MAXIMAL FUNCTIONS AND ALMOST-EVERYWHERE CONVERGENCE

HONORS THESIS IN MATHEMATICS
SAM THIEL

Abstract. In this paper, we will discuss the general framework of maximal functions as a method for proving almost-everywhere convergence of sequences of functions and series of functions. We will present the topic in a great amount of generality, which will allow us to prove independent results from many areas of analysis, including Real/Classical Analysis, Measure Theory, Functional Analysis, Ergodic Theory, Probability, and Harmonic Analysis.

Contents

1. Introduction

In this paper, we will discuss the general framework of maximal functions as a method for proving almost-everywhere convergence of sequences of functions. We will consider sequences of functions of the form $T_n f$ where T_n are (typically) linear operators mapping into a space of measurable functions, and f belongs to a Banach space. Our central question will be to ask if

$$\lim_{n\to\infty} T_n f =: T_\infty f$$

exists almost everywhere (a.e.) with respect to some measure, μ. Rather than work with the sequence directly, we will work with the sequence's associated *maximal function*,

$$T^* f := \sup_n |T_n f| \ .$$

It turns out, that if we know that T^* is measurable, and that $T_\infty f$ exists for a dense subset of our Banach space, we can state the following.

Theorem 1.1. *If, for all $C > 0$, T^* satisfies*

$$\lim_{\delta\to 0} \sup_{f;\|f\|\leq\delta} \mu(\{x \mid T^*f(x) > C\}) = 0 \tag{1.1}$$

then $T_\infty f$ exists μ-almost everywhere for each f.

We will refer to this as the C–δ hypothesis (See also Definition 2.3). This is a technical condition, and we will typically avoid proving this directly; rather, the C–δ hypothesis is satisfied if we can find a bound of the form

$$\mu(\{x \mid T^*f(x) > C\}) \leq Q\left(\frac{\|f\|}{C}\right) \tag{1.2}$$

where Q is such that $\lim_{x\to 0} Q(x) = 0$. Bounds of this form are called maximal inequalities, and proving maximal inequalities will be the core way that we satisfy the C–δ hypothesis in these proofs.

There are surprisingly many applications of Theorem 1.1, and we will discuss many of them throughout this paper. One of the most familiar applications is that of the Lebesgue differentiation theorem, which we discuss in Section 3 (Theorem 3.4). The Hardy–Littlewood maximal function

$$M_{\text{HL}}f(x) := \sup_{\epsilon>0} \frac{1}{2\epsilon}\int_{x-\epsilon}^{x+\epsilon} |f(y)|dy \tag{1.3}$$

and the associated Hardy–Littlewood maximal inequality (λ represents the Lebesgue Measure)

$$\lambda(M_{\text{HL}}f > C) \leq \frac{3}{C}\|f\|_1 \tag{1.4}$$

are familiar objects in measure theory, and we will see just how quickly the Lebesgue differentiation theorem follows from these two objects under our general maximal functions framework.

This framework is of central importance in many key results in analysis. For example, proofs of Carleson's Theorem, first proven in 1966 [3], proves the a.e. convergence of Fourier Series for L^p $(1 < p < \infty)$ functions using this maximal functions framework. The Strong Law of Large Numbers is another result we will prove under this framework (specifically, as a special case of the Birkhoff Ergodic Theorem).

To provide some additional motivation for the framework, consider the following. In an introductory analysis course, one of the first theorems students learn is that all convergent sequences are bounded. More importantly, they learn that the converse is false: not all bounded sequences are convergent. This is unfortunate, since boundedness is an easy and intuitive condition to check, and it being necessary but not sufficient is disappointing.

Some relief comes in the form of partial converses, which provide convergence results for strictly bounded sequences. The monotone sequence theorem is the classic example; that every bounded, monotone sequence is convergent. Similarly, the limit superior (and inferior) allows for limit-like results for bounded sequences without requiring convergence. Even series results, like the comparison and M-tests, have a heuristically similar character; proving a convergence result by saying something about the boundedness of the respective sequences.

In spirit, the maximal functions framework is another such converse. By making statements about the "bounds" of our sequence of functions, we can produce convergence results.

1.1. Structure of the Paper.

We will begin in Section 2 by formally defining maximal functions, before introducing the central theorem of this paper, Theorem 2.4 ("The Magic of Maximal Functions"). We finish Section 2 by discussing maximal inequalities and proving that they imply the C–δ hypothesis. The rest of the paper is dedicated to the applications of Theorem 2.4 to various areas in analysis.

In Section 3 we begin by discussing the Hardy–Littlewood maximal function, and associated maximal inequality, which provides a more familiar example of this framework in action to prove the Lebesgue differentiation theorem (Theorem 3.4). We then generalize the Lebesgue differentiation theorem to all σ-finite measures (Theorem 3.8), and then utilize this generalization to discuss Radon–Nikodym derivatives (Theorem 3.9). In Section 4 we continue to relate results to the Hardy–Littlewood maximal function, in particular we prove approximate identity results for convolutions on $\mathbb{R}$ (Theorem 4.6) and on the circle $\mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$

(Theorem 4.7). From this we prove the Lebesgue–Fejer theorem (Theorem 4.9), and the Uniqueness Property of Fourier series (Corollary 4.11).

In Section 5 we turn to proving the Strong Law of Large Numbers (Theorem 5.19). We approach this using our maximal functions approach to first prove the Birkhoff Ergodic Theorem (Theorem 5.14), and then combining the Ergodic Theorem with a special ergodic system, the Bernoulli Shift (Example 5.9), to obtain the Strong Law as a corollary.

In Section 6, we return to the topic of the C–δ hypothesis, and motivate the introduction of the distribution function (of a measurable function), as well as the space of weak-L^p functions. We show that proving that a maximal operator is bounded into weak-L^p constitutes a maximal inequality (Theorem 6.6), and thus the C–δ hypothesis, and reformulate some of our previous results in this new terminology. Then we use this alternative structure to generalize results from Sections 3 and 4 by utilizing a key interpolation theorem.

We conclude the paper in Section 7 by discussing the convergence of Fourier series. We discuss the lack of convergence for continuous functions and L^1 functions (Theorem 7.3), and then finish off the section by discussing Carleson's theorem (Theorem 7.5), which proves almost everwhere convergence of L^p functions for all other $p < \infty$.

2. Maximal Functions and the C–δ Hypothesis

In this section, we will present the core framework for the entire paper. We will begin with some definitions:

Definition 2.1 (Operators)**.** Let X be a Banach space and let (Ω, σ, μ) be a measure space. Let $\mathcal{M}(\Omega)$ denote the set of $\mathbb{R}$ or $\mathbb{C}$-valued measurable functions on Ω. Furthermore, let $\langle I, \prec \rangle$ be a directed system. For each $\alpha \in I$, define a map $T_\alpha : X \to \mathcal{M}(\Omega)$. We will call these maps *operators*. The T_α then forms a net of operators (from X to Ω).

By convention, the term "operator" will always describe a map of this form unless specifically designated otherwise. Note that as we have defined them, the term operator does not mean *linear* operator or *bounded* operator. While both of these special kinds of operators will be of importance to us, we have chosen to use the term operator in a more general sense to emphasize that Theorem 2.4 does not require either condition. The use of nets here as opposed to sequences allows us to index our operators using uncountable parameters, as opposed to only countable ones. For our purposes, this allows us to avoid distinguishing between limits of sequences, and continuum limits.

With our definition of operators, we can introduce the central object of our discussion, maximal functions:

Definition 2.2 (Maximal Functions)**.** Let T_α be a net of operators (Definition 2.1). For each $f \in X$ and $x \in \Omega$, we define

$$T^* f(x) := \sup_{\alpha \in I} |T_\alpha f(x)| \, . \tag{2.1}$$

If $T^* f$ is measurable, we call it a *maximal function* (on Ω). The operator T^* is called a *maximal operator* (from X to Ω).

The detail of note here is that we only define these objects if the maximal function is a measurable. If our indexing set is countable, we obtain measurability of $T^* f$ for free by the properties of the supremum. Otherwise, we have to do a bit more work. One way of showing that the maximal functions are measurable is to show that the map $\alpha \mapsto T_\alpha f$ is continuous for each f, and that the indexing set I has a countable dense subset. Our purposes only require doing this for an indexing set of the real numbers, for which we have the rational numbers as a dense set.

Maximal functions inherit some key properties from the net of operators that they are defined by. For example, if T_α are subadditive ($T_\alpha(f+g) \le T_\alpha f + T_\alpha g$), or positively homogeneous ($T_\alpha(\lambda f) = |\lambda| T_\alpha f$, $\lambda \in \mathbb{C}$), T^* will be as well. In the same vein, if T_α are linear, while T^* will not be linear, it will be subadditive and positively homogeneous.

Unfortunately, there are also useful properties that the net of operators T_α may have that are not inherited by T^*. For instance, if T_α is a net of bounded operators (into some Banach subspace of $\mathcal{M}(\Omega)$), T^* may not be a bounded map into the same subspace. We illustrate specific examples of this behavior in the discussion at the beginning of Section 6.

We have one final definition before we present our main theorem.

Definition 2.3 (The C–δ Hypothesis)**.** Let T^* be a maximal operator from X to Ω. Define for $C, \delta > 0$,

$$M(\delta, C) := \sup_{f; \|f\| \leq \delta} \mu(T^* f(x) > C) \ . \tag{2.2}$$

We say that T^* satisfies *the C–δ hypothesis* if, for all $C > 0$, we have that

$$\lim_{\delta \to 0^+} M(\delta, C) = 0 \ . \tag{2.3}$$

As we will see in Theorem 2.4, the C–δ hypothesis is the critical hypothesis for proving a.e. convergence. Our work in Sections 3, 4, 5 and 7 will focus on proving that specific maximal operators satisfy the C–δ hypothesis, and our work in Section 2.2 and Section 6 will focus on methods for showing that a maximal operator satisfies the C–δ hypothesis.

2.1. **The Magic of Maximal Functions.**

The following is the main result of this paper, which makes precise Theorem 1.1 and the surrounding discussion from the introduction.

Theorem 2.4 (The Magic of Maximal Functions)**.** *Let T_α be a net of subadditive operators from X to Ω (Definition 2.1) and T^* be the associated maximal operator (Definition 2.2). Suppose that*

- *(H1) There exists a dense subset $X_0 \subseteq X$ where for each $f \in X_0$, the net $T_\alpha f$ has a measurable limit for a.e. $x \in \Omega$.*
- *(H2) The operator T^* satisfies the C–δ Hypothesis (Definition 2.3).*

Then, we have:

- *(i) For every $f \in X$, $T_\alpha f(x)$ has a measurable limit, $T_\infty f(x)$, for a.e. $x \in \Omega$. We have that T_∞ is an operator from X to Ω.*
- *(ii) Moreover, if $\{f_n\}_{n\in\mathbb{N}}$ satisfies $\|f_n - f\|_X \to 0$, then $T_\infty f_n(x) \to T_\infty f(x)$ in measure, i.e., for every $\epsilon > 0$, we have that*

$$\lim_{n\to\infty} \mu(|T_\infty f_n(x) - T_\infty f(x)| > \epsilon) = 0 \ . \tag{2.4}$$

Remark 2.5. Theorem 2.4 also applies to complex-valued operators; if $T_\alpha f$ are $\mathbb{C}$-valued, then we require additivity instead of subadditivity with no change in proof.

Proof. We will dedicate this proof to showing that the limit exists. The limit of measurable functions is measurable, and so we obtain the limit operator T_∞ as a consequence of that.

Fix an arbitrary $f \in X$. For notational convenience, define

$$T_+ f := \limsup_\alpha (T_\alpha f) \ ,$$
$$T_- f := \liminf_\alpha (T_\alpha f) \ ,$$
$$\delta T := T_+ - T_- \ .$$

If we prove that $\delta T = 0$ for a.e. $x \in \Omega$, then the limit exists. Note a-priori that

$$|T_+| \leq T^*, \ |T_-| \leq T^*,$$

and so

$$\delta T \leq |\delta T| \leq |T_+| + |T_-| \leq 2T^* \ . \tag{2.5}$$

By the subadditivity of T_α, and the ordering properties of the supremum, we have that for each $f, g \in X$ that

$$T_+(f) = T_+(f - g + g) \leq T_+(f - g) + T_+(g) \ ,$$

and similarly,

$$T_-(f) \geq T_-(f - g) + T_-(g) \ .$$

This gives us that

$$\delta T(f) \leq T_+(f - g) + T_+(g) - T_-(f - g) - T_-(g) = \delta T(f - g) + \delta T(g) \ .$$

So, if g is such that $\delta T(g) = 0$ (which is equivalent to our desired limit existing for g), we have that

$$\delta T(f) \leq \delta T(f - g) \ . \tag{2.6}$$

With this in mind, by density pick $f_n \in X_0$ such that $f_n \to f$ in X. Note that by the hypothesis (H1), we have (2.6) for $g = f_n$ since we know that our net has an a.e.-limit for each element of X_0. Combining this with (2.5), we find that for each $\epsilon > 0$

$$\begin{aligned}\mu(\delta T(f) > \epsilon) &\leq_{(2.6)} \mu(\delta T(f - f_n) > \epsilon)\\ &\leq_{(2.5)} \mu(T^*(f - f_n) > \epsilon/2)\\ &\leq \sup_{\|g\| \leq \|f - f_n\|} \mu(T^* g > \epsilon/2) = M(\|f - f_n\|, \epsilon/2)\ .\end{aligned}$$

Now, hypothesis (H2) ensures that T^* satisfies the C–δ hypothesis, so we know that $M(\delta, C) \to 0$ as $\delta \to 0$. Since by construction we have that $\|f - f_n\| \to 0$, we know that

$$\lim_{n\to\infty} M(\|f - f_n\|, \epsilon/2) = 0\ ,$$

so that

$$\mu(\delta T(f)(x) > \epsilon) = 0\ . \tag{2.7}$$

for every $\epsilon > 0$.

