MAXIMAL FUNCTION THEORY

MARK H. KIM

ABSTRACT. We study the basic theory of the Hardy-Littlewood maximal function and its variants. Three oft-used decomposition techniques—dyadic decomposition, truncation, and layercake decomposition—are introduced in the context of proving various maximal estimates. We also give applications to pointwise-convergence phenomena in various context, as well as a brief survey of open questions surrounding the maximal function conjectures.

CONTENTS

1.	Motivation: The Lebesgue Differentiation Theorem	1
2.	Weak-type (1,1) Bound of the Hardy-Littlewood Maximal Function	4
3.	Maximal Functions and Pointwise Convergence	7
3.1.	Approximations to the identity revisited	8
3.2.	The strong law of large numbers; martingales	12
3.3.	Birkhoff's ergodic theorem	16
3.4.	Stein's maximal principle	21
4.	An Interpolation Result	21
5.	Further Results: The Maximal Function Conjectures	24
	References	29

Notational remark. Given a metric space (X, d_X), the closed ball of radius $r > 0$ centered at $x \in X$ is denoted by $B(x; r)$.

1. MOTIVATION: THE LEBESGUE DIFFERENTIATION THEOREM

Let $f : [a, b] \to \mathbb{R}$ be continuous. The *fundamental theorem of calculus* states that

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

is differentiable in (a, b) and $F'(x) = f(x)$ for all $x \in (a, b)$. Indeed, the continuity of f furnishes, for each $\varepsilon > 0$ and every $x \in (a, b)$, a $\delta > 0$ such that $|r| < \delta$ implies $|f(x + r) - f(x)| < \varepsilon$. Therefore,

$$f(x) - \varepsilon = \frac{1}{r} \cdot (f(x) - \varepsilon) \cdot h \leq \frac{F(x + r) - F(x)}{r} \leq \frac{1}{r} \cdot (f(x) + \varepsilon) \cdot h = f(x) + \varepsilon$$

for all $|r| < \delta$, and so

$$f(x) - \varepsilon \leq \lim_{h \to 0} \frac{F(x + r) - F(x)}{r} \leq f(x) + \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we conclude that F is differentiable and $F'(x) = f(x)$.

Could we generalize this result to a larger class of functions? Note that a simple modification of the above argument yields the identity

$$\lim_{r\to 0} \frac{1}{2r}\int_{x-r}^{x+r} f(y)\,dy = f(x) \tag{1.1}$$

for all continuous functions f. In this form, the fundamental theorem of calculus is a statement about the behavior of the integral mean value

$$(\mathcal{A}_r f)(x) = \frac{1}{2r}\int_{x-r}^{x+r} f(y)\,dy \tag{1.2}$$

as the length of the interval $(x-r, x+r)$ centered at x decreases to 0. Since $(\mathcal{A}_r f)(x)$ is well-defined for all $f \in L^1([x-r, x+r])$, it makes sense to try and prove (1.1) for all $f \in L^1_{\mathrm{loc}}(\mathbb{R})$, the space of measurable functions on $\mathbb{R}$ whose integral is finite on each compact subset of $\mathbb{R}$.

We could also consider d-dimensional generalizations of (1.1). To this end, we must determine what we wish to take as n-dimensional generalizations of intervals. We abstract three properties of intervals: compactness, convexity, and symmetry.

Definition 1.3. A nonempty set $B \subseteq \mathbb{R}^d$ is *convex* if, for each pair of points x_1 and x_2 in B, the *convex combination* $(1-\lambda)x_1 + \lambda x_2$ is in B for all $0 \leq \lambda \leq 1$.

Definition 1.4. A nonempty set $B \subseteq \mathbb{R}^d$ is *centrally symmetric* with respect to $p \in B$ if B is invariant under the affine transform $x \mapsto 2p - x$. This is equivalent to saying that $p + h \in B$ if and only if $p - h \in B$.

From here on, we shall write *centrally symmetric convex body* to refer to a subset of $\mathbb{R}^d$ that is compact, convex, and centrally symmetric whose center p is the origin.

Since we can rewrite (1.2) as

$$(\mathcal{A}_r f)(x) = \frac{1}{2r}\int_{-r}^{r} f(x+y)\,dy$$

it suffices to consider centrally symmetric convex bodies with respect to the origin in asking the following question:

Question 1.5. *Given a centrally symmetric convex body $B \subseteq \mathbb{R}^d$, does the integral mean value*

$$(\mathcal{A}_{rB} f)(x) = \frac{1}{m(rB)}\int_{rB} f(x+y)\,dy$$

of $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ converge pointwise to $f(x)$ as $r \to 0$? Here we have used rB to denote the scaled set $rB = \{ry : y \in B\}$.

Whenever f is continuous, the argument for the one-dimensional fundamental theorem of calculus can be applied with minor modification to answer Question 1.5 in the affirmative. If f is merely in $L^1(\mathbb{R})$, then, for each $\varepsilon > 0$, we can find a $g \in \mathscr{C}_c(\mathbb{R}^d)$ such that $\|f - g\|_1 < \varepsilon$. We can then rewrite $(A_{rB} f)(x) - f(x)$ as

$$\mathcal{A}_{rB}(f-g)(x) + (\mathcal{A}_{rB} g)(x) - g(x) + g(x) - f(x).$$

By continuity, we have $(\mathcal{A}_{rB} g)(x) \to g(x)$ as $r \to 0$, whence we have the estimate

$$\begin{aligned}
\limsup_{r\to 0} |(\mathcal{A}_{rB} f)(x) - f(x)| &\leq \limsup_{r\to 0} |\mathcal{A}_{rB}(f-g)(x)| + \lim_{r\to 0} |(\mathcal{A}_{rB} g)(x) - g(x)| \\
&\quad + |f(x) - g(x)| \\
&= \limsup_{r\to 0} \mathcal{A}_{rB}(|f-g|)(x) + |f(x) - g(x)|.
\end{aligned}$$

Therefore, the study of the integral mean values of f at $x \in \mathbb{R}^d$ depends crucially on the quantity

$$\limsup_{r\to 0} \mathcal{A}_{rB}(|f-g|)(x)$$

which is bounded from above by the "maximal function"

$$\sup_{r>0} \mathcal{A}_{rB}(|f-g|)(x).$$

This motivates us to introduce our main object of study:

Definition 1.6. The *Hardy-Littlewood maximal function* of $h \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ over a centrally symmetric convex body $B \subseteq \mathbb{R}^d$ is

$$(\mathcal{M}_B f)(x) = \sup_{r>0} (\mathcal{A}_{rB}|f|)(x) = \sup_{r>0} \frac{1}{m(rB)} \int_{rB} |f(x+y)|\, dy.$$

Let us return to the problem at hand. What we have discussed so far assures us that if $f \in L^1(\mathbb{R}^d)$, then, for each $\varepsilon > 0$, there exists a $g \in L^1(\mathbb{R}^d)$ that yields the estimate

$$\limsup_{r\to 0} |(\mathcal{A}_{rB} f)(x) - f(x)| \leq \mathcal{M}_B(f-g)(x) + |f(x) - g(x)|. \tag{1.7}$$

Ideally, we would be able to show that

$$\mathcal{M}_B(f-g)(x) + |f(x)-g(x)| < C\varepsilon$$

for some constant C independent of ε, thus proving that $\mathcal{A}_{rB} f \to f$ pointwise *everywhere*. But this is too much to hope for, as

$$\left(\mathcal{A}_{(-r,r)}\chi_{(0,1)}\right)(0) = \frac{1}{2r}\int_0^r dy = \frac{1}{2}$$

for all $0 < r < 1$, which does not converge to $\chi_{(0,1)}(0) = 0$ as $r \to 0$.

So then, if we hope to obtain an affirmative answer to Question 1.5, then we must settle for an almost-everywhere statement, which is equivalent to the statement that the set

$$\left\{ x \in \mathbb{R}^d : \limsup_{r\to 0} |(\mathcal{A}_{rB} f)(x) - f(x)| > 0 \right\}$$

is of Lebesgue measure zero. Since the above set is the intersection of the sets

$$E_k = \left\{ x \in \mathbb{R}^d : \limsup |(\mathcal{A}_{rB} f)(x) - f(x)| > \frac{1}{k} \right\}, \tag{1.8}$$

it suffices to show that $m(E_k) = 0$ for all $k \in \mathbb{N}$.

Now, estimate (1.7) implies that

$$\begin{aligned} m(E_k) \quad &\leq \quad m\left(\left\{x : \mathcal{M}_B(f-g)(x) > \frac{1}{2k}\right\}\right) \\ &\quad + m\left(\left\{x : |f(x)-g(x)| > \frac{1}{2k}\right\}\right). \end{aligned} \tag{1.9}$$

Observe that[1]

$$m\left(\left\{x : |f(x)-g(x)| > \frac{1}{2k}\right\}\right) = \int_{\left\{x:|f(x)-g(x)|>\frac{1}{2k}\right\}} 1\,dy$$
$$\leq \int_{\left\{x:|f(x)-g(x)|>\frac{1}{2k}\right\}} \frac{|f(y)-g(y)|}{1/2k}\,dy \leq \int_{\mathbb{R}^d} \frac{|f(y)-g(y)|}{1/2k}\,dy$$
$$= 2k\|f-g\|_1.$$

It is therefore natural to hope for a bound of the form

$$m\left(\left\{x : \mathcal{M}_B(f-g)(x) > \frac{1}{2k}\right\}\right) \leq 2kA\|f-g\|_1 \tag{1.10}$$

for some constant A independent of $(f-g)$, so that (1.9) can be written as

$$m(E_k) \leq 2k(A+1)\|f-g\|_1 < 2k(A+1)\varepsilon.$$

Since ε was arbitrary, we can then conclude that $m(E_k) = 0$. Assuming that (1.10) holds true, we now have the proof of the following result for the L^1 case.

Theorem 1.11 (Lebesgue differentiation theorem)**.** *If $f \in L^1_{\text{loc}}(\mathbb{R}^d)$, and if $B \subseteq \mathbb{R}^d$ is a centrally symmetric convex body, then*

$$\lim_{r\to 0}(\mathcal{A}_{rB}f)(x) = f(x)$$

for almost every $x \in \mathbb{R}^d$.

The general case of the theorem follows easily from the L^1 case by considering the "compact cutoff" $f\chi_{\overline{B(0;k)}}$ of $f \in L^1_{\text{loc}}(\mathbb{R}^d)$ with respect to closed balls $B(0;k)$ of radius $k \in \mathbb{N}$ centered at the origin. Indeed, each compact cutoff $f\chi_{\overline{B(0;k)}}$ is in L^1, and so the Lebesgue differentiation theorem for L^1 functions furnishes a measure-zero set

$$\begin{aligned} F_k &= \left\{x \in B(0;k) : \lim_{r\to 0}(\mathcal{A}_{rB}f\chi_{\overline{B(0;k)}})(x) = f\chi_{\overline{B(0;k)}}(x)\right\} \\ &= \left\{x \in B(0;k) : \lim_{r\to 0}(\mathcal{A}_{rB}f)(x) = f(x)\right\}. \end{aligned}$$

Since

$$\bigcup_{k=1}^{\infty} F_k = \left\{x \in \mathbb{R}^d : \lim_{r\to 0}(\mathcal{A}_{rB}f)(x) = f(x)\right\},$$

it follows that $\mathcal{A}_{rB}f \to f$ almost everywhere, thereby establishing the theorem for locally integrable functions.

2. Weak-type (1,1) Bound of the Hardy-Littlewood Maximal Function

In order to complete the proof of Theorem 1.5, we must establish (1.10)—and we do so now.

Theorem 2.1 (Weak-type $(1,1)$ bound on the maximal function)**.** *If $B \subseteq \mathbb{R}^d$ is a a centrally symmetric convex body, then there exists a constant $A_{d,1,B}$ such that*

$$m\left(\left\{x \in \mathbb{R}^d : \mathcal{M}_Bf(x) > \alpha\right\}\right) \leq \frac{A_{d,1,B}}{\alpha}\|f\|_1$$

[1] This is Chebyshev's inequality.

for each $\alpha > 0$ and every $f \in L^1(\mathbb{R}^d)$. The constant $A_{d,1,B}$ depends only on the dimension d and the body B.

