Chapter 2

On Oscillation and Convergence

A common tension that we are often faced with is the negotiation between accuracy and computability: well approximating complicated pieces of data by simpler data sets. For instance, when faced with an overwhelming collection of information, a common – and fundamental – question typically concerns the average behavior. The trade off is a total loss of fine scale approximation (a potentially large trade-off in accuracy), for a single piece of information which approximates the whole collection. Sometimes, this single piece of information is insufficient, in which case a natural way to proceed is to ask for *two* averages: over the first and second halves of the data sets. A slightly finer scale analysis is introduced, at the price of doubling the amount of information we are forced to record.

This process iterates naturally, with the expectation that sufficiently refined sampling will recover the initial data. Mathematically, these circle of ideas are captured by differentiation/integration theory.

For continuous functions on $f : [0,1] \to \mathbb{R}$, define the so-called *conditional expectation operators*, see Definition 3.39 below,

$$\mathbb{E}_{-k} f(x) := \sum_{|I|=2^{-k}} \left(2^k \int_I f(t)\, dt \right) \cdot \mathbf{1}_I, \tag{2.1}$$

where the sum runs over *dyadic* $I \in \mathcal{D}$ contained in $[0,1]$, see (1.38), i.e. intervals of the form

$$\{[\frac{a-1}{2^k}, \frac{a}{2^k}] : 1 \leq a \leq 2^k,\ a \in \mathbb{Z}\} = \{I \subset [0,1] : I \in \mathcal{D}(-k)\},$$

see (1.38). Note that we are now working with closed dyadic intervals, so the nesting property of dyadic intervals $P, Q \in \mathcal{D}$ instead becomes

$$P \cap Q \in \{\emptyset, P, Q\}$$

up to sets of measure zero, but the changes are otherwise formal. Each $\mathbb{E}_{-k}$ replaces the initial function, f, with its average over each dyadic interval of length 2^{-k}, "projecting" f onto the σ-algebra generated by

$$\{I \subset [0,1] : I \in \mathcal{D}(-k)\}$$

in the sense of (3.41).

Then, the fundamental theorem of calculus implies

$$\mathbb{E}_{-k} f \to f$$

pointwise for continuous $f \in \mathcal{C}([0,1])$.

The question presents: for which other classes of functions, f, can we make sense of the approximation

$$\mathbb{E}_{-k} f \quad \text{“=”} \quad f$$

for k sufficiently large? One way to make sense of this question is by approximation: we are free to compare f to any continuous g we wish:

$$|\mathbb{E}_{-k} f - f| \leq |\mathbb{E}_{-k} f - \mathbb{E}_{-k} g| + |\mathbb{E}_{-k} g - g| + |g - f|. \tag{2.2}$$

The second term, $|\mathbb{E}_{-k} g - g|$ is asymptotically negligible for continuous g. We also hope that we can approximate f by g well enough that $\varphi := (g - f)$ is "small" in an appropriate sense. The question presents:

Question 2.3. If φ is "small," can one ensure that $\{\mathbb{E}_{-k}\varphi\}$ are *all* small?

It turns out that this question is more naturally posed in slightly greater generality, in terms of *averaging* operators. If we let

$$\mathrm{Avg}_k f(x) := 2^{-k} \cdot \int_{-2^{k-1}}^{2^{k-1}} f(x-t)\, dt, \tag{2.4}$$

so $\mathrm{Avg}_k f$ is just given by $\left(2^{-k} \cdot \mathbf{1}_{B(2^{k-1})}\right) * f$, convolution with the L^1-normalized ball of radius 2^{k-1}, then

$$\mathrm{Avg}_{-k} \quad \text{“=”} \quad \mathbb{E}_{-k},$$

in that both operators "blur" at spatial scales 2^{-k}, discarding "fine scale" information below this threshold, and preserving "coarse scale" properties that can be detected above this spatial threshold, see (2.4) and (2.1). We explore this phenomenon in Examples 2.5 and 2.7, and in Exercise 2.9 below.

Example 2.5. Suppose that $|f'(x)| \leq 2^j$. By the Mean-Value Theorem, if $|x - x'| \leq c \cdot 2^{-j}$,

$$|f(x) - f(x')| \leq c,$$

so that if $k \geq j + O(1)$, both $\text{Avg}_{-k}, \mathbb{E}_{-k}$ "preserve" $f(x)$:

$$\text{Avg}_{-k} f(x),\ \mathbb{E}_{-k} f(x) = f(x) + O(2^{j-k}).$$

In particular,

$$\|\text{Avg}_{-k} f - \mathbb{E}_{-k} f\|_{L^\infty(\mathbb{R})} \lesssim 2^{j-k}. \tag{2.6}$$

Example 2.7. Suppose that $|\int_I f(x)\ dx| \leq c$ whenever $|I| \geq 2^{-j}$. Then whenever $k \leq j - O(1)$

$$\|\text{Avg}_{-k} f\|_{L^\infty(\mathbb{R})},\ \|\mathbb{E}_{-k} f\|_{L^\infty(\mathbb{R})} \leq 2^k \cdot c. \tag{2.8}$$

The following exercise is designed to reinforce the importance of Examples 2.5 and 2.7.

Exercise 2.9. Consider $f_j(x) := \sin(2\pi \cdot 2^j x) : [0,1] \to \mathbb{R}$. Evaluate $\mathbb{E}_{-k} f_j$ and $\text{Avg}_{-k} f_j$, and estimate $\|f_j\|_{L^2(\mathbb{R})}$, $\|(\mathbb{E}_{-k} - \text{Avg}_{-k}) f_j\|_{L^2(\mathbb{R})}$ for $|j - k| \gg 1$.
Hint: Try Taylor expanding, taking into account (2.6) *and* (2.8).

