Unreal Analysis 2 - Lecture Notes

as taught by Prof. John Toth, McGill Winter 2020 (well, mostly)

Viet Nguyen

Contents

Foreword

One can consider the theory of weights of one's balls, and deduce (after months of tedious calculations) that the weight of the balls of the author of this series of notes is not of bounded variation.

1 Introduction

This section serves as a brief display of preliminary materials to a thorough understanding of this course's bs material. We briefly introduce the preliminaries to measure theory and integration, strong requisites for the rest of the course. Note that this section is not part of the course.

1.1 Notation

The Lebesgue measure will be denoted μ instead of some sketchy λ or m, because I like Greek letters, and $d\lambda$ doesn't look aesthetic (unless when truly necessary, like in the subsection about BV functions).
C^0 is the class of continuous functions, C^k is the class of functions such that every derivative up to the k-th derivatives exist and are continuous.

1.2 Measure Theory

Definition 1.1. A **measurable space** $(X, \mathcal{M})$ is a set X together with a family $\mathcal{M}$ of subsets of X, called a σ-algebra or σ-field, satisfying the following axioms:

1. $\emptyset \in \mathcal{M}$
2. $A \in \mathcal{M} \implies A^c \in \mathcal{M}$
3. if $\{A_i \in \mathcal{M}\}_{i \in I}$ is a countable family, then $\bigcup\limits_{i \in I} A_i \in \mathcal{M}$

Requiring only finite additivity, we get an **algebra** or **field**.

Proposition 1.2. The intersection of an arbitrary collection of σ-algebras on a set X is a σ-algebra on X.

Corollary 1.3. Given any $\mathcal{B} \subset \mathcal{P}(X)$, there exists a least σ-algebra containing $\mathcal{B}$.

We will denote this least σ-algebra generated by $\mathcal{B}$ by $\sigma(\mathcal{B})$.

Definition 1.4. Given 2 measurable spaces $(X, \mathcal{M}_X)$ and $(Y, \mathcal{M}_Y)$, a function $f : \mathcal{M}_X \longrightarrow \mathcal{M}_Y$ is said to be measurable if $f^{-1}(B) \in \mathcal{M}_X$ whenever $B \in \mathcal{M}_Y$

Definition 1.5. Given a family of functions $\{f_n : X \longrightarrow Y\}_{n \in \mathbf{N}}$, we say that the family **converges pointwise** to f if $\forall x \in X, \lim_{n \to \infty} f_n(x) = f(x)$.

Theorem 1.6. If a family of measurable functions $\{f_n : X \longrightarrow Y\}_{n \in \mathbf{N}}$ converges pointwise to to f, then f is also measurable.

Definition 1.7. A measurable function is called **simple** if its range is finite.

For real-valued functions, we have the following very important result:

Theorem 1.8. Let $f : X \longrightarrow \mathbf{R}$ be a non-negative measurable function. Then, there exists a family of simple functions $\{s_i\}_{i \in \mathbf{N}}$ such that $\forall i \in \mathbf{N}, s_i \leq s_{i+1} \leq f$ and the $s_i \to f$ pointwise.

We now get to several core definitions.

Definition 1.9. A (positive) **measure** μ on a measurable space $(X, \mathcal{M})$ is a function from $\mathcal{M}$ to $[0, +\infty]$ such that if $\{A_i\}_{i \in I}$ is a countable family of pairwise disjoint sets, then

$$\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i)$$

A set equipped with a σ-algebra and a measure defined on it is called a **measure space**. Below are several properties of measures, namely monotonicity and continuity.

Proposition 1.10.

1. If $A \subseteq B$, then $\mu(A) \leq \mu(B)$.
2. IF $\{A_n\}$ is an increasing sequence of sets, and $\bigcup_n A_n = A$, then $\lim_{n \to \infty} \mu(A_n) = \mu(A)$.
3. IF $\{A_n\}$ is a decreasing sequence of sets, and $\bigcap_n A_n = A$, then $\lim_{n \to \infty} \mu(A_n) = \mu(A)$, provided $\mu(A_1) < \infty$.

We now have the notion of a measure completion, given by the following definition and theorem.

Definition 1.11. A measure space $(X, \mathcal{M}, \mu)$ is **complete** if every subset of a negligible set is in $\mathcal{M}$.

Theorem 1.12. (Extension of measure to a complete measure) Given $(X, \mathcal{M}, \mu)$, there is a σ-algebra $\mathcal{M}' \supseteq \mathcal{M}$ and a measure μ' on $\mathcal{M}'$ such that $(X, \mathcal{M}', \mu')$ is complete and $\mu(A) = \mu'(A)$ for any $A \in \mathcal{M}$.

Remark 1.13. The above completion is not unique.

1.3 Integration

We briefly introduce the notion of Lebesgue integration, a ge'1neralization of Riemann integration. As we shall see, the real reason for the predominance of the Lebesgue integral is that is has a much smoother theory. For instance, under some general conditions, one can interchange limits and integration.
Suppose we have a simple function $s : \mathbf{R} \to \mathbf{R}$, whose range is the set $\{a_1, \ldots, a_n\}$. We define $\forall i \in \{1, \ldots, n\}, A_i \coloneqq s^{-1}(a_i)$. The A_i are measurable sets if s is a measurable function. Thus, we define the **Lebesgue integral**:

$$\int s d\mu = \sum_i a_i \mu(A_i)$$

where μ is the Lebesgue measure.

Definition 1.14. We say that a simple function s is **integrable** if whenever a is in the range of s, $a \neq 0 \implies \mu(s^{-1}(a)) < \infty$.

Definition 1.15. Suppose $(X, \mathcal{M}, \mu)$ is a measure space, and that $s : X \to \mathbf{R}$ is an integrable simple function with range $\{a_1, \ldots, a_n\}$. We say that the **integral** of s over X w.r.t. μ is

$$\int_X s d\mu = \sum_{i=1}^{n} a_i \mu(s^{-1}(a_i))$$

Now, given that we can approximate any measurable function f with a sequence of non-decreasing simple functions $\{s_i\}$ such that $s_i \uparrow f$, we have the following definition motivated:

Definition 1.16. Suppose f is an everywhere nonnegative real-valued function. We say that f is **integrable** if the everywhere nonnegative simple functions less than f are integrable and their integrals are bounded. If f is integrable, we define:

$$\int_X f d\mu = \sup_s \int_X s d\mu$$

where the supremum is over all nonnegative simple functions below f. Given a measurable function g taking both positive and negative values, we say that g is integrable if both g_+ and g_- are integrable, and we set

$$\int_X g d\mu = \int_X g_+ d\mu + \int_X g_- d\mu$$

Example 1.17. We take as our measure space $(X, \mathcal{M}, \delta_x)$ where δ_x is the Dirac measure concentrated at the point $x \in X$. Let f be any nonnegative real-valued function. We claim that

$$\int_X f d\delta_x = f(x)$$

Note that the simple function $s(x) = f(x)$ and 0 everywhere else is a simple function below f. The integral of s w.r.t. δ_x is $f(x)$. Furthermore, any simple function t below f has the integral $t(x) \leq f(x)$. Thus, the supremum of the integrals of all the simple functions below f is precisely $f(x)$.

The next example is the standard advertisement for the superrior generality of the Lebesgue integral.

Example 1.18. Let $f : [0, 1] \to \mathbf{R}$ be given by $f(x) = 0$ if x is rational and $f(x) = 1$ if x is irrational. This f is in fact a simple function, in fact, it is even the characteristic function $\chi_{(\mathbf{R} \setminus \mathbf{Q}) \cap [0,1]}$. Thus, its integral is just the measure of the irrationals between 0 and 1, which is 1.

The following are some Lebesgue integral calculus that one encounters on a daily basis.

Proposition 1.19.

1. If $0 \leq f \leq g$, then $\displaystyle\int f d\mu \leq \int g d\mu$.
2. If $0 \leq f$ and $0 \leq c$ is a constant, then $\int cf d\mu = c \int f d\mu$.
3. $\displaystyle\int_A f d\mu = \int_X f \chi_A d\mu$, where χ_A is the characteristic function of A.

Proposition 1.20. Let s, t be simple integrable functions. Then

$$\int (s+t) d\mu = \int s d\mu + \int t d\mu$$

The following is a pretty broken theorem, that when spammed correctly, brings joy to the life of the mathematician practicing the craft μ-almost everywhere.

Theorem 1.21. (Monotone convergence) Let $\{f_n\}$ be a sequence of measurable functions on X, and suppose that

1. $\forall x \in X, 0 \leq f_1(x) \leq f_2(X) \leq \cdots \leq \infty$
2. $\forall x \in X, \sup_n f_n(x) = f(x)$

Then, f is measurable, and $\displaystyle\sup_n \int_X f_n d\mu = \int_X f d\mu$, i.e. $\displaystyle\lim_{n \to \infty} \int_X f_n d\mu = \int_X \lim_{n \to \infty} f_n d\mu$ (we can write lim instead of sup since the supremum is actually attained).

Now, with this pretty broken theorem, the same statements about Lebesgue integral calculus that applied for simple functions can now be stated in terms of measurable functions, since we can slap some monotone convergence on the little simple functions that approximate it.
A slightly more dank convergence theorem is the following:

Theorem 1.22. (Dominated convergence) Let $\{f_n\}$ and g be integrable functions such that $\forall n \in \mathbf{N}, |f_n| \leq g$ (i.e. the f_n's are dominated by g) and $\lim_{n \to \infty} f_n = f$, then f is integrable and

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

The following is your classic analysis 3 student's pocketknife.

Lemma 1.23. (Fatou's lemma) If $\{f_n\}$ is a sequence of nonnegative real-valued measurable functions, then

$$\int \liminf_{n \to \infty} f_n d\mu \leq \liminf_{n \to \infty} \int f_n d\mu$$

2 Signed Measures and Differentiation

We build the theory of differentiation in the abstract setting first, then obtain a more refined result when μ is the Lebesgue measure in $\mathbf{R}^n$.

2.1 Signed Measures

Let $(X, \mathcal{M})$ be a measurable space. a **signed measure** on $(X, \mathcal{M})$ is a function $\nu : \mathcal{M} \to \bar{\mathbf{R}}$ such that:

- $\nu(\emptyset) = 0$;
- ν assumes at most one of the values $\pm\infty$;

- if $\{E_j\}$ is a sequence of disjoint sets in $\mathcal{M}$, then $\nu\left(\bigcup_1^\infty E_j\right) = \sum_1^\infty \nu(E_j)$, where the latter sum converges absolutely if $\nu\left(\bigcup_1^\infty E_j\right)$ is finite;

Every measure is a signed measure.

Example 2.1. Let μ_1, μ_2 measures on $\mathcal{M}$ and at least one of them is finite. Then, $\nu = \mu_1 - \mu_2$ is a signed measure.

Example 2.2. If μ is a measure on $\mathcal{M}$ and $f : X \longrightarrow \bar{\mathbf{R}}$ is a measurable function such that at least one of $\displaystyle\int f^+ d\mu$ and $\displaystyle\int f^- d\mu$ is finite (we call such an f an **extended μ-integrable** function), then the set function ν defined by $\nu(E) = \displaystyle\int_E f d\mu$ is a signed measure.

Remark 2.3. Every signed measure can be represented in either of these two forms.

Proposition 2.4. Let ν be a signed measure on $(X, \mathcal{M})$. If $\{E_j \in \mathcal{M}\}$ such that $E_j \uparrow E$, then $\nu(\bigcup_1^\infty E_j) = \lim_{j \to \infty} \nu(E_j)$. If instead, $E_j \downarrow E$ and $\nu(E_1)$ is finite, then $\nu(\bigcup_1^\infty E_j) = \lim_{j \to \infty} \nu(E_j)$.

Definition 2.5. If ν is a signed measure on $(X, \mathcal{M})$, $E \subset M$ is **positive** for ν if $\forall F \in \mathcal{M}, F \subset E \implies \nu(F) \geq 0$. E is **negative** if the opposite holds, and **null** if the condition is $\nu(F) = 0$.

Thus, in the example $\nu(E) = \displaystyle\int_E f d\mu$, E is positive, negative, or null precisely when $f \geq 0, f \leq 0$, or $f = 0$ μ-a.e. on E.

Lemma 2.6. Any measurable subset of a positive set is positive, and the union of any countable family of positive sets is positive.