Proof. For the sake of simplicity, let us first consider the Euclidean unit ball $B = B_2 = B(0;1) = \{x \in \mathbb{R}^d : |x| < 1\}$. We let

$$E_\alpha = \left\{x \in \mathbb{R}^d : \mathcal{M}_{B_2} f(x) > \alpha\right\}$$

for each $\alpha > 0$ and note that each $x \in E_\alpha$ furnishes an $r_{x,\alpha} > 0$ such that

$$\frac{1}{m(r_{x,\alpha}B_2)} \int_{r_{x,\alpha}B_2} |f(x+y)|\, dy > \alpha.$$

Setting $B_{x,\alpha} = r_{x,\alpha}B_2 + x = \{y + x : y \in r_{x,\alpha}B_2\}$, we see that

$$\int_{B_{x,\alpha}} |f(y)|\, dy = \int_{r_{x,\alpha}B_2} |f(x+y)|\, dy > \alpha\, m(r_{x,\alpha}B_2) = \alpha\, m(B_{x,\alpha}) \tag{2.2}$$

by the translation invariance of the Lebesgue measure.

Since $B_{x,\alpha}$ is a Euclidean ball centered at x, we see that $E_\alpha \subseteq \bigcup_{x \in E_\alpha} B_{x,\alpha}$. This, in particular, yields the estimate.

$$m(E_\alpha) \le m\left(\bigcup_{x \in E_\alpha} B_{x,\alpha}\right). \tag{2.3}$$

A naïve application of (2.3) to (2.2) yields the estimate

$$m(E_\alpha) \le m\left(\bigcup_{x \in E_\alpha} B_{x,\alpha}\right) \le \sum_{x \in E_\alpha} m(B_{x,\alpha}) \le \frac{1}{\alpha} \sum_{x \in E_\alpha} \int_{B_{x,\alpha}} |f(y)|\, dy,$$

but it is not entirely clear how to relate the rightmost quantity in the above display to $\|f\|_1$. If $\mathscr{B}_\alpha = \{x \in E_\alpha : B_{x,\alpha}\}$ were a pairwise-disjoint collection, then we would have the further estimate

$$\frac{1}{\alpha} \sum_{x \in E_\alpha} \int_{r_{x,\alpha}B_2} |f(y)|\, dy = \frac{1}{\alpha} \int_{\bigcup_{x \in E_\alpha} B_{x,\alpha}} |f(y)|\, dy \le \frac{1}{\alpha} \|f\|_1.$$

Nevertheless, the collection $\mathscr{B}_\alpha$ is almost never pairwise disjoint, and so we seek a way to "disjointify" $\mathscr{B}_\alpha$. This is the content of the following lemma.

Lemma 2.4 (Infinitary Vitali covering lemma)**.** *If $\{B(x_\beta, r_\beta)\}_\beta$ is a collection of Euclidean balls in $\mathbb{R}^d$ whose radii are uniformly bounded, then there exists a pairwise-disjoint countable subcollection $\{B(x_n, r_n)\}_n$ such that*

$$\bigcup_\beta B(x_\beta, r_\beta) \subseteq \bigcup_n B(x_n, 5r_n).$$

Proof of lemma. Let $\mathscr{B} = \{B(x_\beta, r_\beta)\}_\beta$. We let $B(x_1, r_1)$ be a ball in $\mathscr{B}$ such that

$$r_1 > \frac{1}{2} \sup_\beta r_\beta,$$

where the supremum is taken over all possible radii of balls in $\mathscr{B}$. Define $\mathscr{B}_1$ to be the collection of all balls in $\mathscr{B}$ that intersect $B(x_1, r_1)$ nontrivially. The condition on r_1 implies that $5r_1 > r_1 + 2\sup_\beta r_\beta$, and so $B(x_1, 5r_1)$ includes every ball in $\mathscr{B}_1$.

For each $n > 1$, we assume inductively that the balls $B(x_1, r_1), \ldots, B(x_{n-1}, r_{n-1})$ and the collections $\mathscr{B}_1, \ldots, \mathscr{B}_{n-1}$ have been defined. If $\mathscr{B} \smallsetminus \mathscr{B}_{n-1}$ is empty, then we terminate the process. If not, then we let $B(x_n, r_n)$ be a ball in $\mathscr{B} \smallsetminus \mathscr{B}_{n-1}$ such that

$$r_n > \frac{1}{2} \sup_\beta r_\beta,$$

where the supremum is taken over all possible radii of balls in $\mathscr{B} \smallsetminus \mathscr{B}_{n-1}$. Define $\mathscr{B}'_n$ to be the collection of all balls in $\mathscr{B}$ that intersect $B(x_n, r_n)$ nontrivially. Similarly as above, $B(x_n, 5r_n)$ contains every ball in $\mathscr{B}'_n$. We set $\mathscr{B}_n = \mathscr{B}'_n \cup \mathscr{B}_{n-1}$.

By construction, $B(x_n, r_n)$ is disjoint from all balls in $\mathscr{B}_{n-1}$. In particular, for a fixed N, the collection $\{B(x_1, r_1), \ldots, B(x_N, r_N)\}$ is pairwise-disjoint. Moreover, for all $1 \leq n \leq N$, the ball $B(x_n, 5r_n)$ contains all balls in $\mathscr{B}_n$, and so

$$\bigcup_{B \in \mathscr{B}_N} B \subseteq \bigcup_{n=1}^{N} B(x_n, 5r_n). \tag{2.5}$$

We now recall that $\mathbb{R}^d$ cannot contain uncountably many disjoint Euclidean balls: indeed, each such ball intersects $\mathbb{Q}^n$, which is countable. This, in particular, implies that the construction described above must terminate in at most countably many steps[2]. It therefore makes sense to index $\mathscr{B}_n$ and $B(x_n, r_n)$ by natural numbers. Carrying out the construction until it terminates, we see that (2.5) yields the inclusion relation

$$\bigcup_{B \in \bigcup_n \mathscr{B}_n} B \subseteq \bigcup_n B(x_n, 5r_n). \tag{2.6}$$

It therefore suffices to show that $\mathscr{B} = \bigcup_n \mathscr{B}_n$. If not, then every element of $\mathscr{B} \smallsetminus (\bigcup_n \mathscr{B}_n)$ is disjoint from every ball in the collection $\{B(x_n, r_n)\}_n$. Therefore, the above construction can be carried out one more time, contradicting the assumption that the construction has already terminated. It follows that $\mathscr{B} = \bigcup_n \mathscr{B}_n$, and (2.6) yields the desired inclusion relation. □

We now return to the proof of the weak-type bound on the maximal function for the Euclidean ball B_2. We apply the Vitali covering lemma to the collection $\{B_{x,\alpha} : x \in E_\alpha\}$ constructed prior to the displayed identity (2.2) to obtain a pairwise-disjoint subcollection $\{B_{x_n,\alpha}\}$ such that

$$\bigcup_{x \in E_\alpha} B_{x,\alpha} \subseteq \bigcup_n B^*_{x_n,\alpha},$$

where $B^*_{x_n,\alpha}$ is the Euclidean ball obtained by enlarging the radius of $B_{x_n,\alpha}$ by 5. By the scaling properties of the Lebesgue measure, we have

$$m(B^*_n) = 5^d\, m(B_{x_n,\alpha}), \tag{2.7}$$

whence

$$m\left(\bigcup B^*_{x_n,\alpha}\right) = \sum_n m(B^*_{x_n,\alpha}) = 5^d \sum_n m(B_{x_n,\alpha}).$$

(2.3) now yields the estimate

$$m(E_\alpha) \leq m\left(\bigcup_{x \in E_\alpha} B_{x,\alpha}\right) \leq m\left(\bigcup_n B^*_n\right) = 5^d \sum_n m(B_n),$$

[2]This does *not* require transfinite induction, as we can establish this restriction *before* we induct.

to which we apply (2.2) to conclude that

$$m(E_\alpha) \leq 5^d \sum_n m(B_n) \leq \frac{5^d}{\alpha} \sum_n \int_{B_{x,\alpha}} |f(y)|\, dy.$$

By disjointness, it follows that

$$\frac{5^d}{\alpha} \sum_n \int_{B_{x,\alpha}} |f(y)|\, dy = \frac{5^d}{\alpha} \int_{\bigcup_n B_{x,\alpha}} |f(y)|\, dy \leq \frac{5^d}{\alpha} \|f\|_1,$$

whence we have the estimate

$$m(E_\alpha) \leq \frac{5^d}{\alpha} \|f\|_1$$

for all $\alpha > 0$, as was to be shown.

We now assume that B is an arbitrary centrally symmetric convex body in $\mathbb{R}^d$. By the boundedness of B, we can find a constant $c > 0$ such that $B \subseteq cB_2$. We also find another constant $c_B > 0$ such that $c_B m(B) = m(cB_2)$, so that

$$m(rB) = r^d\, m(B) = c_B^{-1} r^d\, m(cB_2) = c_B^{-1} m(c(rB_2))$$

for all $r > 0$. Then we have the estimate

$$\begin{aligned} \frac{1}{m(rB)} \int_{rB} |f(x+y)|\, dy &\leq \frac{1}{m(rB)} \int_{r(cB_2)} |f(x+y)|\, dy \\ &= \frac{1}{c_B^{-1}\, m(rcB_2)} \int_{r(cB_2)} |f(x+y)|\, dy \end{aligned}$$

for all $r > 0$, which implies that

$$\{x \in \mathbb{R}^d : \mathcal{M}_B f(x) > \alpha\} \subseteq \{x \in \mathbb{R}^d : \mathcal{M}_{B_2} f(x) > c_B^{-1}\alpha\}.$$

It now follows from the weak-type bound on $\mathcal{M}_{B_2}$ that

$$m\left(\{x \in \mathbb{R}^d : \mathcal{M}_B f(x) > \alpha\}\right) \leq m\left(\{x \in \mathbb{R}^d : \mathcal{M}_{B_2} f(x) > c_B^{-1}\alpha\}\right) \leq \frac{c_B 5^d}{\alpha} \|f\|_1,$$

which establishes the weak-type bound on $\mathcal{M}_B$. □

3. Maximal Functions and Pointwise Convergence

The strategy of proof of the Lebesgue differentiation theorem (Theorem 1.5) is twofold:

(1) we established the pointwise convergence result $\mathcal{A}_{rB} f \to f$ for a dense subclass ($\mathscr{C}_c(\mathbb{R}^d)$) of the function space in question ($L^1(\mathbb{R}^d)$);

(2) we then established the weak-type bound

$$m\left(\left\{x \in \mathbb{R}^d : \sup_{r>0} \mathcal{A}_{rB}(|f|) > \alpha\right\}\right) \leq \frac{A_{d,1,B}}{\alpha} \|f\|_1$$

on the maximal function $\mathcal{M}_B f = \sup_{r>0} \mathcal{A}_{rB}(|f|)$, which allowed us to upgrade the result on the dense subclass to the whole function space.

The maximal-estimate approach to establishing pointwise convergence has proven to be extremely useful. In what follows, we present a few examples that illustrate the wide range of applicability of the maximal-estimate approach.

3.1. **Approximations to the identity revisited.** We recall that *approximations to the identity* are a collection of integrable functions $(\rho_r)_{r>0}$ on $\mathbb{R}^d$ such that $\rho \in L^1(\mathbb{R}^d)$, $\int \rho = 1$, and $\rho_r(x) = r^{-d}\rho(r^{-1}x)$. Given a centrally symmetric convex body $B \subseteq \mathbb{R}^d$, we set $\rho(x) = \frac{1}{m(B)}\chi_B(x)$ to see that

$$\begin{aligned}\mathcal{A}_{rB}f(x) &= \int_{\mathbb{R}^d} f(x+y)\frac{\chi_B(y)}{m(rB)}\,dy = \int_{\mathbb{R}^d} f(x-y)\frac{\chi_B(y)}{r^d m(B)}\,dy \\ &= \left(f * \frac{1}{r^d m(B)}\chi_B\right)(x) = (f*\rho_r)(x)\end{aligned}$$

for all $r > 0$. Since $\int \rho = 1$, it follows that the kernels $(\rho_r)_{r>0}$ form approximations to the identity.

As per the definition of the Hardy-Littlewood maximal function, we have the maximal bound

$$\sup_{r>0} |(f*\rho_r)(x)| \le C\,\mathcal{M}f, \tag{3.1}$$

with constant $C = \frac{m(\widetilde{B_2})}{m(B)}$, where $\widetilde{B_2}$ is some dilate of the Euclidean ball B_2 that contains the convex body B.

Let us run through the proof of the Lebesgue differentiation theorem one more time—this time, using estimate (3.1).

