \alpha \in A\}$ be an orthonormal basis for $\mathscr{H}$. Then the map $f : \mathscr{H} \to \ell_2(A)$ given by $x \to \hat{x}$, where $\hat{x}(\alpha) = \langle x, e_\alpha \rangle$, is a unitary map.*

Proof. The map f is clearly linear. We show that it is an isometry. Note that $||x||^2 = \sum_{\alpha\in A} |\hat{x}(\alpha)|^2 < \infty$ by Parseval's identity (Theorem 4.38), so f is well-defined ($\hat{x}$ is actually an element of $\ell_2(A)$) and an isometry (since the norm on $\ell_2(A)$ is given by $\sum_{\alpha\in A} g(\alpha)^2$). Now given $g \in \ell^2(A)$, we have that $\sum_{\alpha\in A} |g(\alpha)|^2 < \infty$ and so as the sum $\sum_{\alpha\in A} |g(\alpha)|^2$ converges, the partial sums $\sum g(\alpha)e_\alpha$ form a Cauchy sequence, which by completeness means that we can write $x = \sum_{\alpha\in A} g(\alpha)e_\alpha$ for some $x \in \mathscr{H}$, from which it follows that $g = \hat{x}$ (as the e_αs form a basis and invoking Theorem 4.38). Thus f is a surjective isometry. Now we claim that a surjective isometry between Hilbert spaces is a unitary map. Note that $||fx - fy||_2 = ||f(x-y)||_2 = ||x-y||_1$, and so $fx = fy$, then $x = y$, proving injectivity. Thus f is bijective and thus invertible. Furthermore, we have the identity $\langle fx, fy\rangle_2 = \frac{1}{4}(||fx+fy||_2^2 - ||fx-fy||_2^2) = \frac{1}{4}(||x+y||_1^2 - ||x-y||_1^2) = \langle x, y\rangle_1$. Thus we conclude that f is unitary. □

Thus any Hilbert space is in some sense "isomorphic" to the space $\ell_2(A)$, where A is the indexing set of its orthonormal basis.

5 L^p Spaces

5.1 Hölder's & Minkowski's inequalities

Let (Ω, Σ, μ) be a measure space. Given a measurable $f : \Sigma \to \mathbb{R}$ and $p \geq 1$, we define its p-norm by

$$\|f\|_p = \left(\int |f|^p \, \mathrm{d}\mu \right)^{1/p}.$$

We shall see that this is indeed a norm; homogeneity and positive-definitiveness are immediate. While we could have also defined this for $p < 1$, this would not have been a norm.

For $p = \infty$ we define

$$\|f\|_\infty = \inf\{a : \mu(\{\omega : |f(\omega)| > a\}) = 0\}.$$

It is immediate to see that $\|\cdot\|_\infty$ satisfies the triangle inequality.

We denote by $L^p(\Omega, \Sigma, \mu)$ (or just L^p) the set of measurable functions $f : \Omega \to \mathbb{R}$, up to agreement a.e., such that $\|f\|_p < \infty$. When we show that the p-norm is indeed a norm, we will also prove that this is a vector space, by the triangle inequality.

Towards this goal, we will first prove Hölder's inequality. It is a consequence of the following claim.

Claim 5.1 (Arithmetic Mean-Geometric Mean (AM-GM) Inequality). *For all $x, y \geq 0$ and $\lambda \in (0,1)$ it holds that*

$$x^\alpha y^{1-\alpha} \leq \alpha x + (1-\alpha) y.$$

Proof. If x or y are zero then the statement is immediate. Otherwise

$$x^\alpha y^{1-\alpha} = \exp(\alpha \log x + (1-\alpha) \log y) \leq \alpha x + (1-\alpha) y,$$

by the convexity of the exponent function. $\square$

We say that $p, q \geq 1$ are **conjugate exponents** if $1/p + 1/q = 1$. For example, $p = 2$ and $q = 2$ are conjugate, as are $p = 3$ and $q = 3/2$ and $p = 1$ and $q = \infty$.

Theorem 5.2 (Hölder's Inequality). *For conjugate $p, q \in [1, \infty]$, it holds for all $f \in L^p$ and $g \in L^q$ that $\|f \cdot g\|_1 \leq \|f\|_p \cdot \|g\|_q$.*

Proof. The case of $p = \infty$ is immediate, so we assume $p, q < \infty$. Suppose that $\|f\|_p = \|g\|_q = 1$; the case of vanishing norm is trivial and the cases of non-unit norms reduce to this one by the homogeneity of p-norms. Then

$$\|f \cdot g\|_1 = \int |f \cdot g| \, \mathrm{d}\mu = \int (|f|^p)^{1/p} (|g|^q)^{1-1/p} \, \mathrm{d}\mu \leq \frac{1}{p} \int |f|^p \, \mathrm{d}\mu + \frac{1}{q} \int |g|^q \mathrm{d}\mu = 1/p + 1/q = 1,$$

where the inequality follows from the AM-GM inequality. $\square$

Using Hölder's inequality we can prove that p-norms are indeed norms, by showing the triangle inequality.

Theorem 5.3 (Minkowski's Inequality)**.** *For all $p \in [1,\infty]$ and $f,g \in L^p$ it holds that $\|f+g\|_p \leq \|f\|_p + \|g\|_p$.*

Proof. The cases $p=1$ and $p=\infty$ are immediate, as is the case $f+g=0$. Otherwise, let q be the conjugate exponent to p. Note that

$$|f+g|^p = |f+g|\cdot|f+g|^{p-1} \leq |f|\cdot|f+g|^{p-1} + |g|\cdot|f+g|^{p-1}.$$

Hence

$$\|f+g\|_p^p = \int |f+g|^p \,\mathrm{d}\mu \leq \int |f|\cdot|f+g|^{p-1}\,\mathrm{d}\mu + \int |g|\cdot|f+g|^{p-1}\,\mathrm{d}\mu.$$

We apply Hölder's Inequality to both of these integrals to get

$$\|f+g\|_p^p \leq \left(\|f\|_p + \|g\|_p\right)\cdot \left\||f+g|^{p-1}\right\|_q.$$

Now,

$$\left\||f+g|^{p-1}\right\|_q = \left(\int |f+g|^{q(p-1)}\,\mathrm{d}\mu\right)^{1/q} = \left(\int |f+g|^p\,\mathrm{d}\mu\right)^{1-1/p} = \|f+g\|_p^{p(1-1/p)} = \|f+g\|_p^{p-1}.$$

We thus have that

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p.$$

□

It follows from Minkowski's Inequality that L^p is a normed space. Moreover, by applying monotone and dominated convergence it can be shown that these are moreover Banach spaces.

Given a set A, we denote by $\ell^p(A)$ the space $L^p(A, 2^A, c)$, where c is the counting measure on A. We write $\ell^p = \ell^p(\mathbb{N})$. It is easy to see that for $p \neq q$ it holds that $\ell^p \neq \ell^q$.

Proposition 5.4. *For any $1 \leq p \leq q \leq \infty$ it holds that $\ell^p \subseteq \ell^q$, and moreover $\|f\|_q \leq \|f\|_p$.*

Proof.

$$\|f\|_q = \left(\int |f|^q\,\mathrm{d}\mu\right)^{1/q} \leq \left(\int \|f\|_\infty^{q-p}|f|^p\,\mathrm{d}\mu\right)^{1/q} = \|f\|_\infty^{1-p/q}\|f\|_p^{p/q} \leq (1-p/q)\|f\|_\infty + (p/q)\|f\|_p,$$

where the last inequality is the AM-GM Inequality, with $\alpha = p/q$. Now,

$$\|f\|_\infty^p = \sup_i |f(i)|^p \leq \sum_i |f(i)|^p = \|f\|_p^p,$$

and so $\|f\|_\infty \leq \|f\|_p$ and we are done. □

Proposition 5.5. *Suppose $\mu(\Omega) = 1$. Then for any $1 \le p \le q \le \infty$ it holds that $L^q \subseteq L^p$, and moreover $\|f\|_p \le \|f\|_q$.*

Proof. This is immediate for $q = \infty$. For $q < \infty$, we note that q/p and $q/(q-p)$ are conjugate exponents, and apply Hölder's Inequality:

$$\|f\|_p^p = \int |f|^p \cdot 1 \, \mathrm{d}\mu \le \left\| |f|^p \right\|_{q/p} \cdot \|1\|_{q/(q-p)} = \left(\int (|f|^p)^{q/p} \, \mathrm{d}\mu \right)^{p/q} = \left(\int |f|^q \, \mathrm{d}\mu \right)^{p/q} = \|f\|_q^p.$$

□

For general measure spaces neither of the above propositions hold. For example, for the Lebesgue measure on $(0, \infty)$, $p < 1/a < q$ and $f(x) = x^{-a}$ we have that $f(x)\mathbb{1}_{x>1}$ is in L^q but not L^p, and $f(x)\mathbb{1}_{x<1}$ is in L^p but not L^q. This example hints of a more general phenomenon.

