MEASURE THEORY AND CONVERGENCE IN L^p

AMAN THAWANI

ABSTRACT. This is an expository paper discussing the foundations of measure theory and Lebesgue integration, the resulting theory of L^p spaces, and applications of these developments in the analysis of convergence of numerical methods. Major theorems covered include the Carathéodory criterion, monotone convergence, Fatou's lemma, dominated convergence, Hölder's inequality, Minkowski's inequality, the completeness of L^p, and approximation results in L^p.

CONTENTS

1. NOTATION

In this section, we develop some essential notation for later use.

Given a set $A \subset \mathbb{R}^n$, we write the *characteristic function* (also known as the indicator function) of A over $\mathbb{R}^n$ as

$$\chi_A(x) = \begin{cases} 1, x \in A \\ 0, x \notin A \end{cases} .$$

Additionally, we denote the power set of A as 2^A.

Given a sequence of sets A_n, we say A_n *increases to* A, denoted as $A_n \uparrow A$, if $A_n \subset A_{n+1}$ for all n and $\cup_{n=1}^{\infty} A_n = A$. Similarly, we say A_n decreases to A, denoted as $A_n \downarrow A$, if $A_{n+1} \subset A_n$ for all n and $\cap_{n=1}^{\infty} A_n = A$.

Given a sequence a_n, we write the *limit superior* as

$$\limsup a_n = \inf_{m \geq 1} \sup_{m \geq n} a_m.$$

Similarly, the *limit inferior* is

$$\liminf a_n = \sup_{m \geq 1} \inf_{m \geq n} a_m.$$

Date: September 3, 2025.

Given a function f, we let $f^+ = f$ where $f > 0$ and $f^+ = 0$ everywhere else. We let $f^- = -f$ where $f < 0$ and $f^- = 0$ everywhere else.

2. Foundational Measure Theory

In this section, we construct Lebesgue measure, one of the fundamental results of measure theory and the basis for our later analyses. In doing so, we also provide a more generalized method for the construction of measures.

It is important to realize that not all sets are well-behaved. When we eventually construct measures, it is undesirable that the functionality of those measures be corrupted by pathological or generally ill-behaved subsets of our target set. Thus, we require some sort of edifice to ensure that our measure is only applied to sets that behave somewhat predictably. This is the logic behind the introduction of the *algebra*, or more typically the σ-algebra.

Definition 2.1. Given a set X, we say that a collection of subsets of X, denoted $\mathcal{A}$, is an *algebra* on X if the following properties are satisfied:

(1) $X \in \mathcal{A}$.
(2) For all $A \in \mathcal{A}$, we have $A^C := X \setminus A \in \mathcal{A}$.
(3) For any finite sequence of sets $A_1, \cdots, A_n \in \mathcal{A}$, we have $\cup_{i=1}^n A_i \in \mathcal{A}$.

Properties (1) and (2) together imply $\emptyset \in \mathcal{A}$. Moreover, we can show that an algebra is also closed under finite intersections. We choose a new sequence, namely $B_i = X \setminus A_i$, and apply property (2) to see that $B_i \in \mathcal{A}\ \forall i$. We can then use property (3) on the B_i, in conjunction with De Morgan's laws, to obtain the result.

Definition 2.2. An algebra $\mathcal{A}$ is called a σ-algebra if Definition 1.1 (3) extends to countable unions (and therefore, by De Morgan's laws, to countable intersections).

We would like to be able to construct a *measure*, that is, a scalar-valued function defined on a σ-algebra providing some notion of the size–such as length in one-dimensional space or area in two-dimensional space–of sets in the σ-algebra.

Definition 2.3. Thus, we aim for a measure m defined on an input space and a corresponding σ-algebra, denoted $(X, \mathcal{A})$, to satisfy these properties:

(1) $m(\emptyset) = 0$.
(2) For all sets $A \in \mathcal{A}$, $m(A) \geq 0$.
(3) For a pairwise-disjoint sequence of sets $\{A_i\} \subset \mathcal{A}$, it holds that

$$\sum_{i=1}^{\infty} m(A_i) = m(\cup_i A_i).$$

Property (3) above is known as *countable additivity.* Note that with the physical notions measure is intended to represent, it is intuitive that the measure of a disjoint union should equal the sum of the individual measures. These three properties alone give measures most of the practicality we desire, with the specific functionality depending on the precise measure one chooses to use.

However, constructing measures is often notably difficult. We cannot be certain that all sets of interest are measurable–that is, there exist pathological sets for which no meaningful measure, satisfying countable additivity, exists. Thus, we postpone the construction of concrete measures (including Lebesgue measure) until after introducing some preliminary notions, which will be used in conjunction with the σ-algebra to solve this problem.

We first introduce the *outer measure*, which is fundamental in the construction of measures.

Definition 2.4. Given a set X, we say a function $\mu : 2^X \to [0, +\infty]$ is an outer measure if:

(1) $\mu(\emptyset) = 0$.
(2) For subsets A and B, $A \subset B$ implies $\mu(A) \leq \mu(B)$.
(3) For an infinite sequence of subsets $\{A_1, A_2, \cdots\} \subset 2^X$, we have

$$\mu(\cup_i A_i) \leq \sum_i \mu(A_i).$$

Property (3) also holds for finite sequences of subsets: to represent a length-n sequence, we set $A_{n+1}, A_{n+2}, \cdots = \emptyset$ and employ property (1). Also, note that the outer measure does not necessarily preserve countable additivity, but is characterized by the less strict *countable subadditivity.*

Remark 2.5. A measure also retains the property of countable subadditivity over any collection of subsets contained in the σ-algebra the measure is defined on.

The conditions for an outer measure are weaker than those for a measure (observe that countable additivity implies Definition 2.4 (2)). However, unlike the measure, the outer measure is defined on *all* subsets of the set X. By restricting this aspect of the outer measure, one can attain a measure.

Definition 2.6. We say a set A is measurable with respect to the outer measure μ, or that A is μ-measurable, if for all subsets $B \subset X$, we have

$$\mu(B) = \mu(B \cap A) + \mu(B \cap A^C).$$

Intuitively, this means that A is μ-measurable if breaking any arbitrary subset B into two disjoint components defined by A yields countably-additive behavior with respect to μ. One may describe this as a certain *well-defined behavior* of A with respect to μ.

As such, this route to constructing a measure is as follows: we define an outer measure over the entire power set 2^X, with relatively weak properties, and then limit that outer measure to the subset of 2^X consisting of well-behaved sets to obtain a measure. The only potential oversights, the astute reader may notice, are that we have not shown that the aforementioned subset (call it $\mathcal{A}$) of 2^X is necessarily a σ-algebra, and that we have not shown that countable additivity will hold when we restrain the outer mesaure to $\mathcal{A}$. A powerful theorem known as the Carathéodory criterion resolves this discrepancy.