As we will see in Chapter 3 below, understanding Examples 2.5 and 2.7 are essentially sufficient to completely understand the behavior of the $\{\text{Avg}_{-k}\}$, the $\{\mathbb{E}_{-k}\}$, and the relationship between the two families of averaging operators.

Another property that is inherited by the Avg_{-k} is the everywhere convergence

$$\text{Avg}_{-k} g \to g \text{ everywhere} \tag{2.10}$$

for continuous g; this is also a consequence of the fundamental theorem of calculus. On the other hand,

$$\sup_{k \geq 1} |\mathbb{E}_{-k} f| \lesssim M_0 f,$$

where

$$M_0 f := \sup_{k \geq 0} \text{Avg}_{-k} |f| \tag{2.11}$$

for some absolute constant C, so that quantitative control over M_0 is the key to answering Question 2.3. In particular, the key to the convergence $\text{Avg}_{-k} f \to f$ is an appropriate estimate on M_0.

To motivate what follows, it is helpful to explore this question in the contrapositive. The following argument, which provides precise testing conditions under which an analogue of (2.10) generalizes, is due to Stein, [**164**].

Theorem 2.1. *Suppose that for each $f \in L^p([0,1])$, $Avg_{-k} f$ converges on a set F of positive measure. Then the maximal function, $M_0 f$, see* (2.11), *is of weak type (p,p): there exists some absolute constant A so that*

$$\|M_0 f\|_{L^{p,\infty}(\mathbb{R})} := \sup_\lambda \lambda \cdot |\{x : M_0 f(x) > \lambda\}|^{1/p} \leq A \cdot \|f\|_{L^p(\mathbb{R})}. \tag{2.12}$$

Although the conclusion of Theorem 2.1 is rather dramatic, it is entirely sharp, see Corollary 1.21.

Lemma 2.2. *Suppose that*

$$\|M_0 f\|_{L^{p,\infty}(\mathbb{R})} \leq A \cdot \|f\|_{L^p(\mathbb{R})}$$

for some $0 < A < \infty$. Then $Avg_{-k} f \to f$ almost everywhere for each $f \in L^p(\mathbb{R})$, $1 \leq p < \infty$.

In particular, the upshot of this discussion is that, by Exercise 2.23 below, one of exactly two conditions hold.

- M_0 is of weak-type (p,p), and $\text{Avg}_{-k} f \to f$ almost everywhere for each $f \in L^p(\mathbb{R})$; or
- M_0 is not of weak-type (p,p), and there exists some $f \in L^p(\mathbb{R})$ so that $\limsup \text{Avg}_{-k} f = \infty$ almost everywhere.

In other words, *quantitative* estimates on M_0 are equivalent to the *qualitative* property of convergence. A large focus of this book will be on establishing similar quantitative estimates; the key to making sense of these estimates (and their proofs) is to understand the deeper qualitative message that our estimates describe, in a personally meaningful way; my perspective on the weak boundedness of M_0 is a statement that $L^p(\mathbb{R})$-functions cannot oscillate at all (fine) scales at (essentially) any location, see Exercise 2.9.

And with these remarks in mind, we turn to the proof of Theorem 2.1.

0.1. The Proof of Theorem 2.1. Theorem 2.1 is established by extracting a family of functions $\{f_n\}$ so that each $\{M_0 f_n\}$ grows increasingly large, and then using the probabilistic method to generate an appropriate amount of constructive interference among the $\{f_n\}$. Accordingly, we begin with the following Euclidean analogue of the Borel-Cantelli Lemma from probability theory.

Lemma 2.3. *Suppose that $\{E_n\} \subset [0,1]$ area collection of measurable sets so that $\sum_n |E_n| = \infty$. Then there exists a collection of shifts, $\{h_n\}$, so that*

$$\lim_{k\to\infty} \bigcap_{n\geq k} \Big(\bigcup_{m\geq n} (E_n + h_n) \Big) = \mathbb{R}. \tag{2.13}$$

Proof. We prove the following weaker claim: there exist shifts $\{h_n\}$ so that

$$[0,1] \subset \bigcup_n (E_n + h_n) \tag{2.14}$$

up to sets of measure zero; this statement will be upgraded to (2.13) in Exercise 2.18 below.

Assuming the result of Exercise 2.19, which asserts that two generic sets can always be significantly "mixed" via an appropriate translation, we use a volume packing argument to establish (2.14).

We do this in a two step process:

Initiate at time 1, and set $F_1 := E_1$.

At time 2, let $A_2 := [0,1] \smallsetminus F_1$, and choose an h_2 so that

$$\Big|\big(E_2 + h_2\big) \cap A_2\Big| \geq 1/2 \cdot |E_2| \cdot |A_2|;$$

set $F_2 := E_2 + h_2$.

Assuming that $\{F_1, \dots, F_{n-1}\}$ have been selected, set

$$A_n := [0,1] \smallsetminus \bigcup_{m<n} F_m,$$

and choose h_n so that

$$\Big|\big(E_n + h_n\big) \cap A_n\Big| \geq 1/2 \cdot |E_n| \cdot |A_n|, \tag{2.15}$$

and set $F_n := E_n + h_n$.

The argument is concluded once we verify that

$$\Big| \bigcup_{m\leq n} F_m \Big| \to 1.$$

But we decompose

$$\begin{aligned} \Big| \bigcup_{m\leq n} F_m \Big| &= \Big| \bigcup_{m<n} F_m \Big| + |F_n \cap A_n| \\ &\geq \Big| \bigcup_{m<n} F_m \Big| + 1/2 \cdot |F_n| \cdot |A_n| \qquad (2.16) \\ &= \Big| \bigcup_{m<n} F_m \Big| + 1/2 \cdot |F_n| \cdot \Big(1 - \Big| \bigcup_{m<n} F_m \Big| \Big), \qquad (2.17) \end{aligned}$$

where we used (2.15) to pass to (2.16). If we set

$$a_n := \Big| \bigcup_{m \le n} F_m \Big|,$$

then by (2.17) we have

$$1 \ge \sum_{n=1}^{N} a_n - a_{n-1} \ge 1/2 \cdot \sum_{n \le N} |F_n| \cdot (1 - a_{n-1}),$$

which forces $a_n \to 1$, by the divergence, $\sum_{n \le N} |F_n| \to \infty$. The proof is complete up to Exercises 2.18 and 2.19 below. □

Exercise 2.18. Extend the conclusion (2.14) to (2.13).
Hint: Try sparsifying the $\{E_n\}$.