Theorem 2.7. (Hahn Decomposition) Let ν a signed measure on $(X, \mathcal{M})$, then there exists a positive set P and a negative set N for ν such that $P \cup N = X$, and $P \cap N = \emptyset$. If P' and N' is another such pair, then $P \triangle P'$ $(= N \triangle N')$ is null for ν.

The decomposition $X = P \cup N$ of X as the disjoint union of a positive set and a negative set is called the **Hahn decomposition** for ν. It is usually not unique (ν-null sets can be transferred from P to N or from N to P), but it leads to a canonical representation of ν as the difference of two positive measures.
We now need a new concept to state the above. We see that two signed measures μ and ν on $(X, \mathcal{M})$ are **mutually singular**, or that ν is **singular w.r.t.** ν, or vice versa, if there exist $E, F \in \mathcal{M}$ such that $E \cap F = \emptyset, E \cup F = X$, E is null for μ, and F is null for ν. Informally speaking, mutual singularity means that μ and ν "live on disjoint sets", i.e. they are supported on different disjoint subsets:

$$\mu \perp \nu$$

Theorem 2.8. (Jordan Decomposition) If ν is a signed measure, there exist unique positive measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$.

Measures ν^+ and ν^- are the **positive** and **negative variations** of ν. We define the **total variation** of ν to be the measure $|\nu|$ defined by:

$$|\nu| = \nu^+ + \nu^-$$

Example 2.9. $E \subset \mathcal{M}$ is ν-null iff $|\nu|(E) = 0$ and $\nu \perp \mu$ iff $|\nu| \perp \mu$ iff $\nu^+ \perp \mu \wedge \nu^- \perp \mu$.

If $\mathrm{rge}(\nu) \subset \mathbf{R}$, then ν is bounded. Also, ν is of the form $\nu(E) = \displaystyle\int_E f d\mu$ where $\mu = |\nu|$ and $f = \chi_P - \chi_N, X = P \cup N$ being a Hahn decomposition for ν.
Integration w.r.t. a signed measure ν is defined as follows, given $\mathcal{L}^1(\nu) = \mathcal{L}^1(\nu^+) \cap \mathcal{L}^1(\nu^-)$:

$$\int f d\nu = \int f d\nu^+ - \int f d\nu^-, f \in \mathcal{L}^1(\nu)$$

A signed measure ν is **finite** if $|\nu|$ is finite (same goes for σ-finite).

2.2 Radon-Nikodym Theorem

Suppose ν is a signed measure, μ is a positive measure on $(X, \mathcal{M})$. We say that ν is **absolutely continuous** w.r.t μ:

$$\nu \ll \mu$$

if $\nu(E) = 0$ for every $E \in \mathcal{M}$ for which $\mu(E) = 0$. Also, $\nu \ll \mu$ iff $|\nu| \ll \mu$ iff $\nu^+ \ll \mu$ and $\nu^- \ll \mu$.
If $\nu \perp \mu$ and $\nu \ll \mu$, then $\nu = 0$, since if E and F are disjoint sets such that $E \cup F = X$, and $\mu(E) = |\nu|(F) = 0$, then the fact that $\nu \ll \mu$ implies that $|\nu|(E) = 0$, whence $|\nu| = 0$ and $\nu = 0$. One can extend the notion of absolute continuity to the case where μ is a signed measure.

Theorem 2.10. Let ν a finite signed measure and μ a positive measure on $(X, \mathcal{M})$. Then, $\nu \ll \mu$ iff $\forall \epsilon > 0, \exists \delta > 0, \mu(E) < \delta \implies |\nu(E)| < \epsilon$.

Remark 2.11. If μ is a measure and f is an extended μ-integrable function, the signed measure ν defined by $\nu(E) = \displaystyle\int_E f d\mu$ is absolutely continuous w.r.t. μ, finite iff $f \in \mathcal{L}^1(\mu)$. For complex-valued $f \in \mathcal{L}^1(\mu)$, the preceding theorem can be applied to Re f and Im f, and we obtain the following useful result:

Corollary 2.12. If $f \in \mathcal{L}^1(\mu)$, $\forall \epsilon > 0, \exists \delta > 0, \mu(E) < \delta \implies \left| \displaystyle\int_E f d\mu \right| < \epsilon$.

We use the following notation to express the relationship $\nu(E) = \displaystyle\int_E f d\mu$:

$$d\nu = f d\mu$$

We can now express an intrinsic connection between signed measures in relation to a given positive measure. (Well, we can only do so after the next lemma...)

Lemma 2.13. Let μ, ν finite measures on $(X, \mathcal{M})$. Either $\nu \perp \mu$, or $\exists \epsilon > 0, E \in \mathcal{M}, \mu(E) > 0$ and $\nu \geq \epsilon \mu$ on E, i.e. E is a positive set for $\nu - \epsilon \mu$.

Theorem 2.14. (Lebesgue-Radon-Nikodym) Let ν be a σ-finite signed measure and μ a σ-finite positive measure on $(X, \mathcal{M})$. There exist unique σ-finite signed measures λ, ρ on $(X, \mathcal{M})$ such that

- $\lambda \perp \mu$
- $\rho \ll \mu$
- $\nu = \lambda + \rho$

Moreover, there is an extended μ-integrable function $f : X \to \mathbf{R}$ such that $d\rho = f d\mu$, and any two such functions are equal μ-a.e.

The decomposition $\nu = \lambda + \rho$ where $\lambda \perp \mu$ and $\rho \ll \mu$ is called the Lebesgue decomposition of ν w.r.t. μ. In the case where $\nu \ll \mu$, we have that $d\nu = f d\mu$ for some f. This result is usually known as the **Radon-Nikodym theorem**, and f is called the **Radon-Nikodym derivative** of ν w.r.t. μ. We denote it by $d\nu/d\mu$:

$$d\nu = \frac{d\nu}{d\mu} d\mu$$

Remark 2.15. In some literature, the Lebesgue decomposition is written as $\nu = \nu_a + \nu_s$ where $\nu_a \ll \mu$ and $\nu_s \perp \mu$.

We should consider $\frac{d\nu}{d\mu}$ as the **class** of functions that are equal to f a.e. The following is the chain rule for Radon-Nikodym derivatives.

Proposition 2.16. Suppose that ν is a σ-finite signed measure and μ, λ are σ-finite measures on $(X, \mathcal{M})$ such that $\nu \ll \mu \ll \lambda$. Then,

- If $g \in \mathcal{L}^1(\nu)$, then $g(\frac{d\nu}{d\mu}) \in \mathcal{L}^1(\mu)$ and

$$\int g d\nu = \int g \frac{d\nu}{d\mu} d\mu$$

- We have $\nu \ll \lambda$ and

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda} \quad \lambda - a.e.$$

Corollary 2.17. If $\mu \ll \lambda \wedge \lambda \ll \mu$, then $\frac{d\mu}{d\lambda} = \frac{d\lambda}{d\mu} = 1$ a.e.

Proposition 2.18. If $\mu_1 \ldots, \mu_n$ are measures on $(X, \mathcal{M})$, there is a measure μ such that $\mu_j \ll \mu$ for all j, namely $\mu = \sum \mu_j$.

T

2.3 Differentiation in $\mathbf{R}^n$

In this section, we analyze the case where $(X, \mathcal{M}) = (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n})$, and $\mu = m$ is the Lebesgue measure. We define $B(x,r) =$ '$\{y \in \mathbf{R}^n : |x-y| < r\}$ as the open ball around x of radius r, and $|E| := \mu(E)$. Let's supposed ν is another measure on the same measurable space. Then, one can define a *pointwise* derivative of ν w.r.t. μ. Consider:

$$F(x) = \lim_{r \longrightarrow 0^+} \frac{\nu(B(x,r))}{\mu(B(x,r))}, \quad x \in \mathbf{R}^n$$

Now suppose we have $d\nu = f d\mu$, $f \in \mathcal{L}^1(\mathbf{R}^n)$, we rewrite F as

$$F(x) = \lim_{r \longrightarrow 0^+} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} f d\mu = \lim_{r \longrightarrow 0^+} \frac{1}{|B|} \int_B f d\mu$$

F is now simply the average of f on $B(x,r)$, and one would hope that $F = f$ μ-a.e. This is the case when $\nu(B(x,r))$ is finite for all x, r and one readily sees that the derivative of the indefinite integral of f is f. This is in a sense a generalization of FTC part 1 which states:

$$F(x) = \int_a^x f(t)dt \implies \frac{d}{dx}F(x) = f(x)$$

Hence, if $\nu(E) = \displaystyle\int_E f d\mu$, the derivative of ν w.r.t. μ at x is:

$$\frac{d\nu}{d\mu} = \lim_{r \longrightarrow 0^+} \frac{1}{|B|} \int_B f d\mu = f, \quad \mu - a.e.$$

General case: We will show that $F(x)$ exists, and that $F(x) = f(x)$ a.e., provided $\nu(B) < \infty, \forall (x,r) \in \mathbf{R}^n \times \mathbf{R}$. We first need a covering lemma:

Lemma 2.19. (Vitali covering) Let $\mathcal{C}$ a collection of open balls (or cubes) in $\mathbf{R}^n$, set $\mathcal{U} = \bigcup_{B \in \mathcal{C}} B$ ($\mathcal{U}$ is open). Let $c < |\mathcal{U}|$. There exist finitely many disjoint balls $B_1, \ldots, B_k \in \mathcal{C}$ such that $\sum \mu(B_j) \geq 3^{-n}c$.

Proof. Recall given $E \in \mathcal{M}$, $\mu(E) = \sup\{\mu(K) : K \subset E, K \text{ compact}\}$, i.e. can approximate the measure of a measurable set with compact subsets. So, $\exists K \subset \mathcal{U}$ compact such that $c < \mu(K)$. Since K is compact, there exists finitely many open balls $A_1, \ldots, A_m \in \mathcal{C}$ such that $K \subset \bigcup_{i=1}^m A_i$. W.l.o.g., assume $\text{rad}(A_i) \geq \text{rad}(A_{i+1})$, meaning the A_i's are arranged in decreasing order. We have the following construction:

- $B_1 = \max\{A_1, \ldots, A_m\}$
- $B_2 = \max\{A_j : A_j \cap B_1 = \emptyset\}$
- $B_3 = \max\{A_j : A_j \cap B_1 = A_j \cap B_2 = \emptyset\}$

- ...

until the list of A_j's are exhausted. Note that the B_j's in this construction control all the volume. If $A_i \neq B_j$ for any j, then $\exists j_0$ such that $A_i \cap B_{j_0} \neq \emptyset$. Let J be the smallest j_0 with property $A_i \cap B_J \neq \emptyset$, i.e. $\text{rad}(A_i) \leq \text{rad}(B_J)$. Then, $A_i \subset B_j^*$ where $B_j^* = 3B_j$ (triple radius). Then, $K \subset \bigcup_{j=1}^k B_j^*$, and hence:

$$c \leq \mu(K) \leq \sum_{j=1}^{k} \mu(B_j^*) = 3^n \sum_{j=1}^{k} \mu(B_j)$$

∎

Definition 2.20. $f : \mathbf{R}^n \to \mathbb{C}$ measurable is **locally integrable**, denoted $f \in \mathcal{L}^1_{loc}(\mathbf{R}^n)$ if $\forall K \in \mathcal{M}$ such that K compact is bounded, $\displaystyle\int_K |f| d\mu < \infty$.

In view of the above, it is of interest to consider the following function: given $f \in \mathcal{L}^1_{loc}(\mathbf{R}^n), x \in \mathbf{R}^n$, we set $A_r f(x)$ to be the average of f on $B(x,r)$ as follows:

$$A_r f(x) = \frac{1}{|B|} \int_B f d\mu$$

for $(x,r) \in \mathbf{R}^n \times (0,\infty)$. Before taking limits in $A_r f(x)$, we need to prove some continuity properties of $A_r f(x)$:

Lemma 2.21. Given $f \in \mathcal{L}^1_{loc}(\mathbf{R}^n), A_r f(x)$ is jointly continuous in $(x,r) \in \mathbf{R}^n \times (0,\infty)$.

Proof. Remark that $\mu(B(x,r)) = cr^n$, where $c = \mu(B(0,1))$. We take the radius to the power of n since $\dim \mathbf{R}^n = n$. Furtheremore, $\mu(S(x,r)) = \mu(\partial B(x,r)) = 0$. If $(x,r) \to (x_0, r_0)$,$\chi_{B(x,r)} \to \chi_{B(x_0,r_0)}$ pointwise on $\mathbf{R}^n \setminus S(x_0,r_0)$. In particular,

- $\chi_{B(x,r)} \to \chi_{B(x_0,r_0)}$ a.e.
- $|\chi_{B(x,r)}| \leq \chi_{B(x_0,r_0+1)} \in \mathcal{L}^1$ if $r < r_0 + \frac{1}{2}$ and $|x - x_0| < \frac{1}{2}$.