Theorem 5.6. *Suppose $1 \le p < q < r \le \infty$, and fix $f \in L^q$. Let $A = \{\omega : |f(\omega)| > 1\}$. Then $f\mathbb{1}_A \in L^p$ and $f\mathbb{1}_{A^c} \in L^r$. Hence $L^q = L^p + L^r$.*

Proof. On A we have that $|f|^p \le |f|^q$, and on A^c we have that $|f|^r \le |f|^q$. □

We end with the following result, which we will not prove (using Hölder's Inequality).

Theorem 5.7. *Suppose $1 \le p < q < r \le \infty$. Then $L^p \cap L^r \subseteq L^q$.*

5.2 The Dual of L^p

In this section we let (Ω, Σ, μ) be any σ-finite measure space, and denote $L^p = L^p(\Omega, \Sigma, \mu)$.

Let p and q be conjugate exponents. Given $g \in L^q$, we can define a linear functional φ_g on L^p by

$$\varphi_g(f) = \int fg.$$

By Hölder's inequality this is indeed finite, and moreover $\|\varphi_g\| \le \|g\|$. The next claim shows that we in fact have equality, as we have already shown for Hilbert spaces (Theorem 4.36). Before that, a simple lemma.

Lemma 5.8. *Let p and $q < \infty$ be conjugate exponents. Given $g \in L^q$, the function $f = |g|^{q-1}/\|g\|_q^{q-1}$ satisfies $\|f\|_p = 1$.*

Proof. Since p and q are conjugate, $(q-1)p = q$, and so

$$\|f\|_p^p = \int |f|^p = \frac{1}{\|g\|_q^{(q-1)p}} \int |g|^{(q-1)p} = \frac{1}{\|g\|_q^q} \int |g|^q = 1.$$

□

Proposition 5.9. *If p and q are conjugate exponents and $q \in [1, \infty]$ then for any $g \in L^q$ it holds that $\|\varphi_g\| = \|g\|$.*

Proof. We prove the case $q \in (1,\infty)$. If $g = 0$ then the result is immediate. Otherwise let

$$f = \frac{|g|^{q-1} \cdot \operatorname{sgn} g}{\|g\|_q^{q-1}},$$

so that $\|f\|_p = 1$ by Lemma 5.8, and

$$\|\varphi_g\| \geq \int fg = \frac{1}{\|g\|_q^{q-1}} \int |g|^q = \frac{\|g\|_q^q}{\|g\|_q^{q-1}} = \|g\|_q.$$

The direction $\|\varphi_g\| \leq \|g\|_q$ follow from Hölder's inequality. ☐

It turns out that for $p \in [1,\infty)$ every bounded linear functional is of the form φ_g, for some $g \in L^q$.

Theorem 5.10. *Let p and q be conjugate exponents, and suppose $p \in [1,\infty)$. Let $\varphi \colon L^p \to \mathbb{R}$ be a bounded linear functional. Then there exists a $g \in L^q$ such that $\varphi = \varphi_g$.*

Before proving this theorem we will state (but not prove) the following lemma.

Lemma 5.11. *The bounded functions are a dense subspace of L^p.*

The proof is a simple application of the dominated or monotone convergence theorem.

Proof of Theorem 5.10. We prove for the case that μ is a finite measure and $p \in (1,\infty)$. Let φ be a bounded linear functional on L^p. We define a map $\eta \colon \Sigma \to \mathbb{R}$ by

$$\eta(A) = \varphi(\mathbb{1}_A).$$

Note that $\mathbb{1}_A \in L^p$ for any $A \in \Sigma$ since $\mu(\Omega) < \infty$.

Suppose that $A_1, A_2, \ldots \in \Sigma$ are disjoint, and that $A = \cup_n A_n$, so that $\mathbb{1}_A = \sum_n \mathbb{1}_{A_n}$. We claim that

$$\lim_n \sum_{k=1}^n \mathbb{1}_{A_k} = \mathbb{1}_A$$

as a sequences of elements of L^p. This holds because

$$\left\| \mathbb{1}_A - \sum_{k=1}^n \mathbb{1}_{A_k} \right\|_p = \left\| \sum_{k=n+1}^\infty \mathbb{1}_{A_k} \right\|_p = \mu(\cup_{k=n+1} A_n)^{1/p},$$

which tends to 0 with n. Since φ is continuous and linear it follows that

$$\eta(A) = \varphi(\mathbb{1}_A) = \varphi\left(\sum_k \mathbb{1}_{A_k}\right) = \lim_n \varphi\left(\sum_{k=1}^n \mathbb{1}_{A_k}\right) = \lim_n \sum_{k=1}^n \eta(A_k) = \sum_k \eta(A_k).$$

We have thus shown that η is a signed measure.

Note that if $\mu(A) = 0$ then $\mathbb{1}_A$ is a.e. equal to zero, and so $\varphi(\mathbb{1}_A) = \varphi(0) = 0$. Hence η is absolutely continuous, and so by Theorem 3.8 there exists a μ-integrable g such that

$$\int f \,\mathrm{d}\eta = \int f g \,\mathrm{d}\mu.$$

By the definition of η and the additivity and linearity of integrals we have that $\varphi(f) = \int f \,\mathrm{d}\eta$ for any simple f. Since bounded functions can be approximated from above and below by simple functions, it follows that the same holds for any bounded f.

Let g_n be a sequence of simple functions that converge pointwise to g and such that $|g_n| \leq |g|$. Then by Fatou's Lemma (Theorem 2.10) we have that

$$\begin{aligned}
\|g\|_q &\leq \liminf_n \|g_n\|_q \\
&= \liminf_n \frac{\|g_n\|_q^q}{\|g_n\|_q^{q-1}} \\
&= \liminf_n \int \frac{|g_n|^q}{\|g_n\|_q^{q-1}} \,\mathrm{d}\mu \\
&\leq \liminf_n \int \frac{|g_n|^{q-1}|g|}{\|g_n\|_q^{q-1}} \,\mathrm{d}\mu \\
&= \liminf_n \int \frac{|g_n|^{q-1}\mathrm{sgn}(g)}{\|g_n\|_q^{q-1}} g \,\mathrm{d}\mu \\
&= \liminf_n \int \frac{|g_n|^{q-1}\mathrm{sgn}(g)}{\|g_n\|_q^{q-1}} \,\mathrm{d}\eta \\
&= \varphi\left(\frac{|g_n|^{q-1}\mathrm{sgn}(g)}{\|g_n\|_q^{q-1}}\right),
\end{aligned}$$

where the last equality is a consequence of the fact that the function being integrated is bounded. Now, by Lemma 5.8, this function has L^p-norm 1. Hence this is at most $\|\varphi\|$. We have thus shown that $g \in L^q$. Since φ agrees with the bounded linear functional φ_g on the bounded functions, and since they are a dense by by the lemma above, these functionals are identical.

□

Note that ℓ^1 is not the dual of ℓ^∞. To see this, let $\varphi_n \colon \ell^\infty \to \mathbb{R}$ be the linear functional $\varphi_n(f) = f(n)$. Note that $\varphi_n(1) = 1$, and hence, if φ is a cluster point of φ_n, then $\varphi(1) = 1$. Such a cluster point must exist, by Theorem 4.28. But $\varphi(f) = 0$ for all finitely supported f, so $\varphi \neq \varphi_g$ for any $g \in \ell^1$.

Proposition 5.12. *For every linear functional φ on L^1 there a $g \in L^\infty$ such that $\varphi = \varphi_g$.*

Proof. We prove for ℓ^1. Let $g(n) = \varphi(\delta_n)$, where $\delta_n(\cdot)$ is the indicator of $\{n\}$. Since φ is bounded, $g \in \ell^\infty$. Given $f \in \ell^1$, denote $f_n = \sum_{k=1}^n \delta_n f(n)$. Then $f_n \to f$ in ℓ^1, and so the finitely supported functions are dense in ℓ^1. Since φ and φ_g agree on these, they are identical. □

5.3 Interpolation (by Eric Ma and Zhaojun Chen)

In this section, our goal is to prove the following theorem.