Finally, since $\delta T \geq 0$, observe

$$\bigcup_{n\in\mathbb{N}} \{x \mid \delta T(f)(x) > 1/n\} = \{x \mid \delta T(f)(x) \neq 0\}\ ,$$

so that by (2.7),

$$\mu(x \mid \delta T(f) \neq 0) = \mu\left(\bigcup_{n\in\mathbb{N}} \{x \mid \delta T(f)(x) > 1/n\}\right) \leq \sum_{n=1}^{\infty} \mu(x \mid \delta T(f)(x) > 1/n) = 0\ .$$

We conclude that $\delta T = 0$ a.e., thus $\lim_\alpha T_\alpha f$ exists a.e. Since $f \in X$ was arbitrary, we may define our limit operator $T_\infty : X \to \mathcal{M}(\Omega)$ by $T_\infty f := \lim_\alpha T_\alpha f$. This proves result (i).

To prove the "moreover" statement, result (ii), observe that

$$T_\infty \leq |T_\infty| \leq T^*\ . \tag{2.8}$$

With this in mind, note that by the order preservation of limits, T_∞ is subadditive. Thus, by subadditivity, we have that

$$T_\infty f \leq T_\infty(f - g) + T_\infty g$$

and

$$T_\infty g \leq T_\infty(g - f) + T_\infty f\ ,$$

which implies that

$$\begin{aligned}|T_\infty f - T_\infty g| &\leq \max(|T_\infty(f - g)|, |T_\infty(g - f)|)\\ &\leq |T_\infty(f - g)| + |T_\infty(g - f)|\\ &\leq_{(2.8)} T^*(f - g) + T^*(g - f)\ .\end{aligned}$$

With this in mind, let $f \in X$ and $f_n \to f$ in X. We then compute that

$$\begin{aligned}\mu(|T_\infty f_n - T_\infty f| > \epsilon) &\leq \mu(T^*(f_n - f) + T^*(f - f_n) > \epsilon)\\ &\leq \mu(T^*(f_n - f) > \epsilon/2) + \mu(T^*(f - f_n) > \epsilon/2) \leq 2M(\|f_n - f\|, \epsilon/2)\ ,\end{aligned}$$

which, by similar argument to (2.7), gives us that

$$\lim_{n\to\infty} \mu(|T_\infty f_n - T_\infty f| > \epsilon) = 0\ ,$$

which proves result (ii). □

On its own, part (i) of Theorem 2.4 only tells us the existence of almost-everywhere limits for every $f \in X$. However, part (ii) allows us to specify precisely what these limits look like, provided that we know the nature of our limits for every element of our dense set, X_0. One might be tempted to appeal to the uniqueness of limits in order to do this, but part (ii) of Theorem 2.4 only yields $T_\infty f_n \to T_\infty f$ in measure, and not almost everywhere. So, we have to be a bit more careful, and develop the uniqueness of limits in measure. This motivates the following:

Lemma 2.6. *Let f_n be a sequence of measurable functions defined on a measure space (Ω, σ, μ). Suppose f, g are functions such that $f_n \to f$ <u>and</u> $f_n \to g$ in measure. Then, $f = g$ μ-almost everywhere.*

Proof. Observe by the continuity of measures that

$$\mu(|f-g|>0) = \mu\left(\bigcup_{k\in\mathbb{N}} \{|f-g|>1/k\}\right) \leq \sum_{k=1}^{\infty} \mu(|f-g|>1/k) \ ,$$

and so it suffices to show that $\mu(|f-g|>1/k)=0$ for each k. By hypothesis, we know that for each $\epsilon > 0$,

$$\begin{aligned} &\mu(|f-f_n|>\epsilon) \to 0 \ , \\ &\mu(|f_n-g|>\epsilon) \to 0 \ , \end{aligned}$$

in n. Thus we may estimate that

$$\begin{aligned} \mu(|f-g|>1/k) &\leq \mu(|f-f_n|+|f_n-g|>1/k) \\ &\leq \mu(|f-f_n|>1/2k) + \mu(|f_n-g|>1/2k) \to 0 \end{aligned}$$

which shows that $\mu(|f-g|>1/k)=0$ for each k, and thus that $f=g$ μ-a.e. ☐

With Lemma 2.6, we can state the result which will allow us to extend whatever behaviors we have on our dense set X_0 to the whole Banach space.

Theorem 2.7 (Pointwise Extension by Density)**.** *Assume the setup and conclusion of Theorem 2.4. Let $f \in X$ and suppose that $X_0 \ni f_n \to f$ in X. If the sequence $T_\infty f_n$ converges to a measurable function g in measure. Then $g = T_\infty f$ almost everywhere.*

Proof. The "moreover" statement of Theorem 2.4 gives us that $T_\infty f_n \to T_\infty f$ in measure. By hypothesis, $T_\infty f_n \to g$, and so by Lemma 2.6, we have that $g = T_\infty f$. ☐

We will see Theorem 2.7 used most commonly in the case that our limit operator is the identity map on L^p. That is, $T_\infty f = f$ for $f \in L^p$. Since convergence in L^p implies convergence in measure, $f_n \to f$ in L^p and $T_\infty f_n = f_n$ is enough to say that $T_\infty f = f$. We will elaborate on this when we discuss specific examples of this in Sections 3 and 4 (e.g. the Lebesgue differentiation theorem, Theorem 3.4).

2.2. **Maximal Inequalities.**

With our main theorem (Theorem 2.4) stated, we connect it to maximal inequalities, which will be our main way of checking that a maximal operator satisfies the C–δ hypothesis throughout this paper.

Definition 2.8 (Maximal Inequalities)**.** Let T^* be a maximal operator from X to Ω. A *maximal inequality* is a bound of the form

$$\mu(T^*f > C) \leq Q\left(\frac{\|f\|_X}{C}\right) \tag{2.9}$$

where $\lim_{x\to 0^+} Q(x) = 0$.

Most commonly, $Q(x)$ will take the form $Q(x) = Kx^\alpha$ for some fixed K, α.

Given the definition of $M(\delta, C)$ in Definition 2.3, its clear to see why such a bound would be helpful. A bound like this on its own guarantees that T^*f is finite a.e., and the right hand side of (2.9) shrinks either with the growth of C or the reduction of $\|f\|_X$.

Heuristically, a.e. finiteness is the condition that the C–δ hypothesis is asking of our maximal operators. In fact, for certain classes of operators, proving that T^* is a.e. finite is enough to satisfy the C–δ hypothesis on its own, for example:

Theorem 2.9. *Let T_n be a sequence of bounded linear operators from X to $L^p(\Omega)$, where Ω is a finite measure space. Furthermore, assume T^*f is finite a.e. for each $f \in X$. Then T^* satisfies the C–δ hypothesis.*

Theorem 2.9 is the result given in Problem 1 of [9, Chapter 2.1]. We omit the proof outlined in [9], which is based on a Baire category theorem argument; when Ω is a finite measure space, Theorem 2.9 is a more general case of Theorem 2.11.

Unfortunately, proving a.e. finiteness is often as difficult proving the C–δ hypothesis itself. Thus, we return to our maximal inequalities, for which we have the following key result.

Theorem 2.10 (The Magic of Maximal Inequalities)**.** *Let T_α be a net of operators from X to Ω and let T^* be the associated maximal operator. Then, if T^* satisfies a maximal inequality, it satisfies the C–δ hypothesis.*

Proof. In view of proving the C–δ hypothesis, fix $C > 0$. If T^* satisfies a maximal inequality, then we have that

$$M(\delta, C) = \sup_{\|f\| \le \delta} \mu(T^* f > C) \le \sup_{\|f\| \le \delta} Q\left(\frac{\|f\|}{C}\right) \ .$$

Taking limits, we obtain

$$\lim_{\delta \to 0^+} M(\delta, C) \le \lim_{\delta \to 0^+} \sup_{\|f\| \le \delta} Q\left(\frac{\|f\|}{C}\right)$$

which, since $Q(x) \to 0$ as $x \to 0^+$, implies

$$\lim_{\delta \to 0^+} M(\delta, C) = 0 \ .$$

. ☐

Maximal inequalities serve as a much more natural objects to work with than the C–δ hypothesis. Much of our work in the following sections will be to prove maximal inequalities for maximal operators in a variety of contexts. We mention as a final note in this section that certain properties of maximal operators imply maximal inequalities in addition to the C–δ hypothesis. Recall that Chebyshev–Markov states that

$$\mu(f > C) \le \frac{\|f\|_p^p}{C^p} = \left(\frac{\|f\|_p}{C}\right)^p$$

if $f \in L^p$. This immediately gives us that

Theorem 2.11. *Let T^* be a maximal operator from X to Ω. Furthermore, assume T^* is a bounded operator from X to $L^p(\Omega)$. Then T^* satisfies a maximal inequality, and therefore, the C–δ hypothesis.*

As we will see in Section 6, this condition is a bit too strong to be of use, but will provide a useful motivation for the developments in Section 6.

3. Differentiation

In this section we will introduce the Hardy–Littlewood maximal function, which we will use in various facets over the next two sections. We illustrate its applications to a.e. convergence by presenting the proof of the Lebesgue differentiation theorem, both in its original and more general form, as well as discussing Radon–Nikodym derivatives.

In the proof of the Lebesgue differentiation theorem, we invite the reader to consider each of the facets of this proof that lead to applying Theorem 2.4. Each of the steps in this proof—linearity of operators, continuity in ϵ, convergence on a dense set, and the maximal inequality—take only a few lines to state and conclude. This is how many of our proofs appear under our maximal functions framework: the hard work of the proof is done in the background.

This is a common theme. The hypotheses of Theorem 2.4 are often known results with their own range of applications, since linear operators, dense sets, and convergence results are core topics of interest in analysis. The Hardy–Littlewood maximal inequality is the only result required in the proof that is special to our framework, which is understandable, since the C–δ hypothesis is a specialized condition to fulfill.

3.1. **The Hardy–Littlewood Maximal Function and Lebesgue Differentiation Theorem.**

For $f \in L^1(\mathbb{R})$, we define the *Hardy–Littlewood maximal function* of f by

$$M_{\text{HL}} f(x) := \sup_{\epsilon > 0} \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} |f(y)| dy \ . \tag{3.1}$$

The maximal function M_{HL} is a familiar object in measure theory. Recall the statement of the associated maximal inequality as well:

Theorem 3.1 (Hardy-Littlewood Maximal Inequality)**.** *Let λ denote the Lebesgue measure. Then, for each $C > 0$, $f \in L^1$,*

$$\lambda(M_{\text{HL}} f(x) > C) \leq \frac{3 \|f\|_1}{C} \ . \tag{3.2}$$

For space reasons, we omit the full proof, which can be found in [9, Chapter 2.3]. However, we mention some key details of the proof. The proof relies on the Vitali covering lemma, stated below.

Lemma 3.2 (Vitali Covering Lemma)**.** *Let $\{B_{r_n}(x_n)\}_{n=1}^k$ be a finite set of of balls in $\mathbb{R}$. Then, there exists a subcollection $V := \left\{B_{r_j}(x_j)\right\}$ such that*

1. *The elements of V are pairwise disjoint, that is $B_{r_j}(x_j) \cap B_{r_\ell}(x_\ell) = \emptyset$ for $j \neq \ell$.*
2. *Scaling this subcollection by a factor of 3 forms an open cover for the original collection. That is,*

$$\bigcup_{j=1}^{k} B_{r_j}(x_j) \subseteq \bigcup_{B_{r_j} \in V} B_{3r_j}(x_j) \ . \tag{3.3}$$

Remark 3.3. The Vitali lemma is useful in the proof of the Hardy–Littlewood maximal inequality due to two key features of the Lebesgue measure. The first is positive homogenity, which allows us to extract the factor of 3 that is created by (3.3). This is the origin of the absolute constant in Theorem 3.1. The proof also relies on the inner regularity of the Lebesgue measure, since the Vitali lemma requires a finite set of open balls by hypothesis; a problem that can be avoided by looking at compact subsets.

We can now state and prove the Lebesgue differentiation theorem.