Proof. We have seen that $f*\rho_r \to f$ pointwise almost everywhere if and only if

$$m\left(\left\{x \in \mathbb{R}^d : |(f*\rho_r)(x) - f(x)| > \frac{1}{k}\right\}\right) \tag{3.2}$$

is zero for all $k \in \mathbb{N}$. We know that

$$\lim_{r\to 0}(g*\rho_r)(x) = g(x) \tag{3.3}$$

for all $x \in \mathbb{R}^d$ whenever $g \in \mathscr{C}_c(\mathbb{R}^d)$. We fix $\varepsilon > 0$ and invoke the density of $\mathscr{C}_c(\mathbb{R}^d)$ in $L^1(\mathbb{R}^d)$ to find $g \in \mathscr{C}_c(\mathbb{R}^d)$ such that $\|f-g\|_1 < \varepsilon$. Let $h = f - g$; (3.3) implies that

$$(3.2) \le m\left(\left\{x : \sup_{r\to 0}\,(h*\rho_r)\,(x) > \frac{1}{2k}\right\}\right) + m\left(\left\{x : |h(x)| > \frac{1}{2k}\right\}\right). \tag{3.4}$$

The maximal bound (3.1) implies that

$$(3.4) \le m\left(\left\{x : C\,\mathcal{M}h(x) > \frac{1}{2k}\right\}\right) + m\left(\left\{x : |h(x)| > \frac{1}{2k}\right\}\right), \tag{3.5}$$

and it now follows from the weak-type (1,1) bound of the maximal function and Chebyshev's inequality that

$$(3.5) \le 2kCA_{1,d,B_2}\|h\|_1 + 2k\|h\|_1 < (2kCA_{1,d,B_2}+1)\varepsilon. \tag{3.6}$$

The pointwise almost-everywhere convergence result is now a direct consequence of (3.6). $\square$

Observe that the above proof applies *verbatim* to any family of approximations to the identity $(\rho_r)_{r>0}$, provided that we have the maximal bound (3.1) and the convergence result (3.3). We first show that the convergence hypothesis is superfluous:

Lemma 3.7. *If $(\rho_r)_{r>0}$ is a family of approximations to the identity on $\mathbb{R}^d$, then, for each $g \in \mathscr{C}_c(\mathbb{R}^d)$,*

$$\lim_{\varepsilon \to 0} (g * \rho_r)(x) = g(x)$$

uniformly on $\mathbb{R}^d$.

Proof. We begin by establishing three properties of $(\rho_r)_{r>0}$. Firstly, we observe that

$$\int \rho_r(y)\, dy = \int \varepsilon^{-d} \rho(\varepsilon^{-1} y)\, dy = \int \rho(y)\, dy = 1 \tag{3.8}$$

for all $r > 0$. Secondly, similar computations yield the uniform bound

$$\sup_{r>0} \int |\rho_r(y)|\, dy = \sup_{r>0} \int |\rho(y)|\, dy = \|\rho\|_1 < \infty. \tag{3.9}$$

Lastly, for each $\varepsilon > 0$, we can find $M_\varepsilon > 0$ such that

$$\int_{|y|>M_\varepsilon} |\rho(y)|\, dy < \varepsilon,$$

Then, given $\delta > 0$, we observe that

$$\int_{|y|>\delta} |\rho_r(y)|\, dy < \varepsilon r^d \leq \varepsilon$$

whenever $r < \min\{\delta M_\varepsilon^{-1}, 1\}$, whence

$$\lim_{r \to 0} \int_{|y|>\delta} |\rho_r(y)|\, dy = 0. \tag{3.10}$$

We are now ready to establish the lemma. (3.8) implies that

$$\begin{aligned} |(g * \rho_r)(x) - g(x)| &= \left| \int g(x-y)\rho_r(y)\, dy - g(x) \int \rho_r(y)\, dy \right| \\ &\leq \int |g(x-y) - g(x)||\rho_r(y)|\, dy. \end{aligned}$$

Fix $\varepsilon > 0$. By continuity of g, we can find $\delta > 0$ such that $|g(x-y) - g(x)| \leq \varepsilon$ for all $|y| \leq \delta$. Combining this with the above estimate, we obtain the upper bound

$$|(g * \rho_r)(x) - g(x)| \leq \int_{|y|\leq\delta} \varepsilon |\rho_r(y)|\, dy + \int_{|y|>\delta} |g(x-y) - g(x)||\rho_r(y)|\, dy.$$

(3.9) implies that

$$\int_{|y|\leq\delta} \varepsilon |\rho_r(y)|\, dy + \int_{|y|>\delta} \leq \varepsilon \int |\rho_r(y)|\, dy = \varepsilon \|\rho\|_1,$$

and (3.10) implies that

$$\int_{|y|>\delta} |g(x-y) - g(x)||\rho_r(y)|\, dy \leq 2\|g\|_\infty \int_{|y|>\delta} |\rho_r(y)|\, dy \to 0$$

as $r \to 0$. The desired result now follows. ◻

Evidently, Lemma 3.7 holds for any family of functions $(\rho_r)_{r>0}$ that satisfies (3.8), (3.9), and (3.10). This leads us to make the following definition:

Definition 3.11. A family of functions $(\rho_r)_{r>0}$ on $\mathbb{R}^d$ is a family of *generalized*[3] *approximations to the identity on* $\mathbb{R}^d$ if

(1) $\int_{\mathbb{R}^d} \rho_r = 1$ for all $r > 0$,
(2) $\sup_{r>0} \|\rho_r\|_1 < \infty$, and
(3) $\lim_{r\to 0} \int_{|y|>\delta} |\rho_r(y)|\, dy$ for each fixed $\delta > 0$.

Lemma 3.7 now implies that the proof of the Lebesgue differentiation theorem goes through for every family of generalized approximations to the identity that satisfies the maximal bound (3.1). In other words, we have established the following theorem:

Theorem 3.12. *If* $(\rho_r)_{r>0}$ *is a family of generalized approximations to the identity that satisfies the maximal bound*

$$\sup_{r>0} |(f * \rho_r)(x)| \leq C\, \mathcal{M}f$$

for some universal constant $C > 0$ *that does not depend on* $f \in L^1(\mathbb{R}^d)$, *then* $f * \rho_r \to f$ *pointwise almost-everywhere as* $r \to 0$.

We defer the study of important examples of approximations to the identity satisfying the maximal bound to a later set of notes, for their proper context is the study of Fourier series and the Fourier transform. Here in this subsection, we content ourselves by furnishing a sufficient condition for the maximal bound on approximations to the identity.

Definition 3.13. We say that a family $(\rho_r)_{r>0}$ of generalized approximations to the identity on $\mathbb{R}^d$ is *radially bounded* if $(\rho_r)_{r>0}$ is in $L^\infty(\mathbb{R}^d)$ and, for each $r > 0$, the function

$$\phi_r(x) = \sup_{|y|\geq|x|} |\rho_r(y)|$$

is integrable on $\mathbb{R}^d$ and $\sup_{r>0} \|\phi_r\|_1 < \infty$.

Theorem 3.14. *If a family* $(\rho_r)_{r>0}$ *of generalized approximations to the identity on* $\mathbb{R}^d$ *is radially bounded, then*

$$\sup_{r>0} |(f * \rho_r)(x)| \leq \sup_{r>0} \|\phi_r\|_1 \mathcal{M}f(x)$$

for each $f \in L^1(\mathbb{R}^d)$ *and every* $x \in \mathbb{R}^d$.

Proof. Fix $r > 0$. Since $\phi_r(y_1) \geq \phi_r(y_2)$ whenever $|y_1| \leq |y_2|$, we see that

$$\int_{|x|/2\leq|y|\leq|x|} \phi_r(y)\, dy \geq \int_{|x|/2\leq|y|\leq|x|} \phi_r(x)\, dy = c|x|^d \phi_r(x)$$

for each $x \in \mathbb{R}^d$, where $c > 0$ is a constant independent of x. Since $\phi_r \in L^1(\mathbb{R}^d)$, the above inequality implies that

$$\lim_{|x|\to 0} |x|^d \phi_r(x) = 0 \quad \text{and} \quad \lim_{|x|\to\infty} |x|^d \phi_r(x) = 0. \tag{3.15}$$

We now claim that

$$|(f * \phi_r)(0)| \leq \|\phi_r\|_1 \mathcal{M}f(0) \tag{3.16}$$

[3] In literature, this family is referred to as *approximations to the identity*, without the *generalized*. Starting in the next set of notes, we shall follow this convention as well.

for all $f \in L^1(\mathbb{R}^d)$ such that $f \geq 0$. If the claim holds, then

$$|(f * \phi_r)(x)| = |(f\tau_x * \phi_r)(0)| \leq \|\phi_r\|_1 \mathcal{M}(f\tau_x)(0) = \|\phi_r\|_1 \mathcal{M}f(x),$$

where $f\tau_x(y) = f\tau_x(x+y)$. It then follows that

$$|(f * \rho_r)(x)| \leq (|f| * |\rho_r|)(x) \leq (|f| * \phi_r)(x) \leq \|\phi_r\|_1 \mathcal{M}f(x),$$

whence taking the supremum over $r > 0$ yields the desired result. It therefore suffices to establish (3.16).

To this end, we fix a nonnegative L^1-function f. We fix $\varepsilon > 0$ and apply a *dyadic decomposition* [4] on ϕ_r by setting

$$E_n = \{x \in \mathbb{R}^d : (1+\varepsilon)^n \leq \phi_r(x) < (1+\varepsilon)^{n+1}\}$$

for each $n \in \mathbb{Z}$. Then

$$\sum_{n\in\mathbb{Z}} (1+\varepsilon)^n \chi_{E_n} \leq \phi_r \leq \sum_{n\in\mathbb{Z}} (1+\varepsilon)^{n+1} \chi_{E_n}. \tag{3.17}$$

$(E_n)_{n\in\mathbb{Z}}$ is pairwise disjoint, and so the above estimate implies that

$$\sum_{n\in\mathbb{Z}} (1+\varepsilon)^n m(E_n) \leq \|\phi_r\|_1 \leq \sum_{n\in\mathbb{Z}} (1+\varepsilon)^{n+1} m(E_n). \tag{3.18}$$

Since $\rho_r \in L^\infty(\mathbb{R}^d)$, we see also that $\phi_r \in L^\infty(\mathbb{R}^d)$, whence $E_n = \varnothing$ for all sufficiently large $n \in \mathbb{Z}$. We let N_0 be the smallest integer such that $E_{N_0} = \varnothing$, and set

$$B_n = \bigcup_{k=N_0-n}^{N_0} E_k.$$

As ϕ_r is radially decreasing, we conclude that each B_n is a ball centered at 0. Note that $B_n \subseteq B_{n+1}$ for all $n \geq 1$. Setting $a_1 = (1+\varepsilon)^{N_0-1}$ and

$$a_n = (1+\varepsilon)^{N_0-n} - a_{n-1}$$

for all $n > 1$, we see that

$$\sum_{n\in\mathbb{Z}} (1+\varepsilon)^n \chi_{E_n} = \sum_{n=1}^{\infty} a_n \chi_{B_n}.$$

It now follows from (3.17) and (3.18) that

$$\phi_r \leq (1+\varepsilon) \sum_{n=1}^{\infty} a_n \chi_{B_n} \quad \text{and} \quad \sum_{n=1}^{\infty} a_n m(B_n) \leq \|\phi_r\|_1.$$

[4] The name "*dyadic* decomposition" derives from the $\varepsilon = 1$ case.

Using what we have established above, we can make the following string of estimates:

$$\begin{aligned}|(f*\phi_r)(0)| &\le (|f|*\phi_r)(0) \le (1+\varepsilon)\sum_{n=1}^{\infty}\left(|f|*a_n\chi_{B_n}\right)(0)\\ &\le (1+\varepsilon)\sum_{n=1}^{\infty} a_n m(B_n)\left(\frac{1}{m(B_n)}\left(|f|*\chi_{B_n}\right)(0)\right)\\ &= (1+\varepsilon)\sum_{n=1}^{\infty} a_n m(B_n)\left(\frac{1}{m(B_n)}\int_{B_n}|f(y)|\,dy\right)\\ &\le \left((1+\varepsilon)\sum_{n=1}^{\infty} a_n m(B_n)\right)\mathcal{M}f(0) = (1+\varepsilon)\|\phi_r\|_1\mathcal{M}f(0).\end{aligned}$$

Since the above estimate holds for all $\varepsilon > 0$, we conclude that

$$|(f*\phi_r)(0)| \le \|\phi_r\|_1\mathcal{M}f(0),$$

which is precisely (3.16). □

We conclude by noting that the use of the Hardy-Littlewood maximal function was merely a convenient way of establishing the weak-type (1,1) bound of the maximal function

$$f \mapsto \sup_{r>0} f*\rho_r$$

with respect to the family $(\rho_r)_{r>0}$ of approximations to the identity.