Exercise 2.19. Prove that for any two sets $E, F \subset [0,1]$, there exists some $|h| \le 1$ so that

$$\Big| (E + h) \cap F \Big| \ge 1/2 \cdot |E| \cdot |F|.$$

Hint: Consider the average behavior of $\mathbf{1}_E * \mathbf{1}_F$*; note that this function is supported inside* $[0,2]$.

With Lemma 2.3 in hand, we turn to the proof of Theorem 2.1. The key property of the $\{\mathrm{Avg}_{-k}\}$ that we will exploit – as opposed to the operators $\{\mathbb{E}_{-k}\}$ – is that the Avg_{-k} are *translation invariant.*

Exercise 2.20. Prove that for any f, and $k \ge 1$, and any shift, $h \in \mathbb{R}$, if we let $T_h g(x) := g(x - h)$, then

$$T_h \mathrm{Avg}_{-k} f = \mathrm{Avg}_{-k}(T_h f).$$

Prove that the $\{\mathbb{E}_{-k}\}$ do *not* satisfy this property.
Hint: For the second point, try shifting $\mathbf{1}_{[0,1]}$ *one unit in the negative direction, see Example* 1.41.

Proof of Theorem 2.1. We assume that (2.12) fails, and construct a function, f, so that $M_0 f = \infty$ almost everywhere, which will imply the result (why?). Notice that since we are considering only functions $f : [0,1] \to \mathbb{R}$, we see that $M_0 f$ is supported on $[-1, 2]$.

Since (2.12) fails, we can extract a sequence of functions, $\{f_k\}$, and altitudes, $\{\lambda_k\}$ so that

$$|\{x \in [-1,2] : M_0 f_k > \lambda_k\}| \ge 2^k \cdot \frac{\|f_k\|_{L^p(\mathbb{R})}^p}{\lambda_k^p}.$$

Set $g_k := k/\lambda_k \cdot f_k$, so that

$$3 \geq |E_k| := |\{x \in [-1,2] : M_0 g_k > k\}| \geq 2^{k/2} \cdot \|g_k\|_{L^p(\mathbb{R})}^p.$$

Let a_k to be a constant multiple of

$$2^{-k/4} \cdot \|g_k\|_{L^p(\mathbb{R})}^{-p},$$

and construct a new sequence,

$$\bigcup_{k \geq 1} \{G_{i_k} : i_k \leq a_k\},$$

so that $G_{i_k} = g_k$. We observe that

$$\sum_i \|G_i\|_{L^p(\mathbb{R})}^p = \sum_k \sum_{i_k} \|G_{i_k}\|_{L^p(\mathbb{R})}^p \tag{2.21}$$

$$= \sum_k a_k \cdot \|g_k\|_{L^p(\mathbb{R})}^p \approx \sum_k 2^{-k/4} \lesssim 1, \tag{2.22}$$

while

$$\sum_k \sum_i |\{x \in [-1,2] : M_0 G_{i_k} > k\}| \geq \sum_k a_k \cdot 2^{k/2} \cdot \|g_k\|_{L^p(\mathbb{R})}^p \approx \sum_k 2^{k/4},$$

which diverges. Set

$$H_{i_k} := \{x \in [-1,2] : M_0 G_{i_k} > k\},$$

and choose now a collection of shifts, $\{h_{i_k}\}$ furnished by Lemma 2.3, so that

$$\bigcap_{m \geq n} \bigcup_{i_k \geq m} H_{i_k} \supset [-1,2]$$

and set

$$F_{i_k}(x) := G_{i_k}(x + h_{i_k}).$$

We will apply M_0 to

$$F_\infty := \sup_{i_k} |F_{i_k}|,$$

which we see is in $L^p(\mathbb{R})$ by (2.22):

$$F_\infty \leq \left(\sum_{i_k} |F_{i_k}|^p \right)^{1/p}.$$

But we bound

$$M_0 F_\infty \geq \sup_{i_k} M_0 F_{i_k} \geq \sup_{i_k} k \cdot \mathbf{1}_{H_{i_k}} = +\infty$$

on $[-1,2]$. □

We conclude this section by showing that local control of M_0 leads to global estimates.

Exercise 2.23. Assuming that

$$\|M_0 f\|_{L^{p,\infty}(\mathbb{R})} \leq A \cdot \|f\|_{L^p(\mathbb{R})}$$

for each $f : [0,1] \to \mathbb{R}$, show that

$$\|M_0 f\|_{L^{p,\infty}(\mathbb{R})} \lesssim A \cdot \|f\|_{L^p(\mathbb{R})}$$

for a general $f \in L^p(\mathbb{R})$.
Hint: Consider functions of the form $\sum_{n \equiv j \mod 10} f \cdot \mathbf{1}_{[n,n+1]}$ *for some* $1 \leq j \leq 10$.

