Real Analysis (Lebesgue) and Basic Probability Theory

Adam Li,
Department of Applied Mathematics & Statistics,
Department of Biomedical Engineering,
Johns Hopkins University,
Baltimore, MD, 21218
ali39@jhu.edu / adam2392@gmail.com

November 11, 2020

Contents

1 Introduction

In these notes, I go over important results in Lebesgue measure theory, Lebesgue integration and then leading up to the theory related in Probability Measure Theory.

I start off with some important notes and results in constructing the Lebesgue measure, exploring the notion of measurability on subsets of the real line.

2 Lebesgue Measure

Lebesgue measure is a notion used to make Lebesgue integration possible, which is a generalization of the Riemann integral.

2.1 Lebesgue Outer (Exterior) Measure

This notion is essentially defined by taking a countable sequence of sets that cover the set of interest. This "approximates" the "volume" of the set from the outside.

What are some basic notions that lead up to our definition of the exterior measure? First, let us examine the problem of **covering**: deconstructing a set into a union of (almost) disjoint structured sets. These structured sets can be balls, cubes, or rectangles possibly. We start off with a basic lemma that allows us to interchange the volume of a rectangle with its covering of rectangles.

Lemma. *If a rectangle R is the almost disjoint union of finitely many other rectangles: $R = \bigcup_{k=1}^{N} R_k$, then the volume of R is equivalent to the summation of the volumes of the R_k's.*

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

What happens when you remove the assumption of **almost disjointness**? Then we have a finite sub-additivity condition.

Lemma. *If $R, R_1, ..., R_N$ are rectangles with $R \subset \bigcup_{k=1}^{N} R_k$, then:*

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

Now we have some understanding of how rectangles and their volumes behave when covered by either almost disjoint sets, or arbitrary sets. This next theorem tells us that every open subset in the real line can be written **uniquely** as a countable union of disjoint open intervals (i.e. rectangles).

Theorem. *Every open subset O of $\mathbb{R}$ can be written uniquely as a countable union of disjoint open intervals.*

This is an interesting result because it allows us to formulate any arbitrary set **uniquely** as a union of much simpler sets. If we extend to $\mathbb{R}^d$, we see however that now we do not have uniqueness, but we can still write any open subset in $\mathbb{R}^d$ as a countable union of **almost disjoint closed cubes**.

Theorem. *Every open subset O of $\mathbb{R}^d$ with $d \geq 1$, can be written as a countable union of almost disjoint closed cubes.*

Before defining the Lebesgue measure, we first start off by defining the **exterior** measure. That is the measure of a set by approximating it from the "outside".

Definition 2.1. If $E \subset \mathbb{R}^d$, the exterior measure of E is:

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

where the infimum is taken over all countable coverings by closed cubes in Theorem 2.1.

Properties of the outer measure:

1. Non-negative: Every closed cube volume has by definition non-negative volume.
2. Monotonic: If $E_1 \subset E_2$, then $m_*(E_1) \leq m_*(E_2)$.
3. Countable sub-additivity
4. Disjoint finite additivity: If $E = E_1 \cup E_2$ with $d(E_1, E_2) > 0$, then:

$$m_*(E) = m_*(E_1) + m_*(E_2)$$

5. Disjoint countable additivity

2.2 Invariance Properties of the Lebesgue Measure

We define the Lebesgue measure, m.

1. m is translation invariant.
2. m is dilation invariant.
3. m is reflection about the origin invariant

2.3 Lebesgue Measurable Functions

We start by defining the characteristic function of a set, E:

$$\chi_E(x) = \begin{cases} 1\text{if } x \in E \\ 0\text{else} \end{cases}$$

We can now remind ourselves of the Riemann integral, which is defined with rectangles.

$$f = \sum_{k=1}^{N} a_k \chi_{R_k}$$

where each R_k is a rectangle and the a_k are constants.

However, let's look at Lebesgue integration where the characteristic function operates on a set of measurable functions with finite measure.

$$f = \sum_{k=1}^{N} a_k \chi_{E_k}$$

where the E_k are measurable sets of finite measure.

2.3.1 Defining Measurability on Functions

Definition 2.2. A function f defined on a measurable subset E of $\mathbb{R}^d$ is measurable if for all $a \in \mathbb{R}$ the set formed by the inverse image:

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

is measurable.

Note that functions are measurable if their one-side bounded preimage is measurable. There are multiple definitions that are equivalent that we will not state here.

Here are some properties that measurable functions have:

Property (Finite valued preimage of open/closed sets)**.** *If f is a finite-valued function. Then it is measurable iff $f^{-1}(O)$ is measurable for every open set and iff $f^{-1}(F)$ is measurable for every closed set F.*

Property (Addition and multiplication)**.** *If f and g are finite-valued functions, then $f + g$ and fg are measurable.*

Property (Integer powers of measurable functions)**.** *If f is measurable, then f^k for $k \geq 1$ are measurable.*

Property (Continuity)**.** *If f is continuous on $\mathbb{R}^d$, then f is measurable.*

Property (Composition of continuous on measurable + finite valued function)**.** *If Φ is continuous and f is measurable and finite valued, then fog is measruable.*

Property (Sequence of measurable functions)**.** *For a sequence of measurable functions, their sup, inf, limsup, and liminf are measurable.*

Property (Limit of sequence of measurable functions)**.** *A limit of sequence of measurable functions is measurable.*

Property (Almost everywhere property)**.** *If f measurable and f(x) = g(x) for a.e. x. Then g is measurable.*

2.3.2 Approximation of Functions with Pointwise Functions

Here, we are interested in seeing that certain classes of functions can be approximated by a sequence of either non-negative or arbitrary point functions.

Approximating Non-negative Measurable Functions We start with the simplest set of functions we might approximate, and these are functions that are non-negative. Note then one can approximate arbitrary functions by approximating the function on the positive support and the function on the negative support.

Theorem. *Suppose f is a non-negative measurable function. Then there exists an increasing sequence of non-negative simple functions* $\{\psi_k\}_{k=1}^{\infty}$ *that converges pointwise to f.*

Monotonicaly increasing sequence:

$$\psi_k(x) \leq \psi_{k+1}(x)$$

Pointwise convergence:

$$\lim_{k\to\infty} \psi_k(x) = f(x), \forall x$$

Proof. We can first truncate the function based on its domain. For $N \geq 1$, Q_N is the cube centered at the origin with side length N. Then we define:

$$F_N(x) = \begin{cases} f(x) \text{if } x \in Q_N \text{and } f(x) \leq N \\ N \text{if } x \in Q_N \text{and } f(x) > N \\ 0 \text{else} \end{cases}$$

This provides us with a sequence of sets that converge to f for all x.

Now, if we also partition this truncated function over its range of [0, N] (remember non-negative function with the maximum value being the truncation). Then for fixed N, and $M \geq 1$, we can define the set:

$$E_{l,M} = \{x \in Q_N : \frac{l}{M} < F_N(x) \leq \frac{l+1}{M}\},\ for\ 0 \leq l \leq NM$$

This set partitions the range of the truncation function into segments of size $\frac{1}{M}$. So the sum over these partitions consists of $F_{N,M}(x)$.

$$F_{N,M}(x) = \sum_{l=1}^{NM} \frac{l}{M} \chi E_{l,M}(x)$$

which note is a linear combination of characteristic functions based on the set $E_{l,M}$.

This $F_{N,M}(x)$ is a simple function since it is the linear combination of characteristic functions. In addition, notice that:

$$0 \leq F_N(x) - F_{N,M}(x) \leq 1/M$$

for all x, since it is partitioned as such. If we now choose $\psi_k = F_{2^k,2^k}$, with N = M = 2^k, then $0 \leq F_N(x) - \psi_k(x) \leq 1/2^k$. These ψ_k are increasing functions and it converges pointwise to f. $\square$

Approximating Measurable Functions Note we can now easily just approximate general measurable functions decomposed into a positive and negative section.

2.3.3 Approximation of Functions with Step Functions

Instead of just simple functions, we can now show approximation using step functions. When using step functions, we only get convergence a.e.

Since we can only get convergence a.e., we simply have to show that f defined on a measurable set with finite measure can be approximated.

Theorem. *Suppose f is measurable on $\mathbb{R}^d$. Then there exists a sequence of step functions $\{\psi_k\}_{k=1}^{\infty}$ that converges pointwise to f(x) for a.e. x.*

Proof. It suffices to show that if E is a measurable set with finite measure, then $f = \chi E$ can be approximated by step functions.

$\square$

2.3.4 Approximation of Functions with Polynomial Functions

Here, we have the famouse Weierstrass theorem. We state the theorem here without going into the proof.

Theorem. *If f is a continuous function supported on a compact interval, $C \subset \mathbb{R}^d$, then given $\epsilon > 0$, we can find p such that:*

$$\sup |f(x) - p(x)| < \epsilon$$

where p(x) is a polynomial function.

2.3.5 Littlewood Principles

1. Every set is nearly a finite union of intervals.
2. Every function is nearly continuous
3. Every convergent sequence is nearly uniformly convergent

The first principle is shown in Theorem 3.4 of Stein, where if E is a measurable set that is finite, then there exists a finite union of closed cubes such that the symmetric differnece between E and F has measure arbitrarily small.

$$m((E-F)\cup(F-E))\leq\epsilon$$

for $F=\bigcup_{j=1}^{N}Q_j$ of closed cubes.

The following theorem gives us conditions for getting **continuity** on a measruable closed subset of a set of finite measure.

Theorem. *Suppose f is measurable and finite valued on E (E has finite measure). Then for every $\epsilon>0$, there exists a closed set $F_\epsilon\subset E$ with:*

$$m(E-F_\epsilon)\leq\epsilon$$

such that f on F_ϵ is continuous.

Proof. First, we construct a sequence of step functions that converge to f a.e. We know this by our approximation theorems earlier.

Then we can find sets E_n so that $m(E_n)<\frac{1}{2^n}$ and f_n is continuous outside E_n.

Since they are finite combination of characteristic functions of rectangles. We can establish continuity outside these sets by tapering off the value in a continuous way. The only area that has discontinuity is the transition from E_n to E_n^c, which we can make have measure zero.

By Egorov's theorem, we can then find a closed set $A_{\epsilon/3}$ such that $f_n\to f$ uniformly and $m(E-A_{\epsilon/3})\leq\epsilon/3$.

Then we can consider a set:

$$F'=A_{\epsilon/3}-\bigcup_{n\geq N}$$

where N is very large so that the sum $\sum_{n\geq N}\frac{1}{2^n}<\epsilon/3$.

Now every $n\geq N$, we have f_n is continuous on F', and so $f=\lim f_n$ is also continuous on F'.

Now, just approximate F' with a closed set F_ϵ that makes the measure of the differences arbitrarily small. □

The following gives conditions for getting **uniform convergence** on a measurable closed subset of a set of finite measure.

Theorem (Egorov's Theorem)**.** *Suppose $\{f_k\}$ is a sequence of measurable functions defined on a measurable set E with finite measure and that $f_k \to f$ a.e. on E.*

Then given $\epsilon > 0$, we can find a closed set $A_\epsilon \subset E$, such that $m(E - A_\epsilon) \leq \epsilon$ and $f_k \to f$ uniformly on A_ϵ.

**If we are willing to throw away an arbitrary small set, then we can get uniform convergence.*

Proof. WLOG, we can assume that $f_k(x) \to f(x)$ for every $x \in E$. Because if we do not have this for some x that has measure zero, then we just redefine $f_k(x)$ as f(x) on that subset of measure zero, hence getting our general situation.

For pairs of non-negative integers, n and k, we construct a set:

$$E_k^n = \{x \in E : |f_j(x) - f(x)| < 1/n, \text{for all j > k}\}$$

We have a $1/n$ error we are okay with and a tail of the sequence wrt f(x) we are interested in.

Given n, we now have that $E_k^n \subset E_{k+1}^n$ and that E_k^n converges to E as k tends to infinity.

By an earlier theorem, we have that $m(E_k^n) \to m(E)$ as $k \to \infty$, so we have that:

$$\lim_{k \to \infty} m(E - E_k^n) \to 0$$

So we can choose k_n such that $m(E - E_{k_n}^n < 2^{-n}$, since we have that the limit converges to 0.

Then we have whenever $j > k_n$ and $x \in E_{k_n}^n$:

$$|f_j(x) - f(x)| < 1/n$$

We now choose N » 1, such that $\sum_{n=N}^{\infty} 2^{-n} < \epsilon/2$, and we make a intersection of all these sets over $n \geq N$:

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

Now, we can compare E to this intersection of sets:

$$m(E - \tilde{A}_\epsilon) \leq \sum_{n=N}^{\infty} m(E - E_{k_n}^n) < \epsilon/2$$

Now if $\delta > 0$, choose $n \geq N$, such that $1/n < \delta$ and note $x \in \tilde{A}_\epsilon$ implies $x \in E_{k_n}^n$, since it is in the intersection of such sets. So then

$$|f_j(x) - f(x)| < 1/n < \delta$$

$\forall j > k_n$ for every $x \in \tilde{A}_\epsilon$.

In order to choose the closed subset, we can use Thm 3.4 to always get a closed subset of a measurable set is arbitrarily small. We can get: $A_\epsilon \subset \tilde{A}_\epsilon$ with $m(\tilde{A}_\epsilon - A_\epsilon) < \epsilon/2$. □

The Lebesgue measurable sets are Borel σ-fields with the added addition of subsets of Borel field of measure zero.