Thus,

$$A_r f(x) = \frac{1}{cr^n} \int_{\mathbf{R}^n} \chi_{B(x,r)}(y) f(y) d\mu(y)$$

and $|\chi_{B(x,r)} f| \leq \chi_{B(x_0,r_0+1)} |f| \in \mathcal{L}^1_{loc}$. By the dominated convergence theorem, it follows that $\displaystyle\int_{B(x,r)} f(y) d\mu(y)$ is continuous in (x,r) and hence, so is $A_r f(x)$. ∎

This motivates the next definition:

Definition 2.22. Suppose $f \in \mathcal{L}^1_{loc}(\mathbf{R}^n)$. Then, the **Hardy-Littlewood maximal function** Hf of f is defined by:

$$Hf(x) := \sup_{r>0} A_r |f|(x) = \sup_{r>0} \frac{1}{|B|} \int_B |f| d\mu$$

Remark 2.23. Hf is measurable, since $(Hf)^{-1}((a,\infty)) = \bigcup_{r>0} (A_r|f|)^{-1}((a,\infty))$ is open for any $a \in \mathbf{R}$, by the previous lemma on continuity.

Essentially, one needs two key tools to prove the Lebesgue differentiation theorem:

1. A Chebyshev-type theorem for the Hardy-Littlewood maximal function
2. Vitali covering

Recall the Chebyshev inequality: given $f \in \mathcal{L}^1(\mathbf{R}^n)$. $\mu(\{x : |f(x)| > \alpha\}) \leq \dfrac{1}{\alpha} \displaystyle\int_{\mathbf{R}^n} |f| d\mu = \frac{1}{\alpha} \|f\|_{\mathcal{L}^1}$.

Proof.
$$\alpha|\{x : |f(x)| > \alpha\}| \leq \int_{\{x:|f(x)|\ \alpha\}} |f|d\mu \leq \int_{\mathbf{R}^n} |f|d\mu$$

■

The Chebyshev-type inequality for Hf is as follows:

Theorem 2.24. (Maximal theorem) There exists a constant $C > 0$ such that $\forall f \in \mathcal{L}^1(\mathbf{R}^n)$ and $\forall \alpha > 0$,
$$|\{x : |f(x)| > \alpha\}| < \frac{C}{\alpha} \|f\|_{\mathcal{L}^1}$$
This is essentially saying that the sets on the left-hand side, dependent on f, can be bounded by some expression in terms of the $\mathcal{L}^1$-norm of f.

Proof. Let $E_\alpha \coloneqq \{x : Hf(x) > \alpha\} = \{x : \sup_{r>0} A_r|f|(x) > \alpha\}$. For each $x \in E_\alpha$, one can choose a radius $r_x > 0$ such that $A_{r_x}|f|(x) > \alpha$. Thus, the collection of balls $\mathcal{C} = \bigcup_{x\in E_\alpha} B(x, r_x) \supseteq E_\alpha$, i.e. E_α is covered by $\mathcal{C}$. By the Vitali covering, given any $c < \mu(E_\alpha)$, we can find finitely many $x_1, \ldots, x_k \in E_\alpha$ such that $\{B(x_j, r_{x_j})\}_{j=1}^k$ are disjoint and satisfy $\sum_{j=1}^k \mu(B_j) \geq \frac{c}{3^n}$. This implies
$$c < 3^n \sum_{j=1}^{k} \mu(B_j) = \frac{3^n}{\alpha} \sum_{j=1}^{k} \int_{B_j} |f(y)|d\mu(y) \leq \frac{3^n}{\alpha} \int_{\mathbf{R}^n} |f(y)|d\mu(y)$$
Note that the right-hand side does not depend on c, and thus tending $c \to \mu(E_\alpha)$, we obtain
$$\mu(E_\alpha) \leq \frac{3^n}{\alpha} \int |f|d\mu$$

■

We are now in a position to prove the Lebesgue differentiation theorem (v.1).

Theorem 2.25. (LDT v.1) Suppose $f \in \mathcal{L}^1_{loc}(\mathbf{R}^n)$. Then, $\lim_{r\longrightarrow 0^+} A_r f(x) = f(x)$ μ-a.e., in other words,
$$\lim_{r\longrightarrow 0^+} \frac{1}{|B|} \int_B (f(y) - f(x))d\mu(y) = 0$$
for a.e. x.

Proof. The proof uses the maximal theorem and approximation of $\mathcal{L}^1$ functions by C^0 functions. Since $A_r f(x)$ is local, given $f \in \mathcal{L}^1$, we can replace f by $f\chi_B$, where B is large enough. It is enough to consider the case where $f \in \mathcal{L}^1(\mathbf{R}^n)$.
Since $C^0(\mathbf{R}^n)$ is dense in $\mathcal{L}^1(\mathbf{R}^n)$, given $\epsilon > 0, \exists g \in C^0(\mathbf{R}^n) \cap \mathcal{L}^1(\mathbf{R}^n)$ such that
$$\int |g - f|d\mu < \epsilon$$
Since g is continuous, $\forall x \in \mathbf{R}^n, \delta > 0$, one can find $r > 0$ such that $|g(y) - g(x)| < \delta$ whenever $|x - y| < r$. Therefore,
$$\begin{aligned} |A_r f(x) - g(x)| &= \frac{1}{|B|} \int_B (g(y) - g(x))d\mu(y) \\ &\leq \frac{1}{|B|} \int_B |g(y) - g(x)|d\mu(y) \\ &\leq \frac{1}{|B|} \int_B \delta d\mu(y) \qquad (*) \end{aligned}$$

if $g \in C^0 \cap \mathcal{L}^1$, $\forall x \in \mathbf{R}^n, A_r g(x) \to g(x)$ as $r \to 0^+$. We need to estimate:

$$\begin{aligned}\limsup_{r \longrightarrow 0^+} |A_r f(x) - f(x)| &= \limsup_{r \longrightarrow 0^+} |A_r(f-g)(x) + (A_r g - g)(x) + (g-f)(x)| \\ &\leq \limsup_{r \longrightarrow 0^+} |A_r(f-g)(x)| + \limsup_{r \longrightarrow 0^+} |(A_r g - g)(x)| + |(g-f)(x)| \\ &\leq H(f-g)(x) + |f-g|(x) \quad (*)\end{aligned}$$

by the Maximal theorem. Now, we set $E_\alpha = \{x : \limsup_r |A_r f(x) - f(x)| > \alpha\}, F_\alpha = \{x : |f-g|(x) > \alpha\}$. From $(*)$, it follows that

$$E_\alpha \subset \{x : H(f-g)(x) > \alpha/2\} \cup \{x : |f-g|(x) > \alpha/2\}$$

Thus, we have that

$$\mu(E_\alpha) \leq \mu(\{x : H(f-g)(x) > \alpha/2\}) + \mu(\{x : |f-g|(x) > \alpha/2\})$$

Thus, the following hold:

$$\begin{aligned}\mu(\{x : |f-g|(x) > \alpha/2\}) &\leq \frac{2}{\alpha} \int |f-g| d\mu < \frac{2}{\alpha}\epsilon \\ \mu(\{x : H(f-g)(x) > \alpha/2\}) &\leq \frac{2C}{\alpha} \int |f-g| d\mu \leq \frac{2C}{\alpha}\epsilon\end{aligned}$$

where the first line follows from Chebyshev, and the second line from the Maximal theorem. Thus, $\forall \epsilon > 0, x \in \mathbf{R}^n, \forall \alpha > 0$

$$\mu(\{x : \limsup_{r \longrightarrow 0^+} |A_r f(x) - f(x)| > \alpha\}) < \frac{2}{\alpha}\epsilon + \frac{2C}{\alpha}\epsilon \to 0$$

Now, since $\epsilon > 0$ is arbitrary, it follows that $\mu(E_\alpha) = 0$. To see this, note that $\forall \alpha' > \alpha, E_{\alpha'} \subset E_\alpha$. Thus, it is enough to estimate $E = \bigcup_{n=1}^{\infty} E_{\frac{1}{n}}$. But $\mu(E) \leq \sum \mu(E_{\frac{1}{n}}) = 0$ as $\forall \alpha > 0, \mu(E_\alpha) = 0$. Thus, the Lebesgue Differentiation Theorem v.1 (LDT v.1) is proved. ∎

Therefore, given some $f \in \mathcal{L}^1_{loc}(\mathbf{R}^n), \lim_{r \longrightarrow 0^+} A_r f(x) = f(x)$ for $x \in \mathbf{R}^n$ a.e. Another way of writing this is:

$$\lim_{r \longrightarrow 0^+} \frac{1}{|B|} \int_B [f(y) - f(x)] d\mu(y) = 0$$

for $x \in \mathbf{R}^n$ a.e. We can make a more refined version of LDT, we call this LDT v.2, in which the term $[f(y) - f(x)]$ can be replaced by $|f(y) - f(x)|$. To develop this idea, we consider the **Lebesgue set of** f:

$$L_f = \{x \in \mathbf{R}^n : \lim_{r \longrightarrow 0^+} \frac{1}{|B|} \int_B |f(y) - f(x)| d\mu(y) = 0\}$$

Theorem 2.26. (LDT v.2) Let $f \in \mathcal{L}^1_{loc}(\mathbf{R}^n)$, and $\mu(L_f^c) = 0$. Then,

$$\lim_{r \longrightarrow 0^+} \frac{1}{|B|} \int_B |f(y) - f(x)| d\mu(y) = 0$$

Proof. $\forall \alpha \in \mathbb{C}$, consider $g_\alpha \in \mathcal{L}^1_{loc}(\mathbf{R}^n)$ given by $g_\alpha(x) = |f(x) - \alpha|$. We apply LDT v.1 to $g_\alpha(x)$:

$$\lim_{r \longrightarrow 0^+} \frac{1}{|B|} \int_B |f(y) - \alpha| d\mu(y) = |f(x) - \alpha|$$

$\forall x \in E_\alpha^c$, where $\mu(E_\alpha) = 0$. Let D be a countable dense subset of $\mathbb{C}$ (e.g. $\mathbf{Q} \times \mathbf{Q}$). Set $E = \bigcup_{\alpha \in D} E_\alpha$ we have $\mu(E) \leq \sum \mu(E_\alpha) = 0$. Consider $x \notin E$: for every $\epsilon > 0$, we can choose $\alpha \in D$ with $|f(x) - \alpha| < \epsilon$ (density). Thus, $|f(y) - f(x)| \leq |f(y) - \alpha| + |f(x) - \alpha| \leq |f(y) - \alpha| + \epsilon$. So, if $x \notin E$,

$$\begin{aligned}\limsup_{r \longrightarrow 0^+} \frac{1}{|B|} \int_B |f(y) - f(x)| d\mu(y) &\leq \frac{1}{|B|} \int_B |f(y) - \alpha| d\mu(y) + \epsilon \\ &= |f(x) - \alpha| + \epsilon < 2\epsilon\end{aligned}$$

this is true $\forall \alpha \in D$. Now, let $\epsilon \to 0^+$ and we obtain the desired result. ∎

Yeah, there's deadass a third version of LDT, called LDT v.3, which deals with the issue of replacing balls by other "nicely shrinking" sets in the theorem.

Definition 2.27. Consider a family $\{E_r\}_{r>0} \subset \mathcal{M}$. We say that E_r's shrink nicely if

1. $E_r \subset B(x,r), \forall r > 0$
2. $\mu(E_r) \approx \mu(B(x,r))$

i.e. $\exists \alpha > 0$ such that $\alpha B(x,r) \leq |E_r| \leq |B(x,r)|$.

Remark 2.28. 1. One can choose $E_r(x)$ to be <u>cubes</u> of size r instead of balls.

2. E_r need not contain x.

Case 1. is very commonly used in harmonic analysis.

We're ready to drop LDT in its final form.

Theorem 2.29. (LDT v.3) Suppose $f \in \mathcal{L}^1_{loc}(\mathbf{R}^n), x \in L_f(|L_f^c| = 0)$. Then,

$$\boxed{\lim_{r \to 0^+} \frac{1}{|E_r|} \int_{E_r} |f(y) - f(x)| d\mu(y) = 0}$$

where $\{E_r\}$ is any nicely shrinking family. Note that this implies the weaker version v.2.