Exercise 2.24. Assuming that

$$\|M_0 f\|_{L^{p,\infty}(\mathbb{R})} \lesssim A \cdot \|f\|_{L^p(\mathbb{R})},$$

prove that M_{HL} is of weak-type (p,p).
Hint: Use monotone convergence to restrict to finitely many scales in the definition of M_{HL} *and then rescale, exploiting dilation symmetry as per Lemma 1.14.*

Chapter 3

The Linear Theory

1. The Pointwise Ergodic Theorem

Pointwise ergodic theory, the motivation for discrete harmonic analysis, has at its roots the classical theorem of Birkhoff [**12**], discussed below, which can be described as follows:

> For every ergodic – that is, "sufficiently randomizing" – measure-preserving transformation, τ, of a probability space, (X, μ), and any integrable function $f \in L^1(X, \mu)$, μ-almost surely, one can recover the mean of f by considering the Cesaro sums
>
> $$\frac{1}{N} \sum_{n \leq N} f(\tau^n x) \to \int f \; d\mu \;\; \mu - \text{a.e.}$$

Informally, this theorem says that one can recover the *spatial mean* of f,

$$\int f \; d\mu,$$

by considering the *temporal means*

$$\{\frac{1}{N} \sum_{n \leq N} f(\tau^n x)\},$$

formed by "sampling" the function f at the "times" $\{\tau^n x\}$ and taking the appropriate average.[1]

[1] Even in the case when τ is not ergodic, the temporal means $\{\frac{1}{N} \sum_{n \leq N} \tau^n f(x)\}$ still converge μ-almost everywhere.

In this chapter, we offer several proofs of this theorem. Along the way, we will introduce and develop many of the tools used in discrete harmonic analysis.

Below, (X, μ, τ) will be a *measure-preserving system*, a σ-finite space, equipped with τ a measure preserving $\mathbb{Z}^D$-action,

$$\tau^y f(x) := f(\tau^y x),$$

see (0.8). For $\{E_n\} \subset \mathbb{Z}^D$, define the averaging operators

$$M_n^\tau f(x) := \frac{1}{|E_i|} \sum_{y \in E_n} (\tau^y f)(x).$$

A major question in ergodic theory is to what extent one can make sense of the limiting behavior of the means $\{M_n^\tau f\}$ as $n \to \infty$. Because we will be interested in the pointwise regime, we introduce the following slightly nonstandard notion.

Definition 3.1. A collection of subsets $\{E_n\} \subset \mathbb{Z}$ is said to be *universally L^p-good* if for every measure-preserving system (X, μ, τ) and every $f \in L^p(X)$, the means $\{M_n^\tau f\}$ converge *pointwise* μ-a.e.

With the above definition in mind, we are prepared to introduce and discuss the classical pointwise ergodic theorem of Birkhoff [**12**]; we encourage the reader to do so while bearing in mind the connection between *quantitative estimates* and *qualitative meaning.*

This chapter will be centered on the one-dimensional theory, but many of the results discussed below extend to the higher dimensional setting. Below, we will sometimes appeal to the higher-dimensional formulations, so when necessary, we will include the details behind the pertaining generalizations.

2. Birkhoff's Theorem

Theorem 3.1 (Pointwise Ergodic Theorem)**.** *Suppose that (X, μ) is a probability space. The collection of intervals $\{[0, n)\}$ is universally L^1-good.*

We first present the standard density-argument approach to this theorem; we will relax the finite measure assumption in Exercise 3.24 below.

Proof. One begins by considering the subset of $L^1(X)$ functions on which pointwise convergence is clear:

(3.2)
$$\{\phi \in L^\infty(X) : \tau\phi = \phi \ \mu\text{-a.e.}\} \oplus \text{Span}\{h - \tau h : h \in L^\infty(X)\} \subset L^1(X).$$

Indeed, if ϕ is fixed by the action, $M_N^\tau \phi = \phi$, while if $f = h - \tau h$ is a so-called *coboundary*, then by telescoping

$$M_N^\tau f = M_N^\tau (h - \tau h) = \frac{1}{N} \sum_{n=0}^{N-1} \tau^n \big(h - \tau h\big)$$
$$= \frac{1}{N}(h - \tau^N h) = O\big(\frac{\|h\|_{L^\infty(X)}}{N}\big).$$

The next claim is that this set is dense in $L^2(X)$, and hence in $L^1(X)$ as well, as Exercise 3.5 will ask you to demonstrate below. So, to complete our density argument, we just need to show that (3.2) is dense in $L^2(X)$:

Exercise 3.3. Use the identity

$$\|f - \tau f\|_{L^2(X)}^2 = 2\text{Re}\ \langle f, (f - \tau f)\rangle,$$

to establish that

$$L^2(X) = \{f \in L^2 : f = \tau f\} \oplus \overline{\text{Span}\{g - \tau g : g \in L^2(X)\}}; \tag{3.4}$$

then conclude the claim.
Hint: Proceed by contradiction and assume that there exist some nonzero h that is perpendicular to the right hand side of (3.4).

Up to Exercise 3.5, the proof is complete. □

Exercise 3.5. Suppose that (X, μ) is a finite measure space. Prove that $L^2(X) \subset L^1(X)$: square integrable functions are integrable. Contrast this to the general case where $\mu(X) = \infty$, see (1.61).

Exercise 3.6. A measure-preserving transformation $\tau : X \to X$ is said to be *ergodic* if the only τ-invariant subsets of X are either null or co-null:

$$\int |\mathbf{1}_E(x) - \mathbf{1}_E(\tau x)|\ d\mu(x) = \mu(E \triangle \tau^{-1} E) = 0 \iff \mu(E) \in \{0, 1\}, \tag{3.7}$$

where

$$A \triangle B := \big(A \setminus B\big) \cup \big(B \setminus A\big)$$

is the symmetric difference. Use this to show that if τ is ergodic, then

$$M_n^\tau f \to \int f\ d\mu$$

almost everywhere.
Hint: If $h = \tau h$ *almost everywhere, then for every altitude* $\lambda > 0$,

$$E_\lambda := \{x : h(x) > \lambda\}$$

is invariant up to null sets; it suffices to show that whenever h is non-constant, there exists some $\lambda > 0$ so that $0 < \mu(E_\lambda) < 1$.