Theorem 2.7. ***Carathéodory Criterion.***

(1) For an outer measure μ defined on 2^X for some set X, the restriction of μ to the collection $\mathcal{A}$ of μ-measurable subsets of X is a σ-algebra.
(2) The restriction of μ to $\mathcal{A}$ is itself a measure.
(3) Any subset $N \subset X$ with $\mu(N) = 0$ is μ-measurable and is therefore an element of $\mathcal{A}$.

Proof. We provide a brief sketch of the proof; the full proof may be found in [1], on pages 26-28.

To prove (1), we first show that $\mathcal{A}$ is an algebra by showing that it satisfies the criteria of Definition 2.1. Note that since $X^C = \emptyset$, we have that for any $B \subset X$,

$$\mu(B \cap X) + \mu(B \cap \emptyset) = \mu(B) + \mu(\emptyset) = \mu(B).$$

Definition 2.1 (1) is thus verified. Confirming Definition 2.1 (2) follows from the symmetry of the definition of a μ-measurable set. Lastly, to verify Definition 2.1 (3), we note that, using induction, it suffices to show that if two sets A_1, A_2 are in $\mathcal{A}$, then their union is in $\mathcal{A}$. This is achievable by writing the arbitrary subset $B \subset X$ as

$$\begin{aligned} B &= (B \cap (A_1 \cup A_2)) \cup (B \cap (A_1 \cup A_2)^C) \\ &= (B \cap A_1) \cup (B \cap A_1^C \cap A_2) \cap (B \cap A_1^C \cap A_2^C), \end{aligned}$$

further decomposing the expression $B \cap A_1^C$ using the definition of μ-measurability, and applying the countable subadditivity of μ.

This allows us to show that $\mathcal{A}$ is closed under countable unions, and is therefore a σ-algebra. Then, applying the countable subadditivity of the measure and the definition of the measurable sets allows us to see that if $A = \cup_i A_i$, then we have

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

This shows that μ is a measure when restricted to $\mathcal{A}$ (as this restriction allows us to invoke the definition of the measurable sets). Lastly, to prove (3), we write a subset E as $(E \cap N) \cup (E \cap N^C)$. We then use the fact that $\mu(N) = 0$ to show that $\mu(E \cap N) = 0$: then, since $(E \cap N^C) \subset E$, we have that

$$\mu(E \cap N) + \mu(E \cap N^C) = \mu(E \cap N^C) \leq \mu(E).$$

However, subadditivity yields the reverse inequality, which means that the two sides are equal. Since E was arbitrary, this means $N \in \mathcal{A}$ by definition, and the proof is thereby complete. □

We can now use this pipeline to construct the Lebesgue-Stieltjes measures, a special case of which is Lebesgue measure, over the real line. Lebesgue measure provides a notion of the "size" of sets and yields a useful integration method, which we will see later on. It is applicable to a wide variety of sets, but, notably, not *all* sets.

Denote the length function over a half-open interval $(a, b]$ as $l((a, b])$. (We use half-open intervals because they "stack" onto each other cleanly–imagine adding $(b, c]$ onto $(a, b]$). We write

$$l((a, b]) = \alpha(b) - \alpha(a)$$

where α is some increasing, right-continuous function. Thus, l is nonnegative.

We define our outer measure on a set X as

$$\mu(E) = \inf \left\{ \sum_i l(A_i) : \cup_i A_i \supset E \right\}$$

where $A_i = (a_i, b_i]$ $\forall i$.

Proposition 2.8. *Given a length function l, the above function μ is an outer measure, and thus if m is its restriction to μ-measurable sets, m is a measure.*

Proof. Observe that $\mu \geq 0$ since $l \geq 0$.

Defintition 2.4 (1) is quickly verified for μ, as we do not need any intervals to cover the empty set, so $\mu(\emptyset) = 0$.

Take subsets $A \subset B \subset X$. Then any group of intervals A_i that covers B also covers A, so $\mu(A)$ cannot exceed $\mu(B)$. Definition 2.4 (2) is thus verified.

To verify Definition 2.4 (3), we first fix $\varepsilon > 0$. We then observe that the half-open intervals of the form $(a, b]$ cover all of $\mathbb{R}$. Therefore, for any set $A_i \subset X$, we can cover A_i with the union of some intervals $I_{i1}, I_{i2}, \cdots$. By the fact that

$$\mu(A_i) = \inf \sum_j l(I_{ij}),$$

where the infimum is over the set of intervals, we can find a set of intervals I_{ij} such that

$$\sum_j l(I_{ij}) \leq \mu(A_i) + \frac{\varepsilon}{2^i}.$$

Since each A_i is contained in $\cup_j I_{ij}$, we have $\cup_i A_i \subset \cup_i \cup_j I_{ij}$. Therefore, by the fact that $\mu(\cup_i A_i)$ is the infimum, over all choices of intervals I_{ij}, of

$$\sum_i \sum_j l(I_{ij}),$$

we can write the following inequality.

$$\begin{aligned}\mu(\cup_i A_i) &\leq \sum_i \sum_j l(I_{ij}) \\ &\leq \sum_i \left[\mu(A_i) + \frac{\varepsilon}{2^i}\right] \\ &= \left[\sum_i \mu(A_i)\right] + \varepsilon.\end{aligned}$$

Since ε was arbitrary, we conclude that Definition 2.4 (3) holds for μ. It follows immediately by the Carathéodory criterion that if m is the restriction of μ to μ-measurable sets, then m is a measure. □

Which measure m we obtain from this process depends on the choice of function α. When α is the identity transformation, m is called Lebesgue measure.

It thus follows that the Lebesgue measure of an interval is its length, and the Lebesgue measure of a single point (and thus of a countable set) is 0.

Going forward, we will use m to denote Lebesgue measure. Additionally, we define a *measure space* $(X, \mathcal{A}, \mu)$ to be a set X equipped with a σ-algebra $\mathcal{A}$ and a measure on $\mathcal{A}$, namely μ.

3. The Lebesgue Integral

Another useful application of measures is to functions. In this section, we first lay out what it means for a function to be *measurable*, and then construct the Lebesgue integral and analyze its useful properties. The Lebesgue integral will be fundamental to the later analysis of L^p functional spaces.

Definition 3.1. Assume we have a σ-algebra $\mathcal{A}$ on a set X. We say a function $f : X \to \mathbb{R}$ is measurable with respect to $\mathcal{A}$ if for any number $a \in \mathbb{R}$, it holds that

$$\{x : f(x) > a\} \in \mathcal{A}.$$

Proposition 3.2. *If the functions f and g are measurable with respect to a σ-algebra $\mathcal{A}$, then so are the functions $cf, f+g, -f, fg, \max(f,g)$, and $\min(f,g)$, where the max and min are taken pointwise.*

The proof of this proposition may be found on pages 44-45 of [1].