Proof.

$$\begin{aligned}
\frac{1}{\mu(E_r)} \int_{E_r} |f(y) - f(x)| d\mu(y) &\leq \frac{1}{\mu(E_r)} \int_B |f(y) - f(x)| d\mu(y) \\
&\leq \frac{1}{\alpha \mu(B)} \int_B |f(y) - f(x)| d\mu(y)
\end{aligned}$$

Take lim sup of both sides and use v.2 to yield the theorem. ∎

2.4 Differentiation for General Measures

Definition 2.30. Let ν a Borel measure on $\mathbf{R}^n$. Then, ν is **regular** if

- $\nu(K) < \infty$ for every compact K
- $\nu(E) = \inf\{\nu(U) : U \text{ open}, E \subset U\}$ for every $E \in \mathcal{M}$
- $\nu(E) = \sup\{\nu(C) : C \text{ closed}, C \subset E\}$ for every $E \in \mathcal{M}$

A regular measure intuitively means that measurable sets can be approximated from the outside by open sets and from the inside by closed sets.

By (i), every regular measure is σ-finite. A signed or complex Borel measure ν is regular if $|\nu|$ is regular.

Example 2.31. If $f \in \mathcal{L}^+(\mathbf{R}^n)$, the measure $f d\mu$ is regular iff $f \in \mathcal{L}^1_{loc}$. One can notice that the condition $f \in \mathcal{L}^1_{loc}$ is equivalent to (i).

We can generalize LDT to ν as follows:

Theorem 2.32. Let ν be a regular signed or complex Borel measure on $\mathbf{R}^n$, and let $d\nu = d\lambda + f d\mu$ be its Lebesgue-Radon-Nikodym representation. Then, for μ-almost every $x \in \mathbf{R}^n$,

$$\lim_{r \to 0^+} \frac{\nu(E_r)}{\mu(E_r)} = f(x)$$

for every nicely shrinking family $\{E_r\}_{r>0} \to x$.

Proof. (Idea) When $d\lambda = 0$, this is just LDT v.3. On the general case, one has to show that for a.e. $x \in \mathbf{R}^n$,

$$\frac{\lambda(E_r)}{\mu(E_r)} \to 0^+$$

which should follow almost directly from the singularity of the meassure λ. ∎

Tangent 2.33. The LDT is a generalization of FTC. From LDT v.3, we observe that the sets

$$E_r = \{x + ry : y \in U \subset B(0,1), U \in \mathcal{M}, \mu(U) > 0\}$$

(i.e. take some x, and imagine some wiggly set around x of maximum radius r, where the wiggliness is determined by how U is shaped) are nicely shrinking to x as $rarr0^+$. Since $E_r \subset B(x,r)$, and $\mu(E_r(x)) \leq cr^n$, by LDT v.3, $\forall f \in \mathcal{L}^1_{loc}(\mathbf{R}^n)$,

$$f(x) = \lim_{r \to 0^+} \frac{1}{|E_r|} \int_{E_r} f d\mu \quad \text{a.e. } x \in \mathbf{R}^n$$

As an example, in 1D ($n = 1$), $y \in (0,1)$,

$$E_r(x) = \{x + ry : |y| < 1\} = (x, x + r)$$

LDT gives

$$f(x) = \lim_{r \to 0^+} \frac{1}{r} \int_x^{x+r} f(y) dy \quad \text{a.e. } x \in \mathbf{R}$$

Since the Lebesgue measure $dy = d\mu(y)$ is invariant under translation by x,

$$\implies f(x) = \lim_{r \to 0^+} \frac{1}{r} \int_0^r f(x + y) dy = \lim_{r \to 0^+} \frac{1}{r} r \int_0^1 f(x + rz) dz \quad \text{(change of variables)}$$

This simply says that if $f \in \mathcal{L}^1_{loc}(\mathbf{R})$, then $f(x) = \lim_{r \to 0^+} \int_0^1 f(x + rz)dz$ a.e. $x \in \mathbf{R}$.

Next, we would like to prove a more elementary version of FTC. Recall that given $f \in C^1([a,b])$, from middle school mathematics, we have the following:

$$f(b) - f(a) = \int_a^b f'(t) dt$$

We now wish to generalize this result to a larger class of functions, using among other things, the LDT version $f(x) = \lim_{r \to 0^+} \int_0^1 f(x + rz)dz$.

2.5 Functions of Bounded Variation

Given $F : \mathbf{R} \to \mathbf{R}$ where F is increasing and right-continuous (F is càdlàg), we can associate to such an F its Lebesgue-Stieltjes measure μ_F (Borel) where on intervals $[a,b)$,

$$\mu_F([a,b)) = F(b) - F(a)$$

Note that the Lebesgue measure μ is the Lebesgue-Stieltjes measure of $F(x) = x$. Extending this measure by Carathéodory, we obtain a measure on $\mathcal{M}$, our Borel algebra. In light of the properties of F, we need several results that apply for monotone functions.

Theorem 2.34. Let $F : \mathbf{R} \to \mathbf{R}$ increasing, and $G(x) = F(x+) = \lim_{y \to x^+} F(y)$ (G is basically an adjustment so that F's domain can be decomposed nicely into half-open intervals of the form $[a,b)$, i.e. G is càdlàg, $G = F$ a.e.). Note that since $F \uparrow$, this limit exists. If F is continuous, $G(x) = F$. Then, we have the following results:

1. The set where F is discontinuous is at most countable.
2. F, G are differentiable a.e., i.e. $F' = G'$ a.e.

Proof. Since $F \uparrow$, the intervals $(F(x-), F(x+))$ are all disjoint. So, for all $|x| < N$, $(F(x-), F(x+)) \subset (F(-N), F(N))$ by monotonicity. Consider

$$S = \sum_{|x|<N} [F(x+) - F(x-)]$$

taking the supremum over all finite subsets yields $S \leq F(N) - F(-N) < \infty$. Therefore, $\{x \in [-N, N] : F(X+) \neq F(x-)\}$ must be countable, otherwise the sum blows up. Part 1. is proved.

Now, since $G(x) \coloneqq F(x+)$, we have that $G \uparrow$. Thus, by definition,

$$\begin{cases} G \text{ is right continuous} \\ G = F \text{ except where } F \text{ is discontinuous} \end{cases}$$

Yes, my boi Lebesgue-Stieltjes pulls up with the dank μ_G:

$$G(x+h) - G(x) = \begin{cases} \mu_G([x, x+h)) & h > 0 \\ -\mu_G([x+h, x)) & h < 0 \end{cases}$$

Also, note that families $\{[x-r, x)\}, \{[x, x+r)\}$ shrink nicely to x as $r \to 0^+$ $(= |h|)$. μ_G are regular (Exercise). Therefore, we an apply LDT (basically, whenever you see regular measures, you pull out LDT and nice things happen) to get that G' exists for a.e. $x \in \mathbf{R}$. We are now left with showing that $G - F$ is also differentiable a.e. and that $G' - F' = 0$ a.e. To prove, note that $H = G - F$ is also increasing and by Part 1., $\{x \in \mathbf{R} : H(x) \neq 0\}$ is at most countable (the points of discontinuities, i.e. we can enumerate them: $\{x_j\}_{j=1}^{\infty}$. Then, $\forall j \in \mathbf{N}, H(x_j) > 0$ and also

$$\sum_{j, |x_j| \leq N} H(x_j) < \infty$$

by Part 1. Let δ_j be the point masses at x_j and consider

$$\mu = \sum_{j=1}^{\infty} H(x_j)\delta_j$$

$\mu(K) < \infty, \forall K$ compact, it then follows from two statements above that μ is finite on compact sets, and hence regular. Also, $\mu \perp \lambda$ where λ is the Lebesgue measure since $\lambda(E) = \mu(E^c) = 0$ where $E = \{x_j\}_{j=1}^{\infty}$. But then,

$$\frac{H(x+h) - H(x)}{h} \leq \frac{H(x+h) + H(x)}{|h|}$$

$h > 0$, since $H \geq 0$. Replacing H by μ, while keeping in mind that $H = 0$ a.e. and μ on the x_j's > 0, we have:

$$\frac{H(x+h) + H(x)}{|h|} \leq \frac{4\mu((x - 2|h|, x + 2|h|))}{4|h|}$$

since $F = 0$ off suppμ. This tends to 0 as $|h| \to 0$ by LDT since $\mu \perp \lambda$. This is true $\forall x \in \mathbf{R}$ a.e., where the singular part in LDT goes to zero because of the δ-masses. Thus, $H' = 0$ a.e. and the desired result is proved. ∎

Thus, by right continuity, we have that $G' = F' = 0$ a.e. The upshot is that monotone functions have derivatives almost everywhere. However, there is a much larger class of functions built up from monootone functions that is also differentiable almost everywhere. To define these, given $F : \mathbf{R} \longrightarrow \mathbf{R}$, we consider the **total variation** $T_F : \mathbf{R} \longrightarrow \bar{\mathbf{R}}^+$:

$$T_F(x) = \sup \left\{ \sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| : n \in \mathbf{N}, -\infty < x_0 < \cdots < x_n < \infty \right\}$$

it follows that $T_F(b) - T_F(a) = \sup \left\{ \sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| : n \in \mathbf{N}, a < x_0 < \cdots < x_n < b \right\}$ Note that $T_F \uparrow$ since the sum increases if one puts in additional partitions.

Definition 2.35. The space of functions of bounded variation over $\mathbf{R}$ is

$$BV(\mathbf{R}) = \{F : \mathbf{R} \longrightarrow \mathbf{R} : T_F(\infty) = \lim_{x \longrightarrow \infty} T_F(x) < \infty\}$$

Similarly, over $[a, b]$,

$$BV([a, b]) = \{F : \mathbf{R} \longrightarrow \mathbf{R} : T_F(x) < \infty \forall x \in [a, b]\}$$

Remark 2.36. If $F \in BV(\mathbf{R})$, we trivially have that $F|_{[a,b]} \in BV([a, b])$. Conversely, if $F \in BV([a, b])$, we can set:

$$\tilde{F}(x) = \begin{cases} F(a) & x < a \\ F(x) & x \in [a, b] \\ F(b) & x > b \end{cases}$$

and thus $\tilde{F}(x) \in BV(\mathbf{R})$. The first statement implies that results for $BV(\mathbf{R})$ can be extended to $BV([a, b])$.

The key idea is that there's an intrinsic connection between the function theory of BV functions and Lebesgue-Stieltjes measures, induced by càdlàg functions. Since we've shown in the previous theorem that monotone (and thus the càdlàgs of them) have nice differentiation properties, it would be ideal that the implication be $F \in BV(\mathbf{R})$ has nice differentiation properties μ-a.e.

Example 2.37.

1. $F : \mathbf{R} \longrightarrow \mathbf{R}$, $F \uparrow$, bounded. Then, $F \in BV(\mathbf{R})$. In this case, since $F \uparrow$, $T_F(x) = F(x) - F(-\infty)$, and thus $T_F(\infty) = F(\infty) - F(-\infty)$ exists.
2. Suppose $F : \mathbf{R} \longrightarrow \mathbf{R}$ differentiable with bounded derivative. Then, $F \in BV([a, b])$ for $a, b \in \mathbf{R}$ by the mean-value theorem.

The following is the key link between monotone functions and $BV(\mathbf{R})$.

Lemma 2.38. Suppose $F : \mathbf{R} \longrightarrow \mathbf{R}, F \in BV(\mathbf{R})$. Then, $(T_F \pm F)$ are increasing. (we're hopping from BV functions to monotone functions).

Remark 2.39. Given $F \in BV(\mathbf{R}), F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F)$, where each part is increasing. This is called the **Jordan decomposition** of F into positive and negative variations.

Proof. Assume $x < y, \epsilon > 0$. Choose $x_0 < \cdots < x_n = x$ such that $\sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| \geq T_F(x) - \epsilon$, i.e. we choose a partition such that the sum above is an approximating sum for the total variation of F at x. By adding in the $[x, y)$ chonk, we have that $\sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| + |F(y) - F(x)|$ is an approximating sum for $T_F(y)$. Now, since $F(y) = (F(y) - F(x)) + F(x)$,

$$T_F(y) \pm F(y) \geq \sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| + |F(y) - F(x)| \pm (F(y) - F(x)) \pm F(x)$$

Since $F(y) - F(x)| \pm (F(y) - F(x)) \geq 0$, we thus have that:

$$T_F(y) \pm F(y) \geq T_F(x) - \epsilon \pm F(x)$$

as ϵ is arbitrary, this concludes the proof. ∎

Now that we've established that $BV(\mathbf{R})$ functions are sums of monotonically increasing functions, we can make deductions on the differentiability of BV functions in the following theorem:

Theorem 2.40.