The upshot is that we have proven pointwise convergence of the ergodic averages on a dense subclass of $L^1(X)$, so to complete the proof we need only show that set of functions for which we have pointwise convergence of ergodic averages is *closed* in $L^1(X)$. By combining Stein's maximal principle [**164**], see Theorem 2.1 in Chapter 2 above for a special case, with a cohomology argument due to Halmos on the weak density of ergodic actions in the class of measure preserving transformations [**79**], the only way to do this is to prove a weak-type one-one estimate on the maximal function,

$$M_\tau f := \sup_n |M_n^\tau f|.$$

This is an abstract problem, and the first proof we present will rely on abstract *semi-group* arguments. The second, more modern argument that we present uses Calderón's transference principle [**34**] to transfer this problem to one on the integer lattice.

Proposition 3.2. *For each $f \in L^1(X)$, we have $\|M_\tau f\|_{L^{1,\infty}(X)} \leq \|f\|_{L^1(X)}$.*

Assuming this proposition, the proof is completed by a density argument, see Lemma 1.20.

Exercise 3.8. Provide the details of this density argument.
Hint: For $f \in L^1(X)$, we may decompose

$$f = \varphi^{(N)} + \big(h^{(N)} - \tau h^{(N)}\big) + \eta^{(N)}$$

where $\varphi^{(N)}$ are fixed by τ, $h^{(N)} \in L^\infty(X)$, and $\|\eta^{(N)}\|_{L^1(X)} \leq 2^{-N}$. Then argue as in the proof of Corollary 1.21: bound

$$\limsup |M_n^\tau f - M_m^\tau f| \leq \limsup |M_n^\tau \eta^{(N)} - M_m^\tau \eta^{(N)}| \leq 2M_\tau \eta^{(N)},$$

and use this to estimate the size of the super-level sets

$$\Big\{ \limsup_{n,m\to\infty} |M_n^\tau f - M_m^\tau f| > \epsilon \Big\}.$$

The first proof of Proposition 3.2 follows from the Hopf-Dunford-Schwartz maximal theorem for semigroups [**58**]; this argument requires that we restrict to probability spaces.

Proposition 3.3. *Suppose that (X, μ) is a probability space, and that $P : L^1(X) \to L^1(X)$ is a linear operator with $P1 = 1$ an $P^*1 = 1$ where P^* is the adjoint operator. If P preserves positivity, then the maximal function*

$$P_* f := \sup_{N \geq 1} \frac{1}{N} \sum_{n=0}^{N-1} P^n f \tag{3.9}$$

obeys the inequality

$$\sup_{\lambda \geq 0} \lambda \cdot \mu(\{P_* f > \lambda\}) \leq \int_{\{P_* f > \lambda\}} f \, d\mu. \tag{3.10}$$

Exercise 3.11. Verify that the condition $P^*1 = 1$ is the same as $\int_X Pf \, d\mu = \int_X f \, d\mu$.

Before beginning the proof proper, we establish a preliminary reduction:

Exercise 3.12. By expressing $P_* f - \lambda = P_*(f - \lambda)$, reduce (3.10) to the case where $\lambda = 0$.

Proof of Proposition 3.3. Our task is to show that

$$0 \leq \int_{\{P_* f > 0\}} f \, d\mu.$$

For each $M \geq 1$, consider the maximal function

$$F_M := \sup_{0 \leq N \leq M} \sum_{n=0}^{N-1} P^n f,$$

where we interpret $\sum_{n=0}^{-1} P^n f = 0$ to be the empty sum; by monotone convergence, it suffices to show that for all M large,

$$0 \leq \int_{\{F_M > 0\}} f \, d\mu.$$

By Exercise 3.14 below, we observe the following recursive structure of the $\{F_M\}$:

$$F_M \leq F_{M+1} = \max\{0, f + PF_M\}. \tag{3.13}$$

Integrating this:

$$\begin{aligned} \int_X F_M = \int_{\{F_M > 0\}} F_M &\leq \int_{\{F_M > 0\}} f + \int_{\{F_M > 0\}} PF_M \\ &\leq \int_{\{F_M > 0\}} f + \int_X PF_M, \end{aligned}$$

using the positivity of P for the last inequality. By Exercise 3.11, we have shown that

$$\int_X F_M \leq \int_{\{F_M>0\}} f + \int_X F_M,$$

and since $\int_X F_M \lesssim_M 1$ is finite, we may subtract to show that

$$0 \leq \int_{\{F_M>0\}} f,$$

completing the proof. □

Exercise 3.14. Verify the equality (3.13).

We leave the proof of Proposition 3.2 to the following exercise.

Exercise 3.15 (Proof of the Proposition 3.2 for Probability Spaces)**.** Specialize Proposition 3.3 to the case where P is given by the shift operator τ to conclude Proposition 3.2.

The second proof we provide is more robust, and uses transference to a "universal" dynamical system: $(X, \tau, \mu) = (\mathbb{Z}, n \mapsto n-1, |\cdot|)$. It holds in the full generality of the σ-finite measure-preserving setting.

Proof of the Proposition 3.2, Take Two. By monotone convergence, it's enough to study the (re-defined) maximal function

$$M_\tau f := \sup_{k \leq N} |M_k^\tau f|,$$

provided that our estimates are all independent of N. Let $R \gg N$ be a large parameter, which will eventually be sent to ∞, and define the auxilary function on $\mathbb{Z}$ by

$$F(x;n) := \begin{cases} f(\tau^n x) & \text{if } n \leq R \\ 0 & \text{else.} \end{cases}$$

If we let M_k denote the analogous discrete averaging operators,

$$M_k g(m) := \frac{1}{k} \sum_{n=0}^{k-1} g(m-n), \tag{3.16}$$

then the key to the argument is the following observation:

Exercise 3.17. Suppose that $n \leq R - N$; prove that

$$M_k^\tau f(\tau^n x) \equiv M_k F(x;n)$$

for all $i \leq N$.