3 Lebesgue Integration

Lebesgue integration is a generalization of Riemann integration. Lebesgue integration can be established in increasing generality:

1. simple functions
2. bounded functions on support of finite measure
3. non-negative functions
4. integrable functions

3.1 Integration of Simple Functions

Simple functions are just finite linear combinations of characteristic functions of measurable sets with finite measure. That is:

$$\psi(x) = \sum_{k=1}^{N} a_k \chi E_k(x)$$

Since there are different ways of writing this sum, one can define the **canonical** form that is the **unique** decomposiiton with a_k distinct and non-zero with disjoint E_k.

The Lebesgue integral of simple functions Now, we can define the Lebesgue integral of ψ as:

$$\int_{\mathbb{R}^d} \psi(x)dx = \sum_{k=1}^{M} c_k m(F_k)$$

where c_k are the coefficient defined in the simple function and $m(F_k)$ is the measure of the disjoint subset indexed by k.

Properties of Simple Function Lebesgue Integral

1. independence of representation
2. linearity of simple functions
3. additivity of disjoint subsets
4. monotonicity of simple functions
5. triangle inequality of simple functions

3.2 Integration of Bounded Functions of Finite Measure

Note that we are able to approximate a function, f that is bounded by M and supported on a measurable set E with $\{\psi_n\}$ of simple functions. Each of these simple functions in turn are bounded by M and supported on E.

Definition 3.1 (Lebesuge integral for bounded functions)**.** For a bounded function that is supported on a set of finite measure, the Lebesgue integral is:

$$\int f(x)dx = \lim_{n\to\infty} \psi_n(x)dx$$

Where ψ_n are a sequence of simple functions that are bounded in absolute value and supported on the support of f and converges uniformly to f(x) for a.e. x.

Compared to simple functions and their Lebesgue integral, bounded functions have all the same properties except for the independence of representation.

Lemma. *f is a bounded function supported on E of finite measure. If $\{\psi_n\}$ is any sequence of simple functions bounded by M, supported on E and converges to f for a.e. x, then:*
i) The limit $\lim_{n\to\infty}\int \psi_n$ *exists*
ii) If f = 0 a.e., then the limit of the integral equals 0.

See below for the bounded convergence theorem that applies to these class of functions.

3.3 Integration of Non-negative Functions

Next, we remove the constraint of bounded functions and consider non-negative functions, which may be extended-valued.

Definition 3.2 (Lebesuge integral for non-negative functions)**.** For a bounded function that is supported on a set of finite measure, the Lebesgue integral is:

$$\int f(x)dx = \sup_g \int g(x)dx$$

Where g is the set of all measurable functions $0 \leq g \leq f$, with g being bounded, and supported on set of finite measure.

The supremum may either be finite or infinite. If it is finite then we say that **f is Lebesgue integrable**.

This definition of Lebesgue integrability satisfies linearity, additivity, monotonicity.

3.4 Convergence Theorems

Convergence theorems in Lebesgue integration theory help us define sufficient and/or necessary conditions for exchanging the integral sign and a limit, which are important in other topics.

Theorem (Bounded convergence theorem)**.** *Suppose that $\{f_n\}$ is a sequence of measurable functions that are all bounded by M and are supported on a set E of finite measure and $f_n(x) \to f(x)$ a.e. on x as $n \to \infty$. Then f is measurable, bounded, supported on E for a.e. x and*

$$\int |f_n - f| \to 0 \text{ as } n \to \infty$$

Proof. From the assumptions, we see that f is bounded by M a.e. and vanishes outside E, except on set of measure zero. By the triangle inequality for the integral implies that is suffices to prove that $\int |f_n - f| \to 0$.

$$|\int f_n - f| = |\int f_n - \int f| \leq \int |f_n - f|$$

By linearity of the integral. So if we prove the greater value goes to zero, then the lesser value must also go to zero by positivity of the integral.

By Egorov's theorem, we can construct a closed set, $A_\epsilon \subset E$, such that, $m(E - A_\epsilon) \leq \epsilon$ with $f_n \to f$ uniformly on this set.

Next we want to upper bound the integral, $\int |f_n - f|$ and show that the upper bound depends on arbitrary ϵ. We do this by decomposing the integrand into two disjoint sets, A_ϵ and $E - A_\epsilon$.

$$\begin{aligned}
\int_E |f_n - f| &= \int_{A_\epsilon} |f_n - f| + \int_{E-A_\epsilon} |f_n - f| && (1) \\
&= m(A_\epsilon)|f_n - f| + m(E - A_\epsilon)|f_n - f| && (2) \\
&\leq m(E)\epsilon + \epsilon 2M && (3)
\end{aligned}$$

Since ϵ is arbitrary, M is finite and E has finite measure, then the integral in question must be zero. □

What if we want no assumptions? How do the limits of integrals compare to the integrals of limits always? Fatou's Lemma tells us the inequality. If we simply assume non-neagtive functions without boundedness, we get Fatou's lemma. It can be extended to the monotone convergence theorem, if we assume convergence and a monotonic sequence of functions.

Theorem (Fatou's Lemma)**.** *Suppose $\{f_n\}$ is sequence of measurable non-negative functions. If $\lim_{n\to\infty} f_n(x) = f(x)$ for a.e. x, then:*

$$\int f \leq \liminf_{n\to\infty} \int f_n$$

This provides a general bound on taking the liminf to the outside of the integral.

Proof. If we have $0 \leq g \leq f$, for some g that is bounded and supported on a set of finite measure. We can set up a sequence of functions $g_n(x) = min(g(x), f_n(x))$, then g_n is measurable, supported on E and $g_n \to g$ a.e.

We can now use the bounded convergence theorem 3.4, which gives us: $\int g_n \to \int g$. Since we have the g_n sequence being a lower bound of the f_n sequence by construction, we have that their integrals are bounded by monotonicity:

$$\int g_n \leq \int f_n$$

and so we have: $\int g_n = \int g \leq \int f_n \leq \liminf_{n\to\infty} \int f_n$. Now, by taking the supremum over g, we have the desired result. □

Note it also bounds the limsup since that is bigger. It is also powerful because it does not even assume existence of the limit of the integral. Fatou's lemma will be useful in proving the Monotone convergence theorem, which is another useful theorem for conditions when one can switch the integrand and the limit.

Theorem (Monotone Convergence Theorem)**.** *If $\{f_n\}$ are a sequence of non-negative functions, they converge to f and are a monotonic sequence, then $\lim_{n\to\infty} \int f_n = \int f$.*

Note: a monotonic sequence is where $f_n(x) \leq f_{n+1}(x)$.

Now, what can we say about the integral of an infinite series? Ideally, we would have a theorem that tells us the conditions when we can switch the summation and the integral sign. This corrollary tells us that.

Corollary. *Suppose $a_k(x) \geq 0$ are a sequence of measurable functions. Then $\int \sum_{k=1}^{\infty} a_k(x)dx = \sum_{k=1}^{\infty} \int a_k(x)dx$*

Lemma. *Suppose we have a non-negative function, f (i.e. $f \geq 0$), that is measurable. In addition, there are a sequence of measurable functions $\{f_n\}$ with $f_n(x) \leq f(x)$ and $f_n(x) \to f(x)$ converges for almost every x (i.e. except for $x \in E$ with measure zero).*

Then we have that $\lim_{n\to\infty} \int f_n = \int f$.

Theorem (Dominated Convergence Theorem)**.** *Suppose $\{f_n\}$ is a sequence of measurable functions such that $f_n(x) \to f(x)$ a.e. x as n tends to infinity. If we have an integrable function that upperbounds the absolute value of the sequence of f_n functions (hence "dominated"), then we can interchange the integrand and limit signs. That is if $|f_n(x)| \leq g(x)$ where g is integrable, then*

$$\int |f_n - f| \to 0$$

and by triangle inequality:

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

Proof. □

3.5 General Lebesgue Integrable Functions

We can integrate simple functions, non-negative functions, and bounded functions. We have the following general definition for Lebesgue integrability:

Definition 3.3. If $f : \mathbb{R}^d \to [-\infty, \infty]$ is measurable, then it is Lebesgue integrable if $|f|$ is integrable:

$$\int |f(x)|dx = \sup_g \int g(x)dx < \infty$$

where g is the set of all measurable functions $0 \leq g \leq f$, with g being bounded and supported on a set of finite measure.

Properties of the General Lebesgue Integral The general integral is linear, additive, monotonic and satisfies the triangle inequality. These are important properties as we begin to speak about Lebesgue integration in vector spaces.

Here the dominated convergence theorem is a useful outcome of general Lebesgue integration.

This next proposition allows us to extend some useful properties of integrable function on $\mathbb{R}^d$. The first part of the lemma tells us that there is always a set of finite measure that outside this set, the integral of the absolute value of the function is arbitrarily small. The second condition tells us about absolute continuity (note the δ, ϵ argument similar to continuity arguments).

Lemma (Absolute continuity)**.** *Suppose f is integrable on $\mathbb{R}^d$. Then for every $\epsilon > 0$, the following are true:*

i) There exists a set of finite measure, B such that:

$$\int_{B^c} |f| < \epsilon$$

ii) There is a $\delta > 0$ such that:

$$\int_E |f| < \epsilon$$

whenever $m(E) < \delta$.

first part. WLOG, assume $f \geq 0$.

If we let B_N denote the ball with radius N, centered at the origin, and $f_N(x) = f(x)\chi B_N(x)$, then $f_N \geq 0$ is measurable, and an increasing function. The limit of $\{f_N\}$ is f(x). Therefore by the MCT, we have:

$$\lim_{N\to\infty} \int f_N = \int f$$

By definition of the limit, for a large N, we have:

$$0 \leq \int f - \int f \chi B_N < \epsilon$$

The inner part is equal to $\int f(1-\chi B_N)$, by linearity of the Lebesgue integral. Since $1 - \chi B_N = \chi B_N^c$, then we have that $\int_{B_N^c} f < \epsilon$. $\square$

second part. $\square$

3.6 Lp Spaces

When we take a collection of complex-valued functions on measurable subsets on $\mathbb{R}^d$ that are integrable, then they form a vector space over $\mathcal{C}$.

3.6.1 L1 spaces

L1 spaces are defined on integrable function where the L^1 norm is defined. That is:

$$||f||_{L^1} = \int_{\mathbb{R}^d} |f(x)| dx$$

is the norm of f. It is still a vector space, in that it is closed under addition, and obeys the triangle inequality. It also can define a metric on the space using the norm:

$$d(f, g) = ||f - g||_{L^1}$$

where the $d(f, g) = 0$ if and only if f = g a.e. It is also commutative (i.e. $d(f, g) = d(g, f)$), positive $d(f, g) = ||f - g||_{L^1} \geq 0$ and it obeys the triangle inequality.

It is important to remind ourselves again about the property of completeness for a metric space. Namely, a metric space (V, d) is complete if for every Cauchy sequence $\{x_k\} \in V$ ($d(x_k, x_l) \to 0$ as k and l go to infinity), there exists an element of the metric space $x \in V$, such that the limit of the Cuachy sequence is x. In other words, there is an element within the space that a Cauchy sequence converges to for all Cauchy sequences.

In the next theorem, we show that the space of L^1 functions is complete. This is known as the Riesz-Fischer theorem. This is a very important theorem because it describes the convergence of Cauchy sequences in L^1 spaces, but can be extended to general L^p spaces.

Theorem. *The vector space L^1 is complete in its metric (d(f, g)).*

Proof. We consider a Cauchy sequence $\{f_n\}$ such that $||f_n - f_m||_{L^1} \to 0$ as n, m go to infinity. We would like to extract a subsequence of $\{f_n\}$ that converges to $f \in L^1$ pointwise a.e. and in norm.

A.e. convergence does not hold for general Cauchy sequences. However, if convergence in norm is rapid enough, then a.e. convergence results. We proceed by taking a subsequence of the original Cauchy sequence.

We consider subsequence $\{f_{n_k}\}_{k=1}^{\infty}$ of $\{f_n\}$ with the property:

$$||f_{n_{k+1}} - f_{n_k}|| \leq 2^{-k}$$

for all $k \geq 1$. Their existence is guaranteed by assumption of the Cauchy sequence being in L^1. That is: $||f_n - f_m|| \leq \epsilon$ whenever n and m are larger then some N_ϵ. So we take $n_k = N_{2^{-k}}$, where $\epsilon = 2^{-k}$.

We now consider a telescoping series for f:

$$f(x) = f_{n_1}(x) + \sum_{k=1}^{\infty}(f_{n_{k+1}}(x) - f_{n_k}(x))$$

and

$$g(x) = |f_{n_1}(x)| + \sum_{k=1}^{\infty}|f_{n_{k+1}} - f_{n_k}(x)|$$

TBD

□

3.6.2 Density Properties of the L1 space

Definition 3.4. We say a set, E, of integrable functions is **dense** in L^1 if for every $f \in L^1$ and $\epsilon > 0$, there exists a $g \in E$ such that $||f - g||_{L^1} < \epsilon$

The denseness definition of a set in L^1 is useful because proving properties of functions in L^1 may be obtained by say first proving those properties on a different set of functions that are simpler, and then using density to match a function in L^1 arbitrarily close.

The next theorem states which sets of functions are dense in $L^1(\mathbb{R}^d)$.

Theorem. *The following sets of functions are dense in $L^1(\mathbb{R}^d)$:*

i) Simple functions ii) Step functions iii) Continuous functions of compact support

Related to 2.3.2, 3.4 and 2.3.5.

Proof. i) is related in 2.3.2. We need to see that if $f \in L^1$, and $\epsilon > 0$, then there exists a simple function such that $\int |f - \psi| < \epsilon$.

Approximating the real and imaginary part of f suffices: f is real-valued, and $f = f^+ - f^-$, which suffices to show that $f \geq 0$. Now we have a non-negative function, $f \geq 0$, $f \in L^1$ and we saw in 2.3.2 that there exists simple ψ such that $\int |f - \psi| < \epsilon$.