(a) $F \in BV(\mathbf{R}) \iff \mathrm{Re}(F), \mathrm{Im}(F) \in BV(\mathbf{R})$.

(b) $F : \mathbf{R} \to \mathbf{R}$. Then, $F \in BV(\mathbf{R})$ iff F is the difference of two bounded increasing functions. For $F \in BV$, these functions may be taken to be $\frac{1}{2}(T_F + F)$ and $\frac{1}{2}(T_F - F)$.

(c) If $F \in BV(\mathbf{R})$, then $F(x+)$ and $F(x-)$ exist for all $x \in reals.$ as do $F(\pm\infty)$.

(d) If $F \in BV(\mathbf{R})$, the set of points at which F is discontinuous is countable.

(e) If $F \in BV(\mathbf{R})$ and $G(x) = F(x+)$, then F' and G' exist and $F' = G'$ a.e.

Proof. Essentially, everything is a consequence of the previous results for monotone $\uparrow$ functions + the previous lemma applied to the Jordan decomposition of $F \in BV(\mathbf{R})$. ∎

Remark 2.41. It is useful to note that if $F \in BV(\mathbf{R})$, then (F is obviously bounded) $T_F \pm F$ are all bounded.

Proof. Suppose $x < y \in \mathbf{R}$, $T_F(y) \pm F(y) \geq T_F(x) \pm F(x)$. This implies that

$$|F(y) - F(x)| \leq |T_F(y) - T_F(x)|$$

since putting $[x, y)$ just adds an extra chonk. Then, $|F(y)-F(x)| \leq T_F(y)-T_F(x) \leq T_F(\infty)-T_F(-\infty) < \infty$ since $T_F(\infty) < \infty$. Finally, since F is bounded, T_F is bounded implies $T_F \pm F$ is also bounded. ∎

Now, given this elementary function theory, we want to connect this back to the Stieltjes measure by refining $BV(\mathbf{R})$ to include right-continuity. To do this, we finna normalize:

$$NBV(\mathbf{R}) = \{F : \mathbf{R} \to \mathbf{R} : F \in BV(\mathbf{R}), F \text{ is right continuous and } F(-\infty) = 0\}$$

We now need a lemma that generalizes the right-continuity of T_F:

Lemma 2.42. Suppose $F \in BV(\mathbf{R})$. Then, $T_F(-\infty) = 0$. If F is also right continuous, then so is T_F.

With this lemma + Jordan decomposition, given $F \in NBV(\mathbf{R})$, one can associate with it a Borel measure μ such that

$$F(x) = \mu((-\infty, x])$$

a consequence of the decomposition.

Proof. Let $\epsilon > 0, x \in \mathbf{R}$, choose $x_0 < \cdots < x_n = x$ such that

$$T_F(x) - T_F(x_0) \geq \sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| \geq T_F(x) - \epsilon$$

Hence $T_F(x) - T_F(x_0) \geq T_F(x) - \epsilon \implies \forall y \leq x_0, T_F(y) < \epsilon$.
Now, F right continuous means that fixing an $x_0 \in [a, b]$, $0 < x - x_0 < \delta \implies |F(x) - F(x_0)| < \frac{\epsilon}{2}$. Consider a partitioning $x_0 < x_1 < \cdots < x_n = b$, we have that:

$$T_F(b) - T_F(x_0) \leq \sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| + \frac{\epsilon}{2}$$

so the sum is an approximation for the total variation of F on $[x_0, b]$. Let $\hat{\delta} = \min\{\delta, x_1 - x_0\}$, if we choose x such that $x_0 < x < x_0 + \hat{\delta}$, then $x - x_0 < \delta$ and $x_0 < x < x_1$, we thus have:

$$\begin{aligned} T_F(b) - T_F(x_0) &\leq |F(x) - F(x_0)| + |F(x_1) - F(x)| + \sum_{j=2}^{n} |F(x_j) - F(x_{j-1})| + \frac{\epsilon}{2} \\ &< \frac{\epsilon}{2} + T_F(b) - T_F(x) + \frac{\epsilon}{2} \\ &= \epsilon + T_F(b) - T_F(x) \end{aligned}$$

Thus, $T_F(x) \leq T_F(x_0) + \epsilon$. Hence, we have that T_F is right-continuous. ∎

Theorem 2.43. Let μ be a complex Borel measure on $\mathbf{R}$, define $F(x) := \mu((-\infty, x])$. Then, $F \in NBV$. Conversely, if $F \in NBV$, then $\exists! \mu_F$ complex Borel, such that $\mu_F((-\infty, x]) = F(x)$; moreover, $|\mu_F| = \mu_{T_F}$.

Proof. Let μ complex Borel: $\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)$, where $\mu_j^{\pm}$ are finite measures. If $F_j^{\pm}(x) = \mu_j^{\pm}((-\infty, x])$, then $F_j^{\pm}$ is increasing and right continuous, zero at $-\infty$, and $\mu_j^{\pm}(\mathbf{R}) < \infty$ at ∞. Therefore, F is bounded, normalized, right continuous, increasing, and since $F_j^+ - F_j^-$ are differences of two bounded increasing functions, $\operatorname{Re} F$ and $\operatorname{Im} F$ are BV, and therefore $F \in NBV$. Conversely, any $F \in NBV$ can be written in the form $F = F_1^+ - F_1^- + i(F_2^+ - F_2^-)$, by reverse engineering Theorem 2.40 (a) to (b). Now, each $F_j^{\pm}$ is càdlàg and thus gives rise to a Stieltjes measure $\mu_j^{\pm}$. The proof that $|\mu_F| = \mu_{T_F}$ is an exercise. ∎

Now, given some complex Borel measure we can get a NBV function, while given a NBV function we can get a complex Borel measure from Stieltjes measures. Now, which functions in NBV correspond to measures μ such that $\mu \perp m$, or $\mu \ll m$?

Proposition 2.44. $F \in NBV \implies F' \in \mathcal{L}^1(m)$. Moreover, $\mu_F \perp m \iff F' = 0$ a.e., and $\mu_F \ll m \iff F(x) = \displaystyle\int_{-\infty}^{x} F'(t)dt$.

Proof. Observe that $F'(x) = \lim_{r \to 0^+} \mu_F(E_r)/m(E_r)$ where E_r shrink nicely to x, and we slap the LDT for abstract measures on this baby. Also, Stieltjes measures are regular by default, making μ_F regular. ∎

The condition $\mu_F \ll m$ can be expressed in terms of F as follows: a function $F : \mathbf{R} \to \mathbb{C}$ is **absolutely continuous** if $\forall \epsilon > 0, \exists \delta > 0$ such that for any finite set of disjoint intervals (a_j, b_j),

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

More generally, F absolutely continuous on $[a, b]$ if the condition is satisfied whenever the intervals all lie in $[a, b]$. Take $N = 1$, absolute continuity implies uniform continuity. Now, if F is everywhere differentiable and F' is bounded, then F is absolutely continuous, since $|F(b_j) - F(a_j)| \leq (\max |F'|)(b_j - a_j)$ by the mean value theorem.

Proposition 2.45. $F \in NBV$, then F absolutely continuous $\iff \mu_F \ll m$.

Corollary 2.46. If $f \in \mathcal{L}^1(m)$, then the function $F(x) = \displaystyle\int_{-\infty}^{x} f(t)dt$ is in NBV and is absolutely continuous, $f = F'$ a.e. Conversely, if $F \in NBV$ is absolutely continuous, then $F' \in \mathcal{L}^1(m)$ and $F(x) = \displaystyle\int_{-\infty}^{x} F'(t)dt$.

Proof. Follows from the previous 2 props. ∎

IF we consider functions on bounded intervals, this result can be refined a bit.

Lemma 2.47. IF F is absolutely continuous on $[a, b]$, then $F \in BV([a, b])$.

Proof. Set $\epsilon = 1$, choose some number of intervals such that T_F is controlled. ∎

Then, given $F : [a, b] \to \mathbf{R}$, we can assume $F(a) = 0$ by substituting a constant. We extend such an F to $\tilde{F}$ by normalizing and $\tilde{F}(x) = F(b) \forall x > b$. Given that F is absolutely continuous on $[a, b]$, we have that by the lemma above, $F \in BV(\mathbf{R}) \implies \tilde{F} \in NBV(\mathbf{R})$. We slap FTC to $\tilde{F}$ to get the following:

Theorem 2.48. Let $a < b \in \mathbf{R}, F : [a, b] \to \mathbb{C}$. TFAE:

(a) F is absolutely continuous on $[a, b]$.

(b) $F(x) - F(a) = \displaystyle\int_{a}^{x} f dt, f \in \mathcal{L}^1([a, b], m)$.

(c) F' exists a.e. on $[a,b]$, $F' \in \mathcal{L}^1([a,b],m), F(x) - F(a) = \displaystyle\int_a^x F' dt$.

3 Functional Analysis

3.1 Motivation: PDEs

The study of functional analysis is basically studying linear algebra in the infinite dimensional case. Why do we do this? One of the uses is in PDEs, because it gives rise to a nice theory of general solutions. Let $\Omega \subset \mathbf{R}^n$ be a bounded domain, with C^∞ boundary. Recall the Laplacian:

$$\triangle = \sum_{j=1}^{n} \frac{\partial^2}{\partial x_j^2}, \quad x \in \mathbf{R}^n$$

Consider the Neumann (Direchlet) equations: $-\triangle u_j = \lambda_j u_j$ in $\Omega, \partial_n u_j = 0$ on $\partial\Omega$, where u_j are Neumann eigenfunctions and λ_j are their corresponding eigenvalues. To estimate the λ_j's (one can think of them as vibration frequencies), one studies the Rayleigh-Ritz ratios (or Rayleigh quotients):

$$\frac{\displaystyle\int_\Omega |\nabla u_j|^2 d\mu}{\displaystyle\int_\Omega |u_j|^2 d\mu}$$

In fact, by max/min, the first non-trivial eigenvalue (the lowest mode of vibration) is given by:

$$\lambda_1 = \inf_u \left(\frac{\displaystyle\int_\Omega |\nabla u_j|^2 d\mu}{\displaystyle\int_\Omega |u_j|^2 d\mu} \right)$$

Theorem 3.1. (Poincaré's Inequality, baby version) Let $\Omega \subset \mathbf{R}^n$ a bounded, convex, connected domain. Then, given $f \in C^\infty(\overline{\Omega}, \mathbf{R})$ with

$$\int_\Omega f d\mu = 0$$

there exists $\kappa = \kappa_n(\Omega) > 0$ such that

$$\kappa \int_\Omega |f|^2 d\mu \leq \int_\Omega |\nabla f|^2 d\mu$$

Corollary 3.2. $\lambda_1 \geq \kappa_n(\Omega)$. Note that any f satisfies this Poincaré's Inequality, an a-priori estimate. In particular, the eigenfunctions will satisfy.

Along with functional analysis where one studies completeness and density arguments, one can extend this a-priori estimate to a much larger class of functions, namely the $\mathcal{L}^p$ Sobolev spaces.

3.2 Normed Linear Spaces

Let X be a vector space over some field.

Definition 3.3. A **pre-norm** is a function $\|\cdot\| : X \to \mathbf{R}_0^+$ satisfying

1. $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in X$,
2. $\|\lambda x\| = |\lambda| \, \|x\|$ for all $x \in X, \lambda \in F$.

Additionally, if $\|x\| = 0 \implies x = 0$, then $\|\cdot\|$ is a **norm**.

A vector space with a norm is called a **normed linear space** (NLS).

Remark 3.4. Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a NLS are equivalent if $\exists c_1, c_2 > 0$ such that $\forall x \in X, c\|x\|_1 \leq \|x\|_2 \leq c_2\|x\|_1$. Equivalent norms define equivalent metrics and hence the same topology and the same Cauchy sequences. Note that norms are always equivalent when $\dim X < \infty$. This is true since in finite dimensions, you have a set of basis that you can express, and what you do with them essentially determines the properties of your space.

A NLS complete w.r.t. the norm metric is called a **Banach space**. Note that every NLS can be embedded in a Banach space as a dense subspace, by mimicking the construction of $\mathbf{R}$ from $\mathbf{Q}$ via Cauchy sequences.