3.2. **The strong law of large numbers; martingales.** We now discuss an application of maximal estimates in probability theory. This particular application does not follow exactly our outline of the maximal-estimate approach to establish pointwise convergence but is nevertheless an important example.

We content ourselves by simply introducing the definitions, for a careful discussion of motivations would lead us too far afield. See, for example, [Var01] or [Wil91] a more detailed discussion. We also remark that we follow in this section the standard notations used in probability theory—for example, X for random variables and Ω for spaces instead of f for measurable functions and X for spaces.

Definition 3.19. A *probability space* is a measure space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(\Omega) = 1$. A *random variable* on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a measurable function $X : \Omega \to \mathbb{R}$. The *expectation* of a random variable X is the integral

$$\mathbb{E}[X] = \int_\Omega X\,d\mathbb{P}.$$

Definition 3.20. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A sequence $(\mathcal{F}_n)_{n=1}^\infty$ of σ-subalgebras of $\mathcal{F}$ is *independent* if

$$\mathbb{P}[E_{n_1}\cap\cdots\cap E_{n_K}] = \prod_{k=1}^{K}\mathbb{P}[E_{n_k}]$$

for each subsequence $(n_k)_{k=1}^K$ of the index sequence $(n)_{n=1}^\infty$ and every $E_{n_k} \in \mathcal{F}_{n_k}$. A sequence of random variables $(X_n)_{n=1}^\infty$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is *independent* if $(\sigma(X_n))_{n=1}^\infty$

is an independent sequence of σ-subalgebras of $\mathcal{F}$. Here $\sigma(X_n)$ is the *σ-subalgebra of $\mathcal{F}$ generated by X_n*, defined to be the σ-subalgebra of $\mathcal{F}$ generated by the sets

$$\{\omega \in \Omega : X_n(\omega) \in E\}, \quad \text{where} \quad E \in \mathcal{F}.$$

A fundamental question in probability theory is to determine the long-term behavior of the average

$$\frac{S_N}{N} = \frac{X_1 + \cdots + X_N}{N}$$

of independent. A result that deals with convergence of the above average is referred to as a *law of large numbers*. We shall be interested in the following form of the law of large numbers:

Theorem 3.21 (L^1 strong law of large numbers)**.** *Let $(X_n)_{n=1}^\infty$ be a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}[X_n] = 0$ for all $n \in \mathbb{N}$ and $\sum \mathbb{E}[|X_n|] < \infty$. Then*

$$\lim_{N \to \infty} \frac{S_N}{N} = 0$$

almost everywhere.

The condition $\mathbb{E}[X_n] = 0$ is referred to as *zero-mean.* It is enough to consider zero-mean random variables, for $(X_n - \mathbb{E}[X_n])_{n=1}^\infty$ is always zero-mean.

The L^1 strong law of large numbers is an immediate consequence of the following maximal estimate:

Lemma 3.22 (Kolmogorov's L^1 inequality)**.** *Let $(X_n)_{n=1}^\infty$ be a sequence of zero-mean independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, for each $M \in \mathbb{N}$,*

$$\mathbb{P}\left[\left\{\omega \in \Omega : \sup_{1 \leq n \leq M} |S_N(\omega)| > \alpha\right\}\right] \leq \frac{1}{\alpha} \sum_{n=1}^M \mathbb{E}[|X_n|].$$

Proof of the strong law via Kolmogorov's inequality. For each fixed $N \in \mathbb{N}$, Kolmogorov's inequality implies that

$$\mathbb{P}\left[\left\{\omega \in \Omega : \sup_{1 \leq N \leq M} |S_N(\omega)| > \alpha\right\}\right] \leq \frac{1}{\alpha} \sum_{n=1}^M \mathbb{E}[|X_n|] \leq \frac{1}{\alpha} \sum_{n=1}^\infty \mathbb{E}[|X_n|]$$

for all $\alpha > 0$. Setting $\alpha = \beta M$, we see that

$$\begin{aligned}
\mathbb{P}\left[\left\{\omega \in \Omega : \sup_{1 \leq N \leq M} \left|\frac{S_N(\omega)}{M}\right| > \beta\right\}\right] &= \mathbb{P}\left[\left\{\omega \in \Omega : \sup_{1 \leq N \leq M} |S_N(\omega)| > M\beta\right\}\right] \\
&\leq \frac{1}{M\beta} \sum_{n=1}^\infty \mathbb{E}[|X_n|].
\end{aligned}$$

Since

$$\sup_{1 \leq N \leq M} \left|\frac{S_N}{M}\right| \leq \sup_{1 \leq N \leq M} \left|\frac{S_N}{N}\right|,$$

it follows from the monotonicity of $\mathbb{P}$ that

$$\begin{aligned}
\mathbb{P}\left[\left\{\omega \in \Omega : \sup_{1 \leq N \leq M} \left|\frac{S_N(\omega)}{N}\right| > \beta\right\}\right] &\leq \mathbb{P}\left[\left\{\omega \in \Omega : \sup_{1 \leq N \leq M} \left|\frac{S_N(\omega)}{M}\right| > \beta\right\}\right] \\
&\leq \frac{1}{M\beta} \sum_{n=1}^\infty \mathbb{E}[|X_n|].
\end{aligned}$$

We now conclude that

$$\mathbb{P}\left[\left\{\omega \in \Omega : \sup_{N \in \mathbb{N}} \left|\frac{S_N(\omega)}{N}\right| > \beta\right\}\right] = \lim_{M \to \infty} \mathbb{P}\left[\left\{\omega \in \Omega : \sup_{1 \leq N \leq M} \left|\frac{S_N(\omega)}{N}\right| > \beta\right\}\right] = 0$$

for all $\beta > 0$, whence $S_N/N \to 0$ almost everywhere. □

It therefore suffices to establish Kolmogorov's inequality. To this end, we introduce *conditional expectations* and *martingales.*

Definition 3.23. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{G}$ a σ-subalgebra of $\mathcal{F}$. The *conditional expectation of a random variable* X *on* $(\Omega, \mathcal{F}, \mathbb{P})$ *with respect to* $\mathcal{G}$ is the (almost-everywhere) unique random variable $\mathbb{E}[X|\mathcal{G}]$ on $(\Omega, \mathcal{G}, \mathbb{P})$ such that

$$\int_E \mathbb{E}[X|\mathcal{G}]\, d\mathbb{P} = \int_E X\, d\mathbb{P}$$

for all $E \in \mathcal{G}$. The existence is guaranteed by the Radon–Nikodym theorem.

The conditional expectation is to be understood as the "best guess" at X, given only the information available in $\mathcal{G}$. The following is an immediate consequence of the uniqueness clause in the definition.

Proposition 3.24. *Let* X *be a random variable on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. *If* $\mathcal{G}$ *is a* σ*-subalgebra of* $\mathcal{F}$*, and if* X *is* $\mathcal{G}$*-measurable, then* $\mathbb{E}[X|\mathcal{G}] = X$ *almost everywhere.*

Given a sequence $(X_n)_{n=1}^\infty$ of zero-mean independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we consider a sequence of σ-subalgebras $(\mathcal{F}_N)_{N=1}^\infty$, where

$$\mathcal{F}_N = \sigma(X_1, \ldots, X_N),$$

the σ-subalgebra of $\mathcal{F}$ generated by $\bigcup_{n=1}^N \sigma(X_n)$. It is clear that $\mathcal{F}_N \subseteq \mathcal{F}_{N+1}$ for all $N \in \mathbb{N}$. Furthermore, we see, for each $E \in \mathcal{F}_N$, that

$$\int_E \mathbb{E}[S_{N+1}|\mathcal{F}_N] = \int_E S_{N+1} = \int_E S_N + \int_E X_{N+1} = 0$$

by the independence of $(X_n)_{n=1}^\infty$, whence

$$\mathbb{E}[S_{N+1}|\mathcal{F}_N] = S_N$$

almost everywhere by the uniqueness clause of the conditional expectation. This motivates us to introduce the following definition:

Definition 3.25. A *martingale sequence* on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a pair $((X_n)_n, (\mathcal{F})_n)$ of a sequence $(\mathcal{F}_n)_{n=1}^\infty$ of σ-subalgebras of $\mathcal{F}$ and a sequence $(X_n)_{n=1}^\infty$ of random variables on $(\Omega, \mathcal{F}_n, \mathbb{P})$ such that, for each $n \in \mathbb{N}$, we have the inclusion relation $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ and the identity

$$X_n = \mathbb{E}[X_{n+1}|\mathcal{F}_n]$$

almost everywhere.

Martingales are to be understood as a model of a in which the past events do not help predicting the future events. The main maximal estimate for martingales is as follows:

Theorem 3.26 (Doob's maximal theorem)**.** *If $((X_n)_n, (\mathcal{F}_n)_n)$ is a martingale sequence, then for each $N \in \mathbb{N}$, we have the weak-type estimate*

$$\mathbb{P}\left(\left\{\omega \in X : \sup_{1 \leq n \leq N} |X_n(\omega)| > \alpha\right\}\right) \leq \frac{1}{\alpha}\|X_N\|_1.$$

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

Proof. For each $\alpha > 0$, we let

$$E_\alpha = \left\{\omega \in \Omega : \sup_{1 \leq n \leq N} |X_n(\omega)| > \alpha\right\}.$$

We define, for each $1 \leq n \leq N$,

$$E_\alpha^n = \left(\bigcap_{k=1}^{n-1} \{\omega \in \omega : |X_k(\omega)| \leq \alpha\}\right) \cap \{w \in \Omega : |X_n(\omega)| > \alpha\},$$

so that $E_\alpha^n \in \mathcal{F}_n$. We note that $(E_\alpha^n)_{n=1}^N$ is a pairwise-disjoint sequence whose union is E_α.

We now claim that

$$|X_n| \leq \mathbb{E}[|X_{n+1}||\mathcal{F}_n] \tag{3.27}$$

almost everywhere for each $1 \leq n \leq N-1$. Assuming the claim for now, we observe that

$$\begin{aligned}
\mathbb{P}[E_\alpha^n] &= \frac{1}{\alpha}\int_{E_\alpha^n} \alpha \, d\mathbb{P} \leq \frac{1}{\alpha}\int_{E_\alpha^n} |X_n| \, d\mathbb{P} \\
&\leq \frac{1}{\alpha}\int_{E_\alpha^n} \mathbb{E}[|X_{n+1}||\mathcal{F}_n] \, d\mathbb{P} \leq \frac{1}{\alpha}\int_{E_\alpha^n} |X_{n+1}| \, d\mathbb{P} \\
&\leq \cdots \leq \frac{1}{\alpha}\int_{E_\alpha^n} |X_N| \, d\mathbb{P}.
\end{aligned}$$

Here we have made use of Proposition 3.24. It then follows from the pairwise-disjointness of $(E_\alpha^n)_{n=1}^N$ that

$$\mathbb{P}[E_\alpha] = \sum_{n=1}^N \mathbb{P}[E_\alpha^n] \leq \sum_{n=1}^N \frac{1}{\alpha}\int_{E_\alpha^n} |X_N| \, d\mathbb{P} = \frac{1}{\alpha}\int_{E_\alpha} |X_N| \, d\mathbb{P} \leq \frac{1}{\alpha}\mathbb{E}[|X_N|],$$

as was to be shown.

It thus remains to establish (3.27), which is a special case of Jensen's inequality for conditional expectations. To do so, it suffices to observe that

$$\begin{aligned}
\mathbb{E}[|X_{n+1}||\mathcal{F}_n] &= \mathbb{E}\left[\max\left(X_{n+1}, -X_{n+1}\right) | \, \mathcal{F}_n\right] \\
&\geq \max\left(\mathbb{E}[X_{n+1}|\mathcal{F}_n], \mathbb{E}[-X_{n+1}|\mathcal{F}_n]\right) \\
&= \max\left(X_n, -X_n\right) \\
&= |X_n|,
\end{aligned}$$

where we have made use of the definition of a martingale sequence in the second to the last equality. ☐

Kolmogorov's inequality is now an easy consequence of the maximal theorem.