Intuitively, measurability of f essentially means that f is well-behaved on X relative to $\mathcal{A}$. One can also interpret it as meaning that f's behavior is compatible with the structure of information that $\mathcal{A}$ allows one to observe on X (whether $f(x) > a$ for some a).

In order to integrate a function f over a set $A \subset X$, we first need a means to approximate f in a simple, feasible way.

Definition 3.3. A *simple function* $\phi : X \to \mathbb{R}$ is a function that can be written in the form

$$\phi(x) = \sum_{i=1}^{n} a_i \chi_{E_i}(x)$$

where χ is the indicator function from Section 1, the E_i are measurable sets, and $a_i \in \mathbb{R}$ for each i.

Proposition 3.4. *If $f \geq 0$ is measurable on X (with σ-algebra $\mathcal{A}$), then there exists a sequence of increasing simple functions $\phi_n(x)$ such that $\phi_n \to f$ pointwise on X.*

Proof. For every n, we assemble the sets A_{in} as follows:

$$A_{in} = \Big\{ x : f(x) \in \Big[\frac{i-1}{2^n}, \frac{i}{2^n}\Big) \Big\}$$

with $n2^n$ of these sets ($1 \leq i \leq n2^n$) for each n. One can imagine the A_{in} as $n2^n$ evenly-sized, adjacent, disjoint baskets, covering all values of f in $[0, n)$. Additionally, construct $B_n := \{x : f(x) \geq n\}$. B_n is equal to $X \setminus (\cup_i A_{in})$.

Then, we let

$$\phi_n(x) = \Big[\sum_{i=1}^{n2^n} \frac{i-1}{2^n} \chi_{A_{in}}(x) \Big] + n\chi_{B_n}.$$

Regardless of the input x, exactly one of the $1 + n2^n$ indicator functions appearing in $\phi_n(x)$ will not be zero. Specifically, the χ_{B_n} will "activate" if $f(x) \geq n$, and otherwise only the χ_{A_in} resulting in the closest downward approximation of $f(x)$ will be nonzero.

So, $\phi_n(x) = n$ if $f(x) \geq n$, and it approximates $f(x)$ with a maximal error of 2^{-n} if $f(x) < n$. Thus, $\phi_n \to f$ as $n \to \infty$. $\square$

We are now ready to define the Lebesgue integral of a simple function, from which the definition of the Lebesgue integral of a generic measurable function f will follow. The convergent approximations shown to exist by Proposition 3.4 will be vital.

Definition 3.5. We define the Lebesgue integral of a simple function ϕ (defined on a measure space $(X, \mathcal{A}, \mu)$) as

$$\int \phi d\mu = \int \sum_{i=1}^{n} a_i \chi_{E_i} d\mu = \sum_{i=1}^{n} a_i \mu(E_i)$$

Conventionally, if $a_i = 0$ and $\mu(E_i) = \infty$, we say $a_i \cdot \mu(E_i) = 0$.

It is intuitive that the integral of a simple function with respect to Lebesgue measure is the (weighted) sum of the "lengths" of the E_i. Considering that the a_i determine the values taken on by ϕ where $\chi_{E_i} \neq 0$, this integral is the area (or equivalent, depending on the dimension) under the function ϕ on the set X.

For the following two definitions, assume we have a measure space $(X, \mathcal{A}, \mu)$.

Definition 3.6. If $f \geq 0$ is measurable, then we write

$$\int f d\mu = \sup \Big\{ \int \phi d\mu : 0 \leq \phi \leq f \Big\}$$

where ϕ is restricted to be a simple function.

For generic measurable f, recall the definitions of f^+ and f^- from Section 1. We write

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

Definition 3.7. A measurable function f is said to be integrable if $\int |f| d\mu < \infty$.

We now show some useful properties regarding the Lebesgue integral.

Proposition 3.8. *The following hold on a measure space $(X, \mathcal{A}, \mu)$ for measurable functions f and g:*

(1) If $f \leq g$, then $\int f d\mu \leq \int g d\mu$.
(2) For $c \geq 0$,

$$\int cf d\mu = c \int f d\mu.$$

(3) If $\mu(N) = 0$, then $\int_N f d\mu = 0$.

Proof. (1) Let $\mathcal{C}_f$ be the collection of all simple functions that are $\leq f$. Let $\mathcal{C}_g$ be the collection of all simple functions $\leq g$. Since $f \leq g$, we have $\mathcal{C}_f \subset \mathcal{C}_g$.

Let $\mathcal{C} = \mathcal{C}_g \setminus \mathcal{C}_f$.

Thus,

$$\int g d\mu = \sup_{\phi \in \mathcal{C}_g} \int \phi d\mu = \sup_{\phi \in \mathcal{C}_f \cup \mathcal{C}} \int \phi d\mu \geq \sup_{\phi \in \mathcal{C}_f} \int \phi d\mu = \int f d\mu.$$

(2) Note that if f is measurable, then so is cf. Moreover, if there is a sequence of simple functions $\{\phi_n\}$ such that $\phi_n \to f$ (as Proposition 3.4 decrees there is), then the sequence $\{\psi_n\}$ defined by $\psi_n = c \cdot \phi_n$ converges to cf. Hence,

$$\begin{aligned} \int cf d\mu &= \Big\{ \sup \int \phi d\mu : 0 \leq \phi \leq f, \phi \text{ simple} \Big\} \\ &= \lim_{n\to\infty} \int \psi_n d\mu \\ &= \lim_{n\to\infty} \int c\phi_n d\mu \\ &= \lim_{n\to\infty} c \int \phi_n d\mu = c \int f d\mu. \end{aligned}$$

(3) For each simple function ϕ, let the A_i be measurable subsets of N.

$$\int_N f d\mu = \sup \Big\{ \int_N \phi d\mu : 0 \leq \phi \leq f, \phi \text{ simple} \Big\}$$

$$= \sup \int_N \Big\{ \sum_{i=1}^{n} a_i \chi_{A_i} \Big\}$$

$$= \sup \Big\{ \sum_{i=1}^{n} a_i \mu(A_i) \Big\} = \sup \Big\{ \sum_{i=1}^{n} 0 \Big\} = 0$$

where $A_i \subset N \implies \mu(A_i) = 0 \; \forall i$ by countable subadditivity. □

We now cover four major theorems fundamental to analysis of the Lebesgue integral, the first of which is the monotone convergence theorem.