Theorem 3.5. (Completeness criterion of NLS) Suppose $(X, \|\cdot\|)$ is a NLS. Then, X is complete if and only if every absolutely convergent series in X converges.

Proof. Assume that X is complete. Suppose $\sum^\infty \|x_n\| < \infty$. Let $S_N = \sum^N x_n (\in X)$. We claim that $\{S_N\}^\infty$ is Cauchy. Assume $N > M$:

$$\begin{aligned}\|S_N - S_M\| &= \left\|\sum^N x_n - \sum^M x_n\right\| \\ &= \left\|\sum_{n=M+1}^N x_n\right\| \\ &\leq \sum_{n=M+1}^N \|x_n\| \to 0 \text{ as } M, N \to \infty\end{aligned}$$

Hence since X is complete, $S_N \to x \in X$. Now, assume every absolutely convergent series converges, let $\{x_n\}^\infty$ be Cauchy. We can choose $n_1 < n_2 < \dots$ such that $\|x_n - x_m\| < 2^{-j}$ for $m, n \geq n_j$. Let $y_1 = x_{n_1}$, and $y_j = x_{n_j} - x_{n_{j-1}}$, for $j > 1$ (we make a telescoping series). Thus, $x_{n_k} = \sum_{j=1}^k y_j$, we have that

$$\sum_{j=1}^\infty \|y_j\| = \|y_1\| + \sum_{j=2}^\infty \|y_j\| \leq \|y_1\| + 1 < \infty$$

Hence, by assumption, $\sum^k y_j \to x \in X$ since X is a vector space and adding stuff remains in X, as $k \to \infty$. Therefore, $\lim_{x\to\infty} x_{n_k} = x \in X$. Note that since $\{x_n\}$ is assumed to be Cauchy, $\lim_{n\to\infty} x_n = x \implies X$ is complete. ∎

Example 3.6. Examples of Banach spaces.k Let $(X, \|\cdot\|)$ be a NLS. Then,

1. $\mathcal{B}(X) = \{f : X \longrightarrow \mathbb{C} \text{ measurable}\}$ with $\|f\|_\infty = \sup_x |f(x)| < \infty$.
2. If $(X, \mathcal{M}, \mu)$ is a measure space, then $\mathcal{L}^1(\mu) = \{f : X \longrightarrow \mathbb{C} \text{ measurable}\}$ with
$$\|f\|_1 = \int |f| d\mu < \infty$$
3. Hilbert spaces.
4. $\mathcal{L}^p$ spaces, $\mathcal{L}^p = \{f \text{ measurable s.t. } \|f\|_p = \left(\int |f|^p d\mu\right)^{\frac{1}{p}} < \infty\}$

We claim that $\mathcal{L}^1$ is Banach. Proving this is a simple application of the completeness criterion outlined by the previous theorem. Let $\{f_n\}^\infty \subset \mathcal{L}^1$ and assume that $\sum \|f_n\|_1 < \infty$. Recall that by a previous theorem, $\sum^\infty f_n \to f$ a.e., and $\int \sum^\infty f_n d\mu = \sum^\infty \int f_n d\mu$. To show $\sum^N f_n \to f$ as $N \to \infty$ in $\mathcal{L}^1$, we want

$\left\|\sum^N f_n - f\right\| \to 0$:

$$\left\|\sum^N f_n - f\right\| = \int \left|\sum^N f_n - f\right| d\mu = \int \left|\sum_{n=N+1}^{\infty} f_n\right| d\mu$$
$$\leq \int \sum_{n=N+1}^{\infty} |f_n|\, d\mu$$
$$= \sum_{n=N+1}^{\infty} \int |f_n|\, d\mu \quad \text{(MCT)}$$
$$= \sum_{n=N+\quad 1}^{\infty} \|f_n\|_1 \to 0 \text{ as } N \to \infty$$

Thus, $\sum^N f_n \to f \in \mathcal{L}^1$, and hence $\mathcal{L}^1$ is complete.

We proceed with some basic constructions. Suppose $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ are two normed spaces. Then, $X \times Y$ inherits a norm:

$$\|(x, y)\| = \max\{\|x\|_X, \|y\|_Y\}$$

Note that this is equivalent to many other norms, namely $\|(x, y)\| = \|x\|_X + \|y\|_Y$, or $\sqrt{\|x\|_X^2 + \|y\|_Y^2}$. Product spaces and their norms arsie by considering graphs of linear maps between 2 NLS:

$$T : X \longrightarrow Y \quad \{(x, Tx) \in X \times Y\}$$

As for quotient spaces, let $(X, \|\cdot\|_X)$ be a NLS, $M \leq X$ a linear subspace. One can construct a quotient space $X/M = \{x + M : x \in X\}$, where the equivalence classes are $[x] = x + M$. X/M is a vector space under the operations $(x + M) + (y + M) = x + y + M$, and $\lambda(x + M) = \lambda x + M, \lambda \in \mathbb{C}$. The quotient space is used to discuss or simulate notions of being " off " a subspace. If our space is a Hilbert space, then we have a stronger notion of orthogonal projection. Now, a natural norm on X/M is:

$$\|x + M\| = \inf_{y \in M} \|x + y\|$$

This is indeed a norm since

$$\|x + z + M\| = \inf_{y \in M} \|x + z + y\|$$
$$= \inf_{y \in M} \|x + z + 2y\|, \text{ since } M \text{ is linear}$$
$$\leq \inf_{y \in M} \|x + y\| + \inf_{y \in M} \|z + y\|$$
$$= \|x + M\| + \|y + M\|$$

Now, consider linear operators on a NLS. Let $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ NLS, assume that $T : X \longrightarrow Y$ is linear, i.e. $T(\alpha x + \beta y) = \alpha Tx + \beta Ty$. Unlike finite dimensional linear algebra, we need to distinguish difficult cases depending on how large T is relative to norms. We thus have the following theories:

1. Compact operators: the most direct analogs of linear maps in finite dimensional linear algebra. They are cool.
2. Bounded operators: they are cooler.
3. Unbounded operators: ?? Hol' up

Definition 3.7. Given $T : X \longrightarrow Y$ linear, we say that T is **bounded** provided that $\exists \kappa > 0$ such that

$$\|Tx\|_Y \leq \kappa \|x\|_X$$

A cool way to think about this is that in norm, T doesn't scale a vector up that much.

There are several equivalent formulations of boundedness.

Proposition 3.8. Let X, Y NLS, $T : X \longrightarrow Y$ linear. TFAE:

(a) T is continuous.

(b) T is continuous at zero.

(c) T is bounded.

Proof. (a) implies (b) is trivial. Assume (b), there exists an open set U, $0 \in U$, such that $T(U) \subset \{y \in Y : \|y\| \leq 1\}$. In particular, $\|Tx\| \leq 1$ provided $\|x\| \leq \delta$ for some $\delta > 0$. Thus, for all non-zero $x \in X$:

$$\|Tx\| = \left\| \frac{\|x\|}{\delta} T\left(\frac{\delta}{\|x\|}x\right)\right\| = \frac{\|x\|}{\delta}\left\|T\left(\delta\frac{x}{\|x\|}\right)\right\| \leq \frac{\|x\|}{\delta}\cdot 1 = \delta^{-1}\|x\|$$

Since $\left\|T\left(\delta\frac{x}{\|x\|}\right)\right\| \leq 1$ provided $\left\|\delta\frac{x}{\|x\|}\right\| \leq \delta$, which is always true since $\left\|\frac{x}{\|x\|}\right\| = 1$. Now suppose $\|Tx\| < C\|x\|$ for all $x \in X$. Hence,

$$\begin{aligned}\|Tx_1 - Tx_2\| &= \|T(x_1 - x_2)\| \\ &\leq \epsilon\end{aligned}$$

whenever $\|x_1 - x_2\| < C^{-1}\epsilon$, so that T is continuous. ∎

Definition 3.9. Let T be a bounded linear operator, i.e. $T \in \mathcal{L}(X, Y)$, where $\mathcal{L}$ is the space of all bounded linear maps. Then, the norm of T is defined as

$$\begin{aligned}\|T\| &= \sup\{\|Tx\| : \|x\| = 1\} \\ &= \sup\left\{\frac{\|Tx\|}{\|x\|} : x \neq 0\right\} \\ &= \inf\{C : \|Tx\| \leq C\|x\|, \forall x \in X\}\end{aligned}$$

Proposition 3.10. Suppose $T \in \mathcal{L}(X, Y)$, where Y is Banach. Then, $\mathcal{L}(X, Y)$ is Banach.

Proof. Let $\{T_n\}^\infty$ be a Cauchy sequence in $\mathcal{L}(X, Y)$. If $x \in X$, then $\{T_n x\}$ is Cauchy in Y since $\|T_n x - T_m x\| \leq \|T_n - T_m\|\|x\|$ (since $\|T_n - T_m\| = \sup\{\frac{1}{\|x\|}\|(T_n - T_m)(x)\| : x \neq 0\}$). Define $T : X \longrightarrow Y$ by $Tx = \lim_{n\to\infty} T_n x$, which exists in Y since Y is complete, T is linear since the T_n's are linear. We claim

1. $\|T\| = \lim_{n\to\infty}\|T_n\|$
2. $\|T - T_n\| \to 0$

2) follows since $\|T\| \leq \|T - T_n\| + \|T_n\| < \epsilon + \|T_N\|$ for $N \geq N(\epsilon)$, thus $T \in \mathcal{L}(X, Y)$.
To prove 1), suppose $\|x\| = 1$, we have

$$\|Tx - T_n x\| \to 0$$

as $n \to \infty$. So for any $\epsilon > 0$, we can find a $N(\epsilon)$ so that $\|Tx - T_n x\| < \epsilon, \forall n \geq N(\epsilon)$. Thus, we have

$$\|T_n x\| - \epsilon \leq \|Tx\| \leq \|T_n x\| + \epsilon$$

$$\Longrightarrow \sup_{\|x\|=1}\|T_n x\| - \epsilon \leq \sup_{\|x\|=1}\|Tx\| \leq \sup_{\|x\|=1}\|T_n x\| + \epsilon$$

The rest is obvious. ∎

3.3 Linear Functionals, Hahn-Banach

Let X be a vector space over K. A linear map from X to K is called a **linear functional** on X. If X is a NLS, the space $\mathcal{L}(X, K)$ of bounded linear functionals on X is called the **dual space** of X, denoted X^*. Since K is trivially Banach, X^* is a Banach space with the operator norm.

Example 3.11. $X = \mathcal{L}^1(\mathbf{R}, \mu), g \in C^0(\mathbf{R}), |g(x)| < M, \forall x \in \mathbf{R}$. Consider the map $f \mapsto \displaystyle\int fgd\mu, f \in \mathcal{L}^1$, this is a bounded linear functional.

Remark 3.12. The dual space X^* is of special interest, since it has nice properties. We say $T \in \mathcal{L}(X, Y)$ is an isomorphism (or invertible) if there exists an inverse mapping $T^{-1} \in \mathcal{L}(Y, X)$ (also bounded), i.e. T is a bijection onto Y and satisfies

$$C\,\|x\| \geq \|Tx\| \geq c\,\|x\|\,, c > 0 \quad \left(\left\|T^{-1}\right\| = \frac{1}{\|T\|}\right)$$

If T is an isomorphism and $\|Tx\| = \|x\|$, we say that T is an isometry.

Now, when X is a NLS, $M \leq X$ a subspace, and a map $f : M \longrightarrow \mathbf{R}$ a linear functional, how can we extend f to a linear mapping F on X such that

1. $F|_M = f$, and
2. $\|F\| = \|f\|$.

When $\dim X < \infty$, this is trivial, since we can extend an orthonormal basis for M to all of X by Gram-Schmidt orthonormalization. How can we do this without a basis? Hahn-Banach answers this question in the affirmative, provided that $f : M \longrightarrow \mathbf{R}$ is controlled.

Definition 3.13. A map $\rho : X \longrightarrow \mathbf{R}$ is **sublinear** if

(i) $\rho(x + y) \leq \rho(x) + \rho(y)$

(ii) $\rho(\lambda x) = \lambda\rho(x)$

for $\lambda > 0$ and $x, y \in X$. For example, every semi-norm is a sublinear functional.

3.4 Hilbert Spaces, Properties

Essentially, one can think of a Hilbert space as a Banach space with additional structure, in particular, the inner product. Let $(\mathcal{H}, \mathbb{C})$ be a vector space over the complex field.