Proof of Kolmogorov's inequality via Doob's maximal theorem. Since $(S_N)_{N=1}^{\infty}$ is a martingale sequence, Doob's maximal theorem implies that

$$\mathbb{P}\left[\left\{\omega \in \Omega : \sup_{1 \leq N \leq M} |S_N(\omega)| > \alpha\right\}\right] \leq \frac{1}{\alpha}\mathbb{E}[|S_M|] \leq \frac{1}{\alpha}\sum_{n=1}^{M}\mathbb{E}[|X_n|]$$

for each $M \in \mathbb{N}$ and every $\alpha > 0$. □

3.3. **Birkhoff's ergodic theorem.** Generalizing the strong law, we obtain the *pointwise ergodic theorem*, a foundational theorem in *ergodic theory*—the study of dynamical systems with invariant structures. Our proof will be more or less a straightforward adaptation of the maximal estimate approach to establishing pointwise convergence: we shall establish convergence for a dense subclass, formulate a maximal estimate, and use the maximal estimate to establish convergence for the entire collection of functions.

We begin by introducing the main objects of study in *measurable* ergodic theory: probability-preserving transformations on a fixed probability space and the ergodic hypothesis on these transformations. Once again, it is not our purpose to study ergodic theory systematically in these notes. See, for example, the 2013 lecture notes [Aus13] of T. Austin for a detailed survey.

Definition 3.28. A *probability-preserving transformation* on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a bijective measurable function $T : \Omega \to \Omega$ such that its inverse T^{-1} is measurable and $\mathbb{P}[T^{-1}(E)] = \mathbb{P}[E]$ for all $E \in \mathcal{F}$.

Given a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ and a probability-preserving transformation $T : \Omega \to T$, we define the *time average of X with respect to T* as

$$\widehat{X}(\omega) = \limsup_{N \to \infty} \frac{1}{N}\sum_{n=0}^{N-1} X(T^n\omega),$$

and the *space average of X* as

$$\overline{X} = \mathbb{E}[X].$$

An *ergodic theorem* spells out the relation between the time average and the space average. The pointwise version of the ergodic theorem, due to G. Birkhoff, is as follows:

Theorem 3.29 (Birkhoff's pointwise ergodic theorem)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $T : \Omega \to \Omega$ a probability-preserving transformation. For each random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$, the time average $\widehat{X}$ with respect to T is finite almost everywhere.*

With some effort, we can also prove that $\mathbb{E}[\widehat{X}] = \overline{X}$. Furthermore, if T is *ergodic*, then $\widehat{X} = \overline{X}$ almost everywhere. Since we are mainly concerned with pointwise convergence results in this section, we do not attempt to discuss these additional results that make Birkhoff's theorem truly an ergodic theorem.

To establish the pointwise ergodic theorem, we study the maximal function

$$X^*(\omega) = \sup_{N \in \mathbb{N}} \frac{1}{N}\sum_{n=0}^{N-1} |X(T^n\omega)|.$$

As expected, the maximal function satisfies the weak-type bound:

Lemma 3.30 (Wiener's maximal ergodic theorem)**.** *If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, then there exists a constant $A > 0$ such that*

$$\mathbb{P}[\{\omega \in \Omega : X^*(\omega) > \alpha\}] \leq \frac{A}{\alpha}\mathbb{E}[|X|]$$

for all $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\alpha > 0$.

We postpone the proof of the maximal inequality and deduce the pointwise ergodic theorem from the maximal ergodic theorem.

Proof of Birkhoff's theorem from Wiener's theorem. We suppose for now that the pointwise ergodic theorem has been established for a dense subclass $\mathcal{D}$ of $L^1(\Omega)$. Arguing as in the proof of the Lebesgue differentiation theorem, we fix $X \in L^1(\Omega)$, set

$$\widehat{X}_N(\omega) = \frac{1}{N}\sum_{n=0}^{N-1} X(T^n\omega),$$

and define

$$E_\alpha = \left\{\omega \in \Omega : \lim_{K \to \infty} \sup_{N,M \geq K} \left|\widehat{X}_N(\omega) - \widehat{X}_M(\omega)\right| > \alpha\right\}$$

for each $\alpha > 0$. It evidently suffices to show that $\mathbb{P}[E_\alpha] = 0$ for all $\alpha > 0$.

Fix $\varepsilon > 0$ and find $Y \in \mathcal{D}$ such that $\|X - Y\|_1 < \varepsilon$. Since

$$\left|\widehat{X}_N - \widehat{X}_M\right| \leq |\widehat{Y}_N - \widehat{Y}_M| + |(\widehat{X - Y})_N - (\widehat{X - Y})_M|,$$

we see that

$$\begin{aligned}\mathbb{P}[E_\alpha] \leq& \mathbb{P}\left[\left\{\omega \in \Omega : \lim_{K \to \infty} \sup_{N,M \geq K} \left|\widehat{Y}_N(\omega) - \widehat{Y}_M(\omega)\right| > \alpha\right\}\right] \\ &+ \mathbb{P}\left[\left\{\omega \in \Omega : \lim_{K \to \infty} \sup_{N,M \geq K} \left|(\widehat{X - Y})_N(\omega) - (\widehat{X - Y})_M(\omega)\right| > \alpha\right\}\right].\end{aligned}$$

By the assumption,

$$\mathbb{P}\left[\left\{\omega \in \Omega : \lim_{K \to \infty} \sup_{N,M \geq K} \left|\widehat{Y}_N(\omega) - \widehat{Y}_M(\omega)\right| > \alpha\right\}\right] = 0$$

for all $\alpha > 0$, whence it suffices to estimate

$$\mathbb{P}\left[\left\{\omega \in \Omega : \lim_{K \to \infty} \sup_{N,M \geq K} \left|\widehat{Z}_N(\omega) - \widehat{Z}_M(\omega)\right| > \alpha\right\}\right],$$

where $Z = X - Y$.

We now observe that

$$\sup_{N,M \geq K} \left|\widehat{Z}_N - \widehat{Z}_M\right| \leq 2 \sup_{N \in \mathbb{N}} \left|\widehat{Z}_N\right|$$

for all $K \in \mathbb{N}$, hence

$$\mathbb{P}[E_\alpha] \leq \mathbb{P}\left[\left\{\omega \in \Omega : 2 \sup_{N \in \mathbb{N}} \left|\widehat{Z}_N(\omega)\right| > \alpha\right\}\right]$$

By the maximal ergodic theorem,

$$\mathbb{P}\left[\left\{\omega \in \Omega : 2 \sup_{N \in \mathbb{N}} \left|\widehat{Z}_N(\omega)\right| > \alpha\right\}\right] \leq \frac{2A}{\alpha}\|Z\|_1 < \frac{2A}{\alpha}\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, we conclude that $\mathbb{P}[E_\alpha] = 0$. The desired result now follows.

It therefore suffices to prove the pointwise ergodic theorem for a suitable dense subclass of $L^1(\Omega)$. Since $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, Hölder's inequality implies that $\|X\|_1 \leq \|X\|_2$ for all $X \in L^2(\Omega)$. This, in particular, implies that every dense subclass of $L^2(\Omega)$ is a dense subclass of $L^1(\Omega)$.

To find a nice dense subclass of $L^2(\Omega)$, we write $\mathcal{T}$ to denote the operator

$$\mathcal{T}X(\omega) = X(T\omega)$$

and take the $\mathcal{T}$-*invariant* subspace $L^2(\Omega)^{\mathcal{T}}$ of $L^2(\Omega)$: namely, the collection of $X \in L^2(\Omega)$ such that $\mathcal{T}X = X$. It follows at once that $\mathcal{T}$ is an orthogonal projection onto $L^2(\Omega)^{\mathcal{T}}$, whence we have the direct-sum decomposition

$$L^2(\Omega) = L^2(\Omega)^{\mathcal{T}} \oplus (I - \mathcal{T})\left(L^2(\Omega)\right).$$

We now claim that

$$\mathcal{D}_1 = \{\mathcal{T}X - X : X \in L^\infty(\Omega)\}$$

is a dense subspace of $(I - \mathcal{T})\left(L^2(\Omega)\right)$, so that

$$\mathcal{D} = L^2(\Omega)^{\mathcal{T}} \oplus \mathcal{D}_1$$

is a dense subspace of $L^2(\Omega)$. To see this, we first recall that $L^\infty(\Omega)$ is a dense subspace of $L^2(\Omega)$. In particular, for each $X \in L^2(\Omega)$, we can find a sequence $(X_n)_{n=1}^\infty$ in $L^\infty(\Omega)$ such that $\|X_n - X\|_2 \to 0$ as $n \to \infty$. Now, T is a probability-preserving transformation, hence $\mathcal{T}$ is an isometry on $L^2(\Omega)$. In particular, $\mathcal{T}$ is a bounded operator on $L^2(\Omega)$ into itself. Therefore $I - \mathcal{T}$ is a bounded operator on $L^2(\Omega)$ into itself, whence $\|(I - \mathcal{T})(X_n - X)\|_2 \leq C\|X_n - X\|_2$ for some constant C. It now follows that $(I - \mathcal{T})(X_n) \to (I - \mathcal{T})(X)$ in $L^2(\Omega)$, and the claim follows.

It therefore suffices to prove the pointwise ergodic theorem for random variables in $\mathcal{D}$. Let $X \in \mathcal{D}$ and set $X_1 = \mathcal{T}X$ and $X_2 = (I - \mathcal{T})X$, so that $X_1 \in L^2(\Omega)^{\mathcal{T}}$, $X_2 \in \mathcal{D}_1$, and $X = X_1 + X_2$. Since $\widehat{(X_1)}_N = X$ for all $N \in \mathbb{N}$, it follows that $\widehat{X_1} = X_1$. As for X_2, we observe that

$$\widehat{(X_2)}_N = \frac{1}{N}\sum_{n=0}^{N-1} \mathcal{T}^n X - \mathcal{T}^{n+1}X = \frac{1}{N}\left(X - \mathcal{T}^N X\right),$$

whence

$$\widehat{X_2} = \limsup_{N\to\infty} \frac{1}{N}|X - \mathcal{T}^N X| \leq \limsup_{N\to\infty} \frac{1}{N}(2\|X\|_\infty) = 0.$$

It follows that

$$\widehat{X} = \widehat{X_1} + \widehat{X_2} = X_1.$$

We have thus completed the proof of the pointwise ergodic theorem. □

It now remains to prove the maximal ergodic theorem. We shall first prove an analogous estimate on $\mathbb{Z}$ with the counting measure, then transfer this result to the the case at hand. This is a special case of the *Calderón transference principle*: see [Cal68].

Proof of the maximal ergodic theorem. We first prove an analogue of the maximal ergodic theorem for the translation operator $Tn = n + 1$ on $\mathbb{Z}$ with the discrete σ-algebra and the counting measure $\#$. While $\mathbb{Z}$ isn't a probability space, it will serve as the model case to which we shall reduce all general cases.

Specifically, we shall show that

$$\#\left(\{m \in \mathbb{Z} : f^*(m) > \alpha\}\right) \leq \frac{12}{\alpha} \sum_{m \in \mathbb{Z}} |f(m)| \tag{3.31}$$

for all constants $\alpha > 0$ and all functions $f : \mathbb{Z} \to \mathbb{R}$, where

$$f^*(m) = \sup_{N \in \mathbb{N}} \frac{1}{N} \sum_{n=0}^{N-1} |f(m+n)|.$$

Given a function $f : \mathbb{Z} \to \mathbb{R}$, we consider the extension $\tilde{f} : \mathbb{R} \to \mathbb{R}$ by setting $\tilde{f}(x) = f(n)$ for all $n \leq x < n-1$. Similarly, for each $E \subseteq \mathbb{Z}$, we consider the extension $\tilde{E} = \bigcup_{m \in E} [n, n+1)$. We then see that $m(\tilde{E}) = \#(E)$, $\int_E |\tilde{f}(x)|\, dx = \sum_{m \in E} |f(m)|$, and $\|\tilde{f}\|_{L^1(\mathbb{R})} = \|f\|_{L^1(\mathbb{Z})}$. In particular,

$$\begin{aligned}
\frac{1}{N} \sum_{n=0}^{N-1} |f(m+n)| &= \frac{1}{N} \int_0^N |\tilde{f}(m+t)|\, dt \\
&\leq \frac{N+1}{N} \left(\frac{1}{N+1} \int_{-1}^{N} |\tilde{f}(m+t)|\, dt \right) \\
&\leq 2 \left(\frac{1}{N+1} \int_{-1}^{N} |\tilde{f}(x+t)|\, dt \right)
\end{aligned}$$

for all $x \in [m, m+1)$. Therefore,

$$\begin{aligned}
f^*(m) &\leq 2 \sup_{N \in \mathbb{N}} \frac{1}{N+1} \int_{-1}^{N} |\tilde{f}(x+t)|\, dt \\
&\leq 2 \sup_{N \in \mathbb{N}} \left(\frac{2N}{N+1} \right) \left(\frac{1}{2N} \int_{-N}^{N} |\tilde{f}(x+t)|\, dt \right) \\
&\leq 4 \mathcal{M}\tilde{f}(x)
\end{aligned}$$

for all $x \in [m, m+1)$. It now follows from the weak-type (1,1) bound of the one-dimensional Hardy-Littlewood maximal function over balls (Theorem 2.1) that

$$\begin{aligned}
\#\left(\{m \in \mathbb{Z} : f^*(m) > \alpha\}\right) &\leq m\left(\left\{ x \in \mathbb{R} : \mathcal{M}\tilde{f}(x) > \alpha/4 \right\} \right) \\
&\leq \frac{3}{\alpha/4} \|\tilde{f}\|_{L^1(\mathbb{R})} = \frac{12}{\alpha} \sum_{m \in \mathbb{Z}} |f(m)|.
\end{aligned}$$

3 is the constant obtained from the one-dimensional Vitali covering lemma.