Earlier we saw $\exists$ ψ_k increasing converging to f. Since f dominates ψ_k, then by the DCT, $||f - \psi_k|| \to 0$. So $\psi = \psi_k$ for k large will get us the desired result.

ii) is related in 2.3.3. Since we have the first part of the result, it suffices to show that for simple functions in L^1, $\exists$ step function ϕ such that $||\phi - \psi||_{L^1} < \epsilon$. Then since ψ is dense already, then ϕ is also dense.

Since simple functions are just linear combinations of characteristic sets: $\psi = \chi_E$, with finite measure. If we can show that this can be approximated with step functions, then we have our desired result. This follows from Littlewood's Principles 2.3.5, where every set E is almost a finite union of non-overlapping rectangles (i.e. step functions).

Now we prove iii). Since step functions are dense, then it suffices to show that for step functions in L^1, then there exists g continuous with compact support, such that:

$$||\phi - g||_{L^1}$$

This follows if when $\phi = \chi_{rectangle}$.

If $g(x) = \begin{cases} 1\text{[a,b]} \\ 0 \ x \leq a - \epsilon \text{ or } x \geq b - \epsilon \\ linear \ [a-\epsilon, a], [b, b-\epsilon] \end{cases}$. where g is continuous. Then $||\chi_{[a,b]} - g|| \leq 2\epsilon$, where the ϵ is contributed on the linear segments from the intervals supporting the linear region. In fact each of those intervals are triangles, so in fact their integral is $\epsilon/2$. □

3.6.3 Invariance Properties of the Lebesgue Integral

We will here see that integrable functions are invariant under translations, dilations and convolutions.

Lemma (translation and dilation invariance)**.** *If $h \in \mathbb{R}^d$ with $f_h(x) = f(x-h)$, then if $f \in L^1$, so is f_h. Furthermore, $\int f(x)dx = \int f(x-h)dx$.*

If $f \in L^1$, then so is its reflection.

If $f \in L^1$, and $\delta > 0$, then so is its dilation $f(\delta x)$ with $\int f(\delta x)dx = \delta^{-d} \int f(x)dx$.

Proof. First, we remember that measurable sets are translation invariant. That is $E \subset \mathbb{R}^d$ measurable with finite measure has the following property: For a characteristic function, $\chi_E(x-h) = \chi_{E_h}(x), \ \forall x \in \mathbb{R}^d$. So we then by definition of the Lebesgue integral for simple functions have:

$$\int \chi_E(x)dx = m(E) = m(E_h) = \int \chi_{E_h}(x)dx$$

Next, we know that for $f \geq 0$, we can approximate it pointwise at every x with a sequence of simple functions. These sequences of step functions are also monotonic, so by the MCT, we have the following for non-negative f:

$$\int f dx = \lim_{k\to\infty} \int \psi_k(x)dx = \lim_{k\to\infty} \int \psi_k(x-h)dx = \int f(x-h)dx$$

By the linearity of the L1 space (it's a vector space), this holds for all functions in L1. $\square$

reflection. This is done in the same way because measurable sets have a measurable reflection with the same measure. $\square$

dilation. This is done by also recognizing that for a measurable set E with finite measure that is a subset of $\mathbb{R}^d$, then dilation of points inside E results in a δ^d factor on the measure of E. $\square$

We now see that for f and g measurable functions, and some fixed $x \in \mathbb{R}^d$, then $f(x-y)g(y)$ is integrable, $f(x+y)g(-y)$ and $f(y)g(x-y)$ are integrable. We now define a convolution for Lebesgue integration.

Definition 3.5 (Convolution)**.** *$f * g(x)$ is a convolution that equals $\int_{\mathbb{R}^d} f(x-y)g(y)dy$.*

Next, let us examine the relationship between translation invariance of the Lebesgue integral and continuity of the function in L1 norm. That is because, there are integrable functions which are discontinuous at all x.

Lemma. *If $f \in L^1$, then $||f_h - f||_{L^1} \to 0$ as $h \to 0$. Meaning that is continuous in the norm.*

Note: if f was a continuous function with compact support, this is automatically true.

3.7 Fubini's Theorem

An important issue arising in integration is how to integrate functions in higher dimensions. Fubini's theorem will tell us conditions that we can break down a higher-dimensional integration into integrals along sub-dimensions making integration simpler.

We can generally write $\mathbb{R}^d$ as slices along its sub-dimensions. Say we have $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, then we have:

$$E_x = \{y \in \mathbb{R}^{d_2} : (x,y) \in E\}$$
$$E_y = \{x \in \mathbb{R}^{d_1} : (x,y) \in E\}$$

Theorem (Fubini's Theorem)**.** *Suppose f(x,y) is integrable on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, then for a.e. $y \in \mathbb{R}^{d_2}$:*

i) The slice f^y is integrable on $\mathbb{R}^{d_1}$

ii) The slice function $\int f^y(x)dx$ is integrable on $\mathbb{R}^{d_2}$

iii) The integrals $\int_{\mathbb{R}^{d_2}}(\int_{\mathbb{R}^{d_1}} f(x,y)dx)dy = \int_{\mathbb{R}} f$.

Note: the theorem is symmetric in x and y, so we can conclude vice-versa for a.e. x.

Proof. The outline of the proof will be as follows: We create a larger class of set of all integrable functions that satisfy i), ii) and iii). We show that this class contains the L1 space of integrable functions.

Then we show that the L1 space contains this set.

1. F is closed under finite linear combinations
2. F is closed under limits of monotonic sequences that are integrable
3. The indicator function of a set E, which is a G_δ set of finite measure is contained in F
4. Same as above with a set of measure zero
5. Same as above for a set of finite measure
6. By the density of simple functions in L1, we show that the L1 space is equivalent to F

□

3.7.1 Fubini-Tonelli Theorem

In Fubini's theorem, the assumptions require that $f(x, y)$ **is integrable**. Then as a result, we get interchangability of the integrands. However, Fubini-Tonelli theorem will get the same consequences, but with a different assumption! This is useful since we are not always guaranteed that $\int |f(x, y)| < \infty$ (i.e. f is integrable).

In Fubini-Tonelli's theorem, we may first consider $|f|$, such that the function is non-negative. Then assuming that the function was measurable, then we can compute $\int f$ by iterated integrals. Then if $|f|$ is integrable (i.e. $\int |f| < \infty$), then we now fall under the assumptions of Fubini's theorem.

Theorem (Fubini-Tonelli Theorem)**.** *Suppose $f(x, y)$ is a non-negative measurable function on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Then for a.e. $y \in \mathbb{R}^{d_2}$, we have the following:*
i) The slice f^y is measurable on $\mathbb{R}^{d_1}$
ii) The function $\int_{\mathbb{R}^{d_1}} f^y(x)dx$ is measurable on $\mathbb{R}^{d_1}$
iii) $\int_{\mathbb{R}^{d_2}}(\int_{\mathbb{R}^{d_1}} f(x, y)dx)dy = \int_{\mathbb{R}^d} f(x, y)dxdy$

Proof. □

4 Differentiation of Lebesgue Integrals

Now that we have examined the Lebesgue measure theory and the integration theory that follows it, specifically L^1 and L^p spaces, we turn our attention to differentiation. Differentiation is understood as an inverse operation compared to integration and we seek to formalize the fundamental theorem of calculus in the Lebesgue setting. There are two fundamental theorems of calculus (FTC) questions that we would like to answer:

1. FTC1: If f is integrable on [a,b] and F is its indefinite integral $F(x) = \int_a^x f(y)dy$, then does this imply that F is differentiable (at least for a.e. x) and that specifically $F' = f$?

2. FTC2: What conditinos on a function F on [a,b] guarantee that the derivative exists for a.e. x, that this function F is integrable and that moreover the fundamental theorem of calculus holds:

$$F(b) - F(a) = \int_a^b F'(x)dx$$

In order to extend the FTC to general Lebesgue-measurable functions, we define the notion of a locally integrable function. That is:

Definition 4.1. A function $f : \mathbb{R}^d \to \mathbb{R}$ is locally integrable if it is Lebesgue measurable and

$$\int_K |f(x)|dx < \infty$$

That is it is integrable over a compact subset of its domain.

To write the FTC1, we usually say:

$$f(x) = \frac{d}{dx}\int_a^x f(y)dy$$

whichcan be written in terms of symmetric differences based on the definition of the derivative. Say we define $F(x) = \int_a^x f(y)dy$, then:

$$f(x) = \lim_{h\to 0}\frac{F(x+h) - F(x)}{h} = \lim_{r\to 0}\int_x^{x+r} f(y)dy$$

which can also be written as:

$$f(x) = \lim_{r\to 0^+}\frac{1}{2r}\int_{x-r}^{x+r} f(y)dy$$

with the nth dimension version just using balls with radius r and centered at $x \in \mathbb{R}^d$. This form of the FTC1 is useful because we can then make use of our notion of Lebesgue measure and the maximal functions.

4.1 Mean Value Theorem

Before diving into Lebesgue theory for differentiation, we should remind ourselves of the important Mean-Value theorem (MVT), which states:

Theorem. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function on a closed interval [a, b] and differentiable on the open interval (a, b) with $a < b$. Then there exists some $c \in (a, b)$ such that:*

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

4.2 Differentiation of the Lebesgue Integral

To define the differentiation of the integral, we define a function f on [a,b] that is integrable. We define:

$$F(x) = \int_a^x f(y)dy, \quad a \leq x \leq b$$

We call F the indefinite integral of f. To define the differentiation of F, we recall the definition of the derivative as the limit of a quotient:

$$F'(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h}$$

The averaging problem in differentiation of integrals We are interested in the following problem: If f is integrable, then is it true that if you take an average of functions over a ball, as you take the limit of the measure of the ball going to zero, that the limit is achieved for a.e. x?

The answer is yes when f is continuous at x.

Lemma. *If f is integrable on $\mathbb{R}^d$, and is continuous at x, then:*

$$f(x) = \lim_{m(B) \to 0} \frac{1}{m(B)} \int_B f(y)dy = f(x)$$

Proof. (*) For every $\epsilon > 0$, there exists a $\delta > 0$ such that if $|x - y| < \delta$, then $|f(x) - f(y)| < \epsilon$.

If say radius of ball B is $< \delta/2$ (worst case any two points have at most distance δ), and $x \in B$, then (*) holds whenever $y \in B$ as well.

We want f(x) to be close to this "averaging" function that we are interested in. Note that:

$$|f(x) - \frac{1}{m(B)} \int_B f(y)dy| = |\frac{1}{m(B)} \int_B (f(x) - f(y)dy|$$

Then we have that:

$$|\frac{1}{m(B)} \int_B (f(x) - f(y)dy| \tag{4}$$

$$\leq \frac{1}{m(B)} \int_B |f(x) - f(y)|dy \quad \text{(By tri-ineq)} \tag{5}$$

$$\leq \frac{1}{m(B)} \epsilon m(B) = \epsilon \tag{6}$$

Since ϵ was arbitrary, then we have our equality. □

Now, we would like to answer this question for not only an x where f(x) is continous, but absolutely at **every x**. This leads us to the Hardy-Littlewood Maximal function, where we will see that the answer is yes.

Note that the definition above can be rewritten as:

$$\frac{1}{h}\int_x^{x+h} f(y)dy = \frac{1}{|I|}\int_I f(y)dy$$

where $I = (x, x+h)$ interval, and $|I|$ is the length of that interval. Essentially the above is saying that we take an **average** value of f over the interval I. Then as you take the limit as $|I|$ goes to zero, then we expect these averages to tend to f(x). Now, we want to know when does th following equation hold? (What constraints are needed on the functions and the points x)

$$\lim_{|I|\to 0,\ x\in I} \frac{1}{|I|}\int_I f(y)dy = f(x)$$

or for higher dimensions in more generality:

$$\lim_{m(B)\to 0,\ x\in B} \frac{1}{m(B)}\int_B f(y)dy = f(x)$$

for a.e. x. The limit is taken as the volume of open balls containing x (in high dimensions) goes towards zero.

4.3 Covering Lemma To Demonstrate Weak-L1 Property

In order to prove the Hardy-littlewood theorem, which will tell us that the maximal functions of an integrable function are weak-L^1, then we need the notion of a covering lemma, which states that any **finite** collection of open balls in $\mathbb{R}^d$ has a **disjoint subcollection** that can cover a fraction of the original collection. The fraction is a exponent based on the dimension of the space.

Lemma (Vitali Covering Argument)**.** *Suppose $B = \{B_1, B_2, ..., B_N\}$ is a finite collection of open balls in $\mathbb{R}^d$. Then there exists a disjoint sub-collection of $B_{i_1}, B_{i_2}, ..., B_{i_k}$ of B such that:*

$$m(\bigcup_{l=1}^{N} B_l) \leq 3^d \sum_{j=1}^{k} m(B_{i_j})$$

That is we may find a disjoint sub-collection of balls that covers a fraction of the region covered by the original collection. This fraction scales exponentially as a function of the dimensionality of the space.

Proof. Suppose B and B' are two balls that intersect, with radius of B' (call it r') being less than or equal to B (call it r). Then $B' \subset \tilde{B} = \{y \in \mathbb{R}^d : |x - y| \leq 3r\}$, where $\tilde{B}$ is the ball centered at x with three-times the radius of B.

Note that we need three times the radius of B to fully cover B'. If B and B' have the same radius and are touching then any radius less then 3r would not fully cover B'.

The measure of $\tilde{B}$ by the dilation property of measure is $m(\tilde{B}) = 3^d m(B)$.