Theorem 3.9. *Monotone Convergence Theorem.*

Suppose our measure space is $(X, \mathcal{A}, \mu)$. Suppose f_n is a sequence of nonnegative μ-measurable functions with $f_1 \leq f_2 \leq \cdots$ pointwise and with

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

Then,

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

Proof. By Proposition 3.8 (1), we know that $\int f_n d\mu$ is increasing. Let L be its limit; then, $L \leq \int f d\mu$ since each $\int f_n d\mu$ is less than or equal to $\int f d\mu$.

Thus, we are done if we can show that $\int f d\mu \leq L$.

Let c be arbitrary in $(0, 1)$ and let $\phi := \sum_{i=1}^{N} a_i \chi_{E_i}$ be a simple function that is $\leq f$ pointwise.

For each value of n, define

$$A_n := \Big\{ x : f(x) \geq c \cdot \phi(x) \Big\}.$$

A_n increases to the whole space X as $n \to \infty$ because f_n increases to f, which means that for each value of n we have

$$\int_X f_n d\mu \geq \int_{A_n} f_n d\mu \geq c \int_{A_n} \phi d\mu = c \int_{A_n} \sum_{i=1}^{N} a_i \chi_{E_i}$$

$$= c \sum_{i=1}^{m} a_i \mu(E_i \cap A_n)$$

where the last expression comes from the fact that the integral is taken over A_n, not E_i, so the $\chi_{E_i}(x)$ only contributes to the integral when $x \in A_n$.

We know that $A_n \uparrow X, n \to \infty$. Therefore, sending $n \to \infty$ we get $(E_i \cap A_n) \to E_i$ for each i. Hence,

$$L = \lim_{n\to\infty} \int f_n d\mu \geq c \sum_{i=1}^{m} a_i \mu(E_i) = c \int \phi d\mu.$$

Since c was arbitrary in $(0, 1)$, we get $L \geq \int \phi d\mu$, and when we take the supremum over all ϕ obeying the constraint we get

$$L \geq \int f d\mu$$

and the equality is thus proven. □

We next show an important property of the Lebesgue integral which has not been proven yet, namely its linearity. Many of the following proofs would be extremely cumbersome, if not impossible, without this property. Additionally, linearity is extremely useful because it allows us to treat Lebesgue integration as an algebraic operation–this will be vital in the theory of L^p spaces.

Theorem 3.10. *If f, g are either (integrable) or (nonnegative and measurable), then the Lebesgue integral is linear. That is, $\int (f+g)d\mu = \int f d\mu + \int g d\mu$ where μ is a measure.*

The proof of this theorem may be found on pages 9-10 of [2].

Theorem 3.11. *Fatou's Lemma.*

Suppose $\{f_n\}$ is a sequence of nonnegative measurable functions on a measure space $(X, \mathcal{A}, \mu)$. Then,

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

Proof. Let $g_n = \inf_{i \geq n} f_i$. By the definition of lim inf, the g_n are increasing and $\lim_{n\to\infty} g_n = \liminf_{n\to\infty} f_n$. By the definition of infimum, $g_n \leq f_i \ \forall i \geq n$. By Proposition 3.8 (1), this implies that $\int g_n d\mu \leq \int f_i d\mu \ \forall i \geq n$. Hence

$$\int g_n d\mu \leq \inf_{i\geq n} \int f_i d\mu.$$

Taking the limit as $n \to \infty$ on both sides and using the monotone convergence theorem on the left side gives the desired result. □

Fatou's lemma, while quite a major result, indeed serves as a lemma for the proof of a variety of other theorems. The first of these is the dominated convergence theorem, which will give us the necessary conditions to be able to conclude that a sequence of integrals converges to the integral of the limit of the sequence of integrands. This will be very useful when we later analyze convergence in L^p.

Theorem 3.12. *Dominated Convergence Theorem.*

Suppose that $\{f_n\}$ is a sequence of measurable functions on a measure space $(X, \mathcal{A}, \mu)$, and that there exists an integrable function g, such that $|f_n(x)| \leq g(x)$ for all n and all x. Additionally suppose that $f_n \to f$ pointwise. Then,

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

Proof. For all n, since $|f_n| \leq g$ for all x, it follows that $g + f_n \geq 0$. Hence Fatou's lemma applies for the sequence $h_n := g - f_n \geq 0$ on $(X, \mathcal{A}, \mu)$. Moreover, since f is the limit of the f_n, we have (using linearity) that

$$\int f d\mu + \int g d\mu = \int (f+g) d\mu \leq \liminf_{n\to\infty} \int (f_n + g) d\mu = \liminf_{n\to\infty} \int f_n d\mu + \int g d\mu.$$

The inequality comes by Fatou's lemma.

We assumed g to be integrable, so its Lebesgue integral with respect to μ is finite. We thus subtract that quantity from the inequality

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

to get

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

Similarly, the constraints of the theorem imply $g - f_n \geq 0$ pointwise for all n. Exploiting this by the same steps as previously (via Fatou's lemma, linearity, and the integrability of g), we attain the inequality

$$\int (-f) d\mu \leq \liminf_{n\to\infty} \int (-f_n) d\mu.$$

Recognizing that the right side is equal to $-\limsup_{n\to\infty} \int f_n d\mu$ and negating the entire inequality yields

$$\int f d\mu \geq \limsup_{n\to\infty} f_n d\mu,$$

which, together with the previous result, proves the theorem. □

4. L^p Spaces

Having developed the Lebesgue integral, we now desire a robust framework to analyze the behavior of functions, integration and integrability, and convergence. The L^p vide useful mechanisms that allow us to perform these functional analyses.

Definition 4.1. Given a measure space $(X, \mathcal{A}, \mu)$, we say a condition holds *almost everywhere* on X, abbreviated as a.e., if it holds on $X \setminus N$ where $N \subset X$ and $\mu(N) = 0$.

Definition 4.2. Given a measure space $(X, \mathcal{A}, \mu)$, we say μ is σ-finite if there exist sets $E_1, E_2, \cdots$ such that $E_i \in \mathcal{A}$ for all i, $\mu(E_i) < \infty$ for all i, and $X = \cup_i E_i$.

Definition 4.3. Moreover, a measure space is called a σ-finite measure space if the measure is σ-finite.

Throughout this section, we assume that any measure μ is σ-finite.

We start by defining the L^p norm, which provides a notion of the overall "size" or "magnitude" of a function–although the precise meaning varies with the exponent p. For example, the L^2 norm has useful applications in physics and mathematics, as it is an example of the notion of an *inner product space*.