Theorem 3.4 (Lebesgue Differentiation Theorem)**.** *Let $f \in L^1(\mathbb{R})$. Then, for a.e. $x \in \mathbb{R}$,*

$$\lim_{\epsilon \to 0^+} \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} f(y) dy = f(x) \ . \tag{3.4}$$

Proof. Define the operators T_ϵ by

$$T_\epsilon f := \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} f(y) dy \ . \tag{3.5}$$

for all $\epsilon > 0$. First, note that by the linearity of integrals, these operators are linear.

Next, our operators have a continuum parameter, so for our maximal function to be measurable, we need to show that these operators are continuous in their parameter (See the discussion after Definition 2.2). Fortunately, $1/2\epsilon$ is continuous while $\epsilon > 0$, and integrals of this form are continuous in their bounds. Therefore $T_\epsilon f$ is continuous in ϵ, and our maximal function is measurable.

The maximal function for this collection is

$$T^* f = \sup_{\epsilon > 0} \left| \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} f(y) dy \right| \leq \sup_{\epsilon > 0} \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} |f(y)| dy = M_{\text{HL}} f \ .$$

Because this maximal operator is dominated by M_{HL}, it inherits the Hardy-Littlewood maximal inequality, since

$$\lambda(T^* f > C) \leq \lambda(M_{\text{HL}} f > C) \leq \frac{3}{C} \|f\|_1 \ .$$

Finally, we need to check that the limit exists for a dense subspace of our Banach space, which in this case is L^1. The continuous, compactly supported functions ($\mathcal{C}_c$) are dense in L^1, and for such functions, our limit exists by the fundamental theorem of calculus. Moreover, the limit has the form we want, recovering the given function in the limit.

Since we have linear operators, convergence on a dense set, and the C–δ hypothesis holds, we can apply Theorem 2.4 and the limit exists Lebesgue a.e.

To identify the limit for a general L^1 function, consider a sequence of functions f_n, continuous and compactly supported, that converge to f in the L^1-sense. Because $f_n \to f$ in L^1, we also have $f_n \to f$ in measure. Finally, since $T_\infty f_n = f_n$, we can apply Theorem 2.7, and therefore $T_\infty f = f$ almost everywhere. ☐

3.2. σ-finite Measures and Radon–Nikodym Derivatives.

Now that we've seen the standard proof, we will spend this section generalizing the proof of the Lebesgue Differentiation Theorem to σ-finite measures, and apply that result to the subject of Radon–Nikodym derivatives; given the Lebesgue–Radon–Nikodym decomposition of a measure ν given by $d\nu = g d\mu + d\nu_{\text{sing}}$, we give an explicit characterization of the Radon–Nikodym derivative g as well as the support of ν_{sing}.

We start by generalizing the Hardy–Littlewood maximal function.

Definition 3.5 (Generalized Hardy–Littlewood Maximal Function)**.** Let μ be a σ-finite Borel measure on $\mathbb{R}$. Then, for each $f \in L^1(\mathbb{R}, d\mu)$ define the maximal operator

$$M_{\text{HL}}^{\mu} f := \sup_{\epsilon > 0} \frac{1}{\mu(B_\epsilon(x))} \int_{B_\epsilon(x)} |f(y)| d\mu(y) \ . \tag{3.6}$$

Our eventual proof of the generalized Lebesgue differentiation theorem will follow the standard proof very closely. We will first prove the following generalization of the Hardy–Littlewood maximal inequality for our generalized maximal operator.

Theorem 3.6 (Generalized Hardy-Littlewood Maximal Inequality)**.** *Let μ be a σ-finite Borel measure on $\mathbb{R}$. Then, for each $C > 0, f \in L^1(\mathbb{R}, d\mu)$,*

$$\mu(M_{HL}^{\mu} f > C) \leq \frac{2 \|f\|_1}{C} \ . \tag{3.7}$$

Observe that if we take μ to be the Lebesgue measure, we've actually improved the constant term in our original inequality from 3 to 2. This results from the need to use a different covering lemma to avoid the requirement of positive homogeneity (see Remark 3.3).

Lemma 3.7 (Croft-Garsia Covering Lemma)**.** *Let $\{I_j \mid 1 \leq j \leq k\}$ be a finite set of intervals in $\mathbb{R}$. Then there exist two subsets of these intervals, S and T, with the following properties:*

1. *Elements of S are pairwise disjoint. That is, if $j \neq \ell$,*

   $$I_j, I_\ell \in S \implies I_j \cap I_\ell = \emptyset. \tag{3.8}$$

   The same holds true for T.
2. *$S \cup T$ is a cover for the original set of intervals. That is,*

   $$\bigcup_{j=1}^{k} I_j = \bigcup_{I_j \in S \cup T} I_j \tag{3.9}$$

Proof. First, we are going to construct a subset of intervals J with two properties. The first is that

$$\bigcup_{j=1}^{k} I_j = \bigcup_{I_j \in J} I_j \tag{3.10}$$

and the second is that each subset in this collection is "essential" to the equivalence in (3.10) in the sense that for each $I_\ell \in J$,

$$I_\ell \not\subset \bigcup_{I_\ell \neq I_j \in J} I_j \ . \tag{3.11}$$

That is, removing I_ℓ from J leaves some point uncovered.

To construct a set J, first check if the full set of intervals $\{I_j\}_1^k$ satisfies (3.11). If it does, then we take J to be our full set of intervals. If the full set of intervals does not satisfy (3.11), there must be some ℓ' such that

$$I_{\ell'} \subset \bigcup_{I_{\ell'} \neq I_j \in J} I_j \ .$$

If this is the case, remove $I_{\ell'}$ from our collection, and repeat the process. Do this until we find a set of intervals which obeys (3.11), or only one interval remains.

Now, take the intervals in J and reindex them such that for each $I_{j_n} \in J$, $\inf I_{j_n} \leq \inf I_{j_{n+1}}$. Letting $I_{j_n} = (a_n, b_n)$, our ordering of the indexes implies that

$$a_1 \leq a_2 \leq \cdots \leq a_n \ .$$

If for some ℓ, $b_\ell \geq b_{\ell+1}$, we violate (3.11) since that would imply $I_{j_{\ell+1}} \subset I_{j_\ell}$. Therefore

$$b_1 \leq b_2 \leq \cdots \leq b_n \tag{3.12}$$

as well. Now, consider $I_{j_n} \cap I_{j_{n+2}}$ for each n. If $I_{j_n} \cap I_{j_{n+2}} \neq \emptyset$, since $a_n \leq a_{n+1}$ and $b_{n+1} \leq b_{n+2}$, we would have that $I_{j_{n+1}} \subseteq I_{j_n} \cup I_{j_{n+2}}$, which violates (3.11). Therefore $I_{j_n} \cap I_{j_{n+2}} = \emptyset$.

Since even indices are pairwise disjoint and odd indices are pairwise disjoint, we can take $S = \{I_{j_1}, I_{j_3}, \dots\}$ and $T = \{I_{j_2}, I_{j_4}, \dots\}$ and get our result. ☐

With the Croft–Garsia lemma, the proof of Theorem 3.6 follows the proof of the original Hardy–Littlewood maximal inequality. The only additional detail is to be sure that our measure is inner-regular, but since we have a σ-finite Borel measure, the measure is inner-regular by assumption.

We can now state and prove the generalized Lebesgue differentiation theorem. Note just how similar the proof is to the original.

Theorem 3.8 (Generalized Lebesgue Differentiation Theorem). *Let μ be a σ-finite Borel measure on $\mathbb{R}$, and let $g \in L^1(\mathbb{R}, d\mu)$. Then we have*

$$\lim_{\epsilon \to 0^+} \frac{1}{\mu(B_\epsilon(x))} \int_{B_\epsilon(x)} g(y) d\mu(y) = g(x) \tag{3.13}$$

for μ.a.e. $x \in \mathbb{R}$.

Proof. Let

$$T_\epsilon g(x) = \frac{1}{\mu(B_\epsilon(x))} \int_{B_\epsilon(x)} g(y) d\mu(y) \ . \tag{3.14}$$

The linearity of integrals implies that these operators are linear. The continuity of measures implies that the operators are continuous in their parameter, so T^* is measurable. The maximal operator T^* for this collection is not quite M^μ_{HL} but is dominated by it. Therefore the generalized Hardy–Littlewood maximal inequality (Theorem 3.6) serves as our maximal inequality. Furthermore we have convergence for every uniformly continuous function by

$$|T_\epsilon g(x) - g(x)| \leq \sup_{y \in B_\epsilon(x)} |g(y) - g(x)| \to 0 \ .$$

Since the uniformly continuous functions are dense in L^1, we have convergence on a dense set. Since we have linear operators, convergence on a dense set, and a maximal inequality, Theorem 2.4 applies and the limit exists for every $g \in L^1$. Since convergence in L^1 implies convergence in measure, and $T_\infty g_n = g_n$ for each uniformly continuous g_n, we can apply Theorem 2.7 to complete the proof. ☐

We conclude this section by applying the generalized Lebesgue differentiation theorem to obtain an explicit description of Radon–Nikodym derivatives. The following result is especially useful in the context of probability; while the Lebesgue–Radon–Nikodym decomposition ensures that any probability measure can be decomposed into a discrete measure and a measure with density, the following result gives explicitly what that density is, as well as the support of the discrete measure.

Theorem 3.9 (Radon–Nikodym Derivatives). *Let μ and ν be two σ-finite Borel measures on $\mathbb{R}$, and let*

$$d\nu = g d\mu + d\nu_{\mathrm{sing}} \tag{3.15}$$

be the Lebesgue-Radon-Nikodym decomposition (so $\nu \ll \mu$, $d\nu_{\mathrm{sing}}$ is $d\mu$-singular, and $g \in L^1(d\mu)$). Then, for μ a.e. $x \in \mathbb{R}$,

$$\lim_{\epsilon \to 0^+} \frac{\nu(B_\epsilon(x))}{\mu(B_\epsilon(x))} = g(x) \ , \tag{3.16}$$

in which case, we will often denote

$$g =: \frac{d\nu}{d\mu} \ . \tag{3.17}$$

Moreover, for

$$E := \left\{ x \;\middle|\; \lim_{\epsilon \to 0^+} \frac{\nu(B_\epsilon(x))}{\mu(B_\epsilon(x))} = \infty \right\} \tag{3.18}$$

we have that

$$\mu(E) = 0 \ , \ \nu_{\text{sing}}(\mathbb{R} \setminus E) = 0 \ . \tag{3.19}$$

The proof relies on constructions from the von Neumann proof of the Radon-Nikodym decomposition theorem (see, for instance, [10, Chapter 4.7]). Here σ-finitness arises as a condition in two ways: once from our proof of Theorem 3.8, and once from the proof of the Radon-Nikodym decomposition.

Proof. Let $\eta = \mu + \nu$. Then $\mu \ll \nu$. By the von Neumann proof, we know the following:

There exists $F \in L^2(d\eta)$ with $d\mu = F d\eta$ and $0 \leq F \leq 1$. Moreover, $g = (F)^{-1} - 1$, μ a.e. Furthermore, letting

$$E_0 = \{x \mid F(x) = 0\}$$

we know that $\mu(E_0) = 0$, and $d\nu_{\text{sing}} = \chi_{E_0} d\nu$.

With this background, the generalized Lebsegue differentiation theorem (Theorem 3.8) implies that

$$\lim_{\epsilon \to 0^+} \frac{\mu(B_\epsilon(x))}{\eta(B_\epsilon(x))} = \lim_{\epsilon \to 0^+} \frac{1}{\eta(B_\epsilon(x))} \int_{B_\epsilon(x)} F(y) d\eta(y) = F(x).$$

η (and therefore, μ), a.e.

By the definition of η, we have that

$$\left(\frac{\mu}{\eta}\right)^{-1} = 1 + \left(\frac{\nu}{\mu}\right)$$

which gives us that

$$\lim_{\epsilon \to 0^+} \frac{\nu(B_\epsilon(x))}{\mu(B_\epsilon(x))} = \left(\lim_{\epsilon \to 0^+} \frac{\mu(B_\epsilon(x))}{\eta(B_\epsilon(x))}\right)^{-1} - 1 = (F(x))^{-1} - 1$$

Now the only issue is that on E_0, (up to sets of measure zero) $F(x) = 0$ and so this limit is ∞. Thus letting $E = E_0$ gives us that $\mu(E) = 0$ and $\nu_{\text{sing}}(\mathbb{R} \setminus E) = 0$, which gives us our "moreover" statement. Finally since we may freely modify functions on sets of zero measure, we can modify $F(x)$ on E to get that

$$\lim_{\epsilon \to 0^+} \frac{\nu(B_\epsilon(x))}{\mu(B_\epsilon(x))} = (F(x))^{-1} - 1 = g(x)$$

μ a.e., which completes the proof. ☐

4. Convolution

In this section we will discuss approximate identity results for convolutions by relating them to the Hardy–Littlewood maximal function. We will talk about convolutions both on $\mathbb{R}$, and on the circle, $\mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$. We'll begin with convolutions on $\mathbb{R}$. In particular, our goal will be to show that if a family of functions $\{g_n\}_{n\in\mathbb{N}}$ satisfies certain properties, then we have that $\lim_{n\to\infty}(g_n * f)(x) = f(x)$ almost everywhere for each $f \in L^1$.