Definition 3.14. An inner product is a bilinear form $\langle\cdot,\cdot\rangle : \mathcal{H} \times \mathcal{H} \longrightarrow \mathbb{C}$ with the following properties:

(i) $\langle ax + by, z\rangle = a\,\langle x, z\rangle + b\,\langle y, z\rangle\,, \forall a, b \in \mathbb{C}, x, y, z \in \mathcal{H}$

(ii) $\langle x, y\rangle = \overline{\langle y, x\rangle}$

(iii) $\langle x, x\rangle \geq 0, \forall x \in \mathcal{H}$

Thus, $\mathcal{H}$ equipped with an inner product is a pre-Hilbert space.

Definition 3.15. A pre-Hilbert space is Hilbert provided it is complete (Banach).

Asssuming the natural norm $\|x\| = \sqrt{\langle x, x\rangle}$, we see that $(\mathcal{H}, \|\cdot\|)$ is a Hilbert space. To prove that the above really is a norm, we first state and prove the Cauchy-Schwarz inequality.

Theorem 3.16. (Cauchy-Schwarz) Given $x, y \in (\mathcal{H}, \langle\cdot,\cdot\rangle), |\langle x, y\rangle| \leq \|x\| \cdot \|y\|$. Equality holds when $x = \lambda y$, for some $\lambda \in \mathbb{C}$.

Proof. Let $x.y \in \mathcal{H}$, and assume $\langle x, y\rangle \neq 0$. Define $\alpha = \operatorname{sgn}\langle x, y\rangle = \dfrac{\langle x, y\rangle}{|\langle x, y\rangle|}$, and let $z = \alpha y$. Then, we observe that:

$$\langle x, z\rangle = \langle x, \alpha y\rangle = \alpha\overline{\langle y, x\rangle} = \frac{\langle x, y\rangle}{|\langle x, y\rangle|}\overline{\langle y, x\rangle} = |\langle x, y\rangle|$$

Hence, $\langle x, z\rangle \in \mathbf{R}$, and $\langle x, z\rangle = \langle z, x\rangle$. Then, for all $t \in \mathbf{R}$, consider $f : \mathbf{R} \longrightarrow \mathbf{R}^+, f(t) = \langle x - tz, x - tz\rangle$, where $z = \alpha y$. Thus,

$$f(t) = \|x\|^2 - 2t\,|\langle x, y\rangle| + t^2\,\|y\|^2$$

Note that since f is an inner product, we have that for all $t \in \mathbf{R}$, $f(t) \geq 0$. Consider the minimum t, since f is a quadratic form in t:

$$f'(t) = 0 \iff 2t\,\|y\|^2 - 2\,|\langle x, y\rangle| = 0 \iff t_c = \frac{|\langle x, y\rangle|}{\|y\|^2}$$

Hence,

$$\begin{aligned} 0 \leq f(t_c) &= \|x\|^2 - \frac{2\,|\langle x, y\rangle|^2}{\|y\|^2} + \frac{|\langle x, y\rangle|^2}{\|y\|^4}\,\|y\|^2 \\ &= \left(\|x\|^2\,\|y\|^2 - |\langle x, y\rangle|^2\right)\|y\|^{-2} \end{aligned}$$

yielding our desired result. We have equality if and only if $x - tz = x - \alpha t y = 0$. ∎

Now, we're in a position to prove that the candidate natural norm on pre-Hilbert spaces is actually a norm.

Proposition 3.17. $\|x\| = \sqrt{\langle x, x\rangle}$ is a norm on $\mathcal{H}$.

Proof. The first two norm axioms are trivial from definition. We now want to show that $\|x + y\| \leq \|x\| + \|y\|$. To this end, consider:

$$\|x + y\|^2 = \langle x + y, x + y\rangle = \|x\|^2 + \|y\|^2 + 2\operatorname{Re}\langle x, y\rangle$$

By Cauchy-Schwarz,

$$\begin{aligned} &\leq \|x\|^2 + \|y\|^2 + 2\,\|x\|\,\|y\| \\ &= (\|x\| + \|y\|)^2 \\ \implies &\|x + y\| \leq \|x\| + \|y\| \end{aligned}$$

∎

From now on, we will refer to $\mathcal{H}$ as a hilbert space in norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot\rangle}$, i.e. $\mathcal{H}$ is complete in $\|\cdot\|$. There are two advantages to Hilbert spaces. One is that they enable the notion of orthogonal projection, that follows because we can take inner products between vectors, and from inner products arise angles and all that meme stuff. Another fundamental characteristic that is *unique* to Hilbert spaces are the fact that they fully characterize bounded linear functionals, i.e. their dual space. This is the Riesz representation theorem, which we will have the pleasure to observe soon.

Example 3.18. Let $(X, \mathcal{M}, \mu)$ be a measure space, and consider $\mathcal{L}^2(\mu) = \left\{f : X \longrightarrow \mathbf{R} : \int |f|^2\,d\mu < \infty\right\}$. The space of equivalence classes of equal a.e. functions is a Hilbert space. Recall now the simplified Young's inequality: $ab \leq \frac{1}{2}\left(a^2 + b^2\right)$, we thus can consider $f\bar{g} \leq \frac{1}{2}\left(|f|^2 + |g|^2\right)$, so if $f, g \in \mathcal{L}^2, f\bar{g} \in \mathcal{L}^1$, and we can define the inner product as follows:

$$\langle f, \bar{g}\rangle = \int f\bar{g}\,d\mu$$

We note that writing $\|f\|_{\mathcal{L}^2}^2 = \displaystyle\int |f|^2\,d\mu = \langle f, f\rangle$ is the norm induced by the above product, making $\mathcal{L}^2$ complete.

The following are certain basic identities of Hilbert spaces.

Proposition 3.19. (Parallelogram Law) $\forall x, y \in \mathcal{H}, \|x + y\|^2 + \|x - y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right)$.

Proof. $\|x + y\|^2 + \|x - y\|^2 = \langle x + y, x + y\rangle + \langle x - y, x - y\rangle = 2\left(\|x\|^2 + \|y\|^2\right)$ (just algebra your way to victory). ∎

Proposition 3.20. (Pythagorean Theorem) For $\times_1, \ldots, x_n \in \mathcal{H}$ with $\langle x_j, x_k \rangle = 0$ for $j \neq k$, we have that $\sum \|x_k\|^2 = \|\sum x_k\|^2$

Proof.

$$\begin{aligned}\left\|\sum x_k\right\|^2 &= \left\langle \sum x_k, \sum x_k \right\rangle \\ &= \sum \langle x_k, x_k \rangle \\ &= \sum \|x_k\|^2\end{aligned}$$

∎

Now, we consider the continuity of the inner product.

Proposition 3.21. Let $\{x_n\}^\infty \subset \mathcal{H}, \{y_n\}^\infty \subset \mathcal{H}$, and $x = \lim x_n, y = \lim y_n$. We have that $\lim \langle x_n, y_n \rangle = \lim \langle x, y \rangle$.

Proof.

$$\begin{aligned}|\langle x_n, y_n \rangle - \langle x, y \rangle| &= |\langle x_n - x, y_n \rangle + \langle x, y_n - y \rangle| \\ &\leq |\langle x_n - x, y_n \rangle| \, |\langle x, y_n - y \rangle| \\ &= \|x_n - x\| \, \|y_n\| + \|x\| \, \|y_n - y\| \to 0\end{aligned}$$

∎

3.5 Orthogonal Projection

Given any $M \leq \mathcal{H}$, one can define the orthogonal complement:

$$M^\perp = \{x \in \mathcal{H} : \langle x, y \rangle = 0, \forall y \in M\}$$

We remark that $M^\perp \leq \mathcal{H}$. When M is a closed subspace (closed under Cauchy sequences as well, which is occasionally non-trivial) one has the following very important decomposition:

Theorem 3.22. Let $M \leq \mathcal{H}$ be a closed subspace. Then, $\mathcal{H} = M \oplus M^\perp$. In other words, every $x \in \mathcal{H}$ can be uniquely written (since cH is a direct sum) in the form $x = y + z$, with $y \in M, z \in M^\perp$ are minimal distance to x.

Proof. Given $x \in \mathcal{H} \setminus M$, define $\delta \coloneqq \inf_{y \in M} \|x - y\| > 0$ (since M is closed), and let $\{y_n\}^\infty \subset M$ with $\|x - y_n\| \to \delta$ as $n \to \infty$. We claim that $\{y_n\}$ is Cauchy in M. By the Parallelogram law, we have that

$$2\left(\|y_n - x\|^2 + \|y_m - x\|^2\right) = \|y_n - y_m\|^2 + \|y_n + y_m - 2x\|^2$$

$$\|y_n - y_m\|^2 = 2\left(\|y_n - x\|^2 + \|y_m - x\|^2\right) - 4\left\|\frac{y_n + y_m}{2} - x\right\|^2$$

since $\frac{1}{2}(y_n + y_m)) \in M, \forall m, n$, we have

$$\left\|\frac{y_n + y_m}{2} - x\right\| \geq \delta$$

since δ is an infimum. Thus,

$$\|y_n - y_m\|^2 \leq 2\left(\|y_n - x\|^2 + \|y_m - x\|^2\right) - 4\delta^2$$

Taking $m, n \to \infty$, we see that $\|y_n - y_m\|^2 \leq 0$, and hence $\{y_n\}$ is Cauchy. Since M is closed (w.r.t. Cauchy sequences), it follows that $\lim_{n\to\infty} y_n = y \in M$. Now, using this fact, consider $x \in \mathcal{H} \setminus M = y + z$. We have $\delta = \lim_{n\to\infty} \|x - y_n\| = \|x - y\|$ by continuity of inner products. We now claim that $z \in M^\perp$.

Let $u \in M$. By multiplying z by a non-zero constant, we can assume that $\langle z, u \rangle \in \mathbf{R}$. We consider $f(t) = \|z + tu\|^2 = \|z^2\| + 2t \langle z, u \rangle + t^2 \|u\|^2$. Note that $\|z + tu\|^2 = \|x - (y - tu)\|^2$, but since $y, u \in M$, we know that $\min f(t) = \|x - y\|^2$. The minimum is indeed attained at $t = 0$, where $f'(0) = 0$. Thus, $f'(t) = 2t \|u\|^2 + 2 \langle z, u \rangle = 0 \iff t_c = -\frac{\langle z,u \rangle}{\|u\|^2} \iff \langle z, u \rangle = 0$ and hence $z \perp u \forall u \in M \implies z \in M^{\perp}$. ∎

Now consider any $y \in \mathcal{H}$. The Cauchy-Schwarz inequality shows that $f_y(x) = \langle x, y \rangle$ is a bounded linear functional on $\mathcal{H}$ such that $\|f_y\| = \|y\|$. The map $y \longrightarrow f_y$ is a conjugate-linear isometry of $\mathcal{H}$ into $\mathcal{H}^*$. This map is also surjective, demonstrated in the next part.

3.6 Riesz Representation, Bases

Theorem 3.23. (Riesz' Representation)
Let $f \in \mathcal{H}^*$. Then, there exists a unique $y \in \mathcal{H}$ such that

$$f(x) = \langle y, x \rangle$$

Proof. (Uniqueness) Suppose $f(x) = \langle y_1, x \rangle = \langle y_2, x \rangle$, $\forall x \in \mathcal{H}$. Thus, $\langle y_1 - y_2, x \rangle = 0$. Set $x = y_1 - y_2 \implies \|y_1 - y_2\| = 0 \implies y_1 = y_2$.

(Existence) Set $M = \{x \in \mathcal{H} : f(x) = 0\}$, assume $f \neq 0$, otherwise there's nothing to do. Since f is continuous, M is therefore a closed subspace, and $M \neq \mathcal{H}$. By orthogonal decomposition, we know that $\mathcal{H} = M \bigoplus M^{\perp}$, and $M^{\perp} \leq \mathcal{H}, \neq \{0\}$. Choose $z \in M^{\perp}$ with $\|z\| = 1$. Given $x \in \mathcal{H}$ and $z \in M^{\perp}$, we construct $u \in M$ as follows:

$$u = f(x)z - f(z)x$$

i.e. $f(u) = f(x)f(z) - f(z)f(x) = 0$. Since $u \in M$, this implies $\langle u, z \rangle = 0$ since $z \in M^{\perp}$. Hence, $f(x) \|z\|^2 - f(z)\langle x, z \rangle = 0 \implies f(x) = f(z)\langle x, z \rangle = \langle x, f(z)z \rangle$. ∎

We now turn the discussion to bases.