Let us now consider the general case. We fix an L^1 random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and consider, for each $M \in \mathbb{N}$, the truncated maximal function of order M:

$$X_M^*(\omega) = \sup_{1 \leq N \leq M} \frac{1}{N} \sum_{n=0}^{N-1} \mathcal{T}^n |f(\omega)|.$$

Observe that $(X_M^*)_{M=1}^{\infty}$ is an increasing sequence whose limit is X^*. Therefore, the task of establishing the desired maximal estimate reduces to proving the maximal inequality

$$\mathbb{P}[\{\omega \in \Omega : X_M^*(\omega) > \alpha\}] \leq \frac{A}{\alpha} \mathbb{E}[|X|] \tag{3.32}$$

with constant A independent of $M \in \mathbb{N}$ and $\alpha > 0$.

To this end, we define a function $F : \Omega \times \mathbb{Z} \to \mathbb{R}$ by setting

$$F(\omega, n) = \begin{cases} \mathcal{T}^n X(\omega) & \text{if } n \geq 0; \\ 0 & \text{if } n < 0; \end{cases}$$

so that

$$\frac{1}{N} \sum_{n=0}^{N-1} F(\omega, n) = \frac{1}{N} \sum_{n=0}^{N-1} \mathcal{T}^n X(\omega)$$

for all $(\omega, N) \in \Omega \times \mathbb{N}$. In particular,

$$\frac{1}{N} \sum_{n=0}^{N-1} \mathcal{T}^n X(T^k \omega) = \frac{1}{N} \sum_{n=0}^{N-1} F(\omega, n+k)$$

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

For each $x > 0$, we take the truncation

$$F_x(\omega, n) = \begin{cases} F(\omega, n) & \text{if } n < x; \\ 0 & \text{if } n \geq x; \end{cases}$$

and set $f_{x,\omega}(n) = F_x(\omega, n)$. Note that $f_{x,\omega}(n)$ is a real-valued function on $\mathbb{Z}$ for fixed $x > 0$ and $\omega \in \Omega$.

Fix $M \in \mathbb{N}$ and $x > M$ and set $y = x - M$. Since

$$\frac{1}{N} \sum_{n=0}^{N-1} \mathcal{T}^n X(T^k \omega) = \frac{1}{N} \sum_{n=0}^{N-1} f_{x,\omega}(n+k)$$

whenever $N \leq M$ and $k < y$, we see that

$$X_M^*(T^k \omega) \leq (f_{x,\omega})^*(k) = (f_{y+M,\omega})^*(k) \tag{3.33}$$

whenever $k < y$.

We are now ready to transfer the maximal estimate on $\mathbb{Z}$ to the general case. To this end, we set

$$E_\alpha = \{\omega \in \Omega : X_M^*(\omega) > \alpha\}$$

for each $\alpha > 0$. Since T is a probability-preserving transformation, we see that

$$\mathbb{P}[E_\alpha] = \mathbb{P}[\{\omega \in \Omega : X_M^*(T^k \omega) > \alpha\}]$$

for all $k \in \mathbb{N}$. The definition of the product measure $\mathbb{P} \otimes \#$ implies that

$$(\mathbb{P} \otimes \#)\left(\left\{(\omega, k) \in \Omega \times \mathbb{Z} : X_M^*(T^k \omega) > \alpha \text{ and } 0 \leq k < y\right\}\right) = y\mathbb{P}[E_\alpha]$$

for each fixed $y > 0$. We now apply the comparison inequality (3.33) to conclude that

$$y\mathbb{P}[E_\alpha] \leq \mathbb{E}_\omega \left[\# \left(\{k \in \mathbb{Z} : (f_{y+M,\omega})^*(k) > \alpha\}\right)\right]. \tag{3.34}$$

The maximal estimate (3.31) implies that

$$\begin{aligned} \# \left(\{k \in \mathbb{Z} : (f_{y+M,\omega})^*(k) \right. & \left. > \alpha\}\right) \leq \frac{12}{\alpha} \sum_{k \in \mathbb{Z}} |f_{y+M,\omega}(k)| \\ & = \frac{12}{\alpha} \sum_{k=0}^{\lfloor y+M-1 \rfloor} |f_{y+M,\omega}(k)| = \frac{12}{\alpha} \sum_{k=0}^{\lfloor y+M-1 \rfloor} |\mathcal{T}^k X(\omega)|. \end{aligned}$$

where $\lfloor y+M-1 \rfloor$ is the largest integer that is smaller than or equal to $y+M-1$. We now integrate the above inequality over Ω and apply (3.34) to obtain the following estimate:

$$y\mathbb{P}[E_\alpha] \leq \frac{12}{\alpha}(\lfloor y+M \rfloor)\mathbb{E}[|X|].$$

Therefore,

$$\mathbb{P}[E_\alpha] \leq \frac{12}{\alpha} \cdot \frac{\lfloor y+M \rfloor}{y} \cdot \mathbb{E}[|X|],$$

and sending $y \to \infty$ yields the truncated estimate (3.32) with constant $A = 12$. This completes the proof of the maximal ergodic theorem. □

3.4. **Stein's maximal principle.** A natural question to ask at this point is if the maximal-estimates approach to establishing pointwise convergence is *necessary*: in other words, does pointwise convergence imply a weak-type bound?

This is false in general: Tao, in [Tao11], offers

$$T_n f(x) = \chi_{[n,n+1]} \int_0^1 f(y)\,dy$$

as an example of a sequence of operators on $L^1(\mathbb{R})$ into itself that converges pointwise almost everywhere (to 0) without satisfying a weak-type (1,1) maximal bound. On the other hand, if we impose a compactness hypothesis on our measure space, then the following result of E. M. Stein implies that pointwise almost-everywhere convergence is, in fact, equivalent to the exsitence of a weak-type bound:

Theorem 3.35 (Stein's L^1 maximal principle, 1961; [Ste61])**.** *Let G be a compact, Hausdorff, abelian topological group equipped with the Haar measure*[5] μ. *If $(\varphi_n)_{n=1}^\infty$ is a sequence of operators in $L^\infty(G)$ such that, for each $f \in L^1(G)$, we have the "pointwise convergence criterion"*

$$\limsup_{n\to\infty} |(f * \varphi_n)(x)| < \infty$$

on a set E_f of positive measure, then the maximal operator

$$Mf(x) = \sup_{n\in\mathbb{N}} |(f * \varphi_n)(x)|$$

satisfies the weak-type (1,1) bound.

This is a special case of Theorem 2 in [Ste61], whose proof is beyond the scope of these notes. The paper contains the fully general statement of the L^1 maximal theorem and its variants on L^p-spaces and their proofs, as well as a detailed discussion of motivating examples.

4. An Interpolation Result

While the utility of the weak-type (1,1) bound of various maximal functions should be clear by now, we note that maximal functions do have one notable deficiency: they, in general, do not map integrable functions to integrable functions[6].

[5] The Haar measure is the unique probability measure μ on a compact Hausdorff abelian topological group that is *translation invariant*, viz., $\mu(\{gx : x \in E\}) = \mu(E)$ for each $g \in G$ and every measurable subset E of G.

[6] As we shall see later, this is unavoidable in many cases. In the words of E. M. Stein: "interesting operators in harmonic analysis are rarely bounded on L^1 and L^∞."

Example 4.1 (The Hardy-Littlewood maximal function does not map L^1 into itself). If $f \in L^1(\mathbb{R}^d)$ has positive L^1-norm, then we can find a ball $B(x_0; r) \subseteq \mathbb{R}^d$ such that $I = \int_{B(x_0;r)} |f| > 0$. We let $k > 0$ such that $k|x| \geq 2(|x| + |x_0| + r)$ for all $x \in \mathbb{R}^d$ with $|x| \geq 1$: in fact, it suffices to set $k = 2 + 2(|x_0| + r)$.

We then have the inclusion relation

$$B(x_0; r) \subseteq B(x; 2(|x| + |x_0| + r)) \subseteq B(x; k|x|),$$

whence

$$\begin{aligned}\mathcal{M}f(x) \geq \frac{1}{m(B(x;k|x|))} \int_{B(x;k|x|)} |f(y)|\, dy &\geq \frac{1}{m(B(x;k|x|))} \int_{B(x_0;r)} |f(y)|\, dy \\ &= \frac{1}{|x|^d} \frac{I}{m(B(x;k))} = \frac{1}{|x|^d} \frac{I}{m(B(x;0))}.\end{aligned}$$

Since $I/m(B(x;0)) > 0$, we conclude that $\mathcal{M}f$ cannot be integrable on $\mathbb{R}^d$ □

This deficiency is resolved on higher-order Lebesgue spaces:

Theorem 4.2 (Strong-type (p,p) bound on the maximal function)**.** *If $1 < p \leq \infty$, and if $B \subseteq \mathbb{R}^d$ is a centrally symmetric convex body, then there exists a constant $A_{p,d,B}$ such that*

$$\|\mathcal{M}_B f\|_p \leq A_{d,p,B} \|f\|_p \tag{4.3}$$

for all $f \in L^p(\mathbb{R}^d)$. The constant $A_{d,p,B}$ depends only on the exponent p, dimension d, and the body B.

Proof. If $p = \infty$, then

$$\frac{1}{m(rB)} \int_{rB} |f(x+y)|\, dy \leq \frac{1}{m(rB)} \int_{rB} \|f\|_\infty\, dy = \|f\|_\infty,$$

and so (4.3) holds with the constant $A_{d,\infty,B} = 1$.

We assume that $1 < p < \infty$. To check that $\mathcal{M}_B f$ is well-defined for all $f \in L^p(\mathbb{R}^d)$, we apply a *layercake decomsposition* on the function f, i.e.,

$$f = f\chi_{\{x:|f(x)|>\alpha\}} + f\chi_{\{x:|f(x)|\leq\alpha\}}$$

for a fixed $\alpha > 0$. Evidently, $f\chi_{\{|f|\leq\alpha\}} \in L^\infty(\mathbb{R}^d)$, whence the strong-type (∞,∞) bound implies that $\mathcal{M}_B f\chi_{\{|f|\leq\alpha\}}$ is finite almost everywhere. On the other hand,

$$\begin{aligned}\|f\|_p^p = \int |f(x)|^p\, dx &\geq \int_{|f(x)|>\alpha} |f(x)|^p\, dx \\ &> \int_{|f(x)|>\alpha} \alpha\, dx = \alpha\, m(\{x : |f(x)| > \alpha\}),\end{aligned}$$

whence Hölder's inequality implies that

$$\begin{aligned}\|f\chi_{|f|>\alpha}\|_1 &\leq \|f\|_p \|\chi_{\{|f|>\alpha\}}\|_{p'} \\ &= \|f\|_p\, m(\{x : |f(x)| > \alpha\})^{1/p'} \\ &\leq \alpha^{1/p'} \|f\|_p^{(p+p')/p},\end{aligned}$$

where $1/p + 1/p' = 1$. Therefore, $f\chi_{\{|f|>\alpha\}} \in L^1(\mathbb{R}^d)$, and the weak-type (1,1) bound (Theorem 2.1) implies that $\mathcal{M}_B f\chi_{\{|f|>\alpha\}}$ is finite almost everywhere. Since

$$\mathcal{M}_B f \leq \mathcal{M}_B(f\chi_{\{|f|>\alpha\}}) + \mathcal{M}_B(f\chi_{\{|f|\leq\alpha\}}),$$

it follows that $\mathcal{M}_B f$ is finite almost everywhere.