Theorem 5.13 (The Riesz-Thorin Interpolation Theorem). *Suppose that* $(\Omega_1, \Sigma_1, \mu_1)$ *and* $(\Omega_2, \Sigma_2, \mu_2)$ *are sigma-finite measure spaces and* $p_0 \le p_1, q_0 \le q_1 \in [1, \infty]$. *For* $t \in (0,1)$ *let*

$$p_t^{-1} = (1-t)p_0^{-1} + tp_1^{-1}$$
$$q_t^{-1} = (1-t)q_0^{-1} + tq_1^{-1}.$$

Let T *be a bounded linear map from* $L^{p_0}(\Omega_1, \Sigma_1, \mu_1) + L^{p_1}(\Omega_1, \Sigma_1, \mu_1)$ *to* $L^{q_0}(\Omega_2, \Sigma_2, \mu_2) + L^{q_1}(\Omega_2, \Sigma_2, \mu_2)$. *Let* M_t *be the operator norm of* T *as a map from* $L^{p_t}(\Omega_1, \Sigma_1, \mu_1) \to L^{q_t}(\Omega_2, \Sigma_2, \mu_2)$. *Then* $\log M_t \le (1-t)\log M_0 + t \log M_1$.

Our motivation is as follows. Recall that for $1 \le p < q < r \le \infty$, $L^p \cap L^r \subset L^q \subset L^p + L^r$. A natural question to ask would be whether a linear operator T on $L^p + L^r$ that is bounded on both L^p and L^r individually is also bounded on L^q. The Riesz-Thorin theorem tells us that the answer is yes. To prove this theorem, we will first introduce the following result from complex analysis.

Lemma 5.14 (Three Lines Lemma). *Let* ϕ *be a bounded continuous function on the strip* $0 \le \Re z \le 1$ *(where* $\Re z$ *means the real part of* z*) that is holomorphic on the interior of the strip. If* $|\phi(z)| \le M_0$ *when* $\Re z = 0$ *and* $|\phi(z)| \le M_1$ *when* $\Re z = 1$, *then* $|\phi(z) \le M_0^{1-t} M_1^t$ *for* $\Re z = t, 0 < t < 1$.

Proof. We skip the proof of this because it uses results from complex analysis. □

We will also need the following characterization of the norm.

Theorem 5.15. *Let* (Ω, Σ, μ) *be a* σ*-finite measure space. Let* p, q *be conjugate exponents. Suppose that* g *is a measurable function on the space, and* $fg \in L^1(\Omega, \Sigma, \mu)$ *for all* $f \in S$ *where* S *is the set of all simple functions with finite support on the space, and*

$$M_q(g) = \sup\left\{\left|\int fg\right| : f \in S, \|f\|_p = 1\right\}$$

is finite. Then $g \in L^q(\Omega, \Sigma, \mu)$ *and* $M_q(g) = \|g\|_q$.

Using this result, we prove the Riesz-Thorin Theorem.

Proof of Riesz-Thorin Interpolation. We leave the case where $p_0 = p_1$ as a (homework) exercise. Thus we may assume that $p_0 < p_1$ — then p_0 must be finite, so $p_t < \infty$.

Let $S_i, i = 1,2$ denote the space of simple functions on Ω_i that have finite support. Then $S_i \subset L^p(\Omega_i, \Sigma_i, \mu_i)$ and S_i is dense in $L^p(\Omega_i, \Sigma_i, \mu_i)$ for $p < \infty$. The bulk of the proof will lie in showing that for $f \in S_1, \|Tf\|_{q_t} \le M_0^{1-t} M_1^t \|f\|_{p_t}$. To show this, we rely on the characterization of the q_t norm that we introduced above:

$$\|Tf\|_{q_t} = \sup\left\{\left|\int (Tf) g \, d\mu_2\right| : g \in S_2, \|g\|_{q_t'} = 1\right\},$$

where q_t' is the conjugate exponent of q_t. Then it remains to show that $\left|\int (Tf)g d\mu_2\right| \le M_0^{1-t}M_1^t \|f\|_{p_t}$. This inequality is trivial if $f=0$. Thus, we may scale f such that $\|f\|_{p_t}=1$, so it remains to show that

$$\left|\int (Tf)g d\mu_2\right| \le M_0^{1-t}M_1^t \text{ for } f\in S_1, \|f\|_{p_t}=1, g\in S_2, \|g\|_{q_t'}=1.$$

Now let $f=\sum_{j=1}^m c_j \mathbb{1}_{E_j}, g=\sum_{k=1}^n d_k \mathbb{1}_{F_k}$, where the E_j (resp. F_k) are pairwise disjoint and $c_j, d_k \ne 0$. Now let

$$\alpha(z)=(1-z)p_0^{-1}+zp_1^{-1}, \beta(z)=(1-z)q_0^{-1}+zq_1^{-1}.$$

where $\alpha:\mathbb{C}\to\mathbb{C}, \beta:\mathbb{C}\to\mathbb{C}$. If we let $t\in(0,1)$, then $\alpha(t)=p_t^{-1}, \beta(t)=q_t^{-1}$.

Now fix a $t\in(0,1)$. Then recall $p_t<\infty$, so $\alpha(t)>0$. Thus we define

$$f_z=\sum_{j=1}^m |c_j|^{\alpha(z)/\alpha(t)}\mathbb{1}_{E_j}.$$

If $\beta(t)<1$, we let

$$g_z=\sum_{k=1}^m |d_k|^{(1-\beta(z))/(1-\beta(t))}\mathbb{1}_{F_k}.$$

If $\beta(t)=1$, we let $g_z=g$ instead. We leave the $\beta(t)=1$ case as an exercise. As a potentially helpful clarifying note, recall that $f_z:\Omega_1\to\mathbb{C}, g_z:\Omega_2\to\mathbb{C}$. Now let $\phi:\mathbb{C}\to\mathbb{C}$ by $\phi(z)=\int(Tf_z)g_z d\mu_2$. Observe that ϕ is holomorphic everywhere, and that it is bounded on the strip $0\le\Re z\le 1$. Because $\phi(t)=\int(Tf)gd\mu_2$, it remains to show that $|\phi(z)|\le M_0$ when $\Re z=0$, and $|\phi(z)|\le M_1$ when $\Re z=1$. We check the former and leave the latter to the reader.

Let $z=is, s\in\mathbb{R}$, so $\Re z=0$. We can factor $\alpha(is)=p_0^{-1}+is(p_1^{-1}-p_0^{-1})$. Then observe

$$|f_{is}|\le|f|^{\Re(\alpha(is)/\alpha(t))}=|f|^{p_t/p_0}, |g_{is}|\le|g|^{\Re((1-\beta(is))/(1-\beta(t)))}=|g|^{q_t'/q_0'}.$$

Then we conclude as follows:

$$|\phi(is)|\le\|Tf_{is}\|_{q_0}\|g_{is}\|_{q_0'}\le M_0\|f_{is}\|_{p_0}\|g_{is}\|_{q_0'}\le M_0\|f\|_{p_t}\|g\|_{q_t'}=M_0,$$

where in the last inequality we use the fact that we scaled f.

Finally, we extend the result for functions in S_1 to all measurable functions on $(\Omega_1,\Sigma_1,\mu_1)$. Take any $f\in L^{p_t}(\Omega_1,\Sigma_1,\mu_1)$ and choose a sequence $f_n\in S_1$ such that $|f_n|\le|f|$ and $f_n\to f$ pointwise, and let $E=\{x\in\Omega_1:|f(x)|>1\}, g=f\mathbb{1}_E, g_n=f_n\mathbb{1}_E, h=f-g, h_n=f_n-g_n$. Now assume $p_0<p_1$ w.l.o.g.; then, it is easy to check that $g\in L^{p_0}, h\in L^{p_1}$. By the Dominated Convergence Theorem, $\|f_n-f\|_{p_t}\to 0, \|g_n-g\|_{p_0}\to 0, \|h_n-h\|_{p_1}\to 0$. By continuity, $\|Tg_n-Tg\|, \|Th_n-Th\|\to 0$, and it is left to an exercise that there is a subsequence of g_n, h_n such that $Tg_n\to Tg, Th_n\to Th$ almost everywhere. Then $Tf_n\to f$ almost everywhere. By Fatou's Lemma,

$$\|Tf\|_{q_t}\le\liminf\|Tf_n\|_{q_t}\le\liminf M_0^{1-t}M_1^t\|f_n\|_{p_t}=M_0^{1-t}M_1^t\|f\|_{p_t}.$$

□

Another very useful theorem is the Marcinkiewicz Interpolation Theorem:

Theorem 5.16. *Suppose that* $(X, \mathscr{M}, \mu)$ *and* $(Y, \mathscr{N}, v)$ *are measure spaces;* p_0, p_1, q_0, q_1 *are elements of* $[1, \infty]$ *such that* $p_0 \le q_1$ *and* $q_0 \neq q_1$; *and* $\frac{1}{p} = \frac{1-t}{p_0} + \frac{t}{p_1}$ *and* $\frac{1}{q} = \frac{1-t}{q_0} + \frac{t}{q_1}$, *where* $0 < t < 1$.