Now, we can state an iterative procedure for choosing balls based on the above: We begin by choosing B_1' that is the ball with the largest radius from our collection. Then we remove all balls in our collection that intersect with B'_1. Then choose the next-largest ball from our remaining collection. In this way, we obtain a subcollection of balls, $\{B'_1, ..., B'_k\}$ that are disjoint by definition. Now if we expand each of these balls as done above by a factor of 3, then we get that they cover the original collection plus more.

$$\bigcup_{i=1}^{N} B_i \subset \bigcup_{i=1}^{k} \tilde{B}'_i$$

Now by monotonicity, disjoint finite additivity and dilation-properties of measure, we have:

$$m(\bigcup_{i=1}^{N} B_i) \leq m(\bigcup_{i=1}^{k} \tilde{B}'_i) = 3^d \sum_{i=1}^{k} m(B'_i)$$

$\square$

4.4 Hardy-Littlewood Maximal Function and Weak-L1 Functions

First, we define the Weak-L1 space.

Definition 4.2. A measurable function, $f : \mathbb{R}^d \to \mathbb{R}$ is in weak-$L^1(\mathbb{R}^d)$ if there exists a constant C_f that depends on the function f, but not on t, such that:

$$m(\{x \in \mathbb{R}^d : |f(x)| > t\}) \leq \frac{C}{t}$$

By the Chebyshev inequality, we have that for integrable functions:

$$m(\{x \in \mathbb{R}^d : |f(x)| > t\}) \leq \frac{1}{t}||f||_{L^1}$$

where C is exactly the L1 norm of f. So, weak-L1 functions are a generalization of L1 functions, where the constant term depends on the function.

If f is integrable on $\mathbb{R}^d$, then we define its maximal function f^* as:

$$f^*(x) = \sup_{x \in B} \frac{1}{m(B)} \int_B |f(y)|dy, \quad x \in \mathbb{R}^d$$

where the supremum is taken over all balls containing the point x. Note, we replace the limit in the averaging problem with a supremum and f by its absolute value. Now, one might ask what properties this f^* function has? The following theorem states that it is measurable, finite for a.e. x, and satisfies a "weak-type" inequality, where it is not "that much larger" then f(x).

We shall also see that f^* is not necessarily integrable, even if f is integrable. The inequality in the theorem part iii) is sort of the "best-substitute" for integrability.

Theorem (Properties of the Maximal Function)**.** *Suppose f is integrable on* $\mathbb{R}^d$*, then the following are true:*

i) f^* *is measurable*

ii) $f^*(x) < \infty$ *for a.e. x*

iii) f is a weak-L1 function: $m(\{x \in \mathbb{R}^d : \ f^*(x) > \alpha\}) \leq \frac{A}{\alpha}||f||_{L^1(\mathbb{R}^d)}$ *for all* $\alpha > 0$ *where* $A = 3^d$ *and* $||f|| = \int |f(x)|dx$*.*

Proof. f^* is measurable. We will just show that the set $E_\alpha = \{x \in \mathbb{R}^d : f^*(x) > \alpha\}$ is open. Note that if $x \in E_\alpha$, then there exists a ball centered at x with radius ϵ, such that $x \in B$ and

$$\frac{1}{m(B)} \int_B |f(y)|dy > \alpha$$

Then if you take any point $y \in \mathbb{R}^d$ such that $|y - x| < \epsilon/2$, then y will also belong to B as well and hence belongs to E_α. Therefore, E_α is open, and hence measurable. □

Proof. If we take set $E_\infty = \{x : f^*(x) = \infty\}$, then it is a subset of E_α. If we take the limit as $\alpha \to \infty$, then by the third property we have that:

$$m(E_\infty) \leq 0$$

and so must have measure zero. Therfore $f^* < \infty$ for a.e. x. □

Proof. We fix $\alpha > 0$ and let:

$$E_\alpha = \{x \in \mathbb{R}^d : f^* > \alpha\}$$

as the set of interest where f^* is bigger then our fixed α.

The measure of that set can be written as the supremum over compact subsets.

$$m(E_\alpha) = \sup\{m(K) : K \subset E_\alpha \text{ compact}\}$$

Since this is a supremum, it suffices to show that:

$$m(K) \leq \frac{C}{t} \int |f(y)|dy$$

for every compact subset K of E_α. Now that we are dealing with a compact subset, by the definition of compactness, there exists a finite subcover. Now that there is a finite subcover, we know by the Vitali Covering lemma that there exists a subcollection of balls that cover a fraction of the original subset up to a factor of 3^d.

We call our finite subcover: $\{B_1, ..., B_N\}$. Then if $x \in K$, then there is an open ball B_x centered at x such that:

$$\frac{1}{m(B_x)} \int_{B_x} |f(y)|dy > \alpha$$

by definition of the E_α set, or in other words:

$$m(B_x) < \frac{1}{\alpha} \int_{B_x} |f(y)|dy$$

By taking a finite sucollection of our subcover that are all disjoint, $\{B'_1, ..., B'_k\}$, we get the series of inequalities:

$$m(K) \leq \sum_{i=1}^{N} m(B_i) \tag{7}$$

$$\leq 3^d \sum_{i=1}^{k} m(B'_i) \tag{8}$$

$$\leq \frac{3^d}{\alpha} \sum_{i=1}^{k} \int_{B'_i} |f(y)|dy \tag{9}$$

$$= \frac{3^d}{\alpha} \int_{\cup_{j=1}^{k}} B'_j |f(y)|dy \quad \text{(by def of integral)} \tag{10}$$

$$\leq \frac{3^d}{\alpha} \int_{\mathbb{R}^d} |f(y)|dy \quad \text{(By monotonicity)} \tag{11}$$

□

4.5 Lebesgue Differentiation Theorem

Now that we have an understanding of the maximal function, we are interested in a solution to the averaging problem. The Lebesgue differentiation theorem tells us that if a function is locally integrable, then we can perform the FTC1, and take the derivative of the integral to get the function inside the integrand.

Theorem. *If $f \in L^1_{loc}(\mathbb{R}^d)$, then:*

$$\lim_{m(B(x)) \to 0} \frac{1}{m(B)} \int_B f(y)dy = f(x)$$

for a.e. x.

Proof. □

Now that we have a theorem stating when we can differentiate an integral (i.e. when it is locally integrable), then we can define a set of points x, for which the limits exists for a suitable function f. This set is called the Lebesgue set.

Definition 4.3. If $f \in L^1_{loc}$, then a point $x \in \mathbb{R}^d$ belongs to the Lebesgue set) of f, if there exists a constant $c \in \mathbb{R}$ such that:

$$\lim_{r \to 0} \frac{1}{m(B_r(x))} \int_B |f(y) - c|dy = 0$$

That is if you have points in the Lebesgue set, then you have the results of the HWL maximal theorem. The Lebesgue set of f, does depend on which f we choose. It has some nice properties that allow us to take a wide variety of averages and recover points of the functions.

4.6 Good Kernels and Approximating the Identity

Now that we have explored the averaging problem and shown various properties of the maximal function in relation to differentiation of an integral, we want to now look at averages of functions given as convolutions. Namely:

$$(f * K_\delta)(x) = \int_{\mathbb{R}^d} f(x-y)K_\delta(y)dy$$

where f is an integrable function, which is **fixed**, while the K_δ varies over a class of fucntions, referred to as **kernels**. We are interested in convolutions and their integrals because they come up in the Fourier transform.

Definition 4.4 (Good Kernels)**.** We specifically call K_δ **good kernels** if they are integrable and satisfy the following conditions for $\delta > 0$:

1. Normalization to 1: $\int_{\mathbb{R}^d} K_\delta(x)dx = 1$
2. Bounded L1 norm: $\int_{\mathbb{R}^d} |K_\delta(x)|dx \leq A$
3. For every $\eta > 0$, we have: $\int_{|x|\geq\eta} |K_\delta(x)|dx \to 0$ as $\delta \to 0$

Definition 4.5 (Approximations to the Identity Kernels)**.** Restricting the class of kernels to be more narrow then "good kernels", we define approximations to the identity kernels as K_δ that are integrable and satisfy the following conditions:

1. Normalization to 1: $\int_{\mathbb{R}^d} K_\delta(x)dx = 1$
2. Bounded at every δ: $|K_\delta(x)| \leq A\delta^{-d}$ for all $\delta > 0$
3. $|K_\delta(x)| \leq A\delta/|x|^{d+1}$ for all $\delta > 0$ and $x \in \mathbb{R}^d$

Note that condition i) is the same as a good kernel, but conditions ii and iii are more restrictive then what is mentioned in good kernels. Therefore, all "approximations to the identity kernels" are also "good kernels".

4.7 Differentiation of Functions

Now, we explore the second Fundamental theorem of calculus in the view of Lebesgue measure. That is, under what general conditions does the following hold?

$$F(b) - F(a) = \int_a^b F'(x)dx$$

There are two problems related to existence and measurability.

First, off the RHS is complicated if $F'(x)$ does not exist. The exemplar is the Weierstrass function that is continuous, but nowhere differentiable. So F being continuous is not enough of a condition to guarantee the FTC2.

Second, if $F'(x)$ exists, it might not be Lebesgue measurable.

4.7.1 Bounded Variation - To Enable Differentiability Almost Everywhere

We study a class of functions that have **bounded variation** that will give us existence of the derivative. We will see that functions of bounded variation are differentiable a.e. But we will see by usage of the Cantor-lebesgue function, that it does not guarantee the validity of the FTC2. After bounded variation, we will explore this next concept of **absolute continuity** that will give us the necessary conditions for the FTC2. First we define a few keywords.

Let γ be a parametrized curve in the plane: $z(t) = (x(t), y(t))$ with $a \leq t \leq b$. x and y are continuous real valued functions supported on [a, b].

Definition 4.6 (Rectifiability of a curve)**.** We say that the curve is rectifiable if there exists M $< \infty$ such that for any partition $P = \{t_0, t_1, ..., t_N\}$ such that $a = t_0 < t_1 < ... < t_N = b$ of [a,b]:

$$\sum_{j=1}^{N} |z(t_j) - z(t_{j-1})| \leq M$$

Definition 4.7 (Length of a curve)**.** The length $L(\gamma)$ of curve, γ is the supremum over all partition of the sum:

$$L(\gamma) = \sup_{P} \sum_{j=1}^{N} |z(t_j) - z(t_{j-1})|$$

or note that $L(\gamma) = \inf_M\{\sum_{j=1}^{N} |z(t_j) - z(t_{j-1})| \leq M\}$. That is the infimum over all M's that satisfy the rectifiability of the curve.

With these two definitions, one is interested in determining when rectifiability occurs. What conditions on the functions x and y are needed?

Suppose F(t) is a complex-valued function defined on [a,b] with a partition P on the interval. We define the **variation** of F on this **particular partition** as:

$$\sum_{j=1}^{N} |F(t_j) - F(t_{j-1})|$$

Definition 4.8 (Bounded Variation of a Function)**.** The function F is of bounded variation if the variations of F over **all partitions** are bounded.

$$\sum_{j=1}^{N} |F(t_j) - F(t_{j-1})| \leq M$$

for all partitions $a = t_0 < t_1 < ... < t_N = b$.

By the triangle inequality, if we take refinements of a partition P, then the variation of F on the refinement is greater then or equal to the variation of F on P. Bounded variation is a useful property as we will see in this next theorem because it guarantees curve rectifiability (a necessary **and** sufficient condition).

Theorem (Functions of Bounded Variation Form Rectifiable Curves)**.** *A curve parametrized by (x(t), y(t)) with $a \leq t \leq b$ is rectifiable if and only if x and y are functions of bounded variation.*

What does it mean to b e a function of bounded variation? It means that a function cannot oscillate too often with very high amplitudes. Examples of bounded variation functions are:

1. real-valued, monotonic and bounded functions
2. differentiable with F' bounded functions

In addition to bounded variation, we also can define the total variation of a function f on [a,x] by:

$$T_F(a,x) = \sup_P \sum_{j=1}^{N} |F(t_j) - F(t_{j-1})|$$

where the sup is taken over partitions of [a,x]. There is also the notion of positive and negative variation when F is real-valued. Note in total and bounded variation, F may be complex-valued, since we take the absolute value.

The positive and negative variation are:

$$P_F(a,x) = \sup \sum_{+} F(t_j) - F(t_{j-1})$$

where the sum is over all j such that $F(t_j) \geq F(t_{j-1})$.

$$N_F(a,x) = \sum \sum_{-} -(F(t_j) - F(t_{j-1}))$$

where the sum takes over all $F(t_j) \leq F(t_{j-1})$. That is positive/negative variation occurs over the "monotonic" regions of [a,x]. Positive and negative variation are useful since they require you to only look at a simplified expression. They are related to the function and the total variation as follows:

Lemma (BV Functions can be written in form of positive and negative variation (or as total variation))**.** *Suppose F is real-valued and of bounded variation on [a,b]. Then for all $a \leq x \leq b$:*

$$F(x) - F(a) = P_F(a,x) - N_F(a,x)$$

and

$$T_F(a,x) = P_F(a,x) + N_F(a,x)$$

Proof. □

Next, we will see that a function of BV is equivalent to saying that the function is the difference of two bounded monotonic functions. We make use of the previous lemma specifically to create those two monotonic functions using the positive and negative variation of the function.

Theorem (BV functions are the difference of two increasing bounded functions)**.** *Real-valued F on [a,b] is of BV if and only if F is the difference of two increasing bounded functions.*

A consequence of this theorem is to see that the total variation of a function and the curve parametrized by a continuous function are closely tied. Namely that the length of the curve between two points, A and B: L(A,B) is equivalent to the $T_F(A, B)$, where F is the curve. Thus $L(A, B)$ is a continuous function of B (and of A). Thus we observe that if a function of BV is continous, then so is its total variation.