If we require instead that the convergence is in L^1, we have the following (see [6, Chapter 1.2]).

Theorem 4.1 (L^1-Approximate-Identities)**.** *Let $f \in L^1(\mathbb{R})$. Suppose $\{g_n\}_{n\in\mathbb{N}}$ is a family of measurable functions satisfying*

(1) $\sup_n \|g_n\|_1 < \infty$,
(2) $\int g_n = 1$,
(3) $\forall r > 0$, $\lim_{n\to\infty} \int_{|y|>r} |g_n| dy = 0$.

Then, we have that $\|(g_n * f) - f\|_1 \to 0$ in n.

Conditions (1)–(3) are similar to the conditions we will need for a.e. convergence. This is because the proof of the above theorem already works by appealing to a dense set; conditions (2) and (3) imply that if f is continuous and compactly supported, then $g_n * f \to f$ *uniformly.* Since our machinery also requires a dense set, we will preserve these two conditions. Condition (1) will be the condition we modify in order to prove a.e. convergence (Theorem 4.6). In particular, we will modify it in such a way that it allows us to use a maximal inequality.

The machinery underlying this maximal inequality will depend on *symmetrically-decreasing* functions, that is, a function $f(x)$ which is decreasing in $|x|$ as opposed to just x. One particularly important property for proving a maximal inequality is that if f is a symmetrically decreasing function, then the preimages of superlevel sets of f are simply balls centered at the origin,

$$\{f > \alpha\} = B_r(0) \tag{4.1}$$

for some $r > 0$ depending on α and f.

When we are unable to work with a symmetric-decreasing function directly, we can instead dominate a function by one. This motivates defining the *symmetric envelope* of a function.

Definition 4.2. Let f be a measurable function. The *symmetric envelope* of f, denoted $\widetilde{f}$, is defined by

$$\widetilde{f}(x) = \sup_{|y| \geq |x|} |f(y)|. \tag{4.2}$$

From the definition, it's easy to see that $\widetilde{f} \geq f$ and that $\widetilde{f}$ is symmetric decreasing.

A common theme when working with convolutions is to try to split the convolution $f * g$ into two terms, one which depends on f, and one which depends on g. This strategy is how we will obtain our maximal inequality, as a consequence of the Hardy–Littlewood maximal inequality (Theorem 3.1) and the following.

Theorem 4.3 (Maximal Convolution Inequality)**.** *Let f, g be measurable functions on $\mathbb{R}$. Then,*

$$|g * f(x)| \leq \|\widetilde{g}\|_1 \, M_{\mathrm{HL}} f(x) \ . \tag{4.3}$$

In particular, for a family of functions g_n,

$$\sup_n |g_n * f(x)| \leq \sup_n \|\widetilde{g_n}\|_1 \, M_{\mathrm{HL}} f(x) \ . \tag{4.4}$$

The proof of this theorem relies on the following general fact. Given a measure space (Ω, σ, μ), we can represent the L^1-norm of a measurable function f by

$$\|f\|_1 = \int |f| d\mu = \int_0^\infty \mu(|f| > t) dt. \tag{4.5}$$

Remark 4.4. We will state this fact in more generality later in Theorem 6.2. The right hand side of this equality is integrating over the *distribution function* of f, which will be an object of importance to us in Section 6.

This fact produces the following lemma by taking $\Omega = \mathbb{R}$, and taking μ to be either the Lebesgue measure, or a point mass probability measure at x, respectively.

Lemma 4.5. *Let $f \in L^1(\mathbb{R})$. Then, we have*

$$\|f\|_1 = \int_0^\infty \lambda(|f| > t) dt \ , \tag{4.6}$$

and,

$$|f(x)| = \int_0^\infty \chi_{\{|f(x)| > \alpha\}}(x) d\alpha \ . \tag{4.7}$$

Equation (4.7) is often called the *wedding-cake* or *layer-cake* representation.

Proof of Theorem 4.3. By the triangle inequality and the definition of the symmetric envelope, we have

$$|g * f| \leq \int_{\mathbb{R}} |g(y - x)||f(y)| dy \leq \int_{\mathbb{R}} |\widetilde{g}(y - x)||f(y)| dy,$$

so that the wedding-cake representation for $\widetilde{g}$ yields,

$$\int_{\mathbb{R}} |\widetilde{g}(y-x)||f(y)|dy = \int_{\mathbb{R}} \left[\int_0^\infty \chi_{\{|\widetilde{g}(y-x)|>\alpha\}}(y-x)d\alpha\right]|f(y)|dy.$$

Now, since $\widetilde{g}$ is symmetric decreasing, (4.1) tells us that $\{\widetilde{g} > \alpha\} = B_{r(\alpha)}(0)$. Given the shifting in our integrand by x, we find that

$$\begin{aligned}\int_{\mathbb{R}} \left[\int_0^\infty \chi_{\{|\widetilde{g}(y-x)|>\alpha\}}(y-x)d\alpha\right]|f(y)|dy &= \int_{\mathbb{R}} \left[\int_0^\infty \chi_{B_{r(\alpha)}(x)}(y)d\alpha\right]|f(y)|dy \\ &= \int_0^\infty \int_{\mathbb{R}} \chi_{B_{r(\alpha)}(x)}(y)|f(y)|dyd\alpha \\ &= \int_0^\infty \int_{B_{r(\alpha)}(x)} |f(y)|dyd\alpha\end{aligned}$$

where the conversion to the double integral and subsequent change of order is justified by Tonneli's theorem. Continuing, we find

$$\begin{aligned}\int_0^\infty \int_{B_{r(\alpha)}(x)} |f(y)|dyd\alpha &= \int_0^\infty \lambda(B_{r(\alpha)}(x)) \frac{1}{\lambda(B_{r(\alpha)}(x))} \int_{B_{r(\alpha)}(x)} |f(y)|dyd\alpha \\ &\leq M_{\mathrm{HL}}f(x) \int_0^\infty \lambda(B_{r(\alpha)}(x))d\alpha \ .\end{aligned}$$

Finally, the translation invariance of the Lebesgue measure combined with our construction of this ball gives, by Lemma 4.5,

$$\begin{aligned}M_{\mathrm{HL}}f(x) \int_0^\infty \lambda(B_{r(\alpha)}(x))d\alpha &= M_{\mathrm{HL}}f \int_0^\infty \lambda(\widetilde{g} > \alpha)d\alpha \\ &= M_{\mathrm{HL}}f(x) \left\|\widetilde{g}\right\|_1 \ ,\end{aligned}$$

which completes the proof. □

With Theorem 4.3, we can state and prove our almost everywhere approximate identity result for convolutions.

Theorem 4.6 ($\mathbb{R}$-Approximate-Identities). *Let $\{g_n\}$ be a family of functions that satisfies the following:*

$$\text{(1) } \sup_{n\in\mathbb{N}} \|\widetilde{g}_n\|_1 < \infty \ , \tag{4.8}$$

$$\text{(2) } \int g_n(x)dx = 1 \ , \tag{4.9}$$

$$\text{(3) For all } r > 0, \lim_{n\to\infty} \int_{|y|>r} |g_n(y)|dy = 0 \ . \tag{4.10}$$

Then, for each $f \in L^1(\mathbb{R})$, we have

$$\lim_{n\to\infty} (g_n * f)(x) = f(x)$$

for Lebesgue a.e. $x \in \mathbb{R}$.

As discussed earlier, the hypotheses of Theorem 4.6 are not very different from Theorem 4.1, which was our result for L^1 convergence. In fact, hypothesis (1) in Theorem 4.6 implies that $\sup_n \|g_n\|_1 < \infty$, which is the weaker hypothesis used in Theorem 4.1. Therefore, a family of functions g_n satisfying the hypotheses of Theorem 4.6 is an approximate identity in both the a.e. and L^1 senses.

Proof. We set up to use Theorem 2.4. Define the operators

$$T_n f := g_n * f.$$

By the linearity of convolutions, these operators are linear. The maximal function for this collection is measurable, because we are working with a countable collection. Additionally, we have that

$$T^* f = \sup_n |g_n * f| \leq \sup_n \|\widetilde{g_n}\|_1 \, M_{\mathrm{HL}} f \tag{4.11}$$

by Theorem 4.3. Hypothesis (1), given in (4.8), makes the right hand side of (4.11) a constant multiple of the Hardy-Littlewood maximal function. Therefore this maximal function inherits the Hardy–Littlewood maximal inequality, and Theorem 2.10 applies.

For our dense set, as mentioned above, if $f \in \mathcal{C}_c$, continuity estimates combined with (4.9) and (4.10) implies that $g_n * f \to f$ uniformly. Since the continuous, compactly supported functions are dense in L^1, we have convergence on a dense set.

Since we have linear operators, convergence on a dense set, and a maximal inequality, Theorem 2.4 applies and we have convergence a.e. for each $f \in L^1$.

Finally, we identify these limits. Since convergence in L^1 implies convergence in measure, and $T_\infty h = h$ for each $g \in \mathcal{C}_c$, Theorem 2.7 applies and $g_n * f \to f$ for all $f \in L^1$. ☐

With our approximate identity result for the real line complete, we turn our attention to the circle, $\mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$. Our first result is the analogue of Theorem 4.6 for $\mathbb{T}$. In fact, once we define the symmetric envelope on $\mathbb{T}$, the result follows directly from Theorem 4.6.

In Corollary 4.7 below, we identify elements of $\mathbb{T}$ with the complex numbers $e^{i\theta}, -\pi \le \theta < \pi$ to emphasize the distinction between the circle and the real line.

Corollary 4.7 (Periodic-Approximate-Identities)**.** *Define the symmetric envelope on* $\mathbb{T}$ *by*

$$\widetilde{g}(e^{i\theta}) := \sup_{\psi \in (\pi-|\theta|, \pi+|\theta|)} g(e^{i\psi}) \ .$$

Let g_n *be a family of functions which satisfy the following.*

$$\text{(1) } \sup_{n\in\mathbb{N}} \|\widetilde{g}_n\|_1 < \infty \ , \tag{4.12}$$

$$\text{(2) } \int g_n(e^{i\theta}) \ , \frac{d\theta}{2\pi} = 1 \tag{4.13}$$

$$\text{(3) For all } 0 < r < \pi, \lim_{n\to\infty} \int_{|\theta|>r} |g_n(e^{i\theta})| \frac{d\theta}{2\pi} = 0 \ . \tag{4.14}$$

Let $f \in L^1(\mathbb{T})$*. Then, for a.e.* $\theta \in [-\pi, \pi)$*,*

$$\lim_{n\to\infty} (g_n * f)(e^{i\theta}) = f(e^{i\theta}) \ . \tag{4.15}$$

Proof. Consider "unwrapping" the circle onto the real line by considering the following functions:

$$F(\theta) := \begin{cases} f(e^{i\theta}), & \theta \in [-2\pi, 2\pi) \\ 0, & \text{otherwise.} \end{cases}$$

$$G_n(\theta) := \begin{cases} \frac{1}{2\pi} g_n(e^{i\theta}), & \theta \in [-\pi, \pi) \\ 0, & \text{otherwise.} \end{cases}$$

One can then verify that $F \in L^1(\mathbb{R})$ and $\{G_n\}_{n\in\mathbb{N}}$ satisfies the conditions of Theorem 4.6. Since $G_n * F = F$ *a.e.* in $\mathbb{R}$, in particular it holds *a.e.* on $[-\pi, \pi)$. Since F and G_n are defined by f and g_n on this region, we must have that $g_n * f = f$ *a.e.* on $\mathbb{T}$. ☐

Remark 4.8. Because $\mathbb{T}$ with the normalized Lebesgue measure is a probability space, this result extends to all L^p spaces with $p \ge 1$. However, using techniques we develop in Section 6, we will be able to prove this theorem for L^p functions directly. The same is true for the following two results in this subsection, Theorem 4.9 and Corollary 4.11.

With almost everywhere approximate identities on the circle, we can prove the Lebesgue–Fejer Theorem:

Theorem 4.9 (Lebesgue–Fejer)**.** *Let* $f \in L^1(\mathbb{T})$*, and let* $S_n[f]$ *denote the nth (symmetric) partial sum of the Fourier series of* f*. Then,*

$$C_n[f] := \frac{1}{n} \sum_{j=0}^{n} S_n[f] \to f \tag{4.16}$$

almost everywhere.

Proof. One can rewrite the Cesaro means in (4.16) as

$$C_n[f] = K_n * f$$

where K_n is the Fejer kernel, defined by

$$K_n(e^{i\theta}) := \frac{1}{n}\left[\frac{\sin(n\theta/2)}{\sin(\theta/2)}\right]^2 .$$

Because the Fejer kernel satisfies the conditions of Corollary 4.7 (see [6, Chapter 1.2] or [7, Chapter 1.2], in addition to Remark 4.10 below), we have that

$$K_n * f \to f \ , \text{ a.e.}$$

□

Remark 4.10. We have brushed over the core detail in this proof: proving that K_n does in fact satisfy the conditions we need to apply Corollary 4.7. Two of the three conditions are well-known results for K_n. However, as discussed in the previous section condition (i) in Corollary 4.7 is unique to our maximal functions approach. However, one can quickly show that $\left\|\widetilde{K}_n\right\|_1$ are uniformly bounded from the known bounds of

$$0 \le K_n \le n$$

and

$$K_n \le \frac{\pi^2}{n\theta^2} \ ;$$

see [9, Chapter 2.4] for more details.