To show that $\mathcal{M}_B$ satisfies the strong-type (p,p) bound, we apply the layercake decomposition to $f \in L^p(\mathbb{R}^d)$ at $\alpha/2$ in place of α:

$$f = f\chi_{\{|f|>\alpha/2\}} + f\chi_{\{|f|\leq\alpha/2\}}.$$

As $f\chi_{\{|f|\leq\alpha/2\}} \leq \alpha/2$, we see that $\mathcal{M}_B f\chi_{\{|f|\leq\alpha/2\}} \leq \alpha/2$, and so

$$\mathcal{M}_B f \leq \mathcal{M}_B(f\chi_{\{|f|>\alpha/2\}}) + \mathcal{M}_B(f\chi_{\{|f|\leq\alpha/2\}}) \leq \mathcal{M}_B(f\chi_{\{|f|>\alpha/2\}}) + \frac{\alpha}{2}.$$

Therefore,

$$\{x \in \mathbb{R}^d : \mathcal{M}_B f(x) > \alpha\} \subseteq \{x \in \mathbb{R}^d : \mathcal{M}_B f\chi_{\{|f|>\alpha/2\}}(x) > \alpha/2\},$$

and so the weak-type (1,1) bound yields the estimate

$$m(\{x \in \mathbb{R}^d : \mathcal{M}_B f(x) > \alpha\}) \leq \frac{A_{d,1,B}}{\alpha}\|f\chi_{\{|f|>\alpha/2\}}\|_1. \tag{4.4}$$

In order to establish the strong-type (p,p) bound via (4.4), we shall relate $\|\mathcal{M}_B f\|_p$ to $m(\{x \in \mathbb{R}^d : \mathcal{M}_B f(x) > \alpha\})$, and $\|f\|_p$ to $\|f\chi_{|f|>\alpha/2}\|_1$. To this end, we introduce the *distribution function*

$$\lambda_g(\alpha) = m\left(\{x \in \mathbb{R}^d : g(x) > \alpha\}\right)$$

for each measurable function $g : \mathbb{R}^d \to \mathbb{C}$. The key fact is as follows:

Lemma 4.5. *If $0 < q < \infty$, then*

$$\|g\|_q^q = \int_0^\infty \lambda_g(\alpha^{1/q})\, d\alpha.$$

Proof of lemma. If $q = 1$, then

$$\begin{aligned}
\int_0^\infty \lambda_g(\alpha)\, d\alpha &= \int_0^\infty \int_{\mathbb{R}^d} \chi_{\{|g|>\alpha\}}(x)\, dx\, d\alpha \\
&= \int_0^\infty \int_{\mathbb{R}^d} \chi_{\{(x,\alpha):|g(x)|>\alpha\}}\, dx\, d\alpha \\
&= \int_{\mathbb{R}^d} \int_0^\infty \chi_{\{(x,\alpha):|g(x)|>\alpha\}}\, d\alpha\, dx \\
&= \int_{\mathbb{R}^d} \int_0^{|g(x)|} d\alpha\, dx \\
&= \int_{\mathbb{R}^d} |g(x)|\, dx \\
&= \|g\|_1
\end{aligned}$$

by the Fubini-Tonelli theorem. If $q \neq 1$, then

$$\|g\|_q^q = \|g^q\|_1^q = \int_0^\infty \lambda_{g^q}(\alpha)\, d\alpha$$

by what we have shown above. But then $g(x)^q > \alpha$ if and only if $g(x) > \alpha^{1/q}$, and so $\lambda_{g^q}(\alpha) = \lambda_g(\alpha^{1/q})$. It now follows that

$$\|g\|_q^q = \int_0^\infty \lambda_{g^q}(\alpha)\, d\alpha = \int_0^\infty \lambda_g(\alpha^{1/q})\, d\alpha,$$

which is the desired result. □

Applying the lemma to $g = \mathcal{M}_B f$, we see that

$$\|\mathcal{M}_B f\|_p^p = \int_0^\infty \lambda_{\mathcal{M}_B f}(\alpha^{1/p})\, d\alpha.$$

(4.4) implies that

$$\lambda_{\mathcal{M}_B f}(\alpha^{1/p}) \leq \frac{A_{d,1,B}}{\alpha^{1/p}} \|f\chi_{\{|f|>\alpha^{1/p}/2\}}\|_1,$$

and so

$$\begin{aligned}
\|\mathcal{M}_B f\|_p^p &\leq A_{d,1,B} \int_0^\infty \alpha^{-1/p} \int_{\mathbb{R}^d} |f\chi_{\{|f|>\alpha^{1/p}/2\}}(x)|\, dx\, d\alpha \\
&= A_{d,1,B} \int_0^\infty \alpha^{-1/p} \int_{\mathbb{R}^d} |f(x)|\chi_{\{(x,\alpha):|f(x)|>\alpha^{1/p}/2\}}(x)\, dx\, d\alpha \\
&= A_{d,1,B} \int_{\mathbb{R}^d} |f(x)| \int_0^\infty \alpha^{-1/p}\chi_{\{(x,\alpha):|f(x)|>\alpha^{1/p}/2\}}(x)\, d\alpha\, dx \\
&= A_{d,1,B} \int_{\mathbb{R}^d} |f(x)| \int_0^{|2f(x)|^p} \alpha^{-1/p}\, d\alpha\, dx
\end{aligned}$$

by the Fubini-Tonelli theorem. Since $1 < p < \infty$, we can evaluate the inner integral explicitly as follows:

$$\int_0^{|2f(x)|^p} \alpha^{-1/p}\, d\alpha = \frac{p}{p-1}\alpha^{(p-1)/p}\Big|_0^{|2f(x)|^p} = \frac{p}{p-1}|2f(x)|^{p-1}.$$

Substituting the result of computation to the above estimate, we see that

$$\|\mathcal{M}_B f\|_p^p \leq \frac{p2^{p-1}A_{d,B}}{p-1}\int_{\mathbb{R}^d} |f(x)|^p\, dx = \frac{p2^{p-1}A_{d,B}}{p-1}\|f\|_p^p.$$

It follows that

$$\|\mathcal{M}_B f\|_p \leq 2^{\frac{p-1}{p}} A_{d,1,B}^{\frac{1}{p}} \left(\frac{p}{p-1}\right)^{1/p} \|f\|_p,$$

which is the desired bound. □

Results of this type are referred to as *real interpolation*: *real*, because they do not make use of complex analysis, and *interpolation*, because they establish intermediate bounds from endpoint bounds. We shall study classical interpolation theorems in the next set of notes.

5. Further Results: The Maximal Function Conjectures

The interpolated constant

$$A_{d,p,B} = 2^{\frac{p-1}{p}} A_{d,1,B}^{\frac{1}{p}} \left(\frac{p}{p-1}\right)^{1/p}$$

given in the proof of the strong-type (p,p) bound (Theorem 4.2) depends crucially on the weak-type (1,1) constant $A_{d,1,B}$, which is far from optimal. Indeed, the 5 factor in the infinitary Vitali covering lemma (Lemma 2.4) can be decreased to any constant $C > 3$ by selecting balls of radii larger than $2/(C-1)$ times the supremum of the radii. Furthermore, $C = 3$ can be achieved if we restrict our attention to finite collections of balls:

Lemma 5.1 (Finitary Vitali covering lemma)**.** *If $\{B(x_n, r_n) : 1 \leq n \leq N\}$ is a finite collection of Euclidean balls in $\mathbb{R}^d$, then there exists a pairwise-disjoint countable subcollection $\{B(x_{n_k}, r_{n_k})\}_k$ such that*

$$\bigcup_{n=1}^{N} B(x_n, r_n) \subseteq \bigcup_k B(x_{n_k}, 3r_{n_k}).$$

Sketch of proof. Without loss of generality, we assume that $r_1 \geq r_2 \geq \ldots \geq r_N$. Let $n_1 = 1$. Since $3r_1 \geq r_1 + 2r_1 \geq r_1 + 2r_n$ for all $1 \leq n \leq N$, we see that any ball that intersects $B(x_1, r_1)$ must be contained in $B(x_1, 3r_1)$. We now proceed as in the proof of the infinitary Vitali covering lemma. □

We can apply the finitary Vitali covering lemma to the proof of the weak-type (1,1) bound by considering compact subsets of

$$E_\alpha = \{x : \mathcal{M}_{B_2} f(x) > \alpha\}.$$

Indeed, we let $\{B_{x,\alpha}\}$ be the collection of balls constructed prior to the displayed identity (2.2) and let K be a compact subset of E_α. By extracting a finite subcover and applying the finitary Vitali covering lemma, we obtain the estimate

$$m(K) \leq \frac{3^d}{\alpha} \|f\|_1.$$

The inner regularity of the Lebesgue measure now implies that

$$m(E_\alpha) \leq \frac{3^d}{\alpha} \|f\|_1.$$

Even so, the improved constant $A_{d,1,B_2} = 3^d$ iAsymptotic Bounds on the Optimal Constants of the Hardy-Littlewood Maximal Functions still exponential and grows quite fast as $d \to \infty$. So does the strong-type constant $A_{d,p,B}$, as it depends on $A_{d,1,B}$. The natural question to ask, therefore, is whether we could better control the constant $A_{d,1,B}$.

We pause for a moment to introduce a notation: $\|\mathcal{M}_B\|_{L^1(\mathbb{R}^d) \to L^{1,\infty}(\mathbb{R}^d)}$ denotes the smallest constant $A_{d,1,B}$ such that the inequality

$$m\left(\left\{x \in \mathbb{R}^d : \mathcal{M}_B f(x) > \alpha\right\}\right) \leq \frac{A_{d,1,B}}{\alpha} \|f\|_1$$

holds for all $f \in L^1(\mathbb{R}^d)$. Compare this notation with $\|\mathcal{M}_B\|_{L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)}$, which denotes the smallest constant $A_{d,p,B}$ such that the inequality

$$\|\mathcal{M}_B f\|_p \leq A_{d,p,B} \|f\|_p$$

holds for all $f \in L^p(\mathbb{R}^d)$. The strange symbol $L^{1,\infty}$ will be explained in the next set of notes.

Let us return to the topic at hand. The Vitali-covering-lemma argument yields the estimate

$$\|\mathcal{M}_B\|_{L^1(\mathbb{R}^d) \to L^{1,\infty}(\mathbb{R}^d)} \leq 3^d.$$

The first major improvement was given by E. M. Stein and J. O. Strömberg in [SS88]:

Theorem 5.2 (Stein–Strömberg, 1988; [SS88])**.** *There exists a constant $c > 0$ such that*

$$\|\mathcal{M}_B\|_{L^1(\mathbb{R}^d) \to L^{1,\infty}(\mathbb{R})^d} \leq c\, d \log d$$

for each $d > 1$ and every centrally symmetric convex body $B \subseteq \mathbb{R}^d$. Over the standard Euclidean ball B_2, we have the improved bound

$$\|\mathcal{M}_{B_2}\|_{L^1(\mathbb{R}^d)\to L^{1,\infty}(\mathbb{R}^d)} \leq c'\, d$$

for all $d \geq 1$, where $c' > 0$ is a constant that does not depend on d.

In the same paper, Stein and Strömberg conjectured the following:

Conjecture 5.3 (Stein–Stromberg maximal function conjecture, 1988; [SS88])**.** *For each fixed centrally symmetric convex body $B \subseteq \mathbb{R}^d$,*

$$\sup_{d\geq 1} \|\mathcal{M}_B\|_{L^1(\mathbb{R}^d)\to L^{1,\infty}(\mathbb{R}^d)} < \infty.$$

We remark that the following stronger conjecture, also by Stein and Strömberg, is false:

Conjecture 5.4 (Stein–Stromberg, 1988; [SS88])**.**

$$\sup_{d,B} \|\mathcal{M}\|_{L^1(\mathbb{R}^d)\to L^{1,\infty}(\mathbb{R}^d)} < \infty,$$

where the supremum is taken over all $d \geq 1$ and centrally symmetric convex bodies $B \subseteq \mathbb{R}^d$.