Next, we let

$$M_+ g := \sup_k |M_k g| \tag{3.18}$$

denote the one-sided discrete Hardy-Littlewood maximal function and collect the super-level sets

$$E_\lambda := \{M_+ f > \lambda\}.$$

The following lemma will allow us to conclude the proof.

Lemma 3.4. *Let $\lambda > 0$ be an arbitrary altitude. Then for any $f \in \ell^1(\mathbb{Z})$,*

$$\lambda \cdot |E_\lambda| < \sum_{n \in E_\lambda} |f(n)|. \tag{3.19}$$

In particular, for any $f \in \ell^1(\mathbb{Z})$,

$$\|M_+ f\|_{\ell^{1,\infty}(\mathbb{Z})} \leq \|f\|_{\ell^1(\mathbb{Z})}.$$

Since we have defined

$$M_\tau f(x) := \sup_{k \leq N} |M_k^\tau f|,$$

we have the pointwise inequality,

$$M_\tau f(\tau^n x) \leq M_+ F(x; n)$$

valid for all $n \leq R - N$.

Now, using the measure-preserving nature of τ, Fubini's theorem and the weak-type one-one boundedness of M_+, we may estimate

$$\begin{aligned}(R - N) \cdot \mu(\{M_\tau f > \lambda\}) &= \int_X \sum_{n \in [0, R-N)} \mathbf{1}_{\{(x,n): M_\tau f(\tau^n x) > \lambda\}} \, d\mu(x) \\ &\leq \int_X \sum_{n \in [0, R-N)} \mathbf{1}_{\{(x,n): M_+ F(x;n) > \lambda\}} \, d\mu(x) \\ &\leq \int_X |\{n : M_+ F(x; n) > \lambda\}| \, d\mu(x) \\ &\leq \frac{1}{\lambda} \int_X \|F(x; n)\|_{\ell^1(n)} \, d\mu(x).\end{aligned}$$

By a further application of Fubini's theorem, and taking into account the measure-preserving nature of τ once again, we obtain the bound

$$\begin{aligned}(R - N) \cdot \mu(\{M_\tau f > \lambda\}) &\leq \frac{1}{\lambda} \int_X \|F(x; n)\|_{\ell^1(n)} \, d\mu(x) \\ &= \frac{1}{\lambda} \sum_n \int_X |F(x; n)| \, d\mu(x) \\ &= \frac{R}{\lambda} \cdot \|f\|_{L^1(X)}.\end{aligned}$$

Dividing through by $R-N$ and sending $R \to \infty$ yields the desired weak-type inequality

$$\mu(\{M_\tau f > \lambda\}) \leq \frac{1}{\lambda} \cdot \|f\|_{L^1(X)}.$$

□

It remains only to prove Lemma 3.4, reproduced below for the reader's convenience:

Lemma 3.5. *Let $\lambda > 0$ be an arbitrary altitude and collect*

$$E_\lambda := \{M_+ f > \lambda\}.$$

Then for any $f \in \ell^1(\mathbb{Z})$,

$$\lambda \cdot |E_\lambda| < \sum_{n \in E_\lambda} |f(n)|,$$

see (3.19)*. In particular, for any $f \in \ell^1(\mathbb{Z})$,*

$$\|M_+ f\|_{\ell^{1,\infty}(\mathbb{Z})} \leq \|f\|_{\ell^1(\mathbb{Z})},$$

see (3.18)*.*

Proof. We may assume that $f \geq 0$ is non-negative; we also will assume that f is finitely supported, as the general case follows from a standard limiting argument.

We decompose $E_\lambda = \bigcup_j [a_j, b_j)$ as a disjoint union of maximal intervals with respect to containment, all of which are finite. Now, suppose we knew that for each j,

$$\lambda \cdot \big|[a_j, b_j)\big| < \sum_{[a_j,b_j)} f;$$

a simple summation over our intervals j would yield the result.

So, seeking a contradiction, suppose that for some k, the reverse inequality held:

$$\sum_{[a_k,b_k)} f \leq \lambda \cdot \big|[a_k, b_k)\big|.$$

For notational ease, relabel $a_k = a$, and $b_k = b$, and for $x \in [a, b)$, define the "look ahead" function

$$l(x) := \max\Big\{t : \sum_{[x,t)} f > \lambda \cdot [x, t)|\Big\}.$$

Exercise 3.20. By proving that $l(a) \leq b-1$, derive the desired contradiction and complete the proof.
Hint: Consider averages from a to $l\big(l(a)\big)$.

□

Remark 3.21. The proof of the above proposition is extendable to proving strong-type L^p-bounds on M_τ as well. Moreover, the argument relied on two – and only two – properties of the operator M_τ. Firstly, that M_τ is sublinear; secondly, that M_τ is *transferrable* to an operator on the integer lattice, M_+, which is *semi-local*, i.e. a pointwise limit of operators which extend the domain of support of finitely-supported functions on the lattice in a bounded way. A further inspection of the above argument reveals that in no place did we exploit the fact that we were working on a space of dimension one: the above argument extends readily to the case of $\mathbb{Z}^D$-actions by transference to the $\mathbb{Z}^D$-lattice; in fact, this principle extends to *amenable* groups, with *Fölner sequences* replacing intervals.

The above claims will be established in the following two exercises.

Exercise 3.22. By first establishing that M_+ is $\ell^p(\mathbb{Z})$ bounded, prove that M_τ is bounded on $L^p(X)$.

Exercise 3.23. Let $E_k := [-k,k]^D \subset \mathbb{Z}^D$ be a sequence of cubes and establish $L^p(X)$ estimates for $M_\tau f$ via transference.