Lebesgue–Fejer is a significant result in its own right, but it does give a critical result about Fourier series as a consequence: the sequence of Fourier coefficients of an L^1 function is unique.

Corollary 4.11 (The Uniqueness Property)**.** *Let $f, g \in L^1(\mathbb{T})$. If $\widehat{f}_n = \widehat{g}_n$ for all $n \in \mathbb{Z}$, then $f = g$ almost everywhere.*

Proof. If $\widehat{f}_n = \widehat{g}_n$, then $C_n[f] = C_n[g]$. However, $C_n[f] \to f$ and $C_n[g] \to g$ almost everywhere. Therefore by the uniqueness of limits, $f = g$ almost everywhere. □

5. The Ergodic Theorem and the Strong Law of Large Numbers

In this section we will prove the Strong Law of Large Numbers (SLLN), but will do so from a more general point of view; we will prove the Strong Law as a special case of the Birkhoff Ergodic Theorem, which will follow from our maximal functions framework.

The SLLN states that given a sequence of independent, identically distributed (iid) random variables f_k, that

$$\frac{1}{n}\sum_{k=1}^{n} f_k \to \mathbb{E}(f_1) \tag{5.1}$$

almost surely in n. The SLLN is such a foundational result that it is how many people view calculating the expectation of a random variable in general.

A more demanding proof of the SLLN commonly proven in introductory probability courses (based on Chebyshev–Markov and the Borel–Cantelli lemma) requires that the random variables f_k have fourth moments—that is, $f_k \in L^4$ (see [7, Chapter 5.2], or [1, Chapter 9.5] for an introductory text). This is a very strong requirement, since a random variable having even a second moment is considered a nice property of the random variable. Ideally, we would like to reduce this to the condition that the expectation itself is finite—that is, $f_k \in L^1$—and nothing more.

To do so, we will take a different perspective on the sum in (5.1). Take your sequence of random variables and consider them all together as a function from the underlying probability space to $\mathbb{R}^\infty$ by considering the map

$$f(\omega) := (f_1(\omega), f_2(\omega), \dots) \ . \tag{5.2}$$

And then consider the left-shift map $T : \mathbb{R}^\infty \to \mathbb{R}^\infty$ defined by

$$[T(x)]_j := x_{j+1} \ . \tag{5.3}$$

Combining the maps in (5.2)–(5.3), we can rewrite the sum on the left side of (5.1) as

$$\frac{1}{n}\sum_{k=1}^{n}\pi_1\circ T^k f \tag{5.4}$$

where T^k is the k-fold composition of T, and π_1 is the first coordinate projection.

Sums of the form $\frac{1}{n}\sum_0^{n-1} f\circ T^k$ are the central objects in ergodic theory, where the map T represents a dynamics acting on the underlying probability space. The typical question in the field is if

$$\frac{1}{n}\sum_{0}^{n-1} f\circ T^k \to \int f d\mu \tag{5.5}$$

a.e. with respect to the underlying probability measure. This question, given our reframing, looks similar to the result of the SLLN. This question is answered by the Birkhoff ergodic theorem, which states that (5.5) does hold a.e. if the map T is *ergodic* for the measure μ. See Definition 5.1 below.

It turns out that the map we have defined above to rewrite the statement of the Strong Law does induce an ergodic system, which will allow us to use the statement of the ergodic theorem in order to prove the SLLN.

5.1. Ergodicity and the Ergodic Theorem.

This subsection serves to provide all the necessary preliminaries to state the Birkhoff ergodic theorem. Throughout the rest of Section 5, we take (Ω,σ,μ) to be a probability space, that is $\mu(\Omega)=1$.

Given a probability space, we can imbue the space with a measurable map $T:\Omega\to\Omega$, which obeys

$$\mu(T^{-1}[A])=\mu(A)$$

for all measurable sets A. In this case, we call T a *measure-preserving map*, and we call the probability space alongside T a *measurable dynamical system.* The map T is also a measure preserving map if

$$\int f\circ T d\mu = \int f d\mu \tag{5.6}$$

for all $f\in L^\infty$. We will also use this fact in the context of L^2 functions, since $L^2\supset L^\infty$ in every probability space.

If A is a set which obeys the stronger condition

$$\mu(T^{-1}[A]\triangle A)=0\ , \tag{5.7}$$

(where $\triangle$ is the symmetric set difference) then we say that A is an *invariant set* of T. The invariant sets of a measurable dynamical system form a σ-algebra, which we will denote $\mathcal{I}$. The $\mathcal{I}$-measurable functions are precisely those for which

$$f(T\omega)=f(\omega),\ \ \mu\text{-a.e.} \tag{5.8}$$

If f is a function satisfying (5.8), we call f an *invariant function* of T. With all of this background, we can define ergodicity and give some equivalent formulations.

Definition 5.1. Let (Ω,σ,μ,T) be a measurable dynamical system. We say that the measure μ is *ergodic* for the map T if

$$A\in\mathcal{I} \text{ implies } \mu(A)=1 \text{ or } \mu(A)=0\ . \tag{5.9}$$

That is, the only invariant sets are either null sets or of full measure. By convention, we will refer to the map T as ergodic when the context of the measure is clear. We will similarly refer to the measurable dynamical system as an *ergodic system* if T is ergodic.

Ergodic Theory arises from questions in statistical mechanics, in which case the definition of ergodicity is natural if you are expecting that the dynamics of your system causes any given element of your system to transform with probability 1. However, there are a number of equivalent characterizations of ergodicity. In the following, we list some which will be the most useful to the developments in this section (see [9, Chapter 2.6] for details):

Theorem 5.2 (Ergodicity Equivalents)**.** *Let (Ω,σ,μ,T) be a measurable dynamical system. The following are equivalent:*

(1) μ is ergodic for T;

(2) $A \in \mathcal{I} \implies \mu(A) = 0$ *or* $\mu(A) = 1$*;*

(3) All invariant functions are constant functions;

(4) T *is invertible, and all* L^2 *invariant functions are constant functions.*

Moreover, if f *is an invariant function of an ergodic system, then* $f = \int f d\mu$ *almost everywhere.*

Before we continue, lets explore one example of an ergodic dynamical system relevant in Fourier analysis:

Example 5.3 (Irrational Rotations)**.** Consider the measurable dynamical system consisiting of the probability space $(\mathbb{T}, \frac{d\theta}{2\pi})$ and the dynamics, T_α, defined by

$$T_\alpha \omega = e^{i\alpha}\omega$$

for a fixed $\alpha \in \mathbb{R}$. That is, we send $\theta \to \theta + \alpha \mod 2\pi$. Geometrically, this corresponds to rotating counter-clockwise on the unit circle by an arclength of α.

Proposition 5.4. *Let* $\alpha \in \mathbb{R} \setminus \mathbb{Q}$*. Then the dynamical system* $(\mathbb{T}, \frac{d\theta}{2\pi}, T_\alpha)$ *is ergodic.*

Proof. To prove this, we will show that every L^2 invariant function is constant, since T_α is bijective on the circle. Let $f \in L^2(\mathbb{T})$ be an invariant function. Since we're working in a probability space, $L^1 \supset L^2$, so our function f is L^1 and its Fourier coefficients, defined by

$$\widehat{f}_n := \int_0^{2\pi} f(\theta) e^{-ni\theta} \frac{d\theta}{2\pi} \ , \ n \in \mathbb{Z} \ , \tag{5.10}$$

exist. Furthermore, since we are working on the circle, and f is invariant, we can say that f is both 2π and $\alpha 2\pi$ periodic. This allows us to use the following:

Lemma 5.5. *Let* $f \in L^1$ *be a function that is both* s*-periodic and* t*-periodic, with* $0 < s < t$*. Then,*

$$n\frac{s}{t} \notin \mathbb{Z} \implies \widehat{f}_n = 0 \ . \tag{5.11}$$

Since f is both 2π and $\alpha 2\pi$ periodic, ns/t can only be an integer if either $n\alpha$ or $\frac{n}{\alpha}$ are integers. Since α is irrational, this only happens when $n = 0$, meaning that the only nonzero Fourier coefficient of f is

$$\widehat{f}_0 = \int_0^{2\pi} f(\theta) \frac{d\theta}{2\pi} \ .$$

However, if we consider the constant function $g = \widehat{f}_0$, we find that $\widehat{g}_0 = \widehat{f}_0$, but moreover that $\widehat{g}_n = \widehat{f}_n$. Therefore, by the Uniqueness Property (Corollary 4.11), $f = g$ a.e. Therefore f is a constant function. Since f was an arbitrary L^2 function, Theorem 5.2 tells us that the Lebesgue measure is ergodic for T_α. ☐

In view of the SLLN, our next example of an ergodic system will be the Bernoulli Shift, which we state below in Example 5.9. In view of proving that the Bernoulli Shift is ergodic, we will introduce a stronger condition that the Bernoulli Shift satisfies, called (strongly) *mixing*.

Definition 5.6 (Mixing)**.** Let (Ω, σ, μ, T) be a measurable dynamical system. We say that T is *mixing* or *strongly mixing* if for all $A, B \in \sigma$,

$$\lim_{n\to\infty} \mu(A \cap T^{-n}(B)) = \mu(A)\mu(B) \ .$$

Intuitively, T is mixing if applying the map to your space repeatedly makes events in the space act if they're independent.

Here are a couple quick facts about dynamics that are mixing. We will start with the one we have already alluded to above:

Theorem 5.7. *Let* (Ω, σ, μ, T) *be a measurable dynamical system. If* T *is mixing, then* μ *is ergodic for* T*.*

We note that the converse of Theorem 5.7 is false; irrational rotations on the circle (Example 5.3) form an example of an ergodic system which is not strongly mixing. We also have the following characterization.

Lemma 5.8. *Let (Ω, σ, μ, T) be a measurable dynamical system. T is mixing if and only if, for every $f, g \in L^2(\Omega)$ we have that*

$$\langle f, g \circ T^n \rangle - \langle f, 1 \rangle \langle 1, g \rangle \to 0 \ .$$

Equivalently, if f, g are real valued,

$$\lim_{n\to\infty} \int (f)(g \circ T^n) d\mu = \left(\int f d\mu \right) \left(\int g d\mu \right) \ .$$

See the notes in [9, Chapter 2.6] for details.

We will finish this subsection with an example of a map that is not only ergodic, but mixing. This example will be especially important to us when proving the SLLN (see Section 5.3).

Example 5.9 (Bernoulli Shifts)**.** Start with your favorite probability space, which we will denote $(\Omega, d\mu(\omega))$ (Here we are using $d\mu$ to imply an underlying σ-algebra). From this probability space, construct the infinite product space

$$\otimes_1^\infty (\Omega, d\mu(\omega)) := \left(\bigtimes_1^\infty \Omega, \ \otimes_{j=1}^\infty d\mu(\omega_j) \right) \ .$$

Constructing the infinite product measure requires some finesse, but we will omit most of those details. The construction involves looking at the cylinder sets as a generating σ-algebra (See [2, Section 24] and [10, Chapter 4.12]), which in particular implies the following fact that we use in the proof of Proposition 5.11 below.

Lemma 5.10. *Let $X = \otimes_1^\infty (\Omega, d\mu)$. Let $\mathcal{F}$ be the family of functions which are continuous, and depend on only finitely many variables. Then $\mathcal{F}$ is dense in $L^2(X)$.*

With this background, we can define our dynamics on $\otimes_1^\infty (\Omega, d\mu)$ by taking, for each $\omega \in \otimes_1^\infty (\Omega, d\mu)$,

$$[T(\omega)]_n = \omega_{n+1} \ .$$

The dynamical system $(\otimes_1^\infty (\Omega, d\mu), T)$ is called the *generalized Bernoulli shift*. The name comes from a special case of having $(\Omega, d\mu)$ be $\{0, 1\}$ with a Bernoulli distribution.

Proposition 5.11. *The generalized Bernoulli shift is mixing, and therefore ergodic.*

Proof. Let $f(\omega) = F(\omega_1, \ldots, \omega_j)$, with F continuous. It is clear that due to the dependence only on a finite number of variables, that $\int f \circ T d\mu = \int f d\mu$, since each of these variables belongs to an extra copy of the same space. By Lemma 5.10 and a limiting argument, we have this property for all L^2, giving us (5.6). Therefore T is measure-preserving, and we have a measurable dynamical system.