This is due to J. M. Aldaz:

Theorem 5.5 (Aldaz, 2011; [Ald11])**.** *Let B_∞ be the unit ball with respect to the l^∞-norm*

$$\|(x_1, \ldots, x_d)\|_\infty = \sup_{1\leq n\leq d} |x_n|.$$

Then[7]

$$\lim_{d\to\infty} \|\mathcal{M}_{B_\infty}\|_{L^1(\mathbb{R}^d)\to L^{1,\infty}(\mathbb{R}^d)} = \infty.$$

As of April 2014, the Stein–Stromberg maximal function conjecture is still unresolved. No significant improvement on the Euclidean-ball Stein–Strömberg bound has been given, and the best constant is only known in dimension 1:

Theorem 5.6 (Melas, 2003; [Mel03])**.**

$$\|\mathcal{M}\|_{L^1(\mathbb{R})\to L^{1,\infty}(\mathbb{R}^d)} = \frac{11+\sqrt{61}}{21}.$$

Indeed, some experts are starting to believe that the Stein–Stromberg maximal function conjecture might be false:

> These results suggest (at least to us) that uniform bounds...may fail to exist if one uses euclidean balls (the original question of Stein and Strömberg) since there seems to be no reason to believe that the maximal operator associated to euclidean balls is substantially smaller than the maximal operator associated to cubes. ([AL13], p.228.)

---

[7]In [Aub09], G. Aubrun established an improved lower bound for Aldaz's result: for each $0 < \varepsilon < 1$,

$$\|\mathcal{M}_{B_\infty}\|_{L^1(\mathbb{R}^d)\to L^{1,\infty}(\mathbb{R}^d)} \geq c_\varepsilon (\log d)^{1-\varepsilon}.$$

Amusingly, Auburn's improvement was published two years before Aldaz's breakthrough. I would assume this is because of the massive backlog at *Annals*.

In another direction, A. Naor and T. Tao showed in [NT10] that the convex-body Stein–Strömberg bound is essentially sharp in a large class of measure spaces. To state the result, we introduce a few notions. A *metric measure space* is a triplet (X, d_X, μ) consisting of a metric space (X, d_X) and a measure μ on X that is Radon with respect to the metric topology induced by d_X. We assume that $0 < \mu(B(x;r)) < \infty$ for all $x \in X$ and $r > 0$. Given $d \geq 1$, we say that a metric measure space (X, d, μ) is *Ahlfors–Davis d-regular with constant* $k > 0$ if

$$r^d \leq \mu(B(x;r)) \leq Cr^d$$

for all $X \in X$ and $r > 0$. This condition abstracts and generalizes the scaling property of the Lebesgue measure. Indeed, if $d = 1$, then we recover the *doubling condition*

$$\mu(B(x;2r)) \leq k(2r) \leq 2k\,\mu(B(x;r)),$$

which was crucial in establishing the Vitali covering lemma.

We now define the *Hardy–Littlewood maximal function on the metric measure space* (X, d, μ) by setting

$$\mathcal{M}f(x) = \sup_{r>0} \frac{1}{\mu(B(x;r))} \int_{B(x;r)} |f(y)|\, d\mu(y)$$

for all $f \in L^1(X, \mu)$. Powered by the theory of ultrametric spaces and probabilistic methods, Naor and Tao estbalished the following:

Theorem 5.7 (Naor–Tao, 2011; [NT10])**.** *If (X, d_X, μ) is an Ahlfors–Davis d-regular space with constant $k \geq 5(1 + \frac{1}{d})^{-d}$, then there exists a constant c_k that depends only on k such that*

$$\|\mathcal{M}\|_{L^1(X,\mu)\to L^{1,\infty}(X,\mu)} \leq c_k\, d \log d.$$

In addition, given $d \geq 2$, we can find an Ahlfors–Davis d-regular space (X_d, d_{X_d}, μ_d) with constant 81 such that

$$\|\mathcal{M}\|_{L^1(X_d,\mu_d)\to L^{1,\infty}(X_d,\mu_d)} \geq c'\, d \log d$$

for some universal constant c'. In particular, c' does not depend on (X_d, d_{X_d}, μ_d).

This, along with the counter-result of Aldaz, suggests that resolving the Stein–Strömberg maximal function conjecture requires carring out an intricate analysis of the Hardy–Littlewood maximal function on the Euclidean space that makes an explicit use of the specifics of the geometric structure of the Euclidean space. No argument that relies purely on the metric-measure structure of the Euclidean space will do.

What of the strong-type bounds? Once again, the first main result in this direction is due to Stein:

Theorem 5.8 (Stein, 1982; [Ste82])**.** *For each $1 < p \leq \infty$ and every centrally symmetric convex body $B \subseteq \mathbb{R}^d$,*

$$\sup_{d\geq 1} \|M_B\|_{L^p(\mathbb{R}^d)\to L^p(\mathbb{R}^d)} < \infty.$$

This result was then extended by J. Bourgain:

Theorem 5.9 (Bourgain, 1986; [Bou86b] & [Bou86c])**.** *For each $p > 3/2$,*

$$\sup_{d,p,B} \|\mathcal{M}_B\|_{L^p(\mathbb{R}^d)\to L^p(\mathbb{R}^d)} < \infty,$$

where the supremum is taken over all $d \geq 1$ and centrally symmetric convex bodies $B \subseteq \mathbb{R}^d$.

In view of Bourgan's result, it is reasonable to conjecture the following:

Conjecture 5.10 (Bourgain's maximal function conjecture, 1986; [Bou86a])**.** *For each $p > 1$,*

$$\sup_{d,p,B} \|\mathcal{M}_B\|_{L^p(\mathbb{R}^d)\to L^p(\mathbb{R}^d)} < \infty,$$

where the supremum is taken over all $d \geq 1$ and centrally symmetric convex bodies $B \subseteq \mathbb{R}^d$.

Two partial results are of note. Firstly, Bourgain completed his program for the l^∞-ball B_∞ in [Bou14]:

Theorem 5.11 (Bourgain, 2014; [Bou14])**.**

$$\sup_{d,p} \|\mathcal{M}_{B_\infty}\|_{L^p(\mathbb{R}^d)\to L^p(\mathbb{R}^d)} < \infty,$$

where the supremum is taken over all $d \geq 1$ and $p > 1$.

This result makes explicit use of the product structure of the cube and thus, in Bourgan's words, "does not immediately carry over to other convex symmetric bodies."

Secondly, D. Müller, in [M9̈0], considered additional geometric hypotheses that furnish dimension-independent bounds. To state the result precisely, we first note that every centrally symmetric convex body $B \subseteq \mathbb{R}^d$ of d-dimensional volume 1 admits an invertible linear transformation $T : \mathbb{R}^d \to \mathbb{R}^d$ that puts B in *isotropic position*, viz., there exists a constant $L(B)$ such that

$$\int_{TB} |\langle x, \xi\rangle|^2 \, dx = L(B)^2$$

for all $\xi \in \mathbb{S}^{d-1} \subseteq \mathbb{R}^d$.

Now, for each $\xi \in \mathbb{S}^{d-1}$, we define

$$\varphi_\xi(u) = \mathrm{Vol}_{d-1}\left(\{x \in TB : \langle x, \xi\rangle = u\}\right)$$

for all $u \in \mathbb{R}$, and let π_ξ be the orthogonal projection of $\mathbb{R}^d$ onto the hyperplane perpendicular in $\mathbb{R}^d$ to ξ. Set

$$\sigma(B) = \left(\max\{\varphi_\xi(0) : \xi \in \mathbb{S}^{d-1}\}\right)^{-1} \text{ and}$$
$$Q(B) = \max\{\mathrm{Vol}_{d-1}(\pi_\xi(TB)) : \xi \in \mathbb{S}^{d-1}\}.$$

The main result in [M9̈0] is as follows:

Theorem 5.12 (Müller, 1990; [M9̈0])**.** *There exists a uniform bound*

$$\|\mathcal{M}_B\|_{L^p(\mathbb{R}^d)\to L^p(\mathbb{R}^d)} \leq c_{p,\sigma(B),Q(B)}$$

for all $d \geq 1$, $p > 1$, and a centrally symmetric convex bodies $B \subseteq \mathbb{R}^d$, where the constant $c_{p,\sigma(B),Q(B)}$ that depend only on p, $\sigma(B)$, and $Q(B)$ In particular, the upper bound is independent of d and grows with σ and Q.

In [M9̈0], Müller shows that $\sigma(B)$ is bounded above and below by constant multiples of $L(B)$. Assuming the following conjecture, we see that the Müller bound depends only on p and $Q(B)$:

Conjecture 5.13 (Slicing conjecture)**.** *$L(B)$ is uniformly bounded, independent of the dimension d and the body B.*

On the other hand, Muüller proves in in Section 3 of [M$\ddot{9}$0] that

$$\|\mathcal{M}_{B_q^d}\|_{L^p(\mathbb{R}^d)\to L^p(\mathbb{R}^d)} \leq c_{p,q}$$

for all $p > 1$ and $q \geq 1$, where B_q^d is the d-dimensional unit ball with respect to the l^q-norm. Note that this does not include the $q = \infty$ case. Nevertheless, it is less clear if there should be a uniform control over $Q(B)$.

References

[AL13] J. M. J. M. Aldaz and Pérez Lázaro, *On high dimensional maximal operators*, Banach Journal of Mathematical Analysis **7** (2013), no. 2, 225–243.

[Ald11] J. M. Aldaz, *The weak type* $(1,1)$ *bounds for the maximal function associated to cubes grow to infinity with the dimension*, Annals of Mathematics **173** (2011), 1013–1023.

[Aub09] Guillaume Aubrun, *Maximal inequality for high-dimensional cubes*, Conuentes Mathematici **1** (2009), no. 2, 169–179.

[Aus13] Tim Austin, *Math-ga 2931.001: Ergodic theory (courant institute, nyu fall 2013)*, http://cims.nyu.edu/~tim/et2013.html, 2013.

[Bou86a] Jean Bourgain, *Geometry of banach spaces and harmonic analysis*, Proceedings of the International Congress of Mathematicians (1986), 871–878.

[Bou86b] ______, *On high dimensional maximal functions associated to convex bodies*, American Journal of Mathematics **108** (1986), 1467–1476.

[Bou86c] ______, *On the l^p-bounds for maximal functions associated to convex bodies in $\mathbb{R}^n$*, Israel Journal of Mathematics **54** (1986), no. 3, 257–265.

[Bou14] ______, *On the hardy-littlewood maximal function for the cube*, Israel Journal of Mathematics (2014).

[Cal68] Alberto P. Calderón, *Ergodic theory and translation-invariant operators*, Proceedings of the National Academy of Sciences of the United States of America **59** (1968), no. 2, 349–353.

[Isr14] Arie Israel, *Personal communication*, April 2014.

[M$\ddot{9}$0] Detlef Müller, *A geometric bound for maximal functions associated to convex bodies*, Pacific Journal of Mathematics **142** (1990), no. 2, 297–312.

[Mel03] Antonios D. Melas, *The best constant for the centered hardy-littlewood maximal inequality*, Annals of Mathematics **157** (2003), 647–688.

[MS13] Camil Muscalu and Wilhelm Schlag, *Classical and multilinear harmonic analysis*, Cambridge University Press, 2013.

[NT10] Assaf Naor and Terence Tao, *Random martingales and localization of maximal inequalities*, Journal of Functional Analysis **259** (2010), 731–779.

[SS88] Elias M. Stein and Jan-Olov Strömberg, *Behavior of maximal functions in $\mathbb{R}^n$ for large n*, Arkiv för Matematik **21** (1988), no. 2, 259–269.

[SS05] Elias M. Stein and Rami Shakarchi, *Real analysis: Measure theory, integration, and hilbert spaces*, Princeton University Press, 2005.

[Ste61] Elias M. Stein, *On limits of sequences of operators*, Annals of Mathematics **74** (1961), no. 1, 140–170.

[Ste70] ______, *Singular integrals and differentiability properties of functions*, Princeton University Press, 1970.

[Ste82] ______, *The development of square functions in the work of a. zygmund*, Bulletin of the American Mathematical Society **7** (1982), no. 2, 359–376.

[Tao] Terence Tao, *Math 247a: Fourier analysis*, http://www.math.ucla.edu/~tao/247a.1.06f/.

[Tao11] ______, *Stein's maximal principle*, http://terrytao.wordpress.com/2011/05/12/steins-maximal-principle/, 2011.

[Var01] S. R. S. Varadhan, *Probability theory*, vol. Courant Lecture Notes in Mathematics, American Mathematical Society, 2001.

[Wil91] David Williams, *Probability with martingales*, Cambridge University Press, 1991.