Before proceeding, we remark that Birkhoff's theorem holds for σ-finite measure spaces:

Exercise 3.24. Prove the pointwise ergodic theorem for σ-finite measure spaces.
Hint: First establish convergence for functions of the form

$$\{f : \tau f = f \ \mu\text{-a.e.}\} \oplus \{h - \tau h : h \in L^2(X) \cap L^\infty(X)\}$$

and then use a density argument.

Exercise 3.25. On σ-finite measure spaces, there is, in general, no containment $L^p(X) \subset L^q(X)$ for $p \neq q$, and so the conclusion of the pointwise ergodic theorem for $L^1(X)$-functions does *not* imply the corresponding conclusion for $L^p(X)$-functions for $p \neq 1$. First, provide an example of a space X, and a collection of functions $\{f_p\}$ so that

$$f_p \in L^p(X) \smallsetminus \bigcup_{1 \leq q \neq p \leq \infty} L^q(X).$$

Then establish the conclusion of the pointwise ergodic theorem for functions in $L^p(X),\ 1 \leq p < \infty$, for σ-finite measure spaces, X.
Hint: For the first point, let $X = \mathbb{R}$ *and consider sums of functions of the form* $\frac{1}{|x|^a \cdot \log^A |x|} \cdot \mathbf{1}_{|x| \leq 1},\ \frac{1}{|x|^b \cdot \log^B |x|} \cdot \mathbf{1}_{|x| \geq 1}.$

The density argument used above is rather slick, but unfortunately not as robust as might be hoped: it relies crucially on the "smoothness" of the intervals $\{[0, n)\}$.[2]

Obtaining pointwise convergence results for $\{M_n^\tau f\}$ for rougher, exotic $\{E_n\} \subset \mathbb{Z}$ does not necessarily follow from quantitative estimates on an appropriate maximal function, since the dense-subclass result is often unavailable in this setting. Perhaps the most famous instance of this difficulty arose in the study of averages along the squares, i.e.

$$E_i := \{1, 4, 9, \dots, i^2\} \subset \mathbb{Z}.$$

Indeed, to prove pointwise convergence of the ergodic averages of $L^2(X)$-functions along the squares, Bourgain [**19**] attacked the issue of oscillation more directly, by showing that an appropriate *oscillation operator* was $L^2(X)$-"controlled" ([**19**, §7]) in the sense outlined in the below exercise.

Exercise 3.26. Suppose that $\{f_n\}$ are a collection of $L^2(X)$-functions defined on some measure space (X, μ) which satisfy $\|f_n\|_{L^2(X,\mu)} \leq 1$. Suppose that for any subsequence $\{N_j\}$

$$\frac{1}{J} \int_X \sum_{j=1}^{J} \sup_{N_j \leq n < N_{j+1}} |f_n - f_{N_j}|^2 \, d\mu = o_{J\to\infty}(1). \tag{3.27}$$

Prove that $\{f_n\}$ converge pointwise.
Hint: Proceed by contradiction: assume that there exists some $\epsilon > 0$ *so that*

$$\mu\big(\{\limsup_{n,m} |f_n - f_m| > \epsilon\}\big) > \epsilon,$$

and use this to extract a sequence $\{N_j\}$ *so that, e.g.*

$$\mu\big(\{\sup_{N_j \leq n < N_{j+1}} |f_n - f_{N_j}| > \epsilon/2\}\big) > \epsilon/2.$$

Then apply Chebyshev's inequality to derive a J*-independent lower bound for the left hand side of* (3.27)*.*

Exercise 3.28. Suppose that $f_n = A_n f = A_n(\tau) f : X \to \mathbb{C}$ is given by a finite linear combination of

$$\{1, \tau, \tau^2, \dots\}$$

for some measure-preserving transformation $\tau : X \to X$. By using Exercise 3.26, show that pointwise convergence of $\{A_n f\}$ can be deduced from a quantitative estimate on $\mathbb{Z}$.

[2]Most precisely: $\{[0, n) : n\}$ form a *Fölner sequence* inside of $\mathbb{Z}$. For a discussion of the role of Fölner sequence in pointwise ergodic theory of *amenable groups*, see [**70**].

In a reworking of his argument [**21**], in which he handled more general polynomial sequences

$$E_i := \{P(1), \ldots, P(i)\} \subset \mathbb{Z}, \ P(x) \in \mathbb{Z}[x],$$

Bourgain did so with the assistance of the *r-variation operators* (below), classically used in probability theory to gain quantitative information on the rates of convergence.

3. Introduction to Variation

Definition 3.29. For a collection of scalars $\{a_i\}$,

$$\mathcal{V}^r(a_i) := \sup \Big(\sum_k |a_{i_k} - a_{i_{k+1}}|^r \Big)^{1/r}$$

is the *r-variation* of the $\{f_i\}$, $0 < r < \infty$; the supremum is taken over finite, increasing subsequences $\{i_k\}$. We define $\mathcal{V}^\infty(a_i) := \sup_{s,t} |a_s - a_t|$.

Definition 3.30. For a collection of functions $\{f_i\}$, define the *r-variation* of the $\{f_i\}$ to be the function evaluated by taking the r-variation of the sequence $\{f_i(x)\}$,

$$\mathcal{V}^r(f_i)(x) := \mathcal{V}^r\big(f_i(x)\big)$$

for $1 \le r \le \infty$.

The first observation concerning the variation operators is contained below.

Exercise 3.31. Prove that $\mathcal{V}^r(f_i)$ is pointwise *decreasing* in r.

Given the close connection between $\mathcal{V}^\infty$ and the maximal function, it is perhaps unsurprising that these variation operators are more difficult to control than the maximal function $\sup_i |f_i|$:

Exercise 3.32. For any j, establish the following inequality:

$$\sup_i |f_i| \le \mathcal{V}^\infty(f_i) + |f_j| \le \mathcal{V}^r(f_i) + |f_j|,$$

where $r < \infty$ is arbitrary.