Definition 3.24. $\{u_\alpha\}_{\alpha \in \mathcal{A}} \subset \mathcal{H}$ is orthonormal if $\langle u_\alpha, u_\beta \rangle = \delta^\alpha_\beta = 1$ if $\alpha = \beta$, else 0.

Given a linearly independent sequence $\{x_n\}^N \subset \mathcal{H}$, one can construct an orthonormal set $\{u_n\}^N$ from x_n's via Gram-Schmidt:

$$u_1 = \frac{x_1}{\|x_1\|}$$

Given $u_1, \ldots, u_{N-1}$, set

$$v_N = x_N - \sum_{n=1}^{N-1} \langle x_N, u_n \rangle u_n$$

Since $x_N \notin \operatorname{span}\{x_1, \ldots, x_{N-1}\}$, $v_N \neq 0$. Moreover, given any $u_m; k \leq m \leq N-1$,

$$\langle v_N, u_m \rangle = \langle x_N, u_m \rangle - \langle x_N, u_m \rangle = 0$$

Classic Gram-Schmidt.

Theorem 3.25. (Bessel Inequality) For any $x \in \mathcal{H}$ and any orthonormal set $\{u_\alpha\}_{\alpha \in \mathcal{A}} \subset \mathcal{H}$,

$$\sum_{\alpha \in \mathcal{A}} | \langle x, u_\alpha \rangle |^2 \leq \|x\|^2$$

Proof.

$$\begin{aligned}
0 &\leq \left\| x - \sum_{\alpha \in \mathcal{A}} \langle x, u_\alpha \rangle u_\alpha \right\|^2 \\
&= \|x\|^2 - 2\,\mathrm{Re}\left\langle x, \sum \langle x, u_\alpha \rangle u_\alpha \right\rangle + \left\| \sum \langle x, u_\alpha \rangle u_\alpha \right\|^2 \\
&= \|x\|^2 - 2\,\mathrm{Re} \sum \overline{\langle x, u_\alpha \rangle} \langle x, u_\alpha \rangle + \sum \|\langle x, u_\alpha \rangle u_\alpha\|^2 \\
&= \|x\|^2 - 2 \sum |\langle x, u_\alpha \rangle|^2 + \sum |\langle x, u_\alpha \rangle|^2 \\
&= \|x\|^2 - \sum |\langle x, u_\alpha \rangle|^2
\end{aligned}$$

∎

Definition 3.26. We say that $\{u_\alpha\} \subset \mathcal{H}$ is a basis for $\mathcal{H}$ if for every $x \in \mathcal{H}$,

$$x = \sum_\alpha \langle x, u_\alpha \rangle u_\alpha$$

This means that

$$\left\| x - \sum_{\alpha \in A_n} \langle x, u_\alpha \rangle u_\alpha \right\| \to 0$$

as $N \to 0$ where $A_N \subset A$ is any finite subset (independent of ordering).

We say that $\mathcal{H}$ is separable if there exists a countable, orthonormal basis (Hilbert basis). In this case, the Bessel inequality improves to a stronger version.

Theorem 3.27. Let $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ be separable, and $\{u_\alpha\}_{\alpha \in \mathcal{A}}$ a Hilbert basis, i.e. $\forall x \in \mathcal{H}, x = \sum_\alpha \langle x, u_\alpha \rangle u_\alpha$.

1. $\forall \alpha \in \mathcal{A}, \langle x, u_\alpha \rangle = 0 \implies x = 0$
2. $\|x\|^2 = \sum_{\alpha \in \mathcal{A}} |\langle x, u_\alpha \rangle|^2$ (Parseval's identity)

Proof. Assume $\mathcal{A}$ is countable with a Hilbert basis $\{u_\alpha\}$.

$$\|x\| - \sum_{\alpha \in \mathcal{A}_N} \langle x, u_\alpha \rangle u_\alpha \to 0$$

as $A_N \uparrow A$. By Bessel,

$$\|x\|^2 \geq \sum_{\alpha \in \mathcal{A}_N} |\langle x, u_\alpha \rangle|^2, \quad \forall N$$

On the other hand, from the first limit,

$$\left\| x^2 \right\| \leq \sum_{\alpha \in \mathcal{A}_N} |\langle x, u_\alpha \rangle|^2 + E_N$$

where $E_N \to 0$. Take limits, $\left\| x^2 \right\| = \sum_{\alpha \in \mathcal{A}} |\langle x, u_\alpha \rangle|^2$ ∎

The following are thicc.

Proposition 3.28. Every Hilbert space has an orthonormal basis.

Proposition 3.29. If a Hilbert space is separable, i.e. it has a countable orthonormal basis, then every orthonormal basis is countable.

Proof. Take some countable dense set, go through each element one by one and discard those that are linear combinations of the previous accumulated ones. Gram-Schmidt the rest. ∎

3.7 Fourier Coefficients

The following will be a really short subsection on the topic above. If $\mathcal{H}_1, \mathcal{H}_2$ are Hilbert spaces with inner products $\langle \cdot, \cdot \rangle_1, \langle \cdot, \cdot \rangle_2$, a unitary map $U : \mathcal{H}_1 to \mathcal{H}_2$ is an invertible linear map that preserves inner products:

$$\langle Ux, U_y \rangle_2 = \langle x, y \rangle_1, \forall x, y \in \mathcal{H}_1$$

Remark 3.30. Every unitary map is an isometry. To see this, take $y = x$. Conversely, every surjective isometry is unitary.

Unitary maps are the true "isomorphisms" in the category of Hilbert spaces, as they preserve not only the linear structure and topology but also the inner product (hence, the norm as well). In a sense, every Hilbert space kind of looks like an ℓ^2 space, as clarified below.

Proposition 3.31. Let $\{u_\alpha\}$ be a Hilbert basis for $\mathcal{H}$, consider the linear map:

$$\hat{\cdot} : \mathcal{H} \longrightarrow \ell^2(\mathcal{A})$$

given by $\hat{x}(\alpha) = \langle x, u_\alpha \rangle$ is unitary onto $\ell^2(\mathcal{A})$.

Proof. $\hat{\cdot} : \mathcal{H} \longrightarrow \ell^2(\mathcal{A})$ is a linear isometry by the Parseval identity. Moreover, if $f \in \ell^2(\mathcal{A})$, then $\sum_\alpha |f(\alpha)|^2 < \infty$, and

$$\left\{ \sum_{\alpha \in F_n} f(\alpha) u_\alpha \right\}, \quad F_n \subset \mathcal{A}$$

is Cauchy, $F_n \uparrow \mathcal{A}$. Then, $x = \sum_{\alpha \in \mathcal{A}} \langle x, u_\alpha \rangle u_\alpha \in \mathcal{H}$ exists. Then, $f = \hat{x}$ and so $\hat{\cdot} : \mathcal{H} \longrightarrow \ell^2(\mathcal{A})$ is onto. ∎

Example 3.32. Consider $\mathcal{L}^2(S^1) = \mathcal{L}^2(\mathbf{R}/[0,1])$, now consider $\mathcal{A} = \{e^{2i\pi nx}\}_{n \in \mathbf{Z}}$ is a Hilbert basis for $\mathcal{L}^2(S^1)$:

$$\langle u_n, u_m \rangle = \int_0^1 e^{2i\pi(m-n)x} dx$$

equals to 0 if $m \neq n$ else 1.

So, when you're given $f \in \ell^2(S^1)$, it has an n^{th} Fourier coefficient:

$$\begin{aligned} \hat{f}(n) &= \langle f, u_n(x) \rangle \\ &= \int_0^1 f(x) e^{-2i\pi nx} dx \end{aligned}$$

where $u_n(x) = e^{2i\pi nx}$. Parseval:

$$\sum_{n \in \mathbf{Z}} \left| \hat{f}(n) \right|^2 = \|f\|_2^2 = \int_0^1 |f| \, dx$$

Remark 3.33. Given $f \in \mathcal{L}^2(S^1)$, the fact that (as yet unproved) the u_n's are a Hilbert basis means that

$$\left\| f - \sum_{|n| < N} \hat{f}(n) e^{2i\pi nx} \right\|_2 \rightarrow 0, N \rightarrow \infty$$

Note that this does NOT in general imply that $f(x) = \sum_{n \in \mathbf{Z}} \hat{f}(n) e^{2i\pi nx}$. This is much more subtle.

4 $\mathcal{L}^p$ Theory

4.1 $\mathcal{L}^p$ spaces

Let $(X, \mathcal{M}, \mu)$ be a measure space, let $p > 0$ and set

$$\|f\|_p = \left(\int |f|^p \, d\mu\right)^{\frac{1}{p}}$$

$\mathcal{L}^p = \{f : X \to \mathbb{C} : f \text{ measurable with } \|f\|_p < \infty\}$. As with Cauchy-Schwarz in the Hilbert setting, there's a convexity bound at the core of $\mathcal{L}^p$ theory: The Hölder Inequality. We revisit Cauchy-Schwarz.
The convexity (it's actually a concavity) bound, or Young's inequality: $ab \leq \frac{1}{2}(a^2 + b^2), \forall a, b \geq 0$. Given $f, g \in \mathcal{L}^2$ with $\|f\|_2 \neq 0, \|g\|_2 \neq 0$, set

$$F(x) = \frac{f(x)}{\|f\|_2}, \quad G(x) = \frac{g(x)}{\|g\|_2}$$

now, $\|F\|_2 = \|G\|_2 = 1$. Apply convexity bounds with $a = |F(x)|, b = |G(x)|$ yields

$$|FG(x)| \leq \frac{1}{2}\left(|F(x)|^2 + |G(x)|^2\right)$$

Integrating both sides yields

$$\|FG\|_1 \leq \frac{1}{2}\left(\|F\|_2^2 + \|G\|_2^2\right) = 2$$

Thus, implies $\|fg\|_1 \leq \|f\|_2 \cdot \|g\|_2$. This actually implies the Cauchy-Schwarz inequality, $|\langle f, g\rangle_{\mathcal{L}^2}| \leq \|f\|_2 \cdot \|g\|_2$, while $|\langle f, g\rangle_{\mathcal{L}^2}| \leq \|fg\|_1$. Note that setting $a^2 = x, b^2 = y$, we rewrite Young's inequality in terms of x, y:

$$x^{\frac{1}{2}} y^{\frac{1}{2}} \leq \frac{1}{2}(x + y)$$

a special case of $GM \leq AM$.

Lemma 4.1. For any $a, b \geq 0$ and index $p > 1$, conjugate index q with $\frac{1}{p} + \frac{1}{q} = 1$, we have the actual Young's inequality:

$$ab \leq \frac{1}{p}a^p + \frac{1}{q}a^q$$

Proof. Set $a = x^{1/p}, b = y^{1/q}$, we have that $x^{1/p}y^{1/q} \leq \frac{x}{p} + \frac{y}{q}$. Since log is increasing, enough to take logs from both sides, yielding

$$\frac{1}{p}\log x + \frac{1}{q}\log y \leq \log\left(\frac{x}{p} + \frac{y}{q}\right)$$

∎

Now, for the most OP thing in $\mathcal{L}^p$ theory.

Theorem 4.2. (Hölder's Inequality) Assume $1 < p < \infty$, q conjugate exponent, and $f, g : X \to \mathbb{C}$ measurable.

$$\|fg\|_1 \leq \|f\|_p \|g\|_q$$

Moreover, there is equality if and only if $\alpha|f|^p = \beta|g|^q$ for some $\alpha, \beta \in \mathbb{C}$.

Proof. By setting $F = \frac{f}{\|f\|_p}, G = \frac{g}{\|g\|_q}$, enough to use

$$|FG(x)| \leq \frac{1}{p}|F(x)|^p + \frac{1}{q}|G(x)|^q$$

and integrate. ∎

Acknowledgments

I would like to thank Prof. Panangaden and his (boosted) COMP 599 - Mathematical Techniques for Machine Learning for giving me the fastest crash course possible on abstract measure theory and integration. To this day I still wonder why the maths covered in that course is denser than the rationals in the reals.
Special kudos to whoever revises math Wikipedia pages. They are truly awesome.

References

Folland, Gerald B. *Real analysis: modern techniques and their applications.* John Wiley & Sons, 2013.
Panangaden, Prakash. *Labelled Markov Processes.* World Scientific, 2009.