If T *is a sublinear map from* $L^{p_0}(\mu) + L^{p_1}(\mu)$ *to the space of measurable functions on* Y *that is weak types* (p_0, q_0) *and* (p_1, q_1), *then* T *is strong type* (p, q). *More precisely, if* $[Tf]_{q_j} \le C_j ||f||_{p_j}$ *for* $j = 0, 1$, *then* $||Tf||_q \le B_p ||f||_p$ *where* B_p *depends only on* p_j, q_j, C_j *in addition to* p; *and for* $j = 0, 1$, $B_p |p - p_j|$ *(resp.* B_p*) remains bounded as* $p \to p_j$ *if* $p_j < \infty$ *(resp.* $p_j = \infty$*).*

The following are two applications of the Marcinkiewicz theorem. The first one concerns the Hardy-Littlewood maximal operator $H : Hf(x) = \sup_{r>0} \frac{1}{m(B(r,x))} \int_{B(r,x)} |f(y)| dy$ ($f \in L^1_{loc}(\mathbb{R}^n)$).

Corollary 5.17. *There is a constant* $C > 0$ *such that if* $1 < p < \infty$ *and* $f \in L^p(\mathbb{R}^n)$, *then* $||Hf||_p \le C \frac{p}{p-1} ||f||_p$.

Our second application is a theorem on integral operators.

Theorem 5.18. *Suppose* $(X, \mathscr{M}, \mu)$ *and* $(Y, \mathscr{N}, v)$ *are* σ*-finite measure spaces, and* $1 < q < \infty$. *Let* K *be a measurable function on* $X \times Y$ *such that, for some* $C > 0$, *we have* $[K(x, .)]_q \le C$ *for a.e.* $x \in X$ *and* $[K(., y)]_q \le C$ *for a.e.* $y \in Y$. *If* $1 \le p < \infty$ *and* $f \in L^p(v)$, *the integral* $Tf(x) = \int K(x, y) f(y) dv(y)$ *converges absolutely for a.e.* $x \in X$, *and the operator* T *thus defined is weak type* $(1, q)$ *and strong type* (p, r) *for all* p, r *such that* $1 < p < r < \infty$ *and* $p^{-1} + q^{-1} = r^{-1} + 1$. *More precisely, there exist constant* B_p *independent of* K *such that* $[Tf]_q \le B_1 C ||f||_1$, $||Tf||_r \le B_p C ||f||_p$ $(p > 1, r^{-1} = p^{-1} + q^{-1} - 1 > 0)$.

6 Representation theorems for continuous function spaces

6.1 Riesz Representation Theorem for the Cantor space

Let X be a compact topological space. Denote by $C(X)$ the space of all real continuous functions on X, endowed with the (uniform/sup) form $\|f\| = \max_{x\in X}|f(x)|$. A linear functional φ on $C(X)$ is said to be **positive** if $\varphi(f) \geq 0$ for all $f \geq 0$.

Recall that a topological space X is **Hausdorff** if for any $x \neq y \in X$ there are disjoint open sets $U_x \ni x$ and $U_y \ni y$.

Proposition 6.1. *Let X be a compact Hausdorff space. Then for every $\varphi \in C(X)^*$ there exist positive linear functionals $\varphi^+, \varphi^- \in C(X)$ such that $\varphi = \varphi^+ - \varphi-$.*

Proof. For every $f \geq 0$ define

$$\varphi^+(f) = \sup\{\varphi(g) : g \in C(X), 0 \leq g \leq f\}.$$

Note that for $f \geq 0$, $0 \leq \varphi^+(f) \leq \left\|\varphi\right\| \cdot \|f\|$. The first inequality follows from the fact that $\varphi(0) = 0$, and the second from $|\varphi(g)| \leq \left\|\varphi\right\| \cdot \|g\| \leq \left\|\varphi\right\| \cdot \|f\|$.

We claim that for $f_1, f_2 \geq 0$ and $c \geq 0$ it holds that $\varphi^+(cf_1 + f_2) = c\varphi^+(f_1) + \varphi^+(f_2)$. Homogeneity and subadditivity follow immediately from the definition. For superadditivity, suppose $0 \leq g \leq f_1 + f_2$. Let $g_1 = \min(g, f_1)$, so that $0 \leq g_1 \leq f_1$, and let $g_2 = g - g_1$, so that $0 \leq g_2 \leq f_2$ and $g_1 + g_2 = g$. Then

$$\varphi(g) = \varphi(g_1) + \varphi(g_2) \leq \varphi^+(f_1) + \varphi^+(f_2)$$

Taking the supremum over all such g yields that

$$\varphi^+(f_1 + f_2) \leq \varphi^+(f_1) + \varphi^+(f_2)$$

Now, given any $f \in C(X)$, let $f = f^+ - f^-$, for $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$, so that $f^+, f^- \geq 0$. Define $\varphi^+(f) = \varphi^+(f^+) - \varphi^+(f^-)$, so that φ^+ is a positive linear functional on $C(X)$, and furthermore $\left\|\varphi^+\right\| \leq \left\|\varphi\right\|$. If we let $\varphi^- = \varphi^+ - \varphi$ then it follows that φ^- is also a positive linear functional and $\varphi = \varphi^+ - \varphi^-$. ☐

This result, which is a kind of Jordan decomposition for linear functionals, implies that to understand $C(X)^*$ it suffices to understand the positive linear functionals.

The **Cantor space** is $\mathscr{X} = \{0,1\}^{\mathbb{N}}$, endowed with the product topology, or the topology of pointwise convergence. It is a compact metric space (see Exercise 1 in §7). This topology has a number of nice properties which we will not prove:

Proposition 6.2. *The following holds for the Cantor space:*

1. *The clopen sets form a countable basis for the topology.*

2. *A countable disjoint union of clopen sets in $\mathscr{X}$ is clopen (if and) only if it is finite.*

3. *The simple functions are dense in $C(\mathscr{X})$.*

To understand (3), note that $f \in C(\mathscr{X})$ is simple if there is some n such that for every x it holds that $f(x)$ is determined by $x(1),\dots,x(n)$. The result then follows from the fact that a continuous map on a compact set is uniformly continuous.

Let μ be a finite measure on $\mathscr{X}$, endowed with the Borel σ-algebra. Then

$$\varphi_\mu(f) = \int f \, \mathrm{d}\mu \tag{6.1}$$

is a bounded positive linear functional on $C(\mathscr{X})$, with $\|\varphi\| = \mu(\mathscr{X})$. Note also that $\varphi_{\mu+\nu} = \varphi_\mu + \varphi_\nu$. The next theorem shows that all positive linear functionals are of this form.

Theorem 6.3 (Riesz Representation Theorem for the Cantor Space)**.** *For every positive $\varphi \in C(\mathscr{X})^*$ there is a finite Borel measure μ on $\mathscr{X}$ such that $\varphi = \varphi_\mu$.*

Proof. Let $\mathscr{A}$ be the algebra of clopen sets on $\mathscr{X}$ (see Exercise 1 in §7). Define $\rho\colon \mathscr{A} \to [0,\infty)$ as follows. Given $A \subset \{-1,+1\}^N$ and

$$U = \{\omega \in \mathscr{X} : (\omega(1),\dots,\omega(N)) \in A\},$$

let $\rho(U) = \varphi(\mathbb{1}_U)$. Then ρ is finitely additive because φ is additive, and ρ is furthermore (vacuously) countably additive because a countable disjoint union of clopen sets in $\mathscr{X}$ is clopen (if and) only if it is finite. Hence ρ is a premeasure. By Carathéodory's Theorem it follows that we can extend ρ to a measure on the Borel σ-algebra of $\mathscr{X}$. This measure is finite because $\rho(\mathscr{X}) = \varphi(\mathbb{1}_X) < \infty$.

Now, for simple continuous functions s it follows from the definition of ρ that $\int s \, \mathrm{d}\mu = \varphi(s)$. Since these are dense, it follows from the continuity of φ and of integration that $\int f \, \mathrm{d}\mu = \varphi(f)$ for all $f \in C(\mathscr{X})$. □

It can be shown that if $\mu \neq \nu$ then $\varphi_\mu \neq \varphi_\nu$. It follows that the map $\mu \mapsto \varphi_\mu$ from the bounded signed measures to $C(\mathscr{X})^*$ is a bijection, and moreover is an isometry, when the bounded signed measures are equipped with the total variation norm.