Similarly, if we take the same $f = F(\omega_1, \ldots, \omega_j)$ from before, and $g \in L^2$, then for $k \geq j$, we have that f and $g \circ T^k$ depend on separate variables. Combining this with the measure-preserving property above gives us that

$$\int (f)(g \circ T^k) d\mu = \left(\int f d\mu \right) \left(\int g d\mu \right).$$

A similar limiting argument due to Lemma 5.10 allows us to conclude that for every $f, g \in L^2$, we have that

$$\lim_{k\to\infty} \int (f)(g \circ T^k) d\mu = \left(\int f d\mu \right) \left(\int g d\mu \right) \ .$$

Therefore we have that T is mixing, and by Theorem 5.7, ergodic. ☐

We define one final object relevant for the Birkhoff ergodic theorem. The object in question is the conditional expectation, which we will define as part of the following theorem.

Theorem 5.12 (Conditional Expectation)**.** *Let (Ω, σ, μ) be a probability space and let $\sigma' \subset \sigma$ be a subalgebra. Let $\mu' = \mu|_{\sigma'}$. For every $f \in L^\infty(\Omega, \sigma, \mu)$, there is a unique $g \in L^1(\Omega, \sigma', \mu')$ so that for all $h \in L^\infty(\Omega, \sigma', \mu')$, we have that*

$$\int h(\omega) f(\omega) \ d\mu(\omega) = \int h(\omega) g(\omega) \ d\mu'(\omega) \ . \tag{5.12}$$

In this case, we will denote

$$g =: \mathbb{E}(f \mid \sigma'). \tag{5.13}$$

and refer to g as the conditional expectation (of f with respect to σ'). Moreover, we have that

$$\|\mathbb{E}(f \mid \sigma')\|_p \leq \|f\|_p \tag{5.14}$$

for each $p \in [1, \infty]$. Thus $\mathbb{E}(\cdot \mid \sigma')$ extends uniquely to a map from L^p to L^p.

The proof of existence for the conditional expectation relies on Hilbert space theory which only works for functions that are at least L^2. In view of proving the SLLN under the weaker assumption that our iid random variables are only L^1 (see Section 5.3), we will want to apply the conditional expectation to L^1 functions. Thus the unique extension is necessary. We will not go into the details of the proof for the unique extension; it relies on showing that the map $f \to \mathbb{E}(f \mid \sigma')$ is *doubly stochastic*, which then implies our unique extension; see [9, Chapter 2.6] for details.

Proof. View $L^p(\Omega, d\mu')$ as a subspace of $L^p(\Omega, d\mu)$. This must be a closed subspace by the completeness of $L^p(\Omega, d\mu')$. In particular, taking $p = 2$, there is an orthogonal projection P satisfying

$$\langle h, f\rangle = \langle h, Pf\rangle \ , \tag{5.15}$$

for all $h \in L^2(d\mu')$ and $f \in L^2(d\mu)$. Taking $g = Pf$ gives us (5.12), showing existence.

To show uniqueness, assume there exist two functions g_1, g_2 that satisfy (5.12). Its easy to see then that

$$\int h(\omega)[g_1(\omega) - g_2(\omega)]d\mu' = 0 \ , \tag{5.16}$$

for every $h \in L^\infty(d\mu')$. Consider the set $E = \{g_1 > g_2\}$. Then, $\chi_E \in L^\infty(d\mu')$, so

$$\int_E g_1(\omega) - g_2(\omega)d\mu' = \int \chi_E(\omega)[g_1(\omega) - g_2(\omega)]d\mu' = 0 \ . \tag{5.17}$$

However, since $g_1 - g_2 > 0$ on E, we must therefore have that $\mu'(E) = 0$ so that $g_2 \geq g_1$ μ'-a.e. Since this argument is symmetrical between g_1 and g_2, we must have also that $g_1 \geq g_2$ a.e., and therefore $g_1 = g_2$. This proves uniqueness. ☐

Because the conditional expectation arises as an orthogonal projection, we immediately have a continuity statement as well.

Corollary 5.13 (Continuity of Conditional Expectation)**.** *Consider the setup of Theorem 5.12. Furthermore, let $f_n \in L^\infty(\Omega)$ be a sequence of functions that converges to f in the L^p sense. Then,*

$$\mathbb{E}(f_n \mid \sigma') \to \mathbb{E}(f \mid \sigma') \tag{5.18}$$

in the L^p sense.

Our particular interest will be looking at the conditional expectation with respect to the invariant set σ-algebra. One detail of note is that $\mathbb{E}(f \mid \mathcal{I})$ will always be an invariant function. Combining this with Theorem 5.2, we have that in an ergodic system,

$$\mathbb{E}(f \mid \mathcal{I}) = \int f d\mu, \ \mu\text{-a.e.}$$

With this in mind, we can state the Birkhoff ergodic theorem, the proof of which we provide in Section 5.3 after discussing preliminaries in Section 5.2

Theorem 5.14 (Birkhoff Ergodic Theorem)**.** *Let (Ω, σ, μ, T) be a measurable dynamical system, and let $f \in L^1(\Omega)$. Then, for a.e. $\omega \in \Omega$,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k = \mathbb{E}(f \mid \mathcal{I}) \ . \tag{5.19}$$

In particular, if μ is ergodic for T, then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k = \int f d\mu \ . \tag{5.20}$$

5.2. **Dense Sets and Maximal Inequalities.** This subsection will be devoted to setting up preliminaries for the proof of the Birkhoff theorem. The proof of the Birkhoff theorem is more involved than previous proofs; in our other results, we could appeal to convergence on a dense set that we had as a prior result, e.g. classes of continuous functions with Lebesgue differentiation (Theorem 3.4) or $\mathbb{R}$-approximate-identities (Theorem 4.6). With the ergodic theorem we are not so lucky, and will have to justify convergence on a dense set ourselves. We will also still have to prove a maximal inequality, which has proven to be a consistently nontrivial task. This section will prove both of those results. Throughout this section we'll denote

$$\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k =: Av_n[f]\ .$$

We will begin with convergence on a dense set. We start by picking some classes of functions where we quickly know what the conditional expectation is. Because $\mathbb{E}(\cdot \mid \mathcal{I})$ is an orthogonal projection onto the invariant functions, we can pick two easy candidates for classes of functions: invariant functions, and functions that lie in the orthogonal compliment of the invariant functions.

Lemma 5.15. *Let (Ω, σ, μ, T) be a measurable dynamical system.*

- *(a) Let f be a T-invariant function. Then $Av_n[f] \to \mathbb{E}(f \mid \mathcal{I}) = f$ in n almost everywhere, and in L^p if f is L^p.*
- *(b) Let $g \in L^\infty$ and let $f = g - g\circ T$. Then $Av_n[f] \to \mathbb{E}(f \mid \mathcal{I}) = 0$ in n almost everywhere, and in L^p if f is L^p.*

Proof. (a) is trivial. For (b), we can quickly see $\mathbb{E}(f \mid \mathcal{I}) = 0$, as f is orthogonal to every invariant function, and so applying our orthogonal projection to it returns 0. As for the almost everywhere and L^p convergence, we will do both at once by showing convergence in L^∞. In this case, $Av_n[f] = Av_n[g - g\circ T]$ is a telescoping sum, giving us that

$$\|Av_n[f]\|_\infty = \left\| \frac{1}{n}[g - g\circ T^n] \right\|_\infty \le \frac{2}{n}\|g\|_\infty \to 0 \tag{5.21}$$

since $g \in L^\infty$. ☐

Since the conditional expectation is linear (arising from the orthogonal projection), it makes sense to add these two kinds of function together; the resulting function will still have its averages converge to the conditional expectation due to linearity. It turns out that functions of this combined type are the dense set we need.

Theorem 5.16. *Let Y be the set of functions $f = f_1 + f_2$, where f_1 is a bounded, T-invariant function, and f_2 is of the form $f_2 = g - g\circ T$ for $g \in L^\infty$. Then Y is dense in L^p for $1 \le p \le 2$.*

Proof. Assume for contradiction that $\overline{Y} \neq L^p$. Hahn-Banach ensures existence of a linear functional which is zero on $\overline{Y}$ and nonzero for at least one function outside of it. Letting q be the dual index of p, the duality of L^p spaces then implies that this linear functional can be represented by a *nonzero* function $h \in L^q$ so that

$$\int hf d\mu = 0 \tag{5.22}$$

for all $f \in Y$. In particular, this includes functions which are either bounded, invariant functions, or of the form $g - g\circ T$ for $g \in L^\infty$. For $g \in L^\infty$, we therefore have that

$$\int hg d\mu = \int h(g\circ T) d\mu\ .$$

Approximating by simple functions, we can extend this to any $g \in L^p$. Since $p \le 2$, and we are working within a probability space, $L^q \subset L^p$, and we can pick $g = \bar{h}$. Using our integral characterization of the invariance of μ, we therefore have that

$$\int |h|^2 d\mu = \int h\overline{h\circ T} d\mu = \int |h\circ T|^2 d\mu\ ,$$

which we may also write as

$$\langle h, h\rangle = \langle h, h\circ T\rangle = \langle h\circ T, h\circ T\rangle\ ,$$

and so

$$\int |h - h\circ T|^2 d\mu = \langle h - h\circ T, h - h\circ T\rangle = \langle h, h\rangle + \langle h\circ T, h\circ T\rangle - 2\langle h, h\circ T\rangle = 0 \ .$$

This implies that $h = h\circ T$ almost everywhere, which implies that h is invariant.

To conclude the proof, recall that (5.22) holds for any invariant $f \in L^\infty$. Since $\mathcal{I}$ is a σ-algebra, we can approximate any invariant $f \in L^p$ by simple functions. Therefore, taking $f = \bar{h}$ in (5.22), we find that

$$\int |h|^2 d\mu = 0$$

and so $h = 0$, which is a contradiction. Therefore $\overline{Y} = L^p$ and this completes the proof. ☐

With our dense set in hand, we turn to the topic of our maximal inequality. To begin, we define our maximal operator. Let

$$M_{\mathrm{E}} f := \sup_n |Av_n[f]| \ . \tag{5.23}$$

Since we are indexing over a countable set, this maximal function is automatically measurable. We will also want to define the following object. For f real-valued, set

$$f^\sharp := \sup_n Av_n[f] \ . \tag{5.24}$$

This is not quite a maximal function, but we will call it one for the sake of the name of the following theorem, which will serve as a key lemma to prove a maximal inequality.

Theorem 5.17 (Hopf–Kakutani–Yoshida Maximal Ergodic Theorem)**.** *Let $f \in L^1$. Then,*

$$\int_{\{\omega \ \mid \ f^\sharp(\omega) > 0\}} f(\omega) d\mu \geq 0 \ . \tag{5.25}$$

Proof. Let $g_0 = 0$. Then, for $n \geq 1$, define the following.

$$\begin{aligned} g_n &:= nAv_n[f] = \sum_{j=0}^{n-1} f\circ T^j \ , \\ h_n &:= \max_{0\leq j\leq n} g_j \ , \\ E_n &:= \{\omega \mid h_n > 0\} \ . \end{aligned}$$

Note that by construction, $h_n/n \to f^\sharp$ in n, so $E_n \to \left\{f^\sharp > 0\right\} =: E$ in n as well. Therefore, by the dominated convergence theorem, it suffices to show that

$$\int \chi_{E_n} f d\mu \geq 0 \tag{5.26}$$

as this will imply (5.25). To prove this, first note that $f + g_n \circ T = g_{n+1}$. This gives us that

$$f + h_n\circ T = \max_{0\leq j\leq n}(f + g_j\circ T) = \max_{1\leq j\leq n+1} g_j \geq \max_{0\leq j\leq n} g_j.$$

Since, on E_n, $h_n > 0$, we must have that some $g_j > 0$ for $j \leq n$, and therefore, on E_n,

$$f + h_n\circ T \geq \max_{0\leq j\leq n} g_j = h_n.$$

Integrating over E_n we thus find that

$$\int_{E_n} f d\mu \geq \int_{E_n} h_n d\mu - \int_{E_n} h_n\circ T d\mu$$

Furthermore, since $g_0 = 0$, $h_n \geq 0$ on the whole space, and since T is measure preserving, we have

$$\int_{E_n} h d\mu = \int h d\mu = \int h\circ T d\mu$$

and so,

$$\int_{E_n} h_n d\mu - \int_{E_n} h_n\circ T d\mu = \int_{\Omega\setminus E_n} h_n\circ T d\mu \geq 0 \ .$$

Therefore,

$$\int_{E_n} f d\mu \geq 0$$

which completes the proof. ☐

With Theorem 5.17, we can prove our maximal inequality.