Definition 4.4. Assume we have a σ-finite measure space $(X, \mathcal{A}, \mu)$. For a finite value of p, namely $1 \leq p < \infty$, we write the L^p norm of a function f as

$$\|f\|_p = \sqrt[p]{\int |f|^p d\mu}.$$

For $p = +\infty$, we write the corresponding norm (L^∞ norm) of f as

$$\|f\|_\infty = \inf\{M \geq 0 : \mu(x : |f(x)| \geq M) = 0\}.$$

We let the functional space L^p be the set of equivalence classes that can be constructed from all functions whose L^p norm is finite. Two functions are defined to be in the same equivalence class if they are equal almost everywhere on the space X. By the definition of "almost everywhere" and the countable subadditivity of the measure, this means that all functions in an equivalence class are equal almost everywhere. One can think about L^p simply as the space of functions whose L^p norm is finite; however, attempting to rigorously construct that functional space results in a fundamental issue with the L^p norm.

Remark 4.5. Note that L^1 is exactly the set of equivalence classes of all Lebesgue integrable functions!

However, a *norm* must satisfy certain properties, so we have not yet shown that the L^p norm is actually a norm on the space L^p.

It is clear that all but one of the norm properties are satisfied for the L^p norm: $\|f\|_p = 0$ when $f = 0$ a.e., $\|f\|_p > 0$ when f is not identically 0 a.e., and for any $\alpha \in \mathbb{R}$ it holds that $\|\alpha f\|_p = |\alpha|\|f\|_p$. These are all quickly verified from the definitions of the L^p norm for finite and infinite p.

Toward showing that the L^p norm obeys the triangle inequality, we first show another useful inequality, namely Hölder's inequality.

Definition 4.6. For a given value of p, we let q be the quantity in $[1, +\infty]$ solving

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Conventionally we say that $1/\infty = 0$. This q is henceforth referred to as the conjugate exponent of our p.

Theorem 4.7. *Hölder's inequality.*

On a measurable space $(X, \mathcal{A}, \mu)$, let $1 \le p \le \infty$ and let q be the conjugate exponent of p. Then for $\mathcal{A}$-measurable functions f and g we have

$$\int |fg| d\mu \le \|f\|_p \cdot \|g\|_q.$$

When $p = 2$ this inequality becomes the well-known Cauchy-Schwarz inequality.

Proof. If $p = \infty$ (and thus $q = 1$), then $|f| \le M := \|f\|_\infty$ almost everywhere. Therefore, we have

$$\int |fg| d\mu \le M \int |g| d\mu = \|f\|_\infty \|g\|_1$$

by Proposition 3.8 (1). The inequality is symmetric with respect to p and q, so the case $p = 1, q = \infty$ is proven the same way.

Now assume $1 < p < \infty$. If $\|f\|_p = 0$ or $\|g\|_q = 0$, then either f or g is equal to 0 almost everywhere, meaning that $\int |fg| d\mu = 0$. Both sides of the inequality thus evaluate to 0.

Therefore, we now assume $\|f\|_p, \|g\|_q > 0$. We also assume both norms are finite, because if either one were infinite the inequality would trivially hold.

Define $F(x) = |f(x)|/\|f\|_p$ and $G(x) = |g(x)|/\|g\|_q$. By construction, showing that $\int FG d\mu \le 1$ is equivalent to proving the theorem.

Since the function $h(x) = e^x$ is everywhere-convex on $\mathbb{R}$, it holds for all $a \le b$ and all $\lambda \in [0, 1]$ that

$$e^{\lambda a + (1-\lambda) b} \le \lambda e^a + (1 - \lambda) e^b.$$

We select $\lambda = \frac{1}{p}$, which makes $1 - \lambda$ equal to $\frac{1}{q}$, and we let $a = p \log F(x), b = q \log G(x)$. Substituting into the convexity inequality above yields

$$F(x)G(x) \le \frac{F(x)^p}{p} + \frac{G(x)^q}{q}.$$

Using Proposition 3.8 (2) and the linearity of the integral, integrating both sides yields

$$\int F(x)G(x) d\mu \le \frac{1}{p} \int F(x)^p d\mu + \frac{1}{q} \int G(x)^q d\mu$$

$$= \frac{1}{p}\int\Big(\frac{|f(x)|}{\|f\|_p}\Big)^p + \frac{1}{q}\int\Big(\frac{|g(x)|}{\|g\|_q}\Big)^q$$
$$= \frac{1}{p\|f\|_p^p}\|f\|_p^p + \frac{1}{q\|g\|_q^q}\|g\|_q^q = \frac{1}{p} + \frac{1}{q} = 1.$$

□

We can now use Hölder's inequality to prove Minkowski's inequality, which is the last step toward establishing the L^p norm as a norm. However, a brief lemma is first required.

Lemma 4.8. *If $p \in [1, +\infty)$ and $a, b \geq 0$ it holds that*

$$(a+b)^p \leq 2^{p-1}(a^p + b^p).$$

Proof. We assume that $a > 0$ and $p > 1$ since the cases where $a = 0$ and where $p = 1$ are trivial.

Let $x = b/a$: then, after dividing both sides by a^p, we realize that showing that

$$f(x) := 2^{p-1}x^p - (1+x)^p \geq 0$$

for all nonnegative x will imply the lemma. Using the assumption that $p > 1$, convexity arguments for the function $g(x) = x^p$ yield the inequality

$$\frac{1^p + x^p}{2} \geq \Big(\frac{1+x}{2}\Big)^p.$$

Multiplying both sides by 2^p shows that $x \geq 0 \implies f(x) \geq 0$. □

Theorem 4.9. *Minkowski's Inequality.*

If $p \in [1, +\infty]$ and f, g are measurable functions on a measure space $(X, \mathcal{A}, \mu)$, then

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof. First let $p = \infty$. Then if $\|f\|_\infty = M_f$ and $\|g\|_\infty = M_g$, let $A_f = \{x : f(x) \geq M_f\}$ and $A_g = \{x : f(x) \geq M_g\}$. Both sets have measure 0 by the definition of "almost everywhere." Then the set $A := \{x : |f(x)+g(x)| \geq M_f + M_g\}$ is contained in the union of A_f and A_g. Therefore,

$$\mu(A) \leq \mu(A_f \cup A_g) \leq \mu(A_f) + \mu(A_g) = 0.$$

The L^∞ norm of $f+g$ is therefore bounded above by $M_f + M_g = \|f\|_\infty + \|g\|_\infty$.

The numerical triangle inequality tells us that $|f+g| \leq |f| + |g|$ pointwise, and so using Proposition 3.8 (1) proves the case where $p = 1$.

We now assume $1 < p < \infty$ and additionally assume that neither $\|f\|_p$ nor $\|g\|_p$ is infinite, and that $\|f+g\|_p > 0$, as otherwise the inequality would trivially hold.

If we let $a = |f(x)|$ and $b = |g(x)|$ in Lemma 4.8 and integrate both sides using Proposition 3.8 (1), we get

$$\int |f+g|^p d\mu \leq 2^{p-1}\Big(\int |f(x)|^p d\mu + \int |g(x)|^p d\mu\Big).$$

Therefore we know that $\|f+g\|_p$ is finite.