Now, we come to a result that guarantees the existence of a derivative. Namely, if a function is of BV, then it is differentiable a.e.

Theorem (F being BV means F is differentiable a.e.)**.** *If F is of BV on [a,b], then F is differentiable on [a,b] a.e.*

A corollary gets us close to the FTC, but with an inequality instead of equality. Here, F is increasing and continous, then we get a derivative a.e. In addition, F will be measurable, non-negative.

Corollary. *If F is increasing and continous, then F' exists a.e. F' is also measurable, non-negative and obeys the following inequality:*

$$\int_a^b F'(x)dx \leq F(b) - F(a)$$

If F is bounded, then F' is integrable.

In order to prove the theorem, there is a useful Rising-sun Lemma by Riesz.

Lemma (Rising-sun lemma)**.** *Suppose G is real-valued and continous on $\mathbb{R}$.*

E is the set of points x such that:

$$G(x + h) > G(x)$$

for some $h = h_x > 0$. That is E is the set of points, where the function G translated to the right by a small amount is greater.

If E is non-empty, then it must be open and hence can be written as a countable disjoint union of open intervals: $E = \bigcup (a_k, b_k)$. If (a_k, b_k) is a finite interval in this union, then we have that $G(b_k) = G(a_k)$

Proof. We assume E is non-empty.

First, we show that E is open.

Since E is open, then it can be written as a countable disjoint union of open intervals.

Next, we show the equality of G at the endpoints of the disjoint intervals. □

Now, if we consider the case when $a = a_k$, then we get the following corollary:

Corollary. *G is as defined in the Rising-sun lemma 4.7.1. If $a = a_k$, then we have $G(a_k) \leq G(b_k)$.*

4.7.2 Why is Bounded Variation Not Enough for Fundamental Theorem of Calculus?

From the results above, we cannot go any further. We will need a stronger notion then Bounded Variation to guarantee the integration of a derivative.

The Cantor-lebesgue function is a continuous function that is increasing and of bounded variation, but:

$$\int_a^b F'(x)dx \neq F(b) - F(a)$$

The Cantor-Lebesgue function is a sequence of continuous increasing functions $\{F_n\}$ such that:

$$|F_{n+1}(x) - F_n(x)| \leq 2^{-n-1}$$

F_n converges uniformly to a continuous limit F. Since m(C) $= 0$ (the measure of the Cantor set is zero), then $F'(x) = 0$ a.e.

Therefore, a function of bounded variation **guarantees the existence of a derivative a.e., but not the Fundamental theorem of calculus**. By taking a stronger notion, we will guarantee existence of derivatives a.e., and ALSO the FTC.

4.7.3 Absolute Continuity - To Enable Second Fundamental Theorem of Calculus

Absolute continuity will allow us to write:

$$F(b) - F(a) = \int_a^x F'(x)dx$$

We define it as:

Definition 4.9 (Absolute continuity)**.** A function F on [a,b] is absolutely continuous if for any $\epsilon > 0$, there exists a $\delta > 0$ such that:

$$\sum_{k=1}^{N} |F(b_k) - F(a_k)| < \epsilon$$

whenever $\sum_{k=1}^{N}(b_k - a_k) < \delta$.

What is given by absolute continuity? We here state some properties that relate to BV and continuity that absolute continuity automatically implies.

1. Absolute continuity implies continuity and uniform continuity
2. If F is absolutely continuous on a bounded interval, then it is of BV. In addition its total variation is absolutely continuous
3. If F is the antiderivative of an integrable function, then F is absolutely continuous.

Here is a proof for why absolute continuity implies BV:

Proof. First, we can see that absolute continuity implies continuity and in fact uniform continuity.

From the definition of F being absolutely continuous, we can fix an $\epsilon > 0$. Then there exists a $\delta > 0$ such that if $[c, d] \subset [a, b]$, with $d - c < \delta$, then $c = t_0 < t_1 < ... < t_N = d$. Then from the definition of AC, we get:

$$\sum_{j=1}^{N} |f(t_j) - f(t_{j-1})| < \epsilon$$

The total variation of f is less then ϵ. So it is trivial to choose a δ that works for all $\epsilon > 0$.

Now, we want to see why this implies BV. We take $\epsilon = 1$. Then we cover our interval [a,b] by at most $M = 2(b-a)/\delta + 1$ sub-intervals, which all of length $< \delta/2$. Now we take the closed interval [a, b] and write it as the finite union of these M almost-disjoint intervals:

$$[a, b] = [s_0, s_1] \cup [s_1, s_2] \cup \ldots \cup [s_{M-1}, s_M]$$

with $s_0 = a, s_M = b$. Each of these intervals is smaller than δ, and we saw earlier that this implies that the sum of the variations on those intervals is less than ϵ.

Now the variation of F on [a,b] is upper-bounded by the sum of the variations, which are all less than or equal to ϵ, as long as the sub-intervals are small enough. This is also upper-bounded by M $*\epsilon$ (= M since $\epsilon = 1$ here) since there is exactly M sub-intervals by construction. These M intervals have length $2(b-a)/\delta + 1 < \infty$, so F is of BV as well. $\square$

This actually implies i) that if F is absolutely continuous, then the total variation (T_F) and the positive/negative variation are also absolutely continuous on $[a, b]$.

Secondly, it implies that F can be written as a difference of 2 increasing absolutely continuous functions. Since if you have a BV variation, then you can write the function as $P_F + F - N_F$, which is a difference of two increasing absolutely continuous functions.

Third, if you form an indefinite integral:

$$F(x) = \int_a^x f(y)dy$$

with f integrable, then F is absolutely continuous. This leads us to the following lemma about the absolute continuity of the integral.

Lemma (Indefinite Integral of an Integrable Function is Absolutely Continuous)**.** *If f is integrable, and $F(x) := \int_a^x f(y)dy$, then F is absolutely integrable.*

Proof. Recall that if $g \in L^1$, then $\forall \epsilon > 0$, there exists a $\delta > 0$ such taht $\int_E |g|dx < \epsilon$ if $m(E) < \delta$. That is the L1 norm of an integrable function can be controlled by the measure of teh set they integrate over.

Now, we let $\epsilon > 0$ and $g = |f|$. We take the set $E = \bigcup_{k=1}^{N}[a_k, b_k]$ as the non-overlapping union of closed intervals in [a,b]. Then we can define the measure of E as: $m(E) = \sum_{k=1}^{N}(b_k - a_k)$. We define E with such sub-intervals such that $m(E) < \delta$. Now, we can make use of our recalled fact.

Now since these intervals are non-overlapping and the corresponding measure is small (i.e. less than δ), the corresponding integral of g over E is:

$$\int_E |f|dy = \sum_{k=1}^{N} \int_{a_k}^{b_k} |f(y)|dy < \epsilon$$

Note that $|F(b_k) - F(a_k)| = |\int_{a_k}^{b_k} f| \leq \int_{a_k}^{b_k} |f|$ by the triangle inequality. This then implies that the sum of $|F(b_k) - F(a_k)|$ over the entire sub-interval is controlled by the sum of $\int_{a_k}^{b_k} |f|$.

Therefore:

$$\sum_{k=1}^{N} |F(b_k) - F(a_k)| < \sum_{k=1}^{N} \int_{a_k}^{b_k} |f| < \epsilon$$

which implies that F is absolutely continuous. □

Note, that absolute continuity is in essence a necessary condition for the FTC.

Finally, we obtain the FTC using absolute continuity as a necessary condition in the following theorem.

Theorem (FTC - integrating a derivative)**.** *If F is absolutely continuous on [a,b], then F' exists a.e. and is integrable. Moreover:*

$$F(x) - F(a) = \int_a^x F'(y)dy$$

for all $a \leq x \leq b$

The Vitali Covering Lemma To Prove FTC In order to prove the FTC, we need to remind ourselves of the Vitali covering lemma from integration theory, where we can take a finite collection of open balls in $\mathbb{R}^d$, then there exists a disjoint sub-collection of those balls such that the measure of the sub-collection covers approximately 3^{-d} of the measure of the entire collection.

We used this to prove the Hardy-Littlewood Maximal theorem, and we can use a variant of this to prove the FTC. We define a variant known as the Vitali covering.

Definition 4.10 (Vitali Covering)**.** For $\mathcal{B} = \{B\}$ a collection of balls is called a Vitali covering of E if for each $\eta > 0$, and $x \in E$, we can find $B \in \mathcal{B}$ with $x \in B$ and $m(B) < \eta$.

I.e. every point is covered by arbitrary small balls.

Note that this collection may be infinite and not necessarily finite.

We now define a modified Vitali Covering Lemma.

Lemma. *Suppose E is a set of finite measure and $\mathcal{B}$ is a Vitali covering of E. Then for any $\delta > 0$, we can find finitely many balls $B_1, B_2, .., B_N \in \mathcal{B}$ that are* ***disjoint*** *and have the following property:*

$$\sum_{i=1}^{N} m(B_i) \geq m(E) - \delta$$

That is, finitely many disjoint balls can cover E up to a δ difference in measure.

Proof. □

Now, we are ready to prove a consequence of absolute continuity that will show that a zero derivative implies a constant function.

Theorem (Absolute Continuity with a zero derivative implies constant function)**.** *If F is absolutely continuous on [a,b], then there exists a derivative a.e. Moreover, if $F'(x) = 0$ for a.e. x, then F is constant.*

Proof. Enough to show that F(b) = F(a). Assumption states that F'(x) exists a.e. and is zero.

So we have that there exists a set $E \subset [a, b]$ such that $m(E) = b - a$ full measure such that:

$$\lim_{x \to 0} \frac{F(x+h) - F(x)}{h} = 0$$

for all $x \in E$. The set that this does not happen has measure zero as we have seen from the BV analysis.

As a consequence of the above, we have the following. For $\epsilon > 0$, for each $\eta > 0$, we can find an open interval $I = (a_x, b_x) \subset (a, b)$ with $x \in I$ such that:

$$\frac{|F(b_x) - F(a_x)|}{|b_x - a_x|} \leq \epsilon$$

and $b_x - a_x < \eta$. Note that all of these $\{(a_x, b_x)\}$ form a Vitali covering of E. Do this for every $x \in E$. Now we can make use of the Vitali covering lemma, we can find finitely many disjoint intervals $I_i = (a_i, b_i)$ for i = 1, ..., N, such that:

$$\sum_{i=1}^{N} m(I_i) \geq m(E) - \delta = b - a - \delta$$

But note that $F(b_i) - F(a_i) \leq \epsilon |b_i - a_i|$ and the fact that these intervals are disjoint and contained in (a,b). Then we can sum over all these disjoint intervals:

$$\sum_{i=1}^{N} |F(b_i) - F(a_i)| \leq \epsilon(b - a)$$

Since everyone of these intervals I_i are in (a,b) and open, then we have:

$$[a, b] - \cup_{i=1}^{N} I_i = \bigcup_{k=1}^{M} [\alpha_k, \beta_k]$$

must be a finite collection of closed intervals that are disjoint. By the earlier argument, we see that:

$$\sum_{k=1}^{M} (\beta_k - \alpha_k) = |\bigcup_{k=1}^{M} [\alpha_k, \beta_k]| \leq \delta$$

Thus if δ is small enough that depends on ϵ, by absolute continuity of F, then we can bound the sum of F evaluated at these endpoints of the subintervals by ϵ. □

We finally can state the Fundamental Theorem of Calculus that tells us precisely when we can integrate a derivative to obtain the value of the function at the integral endpoints.

Theorem (Fundamental Theorem of Calculus 2)**.** *Suppose F is absolutely continuous on [a,b]. Then F' exists a.e. and is integrable. Moreover:*

$$F(x) - F(a) = \int_a^x F'(y)dy$$

for all $a \leq x \leq b$. By selecting $x = b$, we get that $F(b) - F(a) = \int_a^b F'(y)dy$.

Conversly if f is integrable on [a,b], then there exists an absolutely continuous function F, such that F'(x) = f(x) a.e. and in fact:

$$F(x) = \int_a^x f(y)dy$$

Proof. Since F' is integrable by the previous result. Then we define $G(x) = \int_a^x F'(y)dy$, $x \in [a, b]$. So we are integrating an L1 function, so therefore G is absolutely continuous. Then that implies that F is absolutely continuous. By Lebesgue differentiation theorem, we can differentiate G(x):

$$G'(x) = F'(x), \ a.e.$$

So we have two AC functions with the derivative equal a.e. Then $G'(x) - F'(x) = 0$ a.e. is an absolutely continuous function. Therefore, $F(x) - G(x)$ is constant by the previous theorem.

Therefore $F(a) - G(a) = F(a)$ (since G(a) = 0 by definition), so $F(x) - F(a) = G(x) = \int_a^x F'(y)dy$.

Since we have that $F(x) := \int_a^x f(y)dy$ is absolutely continuous (forming indefinite integral of an L1 function is AC). Then by Lebesgue differentiation theorem, we show that:

$$F'(x) = f(x)$$

a.e. □

4.8 Conclusions Thus Far

We have started out with proving the Lebesgue differentiation theorem. We showed that if a function, f, is L1, then a.e. we have that if you differentiate an integral, then you get the integrand back. This showed us the Fundamental Theorem of Calculus 1 in Lebesgue setting.

Specifically, f(x) is equal to the average over smaller and smaller balls around x. Even stronger, we showed that if you take the limit of the average difference in f(x) and f(y) around smaller and smaller balls that contain x, this limit tends to zero. These points specifically are called Lebesgue points. Here, we do not need that f is an L1 function, but rather just that f is locally integrable.