Theorem 5.18 (Ergodic Maximal Inequality)**.** *Let* (Ω, σ, μ, T) *be a measurable dynamical system. For any* $\alpha > 0$ *and* $f \in L^1(\Omega)$*, we have that*

$$\mu(M_{\mathrm{E}} f > \alpha) \leq \frac{\|f\|_1}{\alpha}. \tag{5.27}$$

Proof. Since $|Av_n[f]| \leq Av_n[|f|]$, we may WLOG take $f \geq 0$. For $\alpha > 0$, let $g(\omega) = f(\omega) - \alpha$, and apply Theorem 5.17 to g. Note that

$$Av_n[g] = Av_n[f] - \alpha$$

which implies that

$$g^\sharp > 0 \iff M_{\mathrm{E}} f > \alpha$$

($f^\sharp = M_{\mathrm{E}} f$ since $f \geq 0$). With this, by the maximal ergodic theorem we have that

$$\int_{\{M_{\mathrm{E}} f > \alpha\}} (f - \alpha) d\mu \geq 0 \ ,$$

thus,

$$\int_{\{M_{\mathrm{E}} f > \alpha\}} f d\mu \geq \alpha \mu(M_{\mathrm{E}} f > \alpha) \ .$$

Finishing up, since $f \geq 0$, we have

$$\int_{\{M_{\mathrm{E}} f > \alpha\}} f d\mu \leq \|f\|_1 \ ,$$

so that

$$\mu(M_{\mathrm{E}} f > \alpha) \leq \alpha^{-1} \int_{\{M_{\mathrm{E}} f > \alpha\}} f d\mu \leq \frac{\|f\|_1}{\alpha} \ .$$

☐

5.3. **Proof of the Ergodic Theorem and Strong Law of Large Numbers.** With the setup from Section 5.2, we can prove the Birkhoff theorem and the SLLN.

Proof of Theorem 5.14. Recall the operators $Av_n[f]$. By the linearity of sums, these operators are linear. The maximal operator for these operators is M_{E}, which is measurable. Theorem 5.16 provides a dense set in which we know all functions in the set converge to $\mathbb{E}(\cdot \mid \mathcal{I})$. Theorem 5.18 provides a maximal inequality, and thus the C–δ hypothesis. Since we have linear operators, a dense set, and a maximal inequality, Theorem 2.4 applies and we have convergence for all $f \in L^1$. Furthermore, by Corollary 5.13 we know that if $f_n \to f$ in L^1, $\mathbb{E}(f_n \mid \mathcal{I}) \to \mathbb{E}(f \mid \mathcal{I})$ in L^1, and therefore in measure. Thus Theorem 2.7 applies and we find that $T_\infty f = \mathbb{E}(f, \mathcal{I})$ for all $f \in L^1$. This completes the proof. ☐

With the ergodic theorem proven, we can finally state and prove the SLLN. Though we will not be using maximal functions directly, this proof follows the general theme of our maximal function proofs: we have already done the hard work in the background. Here we will rely on the ergodic theorem, and Example 5.9, the generalized Bernoulli shift.

Theorem 5.19 (The Strong Law of Large Numbers)**.** *Let* $(\Omega, d\mu)$ *be a probability space, and let* f_n *be iidrv's with the additional condition that* $\mathbb{E}(f_n) < \infty$ *for each* n*. Then, for almost every* $\omega \in \Omega$*,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} f_j(\omega) = \mathbb{E}(f_1) \ . \tag{5.28}$$

Proof. Let $f_n : \Omega \to \mathbb{R}$ be iid random variables. WLOG, assume that $f_n \geq 0$. Otherwise, since each f_n is integrable, its positive and negative components must be integrable. Moreover, f_n^+ will be iid random variables, and f_n^- will be iid random variables. Thus we can safely assume positive, and thus L^1, random variables.

Imbue $\mathbb{R}$ with the probability measure given by $\nu(E) = \mu(f_n^{-1}(E)) = \mu(f_1^{-1}(E))$ for each Borel set E, and let T be the generalized Bernoulli shift (See Example 5.9) on $\otimes_1^\infty(\mathbb{R}, d\nu)$. Furthermore, let $\pi_1 : \mathbb{R}^\infty \to \mathbb{R}$ be the first-coordinate projection, that is $\pi_1(x_1, x_2, \dots) = x_1$. Note that $\pi_1 \in L^1(\otimes_1^\infty(\mathbb{R}, d\nu))$ since

$$\int_{\mathbb{R}^\infty} |\pi_1| \otimes_1^\infty d\nu = \int_{\mathbb{R}} |x_1| d\nu(x_1) = \int_{\mathbb{R}} |x_1| d(\mu \circ f_1^{-1})(x_1) = \int_{\Omega} |f_1(\omega_1)| d\mu(\omega_1) = \|f_1\|_1 \ .$$

Letting $f(\omega) = (f_1(\omega), f_2(\omega), \dots)$, we therefore have that

$$\frac{1}{n}\sum_{j=1}^{n} f_j(\omega) = \frac{1}{n}\sum_{j=0}^{n-1} [\pi_1 \circ T^j \circ f](\omega) \ .$$

Since $\pi_1 \in L^1$, and the Bernoulli shift is ergodic, we may apply the ergodic theorem, which gives us that

$$\frac{1}{n}\sum_{j=0}^{n-1} [\pi_1 \circ T^j \circ f](\omega) \to \int \pi_1 \circ f d\mu = \int f_1 d\mu = \mathbb{E}(f_1)$$

μ a.e. as $n \to \infty$. □

6. Bounded Operators and Maximal Inequalities

The goal of this section will be to extend many of our results from Sections 3 and 4, by expanding the class of functions they work for from L^1 to L^p. To do so, we will discuss the topic of distribution functions, and the space of weak-L^p functions.

Recall from our discussion in Section 2 that if a maximal operator T^* is a bounded operator into L^p, then Chebyshev–Markov implies a maximal inequality of the form

$$\mu(T^* f > C) \leq \frac{\|f\|_p^p}{C^p} \ . \tag{6.1}$$

If we want to prove a result for all p, we can do so by showing that T^* is a bounded operator from L^1 to L^1 and from L^∞ to L^∞, and then using an interpolation result (e.g., Riesz–Thorin) to get that T^* is a bounded operator for all L^p. In fact, this will be our eventual strategy (though Riesz–Thorin specifically will not suffice).

Unfortunately, proving that T^* is a bounded operator is not as easy as one might hope; one of the properties that T^* does not retain from our net of operators T_α is boundedness. For a simple example, consider the translation operators

$$T_n f := f(\cdot + n) \ . \tag{6.2}$$

If we take $f \in L^p(\mathbb{R})$, these operators are bounded operators into L^p. However, it easy to see that T^* for this collection is not a bounded operator; taking $f = \chi_{[0,1]}$ gives one such counterexample.

In view of attempting to extend our results that were based on the Hardy–Littlewood maximal function, we have an even worse situation. It is well known that the Hardy–Littlewood maximal function is actually *never* in L^1 if $f \in L^1$ [9, Chapter 2.3, Problem 1]. This means that, as stated, we cannot use an interpolation result to extend our results for even a special subclass of L^1, which may have let us use a density argument to extend our results.

In spite of this, M_{HL} does still satisfy a maximal inequality, and does so in a very similar form to the one that Chebyshev-Markov gives us (with an additional absolute constant). This motivates the idea of a "weak" L^p space; a function which, while not being L^p, satisfies similar a.e. finiteness conditions to L^p functions, which can serve to satisfy the C–δ hypothesis.

6.1. **Distribution Functions and Weak-L^p.** In this subsection we will build up the necessary background and provide some additional motivation for weak-L^p spaces. In particular, we will define the concept of a distribution function, and share some key results about it.

Given a measure space (Ω, σ, μ), one of the objects we care most about, given the formulation of the C–δ hypothesis is the measure of a superlevel set of our maximal function, $\mu(T^* f > C)$. More generally, we want to consider the superlevel set of the modulus of a function.

Definition 6.1 (Distribution Functions)**.** Let (Ω, σ, μ) be a measure space and let $f \in \mathcal{M}(\Omega)$. Define the *distribution function* of f by

$$m_f(t) := \mu(|f| > t)\ , t \in [0, \infty)\ . \tag{6.3}$$

The distribution function has a number of fascinating properties. For example, if m_f is finite a.e., then it is right continous, monotone decreasing, and asymptotically approaches 0 as $t \to \infty$. A function which satisfies these properties is called *right–continuous monotone* (rcm).

We have already used the distribution function in several contexts without describing it as such. Chebyshev–Markov, for one, is simply a bound for the distribution function of an L^p function. The C–δ hypothesis can be rephrased as requiring for each $C > 0$ that

$$\lim_{\delta \to 0^+} \sup_{\|f\| \le \delta} m_{T^* f}(C) = 0\ . \tag{6.4}$$

We even mentioned the distribution function previously, in Remark 4.4, where we used a special case of the following important theorem.

Theorem 6.2. *Let (Ω, σ, μ) be a measure space and let $f \in L^p(\Omega)$ for $1 \le p < \infty$. Then,*

$$\|f\|_p^p = p \int_0^\infty t^{p-1} m_f(t) dt\ . \tag{6.5}$$

See [9, Chapter 2.2] for details.

This equality gives us some insight to how the distribution function interacts with wether or not a function is L^p. For the integral on the right hand side of (6.5) to be finite, as t get large, m_f has to decay significantly faster than t^{p-1} grows. If m_f looks like $1/t^p$ eventually in t, then this integral diverges logarithmically, and f fails to be L^p. However, if this is the case, we would have a bound for the distribution function which is akin to Chebyshev–Markov. This is our motivation for our definition of the space of weak-L^p functions; the functions which fail to be L^p, but only akin to growing logarithmically.

Definition 6.3 (Weak-L^p)**.** Let f be a measurable function and let $p \in [1, \infty)$. define the quantity

$$\|f\|_{p,w} := \sup_{t>0} [t^p m_f(t)]^{1/p}\ . \tag{6.6}$$

If $\|f\|_{p,w} < \infty$, then we say f is a *weak-L^p function*, denoted $f \in L^p_w$. Additionally, we define the space $L^\infty_w := L^\infty$.

While styled like a norm, $\|f\|_{p,w}$ actually fails to be a norm. Lemma 6.5 below indicates that $\|\cdot\|_{p,w}$ fails the triangle inequality. There is a proper norm for L^p_w, which actually turns L^p_w into a Banach space, but that norm is equivalent (in the sense of convergence) to the quantity we have defined, and we otherwise will not need it.

As expected, this construction gives us an inequality akin to Chebyshev–Markov:

Lemma 6.4. *Let $f \in L^p_w$, $p < \infty$. Then, for all $C \in (0, \infty)$,*

$$m_f(C) = \mu(|f| > C) \le \frac{\|f\|_{p,w}^p}{C^p}\ . \tag{6.7}$$

We mention some miscellaneous properties of L^p_w as well. Details in [9, Chapter 2.2]

Lemma 6.5 (Properties of L^p_w)**.** *The space L^p_w and $\|\cdot\|_{p,w}$ enjoy the following properties.*

(1) $L^p \subsetneq L^p_w$; $\|f\|_{p,w} \le \|f\|_p$.

(2) $\|\lambda f\|_{p,w} = |\lambda|\, \|f\|_{p,w}$, *and* $\|f+g\|_{p,w} \le 2(\|f\|_{p,w} + \|g\|_{p,w})$. *In particular, L^p_w is a vector space.*

(3) Let $1 \leq p_1 < p < p_2 < \infty$. If $f \in L_w^{p_1} \cap L_w^{p_2}$ then $f \in L^p$ with

$$\|f\|_p^p \leq \left(\frac{p}{p_2 - p}\right) \|f\|_{p_2,w}^{p_2} + \left(\frac{p}{p - p_1}\right) \|f\|_{p_1,w}^{p_1} \ .$$

Note that this third property is an interpolation result which allows us to interpolate between weak-L^p spaces and yield a true L^p space! While this result on its own won't be powerful enough for our needs (in particular, we would like to let $p_2 = \infty$), it gives some additional motivation towards our eventual interpolation strategy for extending results.

We conclude this section with a generalization of Theorem 2.11, which stated that if T^* was a bounded operator into L^p, then it satisfied the C–δ hypothesis. Here we state the same thing for a bounded operator into L_w^p. Since L^p is a subspace of L_w^p, this result encapsulates our result in Section 2 based on Chebyshev–Markov. However, note that this extension to L_w^p is special, in that it is both necessary, and sufficient.

Theorem 6.6 (Weak-L^p maximal functions). *Let T^* be a maximal operator. Then, T^* is a bounded operator from X to L_w^p, $p < \infty$ if and only if T^* satisfies a maximal inequality of the form*

$$\mu(T^* f > C) \leq \left(\frac{K \|f\|_X}{C}\right)^p \tag{6.8}$$

where K is a constant. In particular, a maximal operator T^* satisfying either of these conditions satisfies the C–δ hypothesis.

Proof. Both directions are quick:

$\Rightarrow$) Suppose T^* is a bounded operator from X to L_w^p. Combining the definition of boundedness with Lemma 6.4, we get that

$$\mu(T^* f > C) = m_{T^* f}(C) \leq \left(\frac{\|T^* f\|_{p,w}}{C}\right)^p \leq \left(\frac{K \|f\|_X}{C}\right)^p \ . \tag{6.9}$$

$\Leftarrow$) Now suppose T^* satisfies a maximal inequality of the form (6.8). Using the definition of weak-L^p combined with (6.8), we find that

$$\|T^* f\|_{p,w} = \sup_{C>0} [C^p m_{T^* f}(C)]^{1/p} \leq \sup_{C>0} \left[C^p \left(\frac{K \|f\|_X}{C}\right)^p\right]^{1/p} = K \|f\|_X \ . \tag{6.10}$$

This completes the proof. ☐

6.2. **Marcinkiewicz Interpolation and $M_{\mathbf{HL}}$.**

In this subsection we state the interpolation theorem which will allow us to prove that M_{HL} is a bounded operator into weak-L^p for all p. This will allow us to extend a great number of results all at once.