Observe that

$$|f+g|^p \leq |f||f+g|^{p-1} + |g||f+g|^{p-1}$$

since we can factor $|f+g|^{p-1}$ out of the right-hand side and apply the numerical triangle inequality. Hence, letting q be the conjugate exponent of p and applying Hölder's inequality gives

$$\begin{aligned}\int |f+g|^p &\leq \|f\|_p\Big(\int |f+g|^{q(p-1)}\Big)^{1/q} + \|g\|_p\Big(\int |f+g|^{q(p-1)}\Big)^{1/q} \\ &= (\|f\|_p + \|g\|_p)\|f+g\|_p^{p/q}\end{aligned}$$

since $q(p-1) = pq - q = p$ by the conjugate exponent equation. Recognizing the left side as $\|f+g\|_p^p$ and dividing through by $\|f+g\|_p^{p/q}$ gives the desired inequality. □

We have now proven that L^p is a normed functional space. Moreover, the following equality condition holds.

Remark 4.10. Equality holds in Minkowski's inequality if and only if $af = bg$ for real numbers $a, b \geq 0$.

A full proof of this condition may be found in pages 8 and 9 of [3].

Next, it will be important to prove that L^p is complete. This will be essential for the later justification of many important convergence arguments.

Theorem 4.11. *If $1 \leq p \leq \infty$, then the functional space L^p, when taken to be a metric space, is complete.*

Prior to embarking upon the proof of this theorem, we observe that it makes sense to view L^p as a metric space, with the L^p norm serving as the distance function: in L^p, $|f-g| = \|f-g\|_p$. Since the L^p norm is nonnegative, symmetric (as $|f-g| = |g-f|$), and $\|f-g\|_p = 0$ when $f = g$ a.e., this setup satisfies the intuitive requirements of a metric space.

Proof. We first prove for a fixed $p < \infty$. Let f_n be a Cauchy sequence in L^p. We are tasked with showing that its limit is in L^p. Since f_n is Cauchy, we know that for every $\varepsilon_j > 0$, there exists $n_j > 0$ such that

$$\|f_n - f_m\| < \varepsilon_j$$

for all $n, m \geq n_j$. Let each $\varepsilon_j = 2^{-j-1}$, and additionally assume WLOG that $n_j \geq n_{j-1}$ (if not, we can redefine $n_j = n_{j-1} + 1$ and the above inequality will still hold).

We set $n_0 = 0$ and f_0 to be the constant 0 function. Then, we propose the limit function of the Cauchy sequence as

$$f = \sum_m \Big[f_{n_m} - f_{n_{m-1}}\Big].$$

To show that this series converges absolutely (a precondition for f being in L^p), we first set $g_j(x) = \sum_{m=1}^{j} |f_{n_m}(x) - f_{n_{m-1}}(x)|$, which clearly increases with j pointwise. We let the limit, if it exists, be denoted by $g(x)$. Then, we have (by Minkowski's inequality) that

$$\|g_j\|_p \leq \sum_{m=1}^{j} \|f_{n_m} - f_{n_{m-1}}\|_p$$

$$\le \|f_{n_1} - f_{n_0}\|_p + \sum_{m=2}^{j} 2^{-(m-1)-1}$$
$$\le \|f_{n_1}\|_p + \frac{1}{2}$$

by our condition that $n_0 = 0$ and thus $f_{n_0}(x) = 0$.

Now Fatou's lemma yields that

$$\int |g(x)|^p d\mu \le \lim_{j\to\infty} \int |g_j(x)|^p d\mu = \lim_{j\to\infty} \|g_j(x)\|_p^p \le \Big(\|f_{n_1}(x)\| + \frac{1}{2}\Big)^p < \infty$$

where we know that $\|f_{n_1}\|$ is finite by the definition of the n_i and their correspondence with our chosen ε_i.

Hence the limit function $g(x)$ is finite almost everywhere, so the series converges absolutely almost everywhere. This means that the proposed limit function f is well-defined almost everywhere. On the measure-zero set where the absolute convergence does not hold (which may or may not be the empty set), set $f = 0$.

From the definition of f,

$$f(x) = \lim_{J\to\infty} \sum_{m=1}^{J} \big[f_{n_m}(x) - f_{n_{m-1}}(x)\big] = \lim_{J\to\infty} f_{n_J}(x)$$

since $f_{n_0} = 0$. Now, Fatou's lemma yields the following: for any j,

$$\|f - f_{n_j}\|_p^p = \int |f - f_{n_j}|^p d\mu \le \liminf_{J\to\infty} \int |f_{n_J} - f_{n_j}|^p d\mu$$
$$= \liminf_{J\to\infty} \|f_{n_J} - f_{n_j}\|_p^p \le 2^{-(j+1)p}.$$

This means that as $j \to \infty$, we have $\|f - f_{n_j}\|_p \to 0$. Thus, the subsequence we built from the n_j values converges under the L^p metric (equivalent to the L^p norm). We aim to show that the whole sequence converges.

Fix $\varepsilon > 0$; then there exists $N \in \mathbb{N}$ such that $\|f_n - f_m\|_p < \varepsilon$ when $n, m \ge N$. Note that this implies that $\|f_{n_j} - f_m\|_p < \varepsilon$ as long as we force j large enough that $n_j \ge N$. Since f has been established to be the limit of f_{n_j}, we once more employ Fatou's lemma and obtain that

$$\|f - f_m\|_p^p \le \liminf_{j\to\infty} \|f_{n_j} - f_m\|_p^p$$
$$\implies \|f - f_m\|_p \le \liminf_{j\to\infty} \|f_{n_j} - f_m\|_p < \varepsilon$$

since we are taking j to ∞. Thus, $f_m \to f$ in the L^p metric as $m \to \infty$.

Now we consider $p = \infty$. Observe that we have already shown, using the ε_j argument, that the absolute convergence of an arbitrary series in the L^p norm implies the completeness of L^p. Thus, let the sequence of functions f_n be such that

$$\sum \|f_n\|_\infty < \infty.$$

In other words, the series converges absolutely in L_∞. Now let

$$G_n = \sum_{k=1}^{n} |f_k|$$

and $G = \lim_{n\to\infty} G_n$.

By Minkowski's inequality,

$$\|G_n\|_\infty = \Big\| \sum_{k=1}^{n} |f_k| \Big\|_\infty \le \sum_{k=1}^{n} \Big\| |f_k| \Big\|_\infty = \sum_{k=1}^{n} \Big\| f_k \Big\|_\infty < \infty.$$

Since $|f_k| \le \|f_k\|_\infty$ almost everywhere by the definition of the L^∞ norm, it follows that $\|G\|_\infty < \infty$ almost everywhere.

Now let $F = \sum_k f_k$. Convergence of F guarantees completeness of L^∞.