6.2 More Riesz Representation Theorems, and regular measures

The Cantor space is in fact a very special space, as the following result indicates:

Theorem 6.4. *For every compact metric space Y there exists a continuous surjection $\pi\colon \mathscr{X} \to Y$.*

We will not prove this theorem, but the idea is this. Suppose that Y has the following property: there exists a sequence $r_1, r_2, \ldots > 0$ such that $\lim_n r_n = 0$, the radius of Y is r_1, and any closed ball of radius r_n is contained in the union of two closed balls of radius r_{n+1}. This holds, for example, for $Y = [0,1]$ and $r_n = 0.5^n$. Now, for any n and any $x \in \{0,1\}^n$ we will construct a ball B_x as follows. Let $B = Y$. Given B_x, let B_{x0} and B_{x1} be two closed balls of radius r_{n+1} that cover B_x. Then the map $\pi \colon X \to Y$ takes x to the singleton at the intersection $B_{x(1)} \cap B_{x(1)x(2)} \cap B_{x(1)x(2)x(3)} \cdots$.

Let Y be a compact topological space. Then every bounded Borel measure μ defines a bounded positive linear functional by (6.1). For compact metric spaces we can prove the same result, using what we have already shown.

Theorem 6.5 (Riesz Representation Theorem for Compact Metric Spaces)**.** *Let Y be a compact metric space. For every $\varphi \in C(Y)^*$ there is a bounded signed Borel measure μ on Y such that $\varphi = \varphi_\mu$.*

Proof. Let $\pi \colon \mathscr{X} \to Y$ be a continuous surjection, which exists by Theorem 6.4. Let $\pi^* : C(Y) \to C(\mathscr{X})$ be given by $\pi^*(f) = f \circ \pi$. Then π^* is an injection, since if $f(y) \neq g(y)$ then $[\pi^*(f)](x) \neq [\pi^*(g)](x)$ for any $x \in \pi^{-1}(y)$. Moreover it is an isometry, and so we can identify $C(Y)$ with its image $\pi^*(C(Y))$, which is a subspace of $C(\mathscr{X})$.

Given a linear functional $\varphi \in C(Y)^*$, we can define a linear functional φ^* on $\pi^*(C(Y))$ by $\varphi^*(\pi^*(f)) = \varphi(f)$. By Hahn-Banach (Theorem 4.6) we can extend φ^* to a linear functional on $C(\mathscr{X})$, which by Theorem 6.3 is equal to some φ_{μ^*} for some bounded signed measure μ^* on $\mathscr{X}$.

Finally, let $\mu = \pi_*(\mu^*)$ be the signed pushforward measure on Y given by $\mu(A) = \mu^*(\pi^{-1}(A))$. Then

$$\varphi_\mu(f) = \int f \, \mathrm{d}\mu = \int \pi^*(f) \, \mathrm{d}\mu^* = \varphi_{\mu^*}(\pi^*(f)) = \varphi^*(\pi^*(f)) = \varphi(f).$$

□

Note that given φ, there is a unique bounded signed measure μ such that $\varphi = \varphi_\mu$. We will not prove this.

A space X is said to be **locally compact** if for every $x \in X$ there is a compact $K \subseteq X$ such that x is in the interior of K. Equivalently, for every $x \in X$ and open $U \ni x$ there is a compact $K \subseteq U$ such that x is in the interior of K.

Examples of locally compact spaces include all compact spaces, $\mathbb{R}^d$, and $\mathrm{GL}(n)$, the space of all invertible n by n matrices. A non-example is ℓ^2, and likewise every infinite dimensional Hausdorff topological vector space.

Denote by $C_c(X)$ the linear space of all real continuous functions on X with compact support. We endow it with the (uniform/sup) norm $\|f\| = \max_{x \in X} |f(x)|$. Note that this is not a Banach space, unless X is compact.

Given a measure μ on $(X, \mathscr{B}_X)$, we would like to define a positive linear functional φ_μ as above by

$$\varphi_\mu(f) = \int f \, \mathrm{d}\mu.$$

However, this might not be finite. To ensure that it is finite, we need to require that $\mu(\operatorname{supp} f) < \infty$. This motivates the first part of the following definition.

A measure μ on $(X, \mathscr{B}_X)$ is said to be **Radon** if

1. $\mu(K) < \infty$ for all compact $K \subseteq X$.

2.
$$\mu(A) = \inf\{\mu(U) : \text{open } U \text{ s.t. } A \subseteq U\} \tag{6.2}$$
for all $A \in \mathscr{B}_X$.

3.
$$\mu(A) = \sup\{\mu(K) : \text{compact } K \text{ s.t. } K \subseteq A\} \tag{6.3}$$
for all open $A \subset X$.

Theorem 6.6 (Riesz Representation Theorem for locally compact spaces)**.** *Let X be a locally compact Hausdorff topological space. For every positive linear functional φ on $C_c(X)$ there is a (unique) Radon measure μ such that $\varphi = \varphi_\mu$.*

We say that a measure is **regular** if (6.2) and (6.3) hold for all Borel sets. It turns out that every σ-finite Radon measure is regular.

The next result shows that for "nice" locally compact Hausdorff spaces the Radon measures have a simple characterization.

Proposition 6.7. *Suppose that X is a locally compact Hausdorff space whose topology has a countable basis (i.e., X is second countable). Then μ is Radon iff $\mu(K) < \infty$ for all compact K.*

6.3 Dual of C_0 (by Shreshth Srivastava and Caden Mikkelsen)

We begin with some preliminaries. Fix an LCH space X. We say a measure μ is a **signed Radon measure** if μ is a signed measure whose positive and negative variations are Radon. Define $M(X)$ to be the set of finite signed Radon measures on X. For $\mu \in M(X)$, define $||\mu|| = |\mu|(X)$. One can show that with the total variation norm, $M(X)$ is a normed vector space.

Recall that a continuous function f is said to vanish at infinity if for all $\varepsilon > 0$, the set $\{x : |f(x)| \geq \varepsilon\}$ is compact. For example, $\mathbb{1}_{[0,1]}$ vanishes at infinity on $[0,1]$ and $g(x) = \frac{1}{e^{x^2}}$ vanishes at infinity of $\mathbb{R}$ and when X is compact, any continuous function vanishes at infinity. On the other hand, $h(x) = x$ and $j(x) = e^x$ do not vanish at infinity on $\mathbb{R}$.

Recall that $C_0(X)$ denotes the space of continuous functions $f : X \to \mathbb{R}$ that vanish at infinity (with the uniform norm), and $C_0(X)^*$ is the space of bounded linear functionals from $C_0(X)$ to $\mathbb{R}$ (with the operator norm).

We also must provide a couple of helpful results which we won't prove here:

Theorem 6.8 (BLT Theorem)**.** *Let X be a normed vector space, Y be a Banach space, and A is a dense linear subspace of X. If $T : A \to Z$ is a bounded linear map, then T can be extended uniquely and continuously to a linear map $T' : X \to Y$.*

Proposition 6.9. *Let ν be a finite measure on $(X, \mathcal{M})$. Then*

(a) $\mathrm{d}\nu/\mathrm{d}|\nu|$ *has absolute value 1 $|\nu|$-a.e.*

(b) if $f \in L^1(\nu)$ then $\left|\int f \mathrm{d}\nu\right| \leq \int |f| \mathrm{d}|\nu|$

Theorem 6.10 (Lusin's Theorem)**.** *Let μ be a Radon measure on an LCH space X, and let $h : X \to \mathbb{R}$ be a measurable function that vanishes outside of a set of finite measure. For any $\varepsilon > 0$, there exists $f \in C_c(X)$ such that $f = h$ except on a set of measure $< \varepsilon$. Further, if h is bounded, f can be taken to satisfy $\|f\|_u \leq \|h\|_u$ where $\|\cdot\|_u$ is the uniform norm.*

Theorem 6.11 (Riesz Representation Theorem)**.** *Let X be an LCH space, and for $\mu \in M(X)$ and $f \in C_0(X)$ let $I_\mu(f) = \int f \mathrm{d}\mu$ ($I_\mu : C_0(X) \to \mathbb{R}$). Then the map $\mu \mapsto I_\mu$ is an isometric isomorphism from $M(X)$ to $C_0(X)^*$.*