We utilized the Hardy-LittleWood-Maximal thm. If we use the HLW maximal function and consider the set where this function is bigger then a constant α, then the measure of this set is smaller then some constant dependent on dimension, α and the L1 norm of f.

We also constructed various covering lemmas that allowed us to approximate the measure of a finite collection of balls with a finite sub-collection of disjoint balls. The approximation is off by a factor of 3^d.

Next, we considered how to integrate derivatives. Specificaly, we wanted to determine conditions for the Second Fundamental Theorem of Calculus. When is $\int_a^b F'(y)dy = F(b) - F(a)$ for the largest class of F functions possible. We need i) at least that F' is integrable.

To ensure that F' is integrable, we introduced functions of BV, which bounds the supremum of the variation of F over all partitions. That is the total variation is bounded by some finite constant M. We then saw that F being of BV is a necessary and sufficient condition for F being written as a difference of 2 bounded and increasing functions (specifically $P_F + F$ and N_F). Now to understand functions of BV, simply need to understand properties of bounded increasing functions.

So when you have a continuous increasing functions, then these functions have:

1. f is differentiable a.e. on (a,b)
2. f' is integrable and $\int_a^b f'(x)dx \leq f(b) - f(a)$

where the second point is the FTC but with an inequality. Note that to get our first condition of having F' being integrable we note that for functions of BV, we get: F is differentiable a.e. and specifically F' is L1 and hence integrable.

Now instead of the inequality, we would like conditions for when equality of $\int_a^b f'(x)dx = f(b) - f(a)$. These class of functions are known as absolutely continuous functions.

Then if you're absolutely continuous, then you are i) differentiable a.e. and ii) derivative $=0$ implies that function is constant. Then we finally saw that absolutely continuity is a sufficient condition for the Second FTC.

Now what else is left? We would like to show that if F is increasing and bounded, then F is differentiable a.e. in (a,b). Note that before we showed this for when F is continuous and bounded.

4.9 Jump Functions and Their Differentiability

We next analyze a class of monotonic functions that are not continuous. This will allow us to show that there exists derivatives a.e. even for non-continuous functions (analagous to Thm 4.7.1). Specifically, we weaken continuity to a class of increasing bounded functions.

Note that if $x \in [a,b]$, then $F(x^-) = \lim_{y \to x_-} F(y)$ and $F(x^+) = \lim_{y \to x_+} F(y)$, which always exist since F is increasing. We will for convenience define $F(a^-) = F(a)$ and $F(b^+) = F(b)$. If F was continuous at x, then this implies that $F(x^-) = F(x^+)$, otherwise there is a "jump" discontinuity at x: $F(x^-) < F(x^+)$. We will see that there are not too many of these. Namely there are at most countably many of these jump discontinuities for an increasing bounded function.

Lemma (Bounded increasing function has at most countable number of discontinuitties)**.** *If F is a bounded increasing function, then it has at most a countable number of discontinuities.*

Proof. If F is discontinuous at x_0, then by definition we have that $\lim_{x \to x_0^-} F(x) = F(x_0^-) < F(x_0^+)$. And if $I_{x_0} = (F(x_0^-), F(x_0^+)) \subset [F(a), F(b)]$. We can pick a rational number in this non-empty open interval $r_{x_0} \in I_{x_0} \cap \mathcal{Q}$ (just intersect our interval with the set of rationals).

If F is also discontinuous at $y_0 > x_0$ a number greater then our current discontinuity, then since F is increasing, we must have the following:

$$\lim_{y \to y_0^+} F(y) = F(y_0^+) > F(x_0^+)$$

and if $F(y_0^-)$ is the left hand limit, then the two intervals defining the discontinuity for x_0 and y_0 form disjoint intervals. $(F(x_0^-), F(x_0^+)) \cap (F(y_0^-), F(y_0^+)) = \phi$. Now if we pick one rational $r \in \mathcal{Q}$ in every interval $(F(x^-), F(x^+))$, such that F is discontinuous at x.

Then there is a 1-1 mapping from a subset of $\mathcal{Q}$ to the set of discontinuities. Therefore, there is a countable number of discontinuities. □

In the middle of a jump at a discontinuity point, the function does not necessarily have to equal any of its endpoints at x_n. So we instead define it as some $\theta_n \alpha_n$, where $\theta_n \in [0, 1]$ and α_n is the height of the jump between its left and right limits.

5 Hilbert Spaces

5.1 The L2 space: square integrable functions

The space of square integrable functions L^2 turns out to be very important. It contains many symmetry properties, namely an inner product, which will help us define the notion of orthogonality. It is very important in Fourier analysis: namely it is complete (compared to the space of Riemann integrals). Since it is complete, it will be very useful compared to the space of square Riemann integrable functions.

Definition 5.1 (L2 - norm)**.** The $L^2(\mathbb{R}^d)$ is all complex-valued functions f, satisfying:

$$\int_{\mathbb{R}^d} |f(x)|^2 dx < \infty$$

We can also define an inner product, which provides a very important structural property. The inner product in L2 is defined as:

$$(f, g) = \int_{\mathbb{R}^d} f(x) g(\bar{x}) dx$$

If $f, g \in L^2(\mathbb{R}^d)$. Note that $(f, f) = ||f||^2_{L^2}$. That is the norm for L2 is derived from the inner product of the space. There are a number of properties for L2 that are important:

Proposition. *The following are properties of the L2 space:*

i) L2 is a vector space. ii) $f(x)g(\bar{x})$ is integrable whenever $f, g \in L^2$. Moreover, the Cauchy-Schwartz inequality is valid:

$$|(f, g)| \leq ||f||_2 ||g||_2$$

iii) If $g \in L^2$ is fixed, then the map $f \mapsto (f, g)$ is linear and also $(f, g) = (\bar{g, f})$.

iv) The triangle inequality is valid, since L2 defines a norm.

Proof. If $f \in L^2$ and $\lambda \in \mathcal{C}$, then $\lambda f \in L^2$. Next, we want to show $f, g \in L^2$, then $f + g \in L^2$. If we consider $|f(x) + g(x)| \leq 2max(|f(x)|, |g(x)|)$, then if we square both sides we maintain the inequality:

$$|f(x) + g(x)|^2 \leq 4(|f(x)|^2 + |g(x)|^2)$$

Then by the definition of the L2 norm:

$$\int |f + g|^2 \leq 4(\int |f|^2 + \int |g|^2) < \infty$$

since both f and g are by themselves L2.

For ii), we recall that $A, B \geq 0$ implies $AB \leq \frac{1}{2}(A^2 + B^2)$. Then we plug in for our f and g:

$$\int |f\bar{g}| \leq \frac{1}{2}(\int |f|^2 + \int |g|^2) = \frac{1}{2}(||f||^2 + ||g||^2) < \infty$$

So therefore $f\bar{g} \in L^1$.

To show that the Cauchy-schwartz inequality holds, we consider if $||f||_2 = 0$, or $||g||_2 = 0$, then it is trivial: fg = a.e. and hence (f,g) = 0. So WLOG, assume that ||f|| and ||g|| are not zero.

We first normalize each f and g, such that:

$$\tilde{f} = f/||f||$$

and the same for g, such that $||\tilde{f}|| = ||\tilde{g}|| = 1$. Then note that we have that $|(f, g)| \leq 1$ following the previous inequality. Now, if we multiply both sides by $||f||||g||$, then we get:

$$|(f, g)| \leq ||f||||g||$$

□

We next seek to define the notion of a limit in the space of $L^2(\mathbb{R}^d)$. The norm on L2 induces a metric, as follows:

$$d(f, g) = ||f - g||_{L^2}$$

So, with a distance metric, we can define the notion of a **Cauchy sequence** $\{f_n\} \subset L^2(\mathbb{R}^d)$, such that:

$$d(f_n, f_m) \to 0$$

as n, m go to infinity. This sequence converges to $f \in L^2$ if $d(f_n, f) \to 0$ as $n \to \infty$.

Next, we will show that the L^2 space is complete with the metric defined by its norm.

Theorem (Completeness of the $L^2(\mathbb{R}^d)$ space)**.** *The space L^2 is complete in its metric.*

Namely this states that every Cauchy sequence in L^2 converges to a function in L^2. This is different compared to the space of square Riemann integrable functions, which does not have completeness.

Proof. We consider a Cauchy sequence in L2, $\{f_n\}_{n=1}^{\infty}$, and consider a subsequence $\{f_{n_k}\}_{k=1}^{\infty}$ of $\{f_n\}$ with the following property:

$$||f_{n_{k+1}} - f_{n_k}|| \leq 2^{-k}$$

for all $k \geq 1$. So the subsequence is Cauchy with an exponential factor.

We then consider the series:

$$f(x) = f_{n_1}(x) + \sum_{k=1}^{\infty}(f_{n_{k+1}}(x) - f_{n_k}(x))$$

and $g(x)$ is the series with absolute values.

Next, we consider the partial sums of each of the series to demonstrate an upper bound. We can then take $K \to \infty$ and apply the MCT to prove that $\int |g|^2 < \infty$.

$$S_K(f)(x) = f_{n_1}(x) + \sum_{k=1}^{K}(f_{n_{k+1}}(x) - f_{n_k}(x)) = f_{n_{K+1}}(x)$$

and

$$S_K(g)(x) = |f_{n_1}(x) + \sum_{k=1}^{K}|(f_{n_{k+1}}(x) - f_{n_k}(x))| = |f_{n_{K+1}}(x)|$$

based on the fact that these are telescoping series. We can then apply the triangle inequality to get:

$$||S_K(g)|| \leq ||f_{n_1}|| + \sum_{k=1}^{K}||f_{n_{k+1}} - f_{n_k}|| \leq ||f_{n_1}|| + \sum_{k=1}^{K}2^{-k}$$

We take the limit as K goes to infinity and apply the MCT to show that since $\int |g|^2 < \infty$ by construction here (f_{n_1} is L1 and hence norm is finite, and the second term is a geometric series). Then since $|f| \leq g$, then $f \in L^2$. Hence the Cauchy sequence converges to a function in L2.

Since the series defining f, converge to f a.e., then that means that $S_K(f(x))$ converges to f a.e. Furhtermore, we note that:

$$|f - S_K(f)|^2 \leq (2g)^2$$

for all K, and since g is integrable, then by the DCT, we have that $||f_{n_k} - f|| \to 0$ as k tends to infinity. So this demonstrates that the subsequence converges in L2. We want that the whole sequence converges, so we next consider that sequence. We know that it is Cauchy, and so given a $\epsilon > 0$, there exists a N such that:

$$||f_n - f_m||_2 < \epsilon, \ \forall m, n > N$$

If n_k is large such that $n_k > N$, then we can have:

$$||f_{n_k} - f||_2 < \epsilon$$

then we can now write using the triangle inequality:

$$||f_n - f|| \leq ||f_n - f_{n_k}|| + ||f_{n_k} - f|| \leq 2\epsilon$$

as long as $n > N$. □

Next, we can state another important property of L2. Namely, it is separable, implying that their linear combinations of countable collections of L2 functions are dense in L^2.

Theorem (Separability of the L2 space)**.** *The space $L^2(\mathbb{R}^d)$ is separable: there exists a countable collection $\{f_k\}$ of elements in L^2, such that linear combinations of these elements are dense in L^2.*

Proof. □

5.2 L2 to Hilbert Spaces

We can now define an abstract L2 space known as the Hilbert space. A Hilbert space is simply taking all the above properties for L2 and combining them into a set of 6 properties:

1.

5.2.1 Orthogonality in Hilbert Spaces

A very important property of Hilbert spaces is the notion of orthogonality, which is a consequence of the inner product defined in this space. Two elements, f and g in hilbert space H are orthogonal if:

$$(f, g) = 0$$

that their inner product is zero.

In this proposition, we will see that orthogonality is explicitly linked to the L2 norm via the Pythagorean theorem.

Proposition (Pythagoran Theorem in Hilbert Spaces)**.** *If $f \perp g$, then*

$$||f + g||^2 = ||f||^2 + ||g||^2$$

.

Proof. When $(f, g) = 0$, then $(g, f) = 0$ as well, so we can write:

$$||f + g||^2 = (f + g, f + g) \tag{12}$$
$$= ||f||^2 + (f, g) + (g, f) + ||g||^2 \tag{13}$$
$$= ||f||^2 + ||g||^2 \tag{14}$$

□

For a finite or countably infinite subset, $\{e_1, e_2, ...\}$ of a hilbert space, we can say that they are orthonormal if their inner product is equal to 1 when same vector and 0 otherwise. That is each e_k has unit norm and is orthogonal to all other vectors.

Orthonormality has four equivalent characterizations in the context of an orthonormal basis for H.

Theorem (Equivalent conditions for an orthonormal set)**.** *The following properties of an orthonormal set $\{e_k\}$ in a Hilbert space are equivalent:*
i) Finite linear combinations of elements in $\{e_k\}$ are dense in H.
ii) If $f \in H$, and $(f, e_j) = 0$ for all j, then $f = 0$.
iii) If $f \in H$ and $S_N(f) = \sum_{k=1}^{N} a_k e_k$ where $a_k = (f, e_k)$ then $S_N(f) \to f$ as N goes to infiniity in the norm.
iv) If $a_k = (f, e_k)$, then $||f||^2 = \sum_{k=1}^{\infty} |a_k|^2$

As a result of this theorem, two important inequalities and identities come out. First, Bessel's inequality holds for any orthonormal family set $\{e_k\}$:

$$\sum_{k=1}^{\infty} |a_k|^2 \leq ||f||^2$$

They are equal when $\{e_k\}$ is an orthonormal basis for the Hilbert space. The identitys is known as **Parsevals identity**.

$$\sum_{k=1}^{\infty} |a_k|^2 = ||f||^2$$

Now, one might ask, when does a basis exist for a Hilbert space, such that Parseval's identity holds? The answer in this theorem is that any hilbert space **has an orthonormal basis**.