We aim to use the Cauchy criterion to show that F converges in the L^∞ norm by showing that the partial sum sequence is Cauchy. Observe that by Minkowski's inequality,

$$0 \le \Big\| F - \sum_{k=1}^{n} f_k \Big\| = \Big\| \sum_{k=n+1}^{\infty} f_k \Big\|_\infty \le \sum_{k=n+1}^{\infty} \|f_k\|_\infty.$$

The last sum must go to zero by the "divergence test" since f_k converges absolutely in L_∞. Hence, F is Cauchy and therefore convergent, which means that L^∞ must be complete.

□

5. Convergence

A particularly helpful use of the L^p spaces is to analyze the convergence of approximation methods. In various applications, mathematicians, scientists, and analysts aim to attain, as closely as possible, the behavior of a given function. Since the L^p distance metric gives a clear notion of the "distance" between functions, it is an excellent method of quantifying the error of approximation. Another application is *quadrature*, or the numerical integration of functions. As we will see later on, particularly intuitive for these approaches are the L^1 and L^2 norms.

As such, in this section, we first motivate the use of our theorems. We then prove the density of simple functions in L^p for $p < \infty$ and develop useful results including the Pythagorean identity in L^p approximation. Then, we end with error bounds in L^p and Parseval's identity.

We first lay out some different forms of convergence which will be useful for further analysis.

Definition 5.1. On a measure space $(X, \mathcal{A}, \mu)$, we say a sequence of measurable functions f_n converges *almost everywhere* to a measurable function f if $f - f_n \to 0$ as $n \to \infty$ everywhere on $X \setminus N$, where $\mu(N) = 0$.

Definition 5.2. Let $1 \le p < \infty$. On a measure space $(X, \mathcal{A}, \mu)$, we say that a sequence of measurable functions f_n converges to a measurable function f *in L^p* if

$$\int |f - f_n|^p d\mu \to 0 \iff \|f - f_n\|_p \to 0$$

as $n \to \infty$.

This form of convergence is intuitive when we consider the argument we used to prove the convergence of the arbitrary Cauchy sequence in L^p under the L^p norm, or metric, in the previous section. It means that "the L^p metric between f and f_n vanishes as $n \to \infty$."

Next, we aim to show that the set of simple functions is dense in $L^p(\mathbb{R})$ (with $p < \infty$). This will validate the numerical approximation of integrals by step functions, a form of quadrature.

Proposition 5.3. *On a measure space $(X, \mathcal{A}, \mu)$ and for $p \in [1, \infty)$, if μ is σ-finite, then the set of simple functions is dense in the space L^p.*

Proof. We let f be a function in L^p. Then, we want to find a sequence of simple functions ϕ_n such that $\|f - \phi_n\|_p \to 0$ as $n \to \infty$, which will show that the set of simple functions is dense in the L^p "metric".

We define a sequence of truncated functions f_n, by the equation

$$f_n = \begin{cases} f(x), |f(x)| \leq n \text{ and } x \in A_n \\ 0, \text{else} \end{cases}$$

where A_n is a measurable subset of X with $\mu(A_n) < \infty$ for all n and $A_n \uparrow X$. Then, it follows that for all n, $|f_n(x)| \leq |f(x)|$, and additionally that $f_n(x) \to f(x)$ almost everywhere on X (since the A_n increase to X).

Note that p is finite. We can thus use the dominated convergence theorem with $|f|^p$ as the dominating function, and taking the pth root yields that $\|f - f_n\|_p \to 0$ almost everywhere. Therefore, it suffices to show that we can accurately approximate each f_n using simple functions with an error that vanishes as $n \to \infty$.

We fix some n. Then, we note that f_n is, by construction, bounded with finite support (zero outside some set A_n of finite measure). We proceed by letting $m \in \mathbb{N}$ be the number of simple functions we want to use to approximate f_n. (The error will vanish as $m \to \infty$.) Then, we partition the range interval $[-n, +n]$ into $K := 2nm$ disjoint intervals of length $1/m$ by the following construction:

$$I_k = \Big[-n + \frac{k-1}{m}, -n + \frac{k}{m} \Big), k = 1, 2, \cdots, K.$$

We add the point $\{n\}$ to I_K, and then the I_k cover $[-n, n]$.

Next, define the set E_k as the preimage of the interval I_k. Since every point in each I_k maps back to the corresponding E_k, we have that $\cup_k E_k = A_n$. Additionally, the E_k are clearly disjoint.

We can now define the simple functions based on n and the fineness of the intervals $1/(2m)$. We let

$$\phi_{nm}(x) = \sum_{k=1}^{K} a_k \chi_{E_k}(x)$$

where each a_k is the midpoint of the corresponding interval I_k. Note that since $\cup_k E_k = A_n$, this function is also zero outside of A_n because all the indicator functions are zero.

Fix some interval width $1/m$. If the point x is in E_j for some j, then it is not in any other E_k since these sets are disjoint. Therefore, $\phi_{nm}(x) = a_j$. Additionally, this means that $f_n(x) \in I_j$ by construction. Since a_j is the midpoint of I_j, we have

$$|f_n(x) - \phi_{nm}(x)| = |f_n(x) - a_j| \leq \frac{1}{2m}$$

for all $x \in A_n$. Moreover, the approximation error is 0 outside of A_n since both functions are equal to 0.

Now we can show that f can be approximated–with respect to the L^p norm–with simple functions, and thus that the set of simple functions is dense in L^p.

We have shown that $\|f - f_n\|_p \to 0$ a.e.. Therefore, given $\varepsilon > 0$, fix n large enough that $\|f - f_n\|_p < \varepsilon/2$. Then, choose $m > \sqrt[p]{\mu(A_n)}/\varepsilon$. It follows by Proposition 3.8

(1) and (2), and the fact that $p \geq 1$, that

$$\|f_n - \phi_{nm}\|_p = \sqrt[p]{\int_{A_n} |f_n(x) - \phi_{nm}(x)|^p d\mu} \leq \sqrt[p]{\int_{A_n} \Big(\frac{1}{2m}\Big)^p d\mu}$$
$$= \frac{1}{2m}\sqrt[p]{\mu(A_n)} < \varepsilon/2.$$

Finally, we apply Minkowski's inequality to see that

$$\|f - \phi_{nm}\|_p \leq \|f - f_n\|_p + \|f_n - \phi_{nm}\|_p < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

□

Given this result, we are now ready to prove a more sophisticated approximation result. This result will give us the notion of a "best" approximation, and is more precise than the previous result, which only informs us that it is possible to approximate a function in L^p with step functions. The result to come is one that will be foundational for tools such as least-squares regression and Fourier series.