Proof. Let $I : M(X) \to C_0(X)^*$ be our map, i.e. $I(\mu) = I_\mu$. It's clear that I is a vector space homomorphism. We now show that I is injective and an isometry.
If $\mu \in \ker I$, then $I(f) = I_\mu(f) = 0$ for all $f \in C_0(X)$. Pick any set $A \subseteq X$. We show $\mu(A) < \varepsilon$ for any ε. Let ε be given, and let $f \in C_c(X)$ be the f provided by Lusin's theorem with $h = \mathbb{1}_A$ (using $\frac{\varepsilon}{2}$). Let E be the set on which $f, \mathbb{1}_A$ differ. Then, we have $\mu(A) = \int_X \mathbb{1}_A \mathrm{d}\mu = \int_X \mathbb{1}_A \mathrm{d}\mu - \int_X f \mathrm{d}\mu = \int_X \mathbb{1}_A - f \mathrm{d}\mu = \int_E \mathbb{1}_A - f \mathrm{d}\mu \leq 2 \cdot \mu(E) < \varepsilon$ (where we use the fact that since $\|f\|_u \leq \|h\|_u = 1$, we have $\sup_{x \in E}(\mathbb{1}_A - f)(x) \leq 2$). Thus, μ is the zero measure, implying that I is injective.
We now prove $||\mu|| = ||I_\mu||$, implying that I is an isometry.
First, we show $||I_\mu|| \leq ||\mu||$. To do this, we argue that $\left|I_\mu(f)\right| \leq ||f||_u ||\mu||$. We have that

$$\begin{aligned} \left|I_\mu(f)\right| &= \left|\int_X f \mathrm{d}\mu\right| \\ &\leq \int_X |f| \, d\left|\mu\right| \quad \text{by } \textbf{Proposition 6.9(b)} \\ &\leq ||f||_u \cdot \left|\mu\right|(X) \\ &= ||f||_u ||\mu|| \end{aligned}$$

Recall that $||I_\mu|| = \sup\{\left|I_\mu(f)\right| : ||f||_u = 1\}$. For any f such that $||f||_u = 1$, we have that $\left|I_\mu(f)\right| \leq ||f||_u ||\mu|| = ||\mu||$. We conclude that $||I_\mu|| \leq ||\mu||$.
Next, we argue that $||\mu|| \leq ||I_\mu||$. Let $h = \mathrm{d}\mu/\mathrm{d}\left|\mu\right|$. Note that $|h| = 1$ and $\mathrm{d}\mu = h \mathrm{d}\left|\mu\right|$. We'll prove that for any $\varepsilon > 0$, $||\mu|| \leq ||I_\mu(f)|| + \varepsilon$. Take $\varepsilon > 0$. By Lusin's theorem, there is some $f \in C_c(X)$ such that $h = f$ except on a set E of measure $< \frac{\varepsilon}{2}$, and $||f||_u = ||h||_u = 1$. Then, we

have

$$
\begin{aligned}
||\mu|| &= \left|\mu\right|(X) \\
&= \int_X 1 \,\mathrm{d}\left|\mu\right| \\
&= \int_X |h|^2 \,\mathrm{d}\left|\mu\right| \\
&= \int_X h^2 \,\mathrm{d}\left|\mu\right| \\
&= \int_X h \cdot h \,\mathrm{d}\left|\mu\right| \\
&= \int_X h \,\mathrm{d}\mu \quad \text{since } \mathrm{d}\mu = h\mathrm{d}\left|\mu\right| \\
&\leq \left|\int_X h \,\mathrm{d}\mu\right| \\
&= \left|\int_X h - f + f \,\mathrm{d}\mu\right| \\
&\leq \left|\int_X h - f \,\mathrm{d}\mu\right| + \left|\int_X f \,\mathrm{d}\mu\right| \\
&\leq \int_E |h| + |f| \,\mathrm{d}\left|\mu\right| + \left|\int_X f \,\mathrm{d}\mu\right| \quad \text{since } h - f \text{ is 0 on } X \setminus E \\
&\leq \int_E 2 \,\mathrm{d}\left|\mu\right| + \left|\int_X f \mathrm{d}\mu\right| \\
&= 2 \cdot \left|\mu\right|(E) + \left|\int_X f \mathrm{d}\mu\right| \\
&< \varepsilon + |I_\mu(f)| \\
&\leq \varepsilon + ||I_\mu||
\end{aligned}
$$

We conclude that our map is an isometry.
We now show that I is surjective. Consider any $I = I^+ - I^- \in C_0(X)^*$. Taking the restriction of the positive and negative variations to $C_c(X)^*$, we can apply the Riesz Representation Theorem for LCH spaces to show that $I^\pm(f)\,|_{C_c(X)} = \int f \,\mathrm{d}\mu^\pm$ for some Radon measures $\mu^\pm$. Namely, $I(f)\,|_{C_c(X)} = \int f \,\mathrm{d}(\mu^+ - \mu^-) = \int f \,\mathrm{d}\mu$ for some $\mu \in M(X)$. By the Bounded Linear Transformation Theorem, this can be continuously extended uniquely to I_μ iff $I(f)\,|_{C_c(X)}$ is bounded, which follows because μ is finite. Thus, I is surjective. ☐

From this result, we immediately get the following corollary.

Corollary 6.12. *If X is a compact Hausdorff space, then $C(X)^*$ is isometrically isomorphic to $M(X)$.*

In fact, in this case $M(X)$ is the space of finite regular Borel measures on X because if X is $\sigma-$compact, all Radon measures are regular, and clearly all finite regular measures are

Radon.

Example: Consider $C[a,b] = C_0[a,b]$, the space of continuous functions on the compact set $[a,b] \subseteq \mathbb{R}$. Given some $I \in C[a,b]^*$, we know from the Riesz Representation Theorem that there exists a finite, regular Borel measure μ such that $I(f) = \int_a^b f\,\mathrm{d}\mu$. Consider the measure's cumulative distribution function $F_\mu(t) = \mu[a,t]$. We can rewrite our functional

$$I(f) = \int_a^b f\,\mathrm{d}\mu = -f(a)\mu\{a\} + \int_a^b f\,\mathrm{d}F_\mu = \int_a^b f\,\mathrm{d}(F_\mu + G_{\mu\{a\}})$$

where $G_{\mu\{a\}} = -\chi_{\mu\{a\}}\mu\{a\}$ and the latter two expressions use the Lebesgue-Stieltjes integral. In this way, we can identify each $I \in C[a,b]^*$ with some right-continuous function F of bounded variation such that $F(a) = 0$. This sounds quite similar to $\mathrm{NBV}[a,b]$, and in fact, it can be shown that under the variation norm (norm of a function is its total variation) $C[a,b]^* \cong \mathrm{NBV}[a,b]$.

As a final application of the Riesz Representation Theorem, we define a topology on $M(X)$ called the **vague topology**. We identify $M(X)$ with $C_0(X)^*$ and define the vague topology on $M(X)$ to be the weak* topology on $C_0(X)^*$. Namely, $\mu_n \to \mu$ if and only if $\int f\,\mathrm{d}\mu_n \to \int f\,\mathrm{d}\mu$ for all $f \in C_0(X)^*$. There is a rather convenient way to determine convergence in $M(\mathbb{R})$ under this topology:

Proposition 6.13. *Suppose $\mu, \mu_1, \mu_2, \cdots \in M(\mathbb{R})$, and let $F_n(x) = \mu_n((-\infty, x])$ and $F(x) = \mu((-\infty, x])$. If $\sup_n \|\mu_n\| < \infty$ and $F_n \to F$ at every x for which F is continuous, then $\mu_n \to \mu$ vaguely. Conversely, if $\mu_n \to \mu$ vaguely, then $\sup_n \|\mu_n\| < \infty$, and if the μ_n are positive, then $F_n \to F$ at every x for which F is continuous.*

Proof. For the first statement, by Theorem 3.23, we know that F has at most countably many discontinuities, meaning $F_n \to F$ a.e. For any continuously differentiable f with compact support, by integration by parts, we see

$$\int f\,\mathrm{d}\mu_n = \int f'(x)F_n(x)\mathrm{d}x \to \int f'(x)F(x)\mathrm{d}x = \int f\,\mathrm{d}\mu$$

where the integrals converge because $F_n \to F$ almost everywhere. This shows the result for all $f \in C_0(\mathbb{R})$ such that f is continuously differentiable with compact support. We state without proof that these f are dense in $C_0(\mathbb{R})$, so the integrals must converge everywhere in $C_0(\mathbb{R})$.