Theorem (Existence of an orthonormal basis in Hilbert spaces)**.** *Any Hilbert space, H has an orthonormal basis.*

5.2.2 Unitary Mappings in Hilbert Spaces

Unitary mappings in Hilbert spaces are similar to the unitary matrices in finite-dimensional vector spaces (i.e. matrices). They preserve norm, are linear, have eigenvalues of absolute value 1 and are bijections (e.g. invertible; one-to-one and onto mappings). Namely a mapping $U : H \to H'$ is called unitary if:

i) U is linear; $U(\alpha f + \beta g) = \alpha U(f) + \beta U(g)$

ii) U is a bijection

iii) $||Uf||_{H'} = ||f||_H$ for all $f \in H$

By definition of being a bijection, there exists an inverse operator of U, which is also unitary. By preservation of the norm in property iii, Unitary operators are linear with respect to inner product:

$$(Uf, Ug)_{H'} = (f, g)_H$$

for all $f, g \in H$.

We can polarize the inner product for F and G that are elements of our space:

$$(F, G) = \frac{1}{4}(||F + G||^2 - ||F - G||^2 + i(||F/i + G||^2 - ||F/i - G||^2))$$

We say that H and H' that are two Hilbert spaces are unitarily equivalent if there is a unitary mapping $U : H \to H'$. As a consequence of the above results, the following corollary states that any two infinite-dimensional Hilbert spaces are unitarily equivalent.

Corollary (Unitary equivalence of infinite-dimensional Hilbert spaces)**.** *Any two infinite-dimensional Hilbert spaces are unitarily equivalent.*

Proof. If H and H' are two infinite-dimensional Hilbert spaces, we can select for each an orthonormal basis. Say these are:

$$\{e_1, e_2, ...\} \in H$$

and

$$\{e'_1, e'_2, ...\} \in H'$$

Then we consider the following mapping for $f = \sum_{k=1}^{\infty} a_k e_k$:

$$U(f) = g$$

where $g = \sum_{k=1}^{\infty} a_k e'_k$. The mapping U is linear and invertible. By Parseval's identity:

$$||Uf||_{H'}^2 = ||g||_{H'}^2 = \sum_{k=1}^{\infty} |a_k|^2 = ||f||_{H'}^2$$

Hence, g is unitarily equivalent to f. □

As a result, all infinite-dimensional Hilbert spaces are unitarily equivalent to $L^2(\mathcal{N})$. In addition, finite-dimensional Hilbert spaces are unitarily equivalent to $\mathcal{C}^d$ if they the same dimension.

Another important notion is a pre-Hilbert space. These are useful for constructing a space H_0 thathas all the properties of a Hilbert space except completeness (e.g. space of Riemann integrable functions). We will see that these pre-Hilbert spaces can be **completed** to form a Hilbert space.

Proposition (Completion of a pre-Hilbert space)**.** *Suppose H_0 is a pre-hilbert space with inner product $(.,.)_0$. Then we can find a Hilbert space H with inner product $(.,.)$ such that:*
i) $H_0 \subset H$
ii) $(f,g)_0 = (f,g)$ whenever $f,g \in H_0$
iii) H_0 is dense in H

Proof. The proof by construction follows Cantor's method of obtaining real numbers as the completion of the rationals in terms of Cauchy sequences of rational numbers. □

5.3 Fourier Series and Fatou's Theorem

When considering Fourier series, we generally consider a larger class of functions, namely the class of integrable functions on $[-\pi,\pi]$. Note that $L^2([-\pi,\pi]) \subset L^1([-\pi,\pi])$ by the Cauchy-Schwartz inequality. We define the nth Fourier coefficient of f as:

$$a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}dx$$

The Fourier series of f is then formally $\sum_{n=-\infty}^{\infty} a_n e^{inx}$, so e^{inx} forms an orthonormal basis for $L^2([-\pi,\pi])$. The next theorem connects integrability with results for the Fourier coefficients and a notion called Abel summability of the Fourier series.

Theorem (Fourier coefficients are unique and Abel summable when f is L1)**.** *Suppose f is integrable on $[-\pi,\pi]$, then:*
i) If $a_n = 0$ for all n, then $f(x) = 0$ for a.e. x
ii) Almost everywhere Abel summability holds: $\sum_{n=-\infty}^{\infty} a_n r^{|n|} e^{inx} \to f(x)$ for a.e. x as $r \to 1$ for $r <$ 1

Proof. We first prove Abel summability, and then demonstrate that the first conclusion follows. In order to prove it, we will make use of the **Poisson kernel**, which is defined on the unit disc in the complex plane as:

$$P_r(y) = \sum_{n=-\infty}^{\infty} r^{|n|}e^{iny} = \frac{1-r^2}{1-2rcos(y)+r^2}$$

Since this is defined for $n \in [-\infty,\infty]$, we simply extend our domain of $f \in L^1([-\pi,\pi])$ to be 2π periodic so that it is defined on the real line. Then we claim that our Abel sum power series is equal to the convolution of the Poisson kernel with our function f.

$$\sum_{n=-\infty}^{\infty} a_n r^{|n|} e^{inx} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-y) P_r(y) dy$$

Note that $P_r(y)$ is a periodic function and bounded, then apply the DCT to pull the infinite sum outside the integral when applying the limit:

$$\sum_{n=-\infty}^{\infty} r^{|n|} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-y) e^{iny} dy$$

Then applying the translation invariance of the Lebesgue integral and the periodicity of f and e^{inx}, we get:

$$\int_{-\pi}^{\pi} f(x-y)e^{iny}dy = \int_{-\pi+x}^{\pi+x} f(y)e^{in(x-y)}dy = e^{inx}\int_{-\pi}^{\pi} f(y)e^{-iny}dy = e^{inx}2\pi a_n$$

where the last equality follows by the definition of the Fourier coefficients. Then combining the two:

$$\sum_{n=-\infty}^{\infty} a_n r^{|n|} e^{inx} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-y) P_r(y) dy$$

Note that in Thm 2.1, we showed that for kernels $\{K_\delta\}$ that is an approximation to the identity and f is integrable, then $(f * K_\delta)(x) \to f(x)$ as $\delta \to 0$. Since the Poisson kernel is an approximation to the identity, we can simply do a change of variables to say $\delta = 1 - r$, then as $r \to 1$, then $\delta \to 0$, and $P_\delta(y)$ will serve as our approximation to the identity.

The first conclusion follows, since now this power series converges to f(x) a.e. $\square$

Since $L^2 \subset L^1$ and Hilbert spaces are L^2, we now seek a similar conclusion for $f \in L^2$ functions.

Theorem (Fourier series in L2 spaces)**.** *Suppose $f \in L^2([-\pi, \pi])$, then we have:*
i) Parseval's relation (essentially Parseval's identity)

$$\sum_{n=-\infty}^{\infty} |a_n|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx$$

ii) The mapping $f \mapsto \{a_n\}$ is a unitary correspondence between $L^2([-\pi, \pi])$ and $l^2(\mathcal{Z})$. That is the mapping from function to fourier coefficients is a unitary map.

iii) The Fourier series of f converges to f in the L^2 norm.

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x) - S_N(f)(x)|^2 dx \to 0$$

as N goes to infinity with $S_N(f) = \sum_n a_n e^{inx}$.

Proof. From our previous theorem, we characterized equivalent properties of an orthonormal set $\{e_k\}$ in a Hilbert space. We can construct an orthonormal set here by taking the exponentials $\{e^{inx}\}$. Since $\{e_k\}$ is defined on the natural numbers, we just have to map $[-\infty, \infty]$ to $\mathcal{N} = \{1, ..., \infty\}$. We simply saay $k = 2n$ for n $>$ 0 and $k = 2|n| - 1$ when n $<$ 0, so k is even when n is positive and odd when n is negative, and then k=1 when n = 0.

Since f is integrable (since $f \in H$), then $a_n = 0$ for all n implies f = 0 a.e. Now $(f, e_j) = a_j = 0$ for all j implies f = 0, which means that all other equivalent characterizations of the orthonormal set we made hold. As a result, we have Parseval's identity, and also convergence in L2 norm of the Fourier series to f. □

5.4 Closed Hilbert Subspaces and Orthogonal Projections

A linear subspace $S \subset H$ is a subset of H that is also a vector space. We say that S is closed if whenever $\{f_n\} \subset S \to f \in H$, then f also is in S.

Note that finite-dimensional Hilbert spaces have the property that all subspaces are closed. However, in infinite-dimensional Hilbert spaces, that can fail. For example, the space of Riemann integrable functions is not closed.

Due to the orthogonality and geometric structure imposed on Hilbert spaces from the inner product, we have the following lemma for closed subspaces of H, our Hilbert space. First, let us restate the **parallelogram law**, which states that in Hilbert spaces we have the following:

$$||f + g||^2 + ||f - g||^2 = 2(||f||^2 + ||g||^2),\ \forall f, g \in H$$

Lemma. *Suppose S is closed subspace of H and $f \in H$, then:*

i) There exists a unique element $g_0 \in S$ that is closest to f in norm:

$$||f - g_0|| = \inf_{g \in S} ||f - g||$$

ii) The element $f - g_0$ is perpendicular to S:

$$(f - g_0, g) = 0,\ \forall g \in S$$

Proof. WLOG we assume $f \notin S$, since otherwise we pick $g_0 = f$. We then let $d := \inf_{g \in S} ||f - g||$ as the minimum distance between f and the set S. Now, we consider a sequence $\{g_n\} \in S$ such that:

$$||f - g_n|| \to d$$

as $n \to \infty$. Now, we claim that this sequence is a Cauchy sequence whos limit will be the desired g_0.

Note that in finite-dimensional cases, S being closed means S is compact, and hence Cauchy sequences converge. In infinite-dimensional case, compactness need not hold, so we proceed by using the parallelogram law in Hilbert space H:

$$A = f - g_n$$

and

$$B = f - g_m$$

then we get: $||f - g_n + f - g_m||^2 + ||g_m - g_n||^2 = 2(||f - g_n||^2 + ||f - g_m||^2)$. Since S is a subspace, then it is a vector space and hence closed under addition, and so $(g_n + g_m) \in S$, so we get that:

$$||2f(g_n + g_m)|| = 2||f - \frac{1}{2}(g_n + g_m)|| \geq 2d$$

by homogeneity and is at least 2 times d. Next, we can bound $||g_m - g_n||^2$ as such:

$$\begin{aligned} ||g_m - g_n||^2 &= 2(||f - g_n||^2 + ||f - g_m||^2) - ||2f - (g_n + g_m)||^2 && (15) \\ &\leq 2(||f - g_n||^2 + ||f - g_m||^2) - 4d^2 && (16) \end{aligned}$$

Since by construction both sequences g_n, g_m have the property that $||f - g_k|| \to d$, then we get that:

$$\lim_{n,m\to\infty} ||g_m - g_n||^2 \leq 2(d^2 + d^2) - 4d^2 = 0$$

and so $\{g_n\}$ is a Cauchy sequence. Since H is complete, then $S \subset H$ is closed, then this Cauchy sequence converges to some element $g_0 \in S$.

Before showing uniqueness, we demonstrate the orthogonality property of the difference. We WTS if $g \in S$ then $g \perp (f - g_0)$. For each ϵ, we consider the perturbation of g_0 as $g_0 - \epsilon g$. Note that this element belongs to S and hence:

$$||f - (g_0 - \epsilon g)||^2 \geq ||f - g_0||^2$$

by the fact that g_0 is the element in S closest to f.

Next, we see that:

$$||f - (g_0 - \epsilon g)||^2 = ||f - g_0||^2 + \epsilon^2||g||^2 + 2\epsilon Re(f - g_0, g)$$

by the linearity of the inner product in the second argument, and the expansion of the norm. By the previous inequality:

$$\epsilon^2||g||^2 + 2\epsilon Re(f - g_0, g) \geq 0$$

So if $Re(f - g_0, g) < 0$, then taking a small $\epsilon > 0$ contradicts the inequality. On the other hand if $Re(f - g_0, g) > 0$, then it is also a contradiction when taking $\epsilon < 0$ but small. So we can conclude that $Re(f - g_0, g) = 0$. Likewise by considering $g_0 - i\epsilon g$, we obtain the same for the imaginary component. This shows that $f - g_0$ is orthogonal to g.

This g_0 chosen is unique since the orthogonality of $(f - g_0)$ and some other $g = g_0 - g'$. Say g' also minimized $||f - g||$. g is in S and hence perpendicular to $(f - g_0)$, and so by the Pythagorean theorem, we'll get:

$$||f - g'||^2 = ||f - g_0||^2 + ||g_0 - g'||^2$$

Since the first two are equal by assumption $||g_0 - g'|| = 0$ and hence $g_0 = g'$. $\square$

With orthogonality of elements with an entire subspace, now one can also define the **orthogonal complement** of S as:

$$S^{\perp} = \{f \in H : (f, g) = 0, \text{for all } g \in S\}$$

Note that the orthogonal complement is also closed, since H is a complete space. The next proposition states that every element of H can be written as a unique sum of an element in S and an element in its orthogonal complement.