Definition 5.4. We say a subset $K \subset L^p$ is *convex* if for any two functions $f, g \in K$ and any $\lambda \in [0, 1]$, it holds that the combination

$$\lambda f + (1 - \lambda)g \in K.$$

This is notably similar to the notion of convexity in finite-dimensional real spaces.

For example, if we fix a finite basis of functions $g_1(x), g_2(x), \cdots, g_n(x) \in L^p$, then the span

$$\Big\{\sum_i a_i g_i(x) : a_i \in \mathbb{R}\Big\}$$

is convex, since taking a linear combination of two elements of the set produces another element of the set (by the definition of a span).

Definition 5.5. We say a subset $K \subset L^p$ is *closed* if it contains all of its limit points. In other words, K is closed if it holds that

$$f_n \in K\ \forall n, \|f - f_n\|_p \to 0 \text{ as } n \to \infty \implies f \in K.$$

These notions are necessary for the following result, which is restricted to closed, convex subsets of L^p.

Theorem 5.6. *Given a fixed $p \in [1, \infty)$, a function $f \in L^p$, and a closed, convex subset $K \subset L^p$, there exists a unique function $g \in K$–known as the best approximation function of f–that minimizes $\|f - g\|_p$.*

Proof. Let

$$d = \inf_{h \in K} \|f - h\|_p.$$

We want to show that d is attained by exactly one function g.

Since d is the infimum, there exists a sequence g_n such that $g_n \in K$ for each n and $\|f - g_n\|_p$ descends to d as $n \to \infty$. We aim to show that $\{g_n\}$ is Cauchy and thus convergent.

Since $\|f - g_n\| \to d$, we know that for every $\varepsilon > 0$, there exists a natural number N such that for all $n \geq m > N$ we have

$$|\|f - g_n\|_p - d| < \varepsilon.$$

Now, we note that by Minkowski's inequality and Remark 4.10, we have

$$\left\|\frac{g_n + g_m}{2} - f\right\|_p = \left\|\frac{g_n}{2} - \frac{f}{2} + \frac{g_m}{2} - \frac{f}{2}\right\|_p$$

$$< \frac{1}{2}\|g_n - f\|_p + \frac{1}{2}\|g_m - f\|_p$$

as long as $g_n \neq g_m$. Since the value of $\|g_i - f\|_p$ is bounded from below by d and we know that $|\|g_m - f\|_p - d|, |\|g_n - f\|_p - d| < \varepsilon$, we must have $g_n \to g_m, N \to \infty$. This means that the sequence $\{g_n\}$ is Cauchy.

Next, since L^p viewed as a metric space is complete (Theorem 4.11), we have that g_n must converge by the Cauchy criterion. We call the limit g, and we then have $\|f - g\|_p = d$.

Moreover, since K is closed, it contains all its limit points (or limit functions, since L^p is a functional space). Since g is the limit of $\{g_n\}$, we have that $g \in K$.

We now show uniqueness of g by a similar argument to before. Assume that there exists another function g^* such that $\|f - g^*\|_p = d$. Then, we have by Minkowski's inequality that

$$\left\|f - \frac{g + g^*}{2}\right\|_p < \frac{1}{2}\|f - g\|_p + \frac{1}{2}\|f - g^*\|_p = \frac{d}{2} + \frac{d}{2} = d$$

as long as $f - g$ and $f - g^*$ are not directly proportional with the same sign. Then, we have generated a function–in K, by convexity–that achieves an L^p distance from f that is less than d, which is a contradiction.

If the functions are directly proportional, it means that $\exists c \geq 0$ such that

$$f - g = c(f - g^*).$$

If $c = 0$, it implies that $g = f$, which means f itself is the best approximating function (which can happen, if $f \in K$). Then $d = 0$, and only a function f^* that is equal to f a.e. (and therefore in the same equivalence class as f) can attain $\|f - f^*\|_p = d$. Therefore, we proceed assuming $c > 0$ and $d > 0$.

Observe that $\|f - g\|_p = \|f - g^*\|_p = d$. Hence, by Proposition 3.8 (2) we have

$$\|f - g\|_p = \sqrt[p]{\int |f - g|^p d\mu}$$

$$= \sqrt[p]{\int |c|^p |f - g^*|^p d\mu}$$

$$= |c| \sqrt[p]{\int |f - g^*|^p d\mu}$$

$$= |c| \cdot \|f - g^*\|_p$$

Hence $d = |c| \cdot d \implies |c| = 1$. Since Remark 4.10 dictates $c \geq 0$, c must be equal to 1, which implies $f - g = f - g^* \implies g = g^*$ a.e.. Uniqueness of g is thus proven.

$\square$

6. Conclusion

We have built out the notion of a *measure* using the notions of *outer measure* and *measurable sets*. The Carathéodory criterion provides a condition to pass from an outer measure to a measure. Following this recipe for constructing measures, we successfully constructed Lebesgue measure.

Next, we built the Lebesgue integral, and showed several useful properties, including approximation by simple functions and monotonicity. Important theorems include the monotone convergence theorem, Fatou's lemma, and the dominated convergence theorem. These theorems are crucial for various other proofs in the following sections.

We then built the L^p functional spaces, with $1 \leq p \leq \infty$, by first defining the L^p norm and then showing several of its properties. We proved that the L^p norm satisfies the norm properties by proving Hölder's inequality and Minkowski's inequality, the latter of which functions as the triangle inequality for L^p. Next, we showed L^p to be a complete metric space under the L^p norm, by an argument relying on Fatou's lemma and Minkowski's inequality.

Finally, we discussed convergence of approximation methods in L^p. We first showed that on a σ-finite measure space, the set of simple functions is dense in L^p. Using this result, we showed that under certain conditions, given a function $f \in L^p$, there is necessarily a unique function g serving as the "best approximation" of f under the L^p norm. These results may be further built upon when discussing the L^p error estimate, convergence of specific numerical methods such as linear interpolation, and Fourier series, the latter of which arise from orthogonal projection from L^p onto a finite-dimensional subspace.

Acknowledgments

I would like to thank my mentor, Sammy Thiagarajan, for the many insights he provided me on the more difficult aspects of measure theory. I would also like to thank fellow REU participants Ian Teo and Carter Fleming for working through many long hours of real analysis content together with me. Lastly, I would like to thank Daniil Rudenko for teaching the Mathematics REU's apprentice program, as well as Peter May for organizing the REU.

References

[1] Ricard F. Bass. *Real Analysis for Graduate Students.* Version 4.3.
[2] Jacob Stump. Introduction to the Lebesgue Integral. https://math.uchicago.edu/ may/REU2023/REUPapers/Stump.pdf.
[3] Elmer Tolsted. An Elementary Derivation of the Cauchy, Hölder, and Minkowski Inequalities from Young's Inequality. https://www.jstor.org/stable/2688239?seq=6.