For the opposite direction, if the $\mu_n \to \mu$ vaguely, then for all $f \in C_0(\mathbb{R})$, $\int f_n\,\mathrm{d}\mu \to \int f\,\mathrm{d}\mu$. $C_0(\mathbb{R})$ is a Banach space, so by the uniform boundedness principle and the Riesz Representation Theorem, $\sup_n \|\mu_n\| = \sup_n \|I_{\mu_n}\| < \infty$. If each μ_n is positive, then clearly μ is positive, Let $F(x)$ be continuous at $x = a$, $N \in \mathbb{N}$, and define f to be 1 on $[-N, a]$, 0 on $(\infty, -N-\varepsilon] \cup [a+\varepsilon, \infty)$, and linear between. We see that

$$F_n(a) - F_n(-N) = \mu_n((-N,a]) \le \int_{\mathbb{R}} f\,\mathrm{d}\mu_n \to \int_{\mathbb{R}} f\,\mathrm{d}\mu \le F(a+\varepsilon) - F(-N-\varepsilon).$$

As $N \to \infty$, we see $F_n(-N) = \mu_n((-\infty, -N]) \to 0$, and $F(-N-\varepsilon) = \mu((-\infty, -N-\varepsilon]) \to 0$. N, n, ε are arbitrary, so $\limsup_{n\to\infty} F_n(a) \le F(a)$. If f is instead defined to be 1 on $[-N+\varepsilon, a-\varepsilon]$, 0 on $(-\infty, N] \cup [a, \infty)$ and linear between, we see

$$F_n(a-\varepsilon) - F_n(-N+\varepsilon) = \mu_n((-N+\varepsilon, a-\varepsilon]) \ge \int_{\mathbb{R}} f \, \mathrm{d}\mu_n \to \int_{\mathbb{R}} f \, \mathrm{d}\mu \ge F(a) - F(-N).$$

Thus, $\liminf_{n\to\infty} F_n(a) \le F(a)$, so $F_n(a) \to F(a)$. $\square$

7 Exercises

1. **Independent fair coin tosses.** Measures are important in probability, where they model the chance of uncertain outcomes. Probability measures are measures that assign unit measure to the entire space.

   For example, let $\Omega = \{-1,+1\}^N$ for some $N \in \mathbb{N}$. We think of elements of Ω as functions $\omega\colon \{1,\ldots,N\} \to \{-1,+1\}$. The i.i.d. (independent and identically distributed) Bernoulli measure $\mu\colon 2^\Omega \to [0,1]$ is given by $\mu(A) = 2^{-N}|A|$. To see why this captures the idea of independent tosses, define for $i \le N$ the map $T_i : \Omega \to \Omega$ by

   $$[T_i(\omega)](j) = \begin{cases} \omega(j) & \text{if } j \neq i \\ -\omega(j) & \text{if } j = i. \end{cases}$$

   The measure μ is T_i-invariant: $\mu(T_i(A)) = \mu(A)$ for all $A \subseteq \Omega$, i.e., the probability of an outcome does not change if we consider the outcome in which the ith coin has the opposite sign.

   Let $\mathscr{X} = \{-1,+1\}^{\mathbb{N}}$. We think of elements of $\mathscr{X}$ as functions $\omega\colon \mathbb{N} \to \{-1,+1\}$. We endow $\mathscr{X}$ with the topology of pointwise convergence: $\lim_n \omega_n = \omega$ if for all $i \in \mathbb{N}$ it holds that $\lim_n \omega_n(i) = \omega(i)$ (note that this implies that the sequence $\omega_n(i)$ eventually stabilizes on $\omega(i)$). With this topology, $\mathscr{X}$ is also known as the Cantor space.

   (a) Prove that this topology is also the topology generated by the metric $D : \mathscr{X} \times \mathscr{X} \to \mathbb{R}_{\geq 0}$ given by

   $$D(\omega,\theta) = \inf\{2^{-N} : (\omega(1),\ldots,\omega(N)) = (\theta(1),\ldots,\theta(N))\}.$$

   (b) Prove that $U \subseteq \mathscr{X}$ is clopen (both closed and open) if and only if there is some $N \in \mathbb{N}$ and $A \subseteq \{-1,+1\}^N$ such that

   $$U = \{\omega \in \mathscr{X} : (\omega(1),\ldots,\omega(N)) \in A\}.$$

   (c) Prove that $U \subset \mathscr{X}$ is open if and only if it is a countable union of clopen sets.

   (d) We would like to extend our i.i.d. measure from the finite setting to the infinite setting. I.e., we would like to find a probability measure $\mu\colon 2^{\mathscr{X}} \to [0,1]$ such that $\mu(T_i(A)) = \mu(A)$ for all $A \subseteq \mathscr{X}$ and $i \in \mathbb{N}$. Prove that this is impossible.

   (e) Let $\mathscr{A}$ be the collection of clopen sets of $\mathscr{X}$. Prove that it is an algebra.

   (f) Define $\rho\colon \mathscr{A} \to [0,1]$ as follows. Given $A \subset \{-1,+1\}^N$ and

   $$U = \{\omega \in \mathscr{X} : (\omega(1),\ldots,\omega(N)) \in A\},$$

   let $\rho(U) = 2^{-N}|A|$. Show that ρ is a premeasure.

   (g) Use Carathéodory's Theorem to show that there is a probability measure on the Borel σ-algebra of $\mathscr{X}$ that is invariant to $\{T_i\}_{i\in\mathbb{N}}$.

2. **Countable additivity from dominance.** Let (Ω, Σ) be a measurable space, and let $\mu\colon \Sigma \to [0,\infty]$ be a measure. We say that μ dominates a map $\nu\colon \Sigma \to [0,\infty]$ if $\mu(A) \geq \nu(A)$ for all $A \in \Sigma$.

   Show that if μ is a finite measure (i.e., $\mu(\Omega) < \infty$), ν is a finitely additive measure ν, and μ dominates ν, then ν is in fact a measure.

3. **Criticality of σ-finite hypothesis to Theorem 1.9.** Let $\mathscr{A}$ be the collection of all subsets in $\mathbb{R}$ that can be expressed as finite unions of half-open intervals $[a,b)$. Let $\mu_0 : \mathscr{A} \to [0,\infty]$ be the function such that $\mu_0(E) = \infty$ if $E \neq \emptyset$ and $\mu_0(\emptyset) = 0$.

   (a) Show that μ_0 is a premeasure.

   (b) Show that $\mathscr{M}(\mathscr{A})$ is the Borel σ-algebra $\mathscr{B}_{\mathbb{R}}$.

   (c) Show that the extension $\mu : \mathscr{B}_{\mathbb{R}} \to [0,\infty]$ of μ_0 defined by 1.1 and Theorem 1.8 assigns an infinite measure to any non-empty Borel set.

   (d) Show that the counting measure $\mu_c(A) = |A|$ is another extension of μ_0 on $\mathscr{B}_{\mathbb{R}}$.

4. **Completing the measure extension.** Let $\mu_0 : \mathscr{A} \to [0,\infty]$ be a premeasure which is σ-finite and let $\mu : \Sigma \to [0,\infty]$ be its (unique!) extension on the σ-algebra Σ of μ^*-measureable sets (recall that Σ may be even larger than $\mathscr{M}(\mathscr{A})$, the σ-algebra generated by $\mathscr{A}$).

   (a) Show that if $E \in \Sigma$, then there is an $F \in \mathscr{M}(\mathscr{A})$ containing E such that $\mu(F \setminus E) = 0$ (thus F consists of the union of E and a null set). Furthermore, show that F can be chosen to be a countable intersection $F = \bigcap_{j=1}^{\infty} F_j$ of sets F_j, each of which is a countable union $F_j = \bigcup_{k=1}^{\infty} F_{j,k}$ of sets $F_{j,k} \in \mathscr{A}$ (i.e., F is an element of $\mathscr{A}_{\sigma\delta}$).

   (b) If $E \in \Sigma$ has finite measure, and $\varepsilon > 0$, show that there exists $F \in \mathscr{A}$ such that $\mu(E \Delta F) \leq \epsilon$, where

   $$E \Delta F := (E \cup F) \setminus (E \cap F) = (E \setminus F) \cup (F \setminus E)$$

   is the symmetric difference of E and F.

   (c) Conversely, if E is a set such that for every $\varepsilon > 0$ there exists $F \in \mathscr{A}$ such that $\mu^*(E \Delta F) \leq \varepsilon$, show that $E \in \Sigma$.

5. **Axioms of integration with countable additivity.** Let (Ω, Σ, μ) be a σ-finite measure space. Recall that L^+ is the set of measurable functions from Ω to $\mathbb{R}_{\geq 0}$.