We will begin with the Marcinkiewicz interpolation theorem. This theorem is especially nice, since it only requires that our "endpoint" operators are bounded into weak-L^p, and allows the upper endpoint to be ∞. We omit the proof for space; see [9, Chapter 2.2] for details.

Theorem 6.7 (Marcinkiewicz Interpolation). *Let $(X, \mu), (Y, \nu)$ be σ-finite measure spaces and let $B_0(X)$ be the set of finitely supported, bounded measurable functions defined on X.*

Additionally, let $1 \leq p_0 < p_1 \leq \infty$. Let $T : B_0(X) \to \mathcal{M}(Y)$ be a subaddive map such that for each $f \in B_0(X)$,

$$\|Tf\|_{p_j,w} \leq C_j \|f\|_{p_j} \ , \ j = 0, 1. \tag{6.11}$$

Then, for $p_0 < p < p_1$, T extends to a bounded map from $L^p(X, \mu) \to L^p(Y, \nu)$. That is, we have for each $f \in L^p(X, \mu)$ that

$$\|Tf\|_p \leq C_p \|f\|_p \ . \tag{6.12}$$

Remark 6.8. We have some details of note:

- As in Remark 2.5, if Tf is complex valued, we require that T is additive.
- The σ-finiteness of the measure space is to ensure that $B_0(X)$ is a dense subset of $L^p(X, \mu)$. Without σ-finiteness, Theorem 6.7 still allows interpolation to $p_0 < p < p_1$, but only for $f \in B_0(X)$.

- We omit the full description of the constant, C_p, in the final bound on L^p. It is longwinded and unimportant to our purposes. However, this constant is one of the main drawbacks of the Marcinkiewicz theorem. In particular, the constant is badly behaved as p approaches either p_0 or p_1; see [9, Chapter 2.2].

The Marcinkiewicz interpolation theorem is our silver bullet for extending our results. In particular, it lets us prove the following.

Theorem 6.9 (M^{μ}_{HL} is L^p)**.** *Let M^{μ}_{HL} denote the Generalized Hardy–Littlewood maximal function (Definition 3.5). We have that M^{μ}_{HL} is a bounded map from L^1 to L^1_w, and from L^p to L^p for $1 < p \leq \infty$.*

Proof. Quick ML estimates immediately show that

$$\|M^{\mu}_{\mathrm{HL}} f\|_{\infty} \leq \|f\|_{\infty} ,$$

so M^{μ}_{HL} is a bounded map from L^{∞} to itself. Furthermore, the (generalized) Hardy–Littlewood maximal inequality is a maximal inequality of the form described in Theorem 6.6 for $p = 1$, and so M^{μ}_{HL} is a bounded map from L^1 to L^1_w.

Finally, since M^{μ}_{HL} is a subadditive map, and μ is defined to be σ-finite, we can apply the Marcinkiewicz interpolation theorem with $p_0 = 1$, $p_1 = \infty$. This completes the proof. □

Since M^{μ}_{HL} is a bounded map in this way, any results we have that relied on the Hardy–Littlewood maximal function to satisfy the C–δ hypothesis immediately extend to L^p for $1 \leq p < \infty$ as long as the dense set is still dense in general L^p spaces on σ-finite measure spaces (e.g. flavors of continuous functions). The case for $p = \infty$ does not inherit a maximal inequality from Theorem 6.6, and thus these theorems do not gain the case of L^{∞} as well. In particular, we have the following.

Corollary 6.10 (Many Extended Results)**.** *The following theorems have the following extensions. Let $1 \leq p < \infty$.*

- *(i) The Lebesgue Differentiation Theorem (Theorem 3.4) and the Generalized Lebesgue Differentiation Theorem (Theorem 3.8) apply to L^p functions on the given measure space.*
- *(ii) $\mathbb{R}$-Approximate-Identities (Theorem 4.6) are still approximate identities if $f \in L^p$. Similarly, Periodic-Approximate-Identities (Theorem 4.7) hold if $f \in L^p$ as well.*
- *(iii) Lebesgue–Fejer (Theorem 4.9) and the Uniqueness Property (Corollary 4.11) both hold for the Fourier series of L^p functions on $\mathbb{T}$.*

As mentioned in Remark 4.8: Theorem 4.7, Theorem 4.9, and Corollary 4.11 can also be proved for general L^p using the fact that L^p spaces are nested when defined on probability spaces. Corollary 6.10 gives another method to proving this fact without relying on the nested behavior of L^p spaces.

7. Convergence of Fourier Series

In this final section, we conclude our discussion on maximal functions by discussing the almost-everywhere convergence of Fourier series. In particular, we will comment on the proof of Carleson's theorem, which relies on the tools we have developed in Section 6 about bounded maximal operators into weak-L^p spaces. The topics in this section are presented as discussion, omitting many details and proofs. Details can be found in [6], [7], and [8].

It is well known that the Fourier series of a function f converges if the function satisfies certain smoothness conditions. If f is $\mathcal{C}^2$, or piecewise $\mathcal{C}^1$, the Fourier series of f converges pointwise to the average of the left and right limits of f at each point. For piecewise $\mathcal{C}^1$ functions, this differs from f only at finitely many points, and thus the Fourier series of f converges to f Lebesgue a.e. [6, Chapter 2.2]. This result cannot be much improved; for every dense, countable subset of $\mathbb{T}$ there is a continuous function whose Fourier series diverges on that set [5].

Thus the question turns to integrability instead of smoothness. We have already seen a necessary condition for the Fourier series of L^p functions to converge almost everywhere. The Lebesgue-Fejer Theorem, Theorem 4.9, implies that Fourier series of L^p functions are Lebesgue a.e. Cesaro-summable, which is necessary (but not sufficient) for convergence almost everywhere. The following two subsections will discuss the convergence of Fourier series based on integrability conditions. We first discuss L^1 functions, and then L^p functions for $1 < p < \infty$.

7.1. **The Carleson Maximal Operator, and the Failure of Convergence of L^1 Functions.** This subsection discusses the failure of convergence of Fourier series for general L^1 functions. We introduce the Carleson maximal operator which is the relevant maximal function for Fourier series.

We define the *Carleson maximal operator* by

$$M_{\mathrm{C}} f := \sup_{n \in \mathbb{N}} |S_n[f]| \ , \tag{7.1}$$

where $S_n[f]$ is the nth symmetric partial sum of the Fourier series of f.

Remark 7.1. The Carleson maximal operator is closely related to the convergence of L^p functions in general, as we will see in the following subsection. We would like to use Theorem 2.4 to prove almost-everywhere convergence, since S_n are linear operators, and the $\mathcal{C}^2$ functions are dense in L^p. So, we want to show that M_{C} satisfies the C–δ hypothesis.

Recall from our discussion in Section 2 that finiteness of our maximal operator is often the key condition underlying the C–δ hypothesis. Maximal inequalities imply a.e. finiteness, but in the case of L^1 functions and the Carleson operator, we even have a stronger reverse implication:

Theorem 7.2. *Let λ represent the (normalized) Lebesgue measure on $\mathbb{T}$. If, for any $f \in L^1(\mathbb{T})$*

$$\lambda(M_C f < \infty) > 0 \ , \tag{7.2}$$

then, there exists a constant $K > 0$ such that

$$\lambda(M_C f > C) \leq \frac{K \|f\|_1}{C} \ . \tag{7.3}$$

That is, if the Carleson maximal operator is finite *somewhere*, then it satisfies a maximal inequality. If this is true, this is enough to use Theorem 2.4 and prove almost-everywhere convergence for L^1 functions. Unfortunately, we have the following result.

Theorem 7.3. *There exists $f \in L^1(\mathbb{T})$ such that*

$$\lambda(M_C f < \infty) = 0 \ . \tag{7.4}$$

This means that we cannot use our maximal functions framework to prove almost-everywhere convergence of L^1 functions. Moreover, the proof of Theorem 7.3 shows that there exists $f \in L^1$ whose Fourier series diverges almost everywhere; see [7, Chapter 6] for details.

7.2. **Carleson's Theorem.** This section discusses the Carleson–Hunt theorem, which proves that the Fourier series of any L^p function, $1 < p < \infty$, converges.

Throughout this paper, we have seen that the C–δ hypothesis is very often where much of the complexity of proving almost-everywhere convergence arises under our maximal functions framework. The developments in Section 2 on maximal inequalities served as our core method of proving the C–δ hypothesis throughout Sections 3, 4 and 5, leading us to prove the Hardy–Littlewood, convolution, and ergodic maximal inequalities (Theorems 3.1, 4.3, and 5.18 respectively). However, the failure of L^1 convergence suggests that a direct maximal inequalities approach may not suffice for proving the convergence of Fourier series for L^p functions.

Thus, we turn to the developments of Section 6. If the Carleson maximal operator is a bounded operator into L^p_w for $p > 1$, then Theorem 6.6 tells us that M_{C} satisfies the C–δ hypothesis. Combining this with the discussion in Remark 7.1 would prove that the Fourier series of L^p functions converge almost everywhere.

Fortunately, we have the following theorem.

Theorem 7.4. *M_{C} is a bounded operator from L^p to L^p_w for $1 < p < \infty$.*

Carleson proved Theorem 7.4 for the case of $p = 2$ [3], which Hunt later extended to the general case [4]. We omit both proofs due to both space and complexity. Details can be found in [8, Chapter 7]. Assuming Theorem 7.4, we conclude this paper by stating and proving the Carleson–Hunt theorem.

Theorem 7.5 (Carleson–Hunt)**.** *Let $f \in L^p(\mathbb{T})$, $1 < p < \infty$. Then,*

$$S_n[f] \to f \ , \ \textit{Lebesgue a.e.} \tag{7.5}$$

Proof. Recall the operators $S_n[f]$. By the linearity of integrals and sums, these operators are linear. The maximal operator for these operators is M_{C}, which is measurable as we are indexing over a countable collection. Since $S_n[f] \to f$ almost everywhere for $f \in \mathcal{C}^2$, we have convergence on a dense set. Theorem 7.4 tells us that M_{C} is a bounded operator into weak-L^p, and so Theorem 6.6 applies. Since we have linear operators, a dense set, and a maximal inequality, Theorem 2.4 applies and we have convergence for all $f \in L^p$. Furthermore, if f_k is a sequence of $\mathcal{C}^2$ functions which converge to f in the L^p sense, and therefore in measure, since $S_n[f_k] \to f_k$, Theorem 2.7 applies, and we know that $S_n[f] \to f$ almost everywhere for every $f \in L^p$. $\square$

ACKNOWLEDGMENT

I want to thank my advisor, Chris Marx, for a multitude of reasons. I am overwhelmingly thankful for the encouragement and support he has provided me over the course of this thesis: from first suggesting the research topic, to the final proofreading of the paper, he has devoted extensive time towards helping me make this paper the best it can be. I have been very lucky to have Chris not only as a research advisor, but also as an academic advisor during my undergraduate years. He is responsible for turning me from "a person who does math", into a mathematician, and I could not be more grateful.

REFERENCES

[1] David Anderson, Timo Seppäläinen, and Benedek Valkó. *Introduction to Probability.* Cambridge University Press, 2018.
[2] Patrick Billingsley. *Probability and Measure, 3rd Ed.* John Wiley Sons, 1995.
[3] Lennart Carleson. On convergence and growth of partial sums of Fourier series. *Acta Mathematica*, 116(none):135 – 157, 1966.
[4] Richard A. Hunt. On the convergence of Fourier series. In *Orthogonal Expansions and their Continuous Analogues (Proc. Conf., Edwardsville, Ill., 1967)*, pages 235–255. Southern Illinois Univ. Press, Carbondale, IL, 1968.
[5] Jean Kahane and Yitzhak Katznelson. Sur les ensembles de divergence des séries trigonométriques. *Studia Mathematica*, 26(3):307–313, 1966.
[6] Yitzhak Katznelson. *An Introduction to Harmonic Analysis, 3rd Ed.* Cambridge University Press, 2003.
[7] Camil Muscalu and Wilhelm Schlag. *Classical and Multilinear Harmonic Analysis, Volume I.* Cambridge University Press, 2013.
[8] Camil Muscalu and Wilhelm Schlag. *Classical and Multilinear Harmonic Analysis, Volume II.* Cambridge University Press, 2013.
[9] Barry Simon. *Harmonic Analysis, A Comprehensive Course in Analysis, Part 3.* Americal Mathematical Society, 2015.
[10] Barry Simon. *Real Analysis, A Comprehensive Course in Analysis, Part 1.* Americal Mathematical Society, 2015.

(S. Thiel) DEPARTMENT OF MATHEMATICS, OBERLIN COLLEGE, OBERLIN, OH 44074
Email address: sthiel@oberlin.edu