Since $H = S \bigoplus S^{\perp}$, then one can define the projection onto S:

$$P_S(f) = g$$

where $f = g + h$, with $f \in H$ and $g \in S, h \in S^{\perp}$. So the projection operator takes an element in our Hilbert space and subtracts the orthogonal complement. The mapping P_S is called an orthogonal projection and has the properties:

1. $f \mapsto P_S(f)$ is a linear operator
2. $P_S(f) = f$ whenever $f \in S$.
3. $P_S(f) = 0$ whenever $f \in S^{\perp}$
4. $||P_S(f)|| \leq ||f||$ for all $f \in H$

5.5 Linear Transformations on Hilbert Spaces

We next look at transformations between two Hilbert spaces, H_1 and H_2. Namely linear transformations / operators $T : H_1 \to H_2$. T is a linear operator if:

$$T(af + bg) = aT(f) + bT(g)$$

for all scalars a,b and $f, g \in H_1$. The following lemma relates the norm of T with the inner product on H_2.

Lemma. *For T a bounded linear operator,* $||T|| = \sup\{|(Tf, g)| : ||f|| \leq 1,\ ||g|| \leq 1\}$.

Proof. Since T is a bounded operator, then $||T|| \leq M$, and so the Cauchy-Schwartz inequality gives:

$$|(Tf, g)| \leq ||Tf|| ||g|| \leq M||f|| ||g||$$

If $||f|| \leq 1$ and $||g||le1$, then the inner product is bounded by M. So then:

$$\sup |(Tf, g)| \leq ||T||$$

Note then we can show that $||Tf|| \leq M||f||$ for any f. We simply normalize f by ||f||. □

Next, we see that linear transformation T is contuous if $T(f_n) \to T(f)$ whenever $f_n \to f$. Next we see that boundedness and continuity are equivalent characterizations for linear operators.

Proposition (Boundedness is equivalent to continuity of linear operators)**.** *A linear operator $T : H_1 \to H_2$ is bounded if and only if it is continuous.*

Proof. If T is bounded, then $||T(f) - T(f_n)||_{H_2} \leq M||f - f_n||_{H_1}$ and so T is continuous, since controlling $||T(f) - T(f_n)$ to be smaller then ϵ is possible by making $||f - f_n|| < \delta$ for some δ.

If T is continuous, we suppose for contradiction that T is not bounded. Then there exists for each N, $f_n \neq 0$ such that $||T(f_n)|| \geq N||f_n||$. If we take the element $g_n = f_n/(N||f_n||)$, then it has norm 1/N, which will converge to zero as n goes to infinity. Since T is continuous, then $T(g_n) \to 0$ also, which contradicts the fact that $||T(g_n)|| \geq 1 = N1/N$. Thus T is bounded. □

5.6 Riesz Representation Theorem

This next section will show the Riesz representation theorem. We first define a linear functional l as a linear transformation from a Hilbert space to the underlying field (i.e. Real or complex numbers).

$$l : H \to \mathcal{C}$$

Note that the inner product on H is a linear functional. In addition, by the Cauchy-Schwartz inequality it is bounded and hence continuous. For a fixed $g \in H$, we can define the linear functional

$$l(f) = |(f, g)| \leq ||f||||g|| = M||f||$$

In addition, $l(g) = M||g|| = ||g||^2$, so the norm of $l(g)$ is M, which is ||g||, since l is a linear transformation. The Riesz representation theorem states that every continuous linear functional on a Hilbert space comes from an inner product.

First, let us define the null space of a linear functional as:

$$S = \{f \in H : l(f) = 0\}$$

Theorem (Riesz representation theorem)**.** *Let l be a continuous linear functional on a Hilbert space H. Then there exists a unique $g \in H$ such that:*

$$l(f) = (f, g)$$

for all $f \in H$. Moreover, $||l|| = ||g||$.

This is the null space of l is the subspace of H such that $(f, g) = 0$ for $g \in S$. This is a nice theorem because it tells us given any continuous linear functional on a Hilbert space, we can uniquely represent it as an inner product with some element of our Hilbert space. In addition, the norm of the linear functional is equal to the norm of that element in our Hilbert space.

5.6.1 Adjoints

Adjoints are similar to transposes in matrix analysis, but a generalization. As a consequence of the Riesz representation theorem, there always exists a unique bounded linear adjoint for a bounded linear transformation.

Proposition. *Let $T : H \to H$ be a bounded linear transformation. There exists a unique bounded linear transformation T^* on H with the following properties:*
i) $(Tf, g) = (f, T^*g)$
ii) $||T|| = ||T^*||$
iii) $(T^*)^* = T$

6 Sample Space and Sigma-Fields

6.1 Sample Space

Ω is our sample space that is a nonempty set of elements, called points.

6.2 Sigma-Fields

6.2.1 General Set Theory Closure Properties

Note:
This is the difference:

$$E\ F = E \cap F^c$$

Whereas this is the symmetric difference:

$$E - F = (E\ F) \cup (F\ E)$$

In Chung's Prob Theory book, there are certain closure properties that any nonempty collection $\mathcal{A}$ of subsets of Ω may have:

1. Closure under complementation: $E \in \mathcal{A}$ implies that $E^c \in \mathcal{A}$
2. Closure under binary union:
3. Closure under binary intersection:
4. Closure under finite union:
5. Closure under finite intersection:

6. Closure under countable increasing union:
7. Closure under countable decreasing intersection:
8. Closure under countable union:
9. Closure under countable intersection:
10. Closure under proper set differences:

6.2.2 General Definitions

Fields

Definition 6.1. A field, F is a subset of Ω that is i) closed under complementation and ii) closed under binary union.

Corollary. *A field, F is closed under binary intersection, finite union and finite intersection.*

Proof. To show that is is closed under binary intersection, we use De Morgan's Law.

For $D, E \in F$, then $D^c, E^c \in F$ by closure under complementation. $D \cup E \in F$, so $(D \cup E)^c = D^c \cap E^c \in F$.

To show that it is closed under finite union and finite intersection, one utilizes induction. It is shown for binary union/intersection, so one needs to then assume it is closed under k-1 union/intersection and prove that it is also closed under k union/intersection. □

Monotone Classes (MC) Monotone classes are sets that obey the closure under countable increasing/decreasing union/intersections. That is each set is either monotonically increasing, or decreasing. Hence the name monotone.

The MC theorem is a way of proving something is a σ-field besides it's original definition. One can start with a field and show that it is also a MC, and then that field is a σ-field!

Theorem. *TBD*

π **Systems** π systems are a weaker notion then fields. They are a class of subsets of Ω that are closed under finite intersection (i.e. closed under binary intersection).

Note that a field also has closure under complementation, which allows them to also have closure under finite unions.

λ **Systems** λ systems are a stronger notion then Monotone classes.

They contain the whole sample space, Ω, are closed under proper set differences and are closed under countable increasing unions.

Note that MC are only closed under monotonically increasing/decreasing union/intersections, whereas a λ system is closed under general countable increasing unions.

To connect π and λ systems to our set of interest, the σ-field, we can utilize another theorem that states if a π system is within a λ system, then the σ-field generated by the π system is within the λ system.

Theorem. *TBD*

6.2.3 Generating Fields, MC, Sigma-Fields

6.2.4 Good Sets Argument

7 Probability Measure

7.1 Introduction

Definition 7.1. A probability measure on a field, $\mathcal{F}$ is a set function that has the properties for a probability measure on a σ-field. Countable additivity need only hold if the countable disjoint union $\sum_{j=1}^{\infty} E_j$ of sets $E_j \in \mathcal{F}$ belongs to $\mathcal{F}$.

These are the

1. positivity: $\forall E \in \mathcal{F}$, we have P(E) ≥ 0
2. countable additivity: If $\{E_j\}$ is a countable collection of pairwise disjoint sets in $\mathcal{F}$, then we get:

$$P(\bigcup_{j=1}^{\infty} E_j) = \sum_j P(E_j)$$

3. unitary: $P(\Omega) = 1$

Definition 7.2. A set function like in a probability measure that has finite additivity instead of countable additivity.

Next, we can talk about traces of σ-fields. Traces are usually defined on sub-sets of the sample space, Ω. So we say $\mathcal{F}$ is a σ-field of subsets of Ω. Let $\Omega_0 \subset \Omega$, that need not belong to $\mathcal{F}$, then we can define the trace of the original σ-field:

$$\mathcal{F}_0 := \mathcal{F} \cap \Omega_0$$

which is a σ-field of subsets of Ω_0.

7.1.1 Axioms of Probability Theory

There are three main accepted axioms of probability theory:

1. positivity: $\forall E \in \mathcal{F} : P(E) \geq 0$
2. countable additivity for pairwise disjoint sets: $P(\bigcup_j E_j) = \sum_j P(E_j)$
3. normalization: $P(\Omega) = 1$

Note that uncountable additivity would not be well defined because i) there is no notion of an uncountable sum and ii) it would imply that all events have probability zero.

Basic Properties of Probability Measure 1. Upper-bounded by 1 2. $P(\Phi) = 0$ 3. $P(E^c) = 1 - P(E)$ 4. 5. Monotonicity: If $E \subset F$, then $P(E) = P(F) - P(F\ E) \leq P(F)$ 6. Monotone sequential property: If $E_n \to E$ from below/above, then $P(E_n) \to P(E)$ 7. Boole's Inequality (countable sub-additivity): $P(\bigcup_j E_j) \leq \sum_j P(E_j)$

Boole's inequality is also known as countable sub-additivity; note that it does not require any disjointness on the sets.

7.1.2 Probability Measures on Fields and Sigma-Fields

Before analyzing the probability measures on a σ-field, we first look at how things change if we only define probability measures on a field. First note that countable additivity need not hold anymore since only σ-fields are closed under countable unions, while fields are simply closed under **finite** unions.

Definition 7.3. A pre-probability measure (prepm) is a probability measure, but instead of countable additivity, we have finite additivity.

How can one extend a prepm on a field to a probability measure on that field? Using the following Lemma, one can extend a weaker notion of a prepm into a pm.

Lemma. *Suppose $\mathcal{F}$ is a field and P is a pre-pm on $\mathcal{F}$. Then the following are equivalent:*

i) P is a pm. on the field, $\mathcal{F}$ ii) Monotone sequential continuity from above (MSCA) at Φ: $A_n \to \Phi, A_n \in \mathcal{F}$, then $P(A_n) \to 0$ iii) MSCA at general A: iv) Monotone sequential continuity from below (MSCB): v) countable additivity:

Note that countable additivity is part of the pm definition! But MSCA at Φ, MSCA and MSCB are ways of taking a prepm and turning it into a pm without directly assuming countable additivity.

ii) implies i). ii) implies iii) via closure under complementation. iii) implies iv) by constructing disjoint unions from increasing unions (and vice versa) iv) implies v) by definition of the pm.

This is interesting because by the "Extension Theorem", one can uniquely and always extend a pm on a field to a pm on a σ-field.

Traces of a field, and σ-fields This is the "process" of tracing down a field to a subset. Can we go from $(\Omega, \mathcal{F}, P)$ to $(\Omega_0, \mathcal{F}_0, P_0)$. To get $\mathcal{F}_0$, you can simply take the intersection of subsets of the original field, $\mathcal{F}$ with the subset of the sample space, Ω_0:

$$\mathcal{F}_0 := \{F \cap \Omega_0 : F \in \mathcal{F}\}$$

Note that $\mathcal{F}_0$ is a σ-field of the subsets of Ω_0 and is called the **trace** of $\mathcal{F}$.

Remember we can "generate" σ-fields, and so we can also trace the sets that generate these σ-fields.

Now, we want to define a probability measure on the sample space and σ-field. We want to define a set function P on $\mathcal{B}_0$, by setting this:

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

This setup ensures that probability obeys our intuition on the unit interval and also makes P well-defined. In addition, it obeys the positivity, unitary axioms and is finitely additive. With finite-additivity, P is a pre-pm on the field $\mathcal{B}_0$.

Next, we want to show that P is a full pm on the field. To do that, all we need to do is check one of the conditions in 7.1.2.

So, P is a pm on a the field, $\mathcal{B}_0$. P will extend uniquely to a probability measure on $\mathcal{B}$ by the Extension Theorem.

7.2 Uniform Probability on (0, 1]

Now that we have defined sample spaces, Borel (sigma) fields, and probability measures, then we want to define the probability density that occurs on the unit interval. We start by building the probability measure on a sub-set of the sigma-field and then "extend" this upwards in size.

7.2.1 Sample Space

Our sample space is defined on the unit interval.

$$\Omega = (0, 1]$$

7.2.2 σ-field

What is our Borel field for $\mathcal{F}$? Should we use the total σ-field? No this ends up being too large of a field. Some remarks on the Borel σ-field.

1. It contains the set of Normal numbers defined by the Rademacher functions
2. It contains the open sets in (0, 1].

Note on Rademacher functions and the set of "normal numbers":

N is the set of normal numbers, where $s_n(\omega)$ is the sum of the first n Rademacher functions. ω is a point that lies in N if and only if $\lim_n n^{-1} s_n(\omega) = 0$.

7.3 Extension Theorem

The extension theorem states that a probability measure on a field has a unique extension to the generated σ-field. This is useful because it allows us to define probability measures on fields, and then automatically extend them and know that the extension is unique!

7.3.1 Probability Measures

The set function, P that we want to be a pm on the $\sigma(F_0)$ can be perhaps measured using our notion of the outer (exterior) measure.

$$P^*(A) = \inf \sum_n P(A_n))$$

By taking an infimum over all countable sequences of the $\mathcal{F}_0$ sets, where the sequence of sets form a covering of A. The measure really can be thought of here as the "volume" of A. Now, we want this set function to obey the rules of probability measures in such a way:

$$P^*(A) = 1 - P^*(A^c)$$

and

$$P^*(A) + P^*(A^c) = 1$$

In order to make this work, people figured out that we can impose the following requirement on P^*:

$$P^*(A \cap E) + P^*(A^c \cap E) = P^*(E)$$

holds for every set E. We will call **M** the class of such sets with the above properties.

Properties of class M