   (a) Suppose that $\Phi\colon L^+ \to [0,\infty]$ satisfies the following axioms of integration for $f, f_1, f_2, \ldots \in L^+$:

   - **Calibration.** If $f = \mathbb{1}_A$ then $\Phi(f) = \mu(A)$.
   - **Homogeneity.** $\Phi(\lambda \cdot f) = \lambda \Phi(f)$ for all $\lambda \geq 0$.
   - **Countable additivity.** $\Phi(\sum_n f_n) = \sum_n \Phi(f_n)$ whenever $\sum_n f_n(\omega) < \infty$ for all ω.

Show that $\Phi(f) = \int f \, \mathrm{d}\mu$. You can use any result stated in these lecture notes, except Theorem 2.3.

(b) Give an example of $\Psi : L^+ \to [0,\infty]$ that satisfies calibration, homogeneity and (finite) additivity such that $\Psi(f) \neq \int f \, \mathrm{d}\mu$ for some $f \in L^+$.

6. **The Fundamental Theorem of Calculus.** Let (Ω, Σ, μ) be a measure space, and consider $f \in L^1$, $f \geq 0$.

   (a) Show that $\mu_f : \Sigma \to [0,\infty]$, $\mu_f(A) = \int_A f \, \mathrm{d}\mu$ is a finite measure.

   (b) Consider the case that μ is the Lebesgue measure on $\mathbb{R}$ and f is continuous. Let $F(x) = \mu_f((-\infty, x])$ and prove that $\frac{\mathrm{d}F}{\mathrm{d}x}(x) = f(x)$.

7. **Uniform Convergence From Pointwise Convergence.** Let $\mu \colon \Sigma \to [0,1]$ be a σ-finite measure. Show that if $f_n \to f$ pointwise a.e. then there exist $A_1, A_2, \ldots \subseteq \Omega$ such that $A_n \subseteq A_{n+1}$, $\cup_n A_n$ is co-null, and on each A_n convergence is uniform.

8. **Comparison of Measures.** Let (Ω, Σ) be a measurable space, and let μ, η be finite measures on it.

   (a) Prove or disprove: If μ and η are equivalent (i.e., have the same null sets) then $f_n \to f$ in measure according to μ iff $f_n \to f$ in measure according to η.

   (b) Show that the following are equivalent:

   - There exists some $C > 0$ such that for all $A \in \Sigma$ it holds that $\mu(A) \leq C\eta(A)$ and $\eta(A) \leq C\mu(A)$.
   - $L_1(\Omega, \Sigma, \mu) = L_1(\Omega, \Sigma, \eta)$.

9. **The Area under a Function.** Let $(\mathbb{R}, \mathscr{B}_{\mathbb{R}}, \lambda)$ be the Borel σ-algebra on $\mathbb{R}$ equipped with the Lebesgue measure. Let $\eta = \lambda \times \lambda$, the product measure defined on $(\mathbb{R}^2, \mathscr{B}_{\mathbb{R}} \otimes \mathscr{B}_{\mathbb{R}})$.

   Let $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be measurable. Then a natural notion of the "areas under f" is

   $$\eta\left(\{(x,y) \in \mathbb{R}^2 : 0 \leq y \leq f(x)\}\right).$$

   Show that this notion coincides with integration, our previous notion of the "area under f":

   $$\eta\left(\{(x,y) \in \mathbb{R}^2 : 0 \leq y \leq f(x)\}\right) = \int f \, \mathrm{d}\lambda.$$

10. **The Cantor space.** Let $\mathscr{X} = \{0,2\}^{\mathbb{N}}$, endowed with the topology of pointwise convergence. As in exercise 1, let μ be the probability measure on $(\mathscr{X}, \mathscr{B})$ such that for every $A \subseteq \{0,2\}^N$ it holds that

    $$\mu(\{\omega \in \mathscr{X} : (\omega(1), \ldots, \omega(N)) \in A\}) = 2^{-N}|A|.$$

Let $\pi\colon \mathscr{X} \to \mathbb{R}$ be given by

$$\pi(\omega) = \sum_{n=1}^{\infty} 3^{-n}\omega(n).$$

Let $\nu = \pi_*\mu$. I.e., $\nu(A) = \mu(\pi^{-1}(A))$. Equivalently,

$$\nu((-\infty, b]) = \mu(\{\omega : \pi(\omega) \le b\}).$$

(a) Show that ν is non-atomic, i.e., $\nu(\{x\}) = 0$ for all $x \in \mathbb{R}$. Equivalently, $b \mapsto \nu((-\infty, b])$ is continuous.

(b) Show that ν and the Lebesgue measure are mutually singular.

(c) Prove that there exists a continuous increasing function $F\colon \mathbb{R} \to \mathbb{R}$ such that $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} = 1$, and F is a.e. differentiable with $F' = 0$.

11. **Total variation of differentiable functions.** Let $F\colon [a,b] \to \mathbb{R}$ be continuously differentiable. Show that $T_F[a,b] = \int_a^b |F'(x)|\,\mathrm{d}x$.

12. **The vector space of finite signed measures.** Let V be the vector space of finite signed measures on $(\mathbb{R}, \mathscr{B})$, endowed with the total variation norm $\|\eta\|_{\mathrm{TV}} = |\eta|(\mathbb{R})$. Prove or disprove that V is a Banach space.

13. **The averaging operator.** Recall that given $f \in L^1(\mathbb{R}^d, \mathscr{B}, \lambda)$, we defined

$$A_r^f(x_0) = C_d r^{-d} \int_{B_r(x_0)} f\,\mathrm{d}x.$$

Let T be the map $f \mapsto A_r^f$. Prove or disprove that T is a bounded linear map $L^1 \to L^1$.

14. **Automatic continuity.** Let X be a compact topological space, and let $V = C(X)$ be the vector space of continuous functions $f\colon X \to \mathbb{R}$, equipped with the norm $\|f\|_\infty = \max_{x\in X} |f(x)|$.

We say that a map $\varphi\colon V \to \mathbb{R}$ is additive if $\varphi(f+g) = \varphi(f) + \varphi(g)$. We say that φ is monotone if $f \ge g$ implies $\varphi(f) \ge \varphi(g)$. Show that every monotone additive map is linear and continuous.

15. **Weak and strong convergence.** Let $\mathscr{H}$ be a Hilbert space and let $v_n \in \mathscr{H}$ converge weakly to a limit $v \in \mathscr{H}$ (i.e., $\lim_n \langle v_n, w\rangle = \langle v, w\rangle$ for all $w \in \mathscr{H}$). Prove or disprove that the following statements are equivalent:

- v_n converges strongly to v (i.e., $\lim_n \|v_n - v\| = 0$).
- $\lim_n \|v_n\| = \|v\|$.

16. **The union of ℓ^q for all q less than p.** Fix $q > 1$.

(a) Show that for any $1 \le p < q$ it holds that $\{f \in \ell^q : \|f\|_p \le n\}$ is a closed subset of ℓ^q that is nowhere dense. You can use Proposition 5.4.

(b) Using the Baire Category Theorem, deduce from this that there exists a $f \in \ell^q$ such that $f \notin \ell^p$ for any $p \in [1, q)$.

17. **Closed linear operators.** Let V, W be Banach spaces, and let $T : V \to W$ be a linear map. The graph of T is $\Gamma(T) = \{(v, w) \in V \times W : w = Tv\}$.

(a) Show that $\Gamma(T)$ is a linear subspace of $V \times W$.

(b) Show that if $\Gamma(T)$ is closed then T is continuous.

18. **A linear functional of L^∞.** Show that there exists a bounded linear functional φ on $L^\infty([0,1])$ such that $\varphi(f) = f(0)$ for every continuous f. You can use Theorem 4.6.

19. **Shift invariant means on ℓ^∞.** Let $V = \ell^\infty(\mathbb{N})$, i.e., V is the space of bounded functions from $\mathbb{N}$ to $\mathbb{R}$. Denote by $\sigma : \mathbb{N} \to \mathbb{N}$ the function $\sigma(z) = z + 1$, and let the shift operator $T : V \to V$ be given by $Tf = f \circ \sigma$. Denote by $1 \in V$ the constant function $z \mapsto 1$. Denote the set of **means** by

$$M = \{\varphi \in V^* : \varphi(1) = 1 \text{ and } \varphi(f) \ge 0 \text{ for all } f \ge 0\}.$$

(a) Show that T is a bounded linear operator.

(b) Show that M is nonempty and closed in the weak* topology.

(c) Show that there exists a shift-invariant mean: a $\varphi \in M$ such that $\varphi \circ T = \varphi$. You can use Theorem 4.28.

(d) Show that for any shift-invariant mean φ it holds that $\varphi(g) = 0$ for all finitely supported g, and that for all $f \in \ell^\infty$

$$\liminf_n f(n) \le \varphi(f) \le \limsup_n f(n).$$