An important example considers the variation of a particular function under the image of the conditional expectation operators, $\{\mathbb{E}_k\}$, i.e. when $f_k := \mathbb{E}_k f$ for some given f; see (2.1). The following computation is accordingly instructive.

Example 3.33. For any $0 < r < \infty$,

$$\sup_{\{k_i\}\subset\mathbb{Z}:k_i\geq K}\big(\sum_i |2^{-k_i}-2^{-k_{i+1}}|^r\big)^{1/r} \leq \sum_{k\geq K}|2^{-k}-2^{-k-1}| \leq 2^{-K},$$

see Exercise 3.31. Consequently, if $f = \mathbf{1}_{[0,1)}$ and

$$f_k := \mathbb{E}_k f$$

then if $x \in [0,1)$

$$\begin{aligned}\big(\sum_i |f_{k_i}-f_{k_{i+1}}|(x)^r\big)^{1/r} &= \big(\sum_{i:k_{i+1}>0}|f_{k_i}-f_{k_{i+1}}|(x)^r\big)^{1/r}\\ &\leq |f| + \sup_{\{k_i\}\subset\mathbb{Z}:k_{i+1}>0}\big(\sum_i |2^{-k_i}-2^{-k_{i+1}}|^r\big)^{1/r}\\ &\lesssim 1,\end{aligned}$$

while if $x \in [0,2^k) \smallsetminus [0,2^{k-1})$,

$$\begin{aligned}\big(\sum_i |f_{k_i}-f_{k_{i+1}}|(x)^r\big)^{1/r} &= \big(\sum_{i:k_{i+1}\geq k}|f_{k_i}-f_{k_{i+1}}|(x)^r\big)^{1/r}\\ &\leq |f_k| + \sup_{\{k_i\}\subset\mathbb{Z}:k_{i+1}\geq k}\big(\sum_i |2^{k_i}-2^{k_{i+1}}|^r\big)^{1/r}\\ &\lesssim |f_k| + 2^{-k}\\ &\lesssim 2^{-k},\end{aligned}$$

so for all $0 < r < \infty$,

$$\mathcal{V}^r(f_k) \lesssim \sum_{I\supset[0,1)\text{ dyadic}}\frac{1}{|I|}\mathbf{1}_I. \tag{3.34}$$

In particular, note that

$$\|\mathcal{V}^r(f_k)\|_{L^p(\mathbb{R})} \lesssim 1$$

and

$$\|\mathcal{V}^r(f_k)\|_{L^{1,\infty}(\mathbb{R})} \lesssim 1,$$

and in fact the majorization (3.34) has essentially appeared in our discussion of M_{D}, see (1.42).

The close relationship between $\mathcal{V}^r(\mathbb{E}_k)$ and M_{D} persists for all $r > 2$, as will be explored in Theorem 3.11. In what follows, we restrict our focus to the case where $r > 2$, see Remark 3.73.

We further explore the variation along the conditional expectation operators by considering an oscillatory version of Example (3.33).

Example 3.35. Let $h = \mathbf{1}_{[0,1/2)} - \mathbf{1}_{[1/2,1)}$, and consider $\mathcal{V}^r(\mathbb{E}_k h)$.

Whenever $k \leq -1$,

$$\mathbb{E}_k h = h \tag{3.36}$$

while whenever $k \geq 0$,

$$\mathbb{E}_k h = 0.$$

Consequently,

$$\begin{aligned}\Big(\sum_i |\mathbb{E}_{k_i} h - \mathbb{E}_{k_{i+1}} h|(x)^r\Big)^{1/r} &= \Big(\sum_{i:k_i \leq -1} |f_{k_i} - f_{k_{i+1}}|(x)^r\Big)^{1/r} \\ &= \Big(\sum_{i:k_i \leq -1,\ k_{i+1} \geq 0} |f_{k_i} - f_{k_{i+1}}|^r\Big)^{1/r} \leq |h|\end{aligned}$$

so

$$|\mathcal{V}^r(\mathbb{E}_k h)| = |h|,$$

after taking a supremum.

Notice the diagonalization that occured in the above calculation: the only contribution to $\mathcal{V}^r(\mathbb{E}_k f)$ derives from the difference in expectations on either side of the scale at which h oscillates.

Given the relationship we have seen between maximal operators and pointwise convergence, Exercise 3.32 suggests that variation operators should also play a role in our study of pointwise convergence phenomenon.

Exercise 3.37. Prove the whenever $\mathcal{V}^r(f_k)(x) < \infty$ for some $r < \infty$, $\{f_k(x)\}$ converge pointwise. On the other hand, prove that there are functions which converge, but which have unbounded r-variation for *any* $r < \infty$. *Hint: For the first case, proceed by contradiction assuming that*

$$\limsup_{k,l} |f_k(x) - f_l(x)| > \epsilon$$

for some $\epsilon > 0$*. Then extract a subsequence as in Exercise* 3.26*. For the second assertion, consider a sequence of (say, constant) functions which converges sub-polynomially.*

The following exercise quantifies the relationship between variation operators and pointwise convergence, see Exercise 3.26.

Exercise 3.38. In the context of Exercise 3.26, prove the pointwise majorization,

$$\Big(\frac{1}{J}\sum_{j=1}^{J} \sup_{N_j \leq n < N_{j+1}} |f_n - f_{N_j}|^2\Big)^{1/2} \leq J^{-1/r} \cdot \mathcal{V}^r(f_n)$$

for $2 \leq r < \infty$; compare with Exercise 3.37.

As far as pointwise ergodic theory is concerned, perhaps the most central result concerning r-variation operators – which in fact appears in Bourgain's argument – is Lépingle's inequality for *martingales* [**115**], a concept from probability theory. Although we will only be interested in one particular type of martingale, we provide a proper definition:

Definition 3.39. Given a collection of nested σ-algebras on a σ-finite measure space,

$$\mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots \subset \mathcal{B} \tag{3.40}$$

define the conditional expectations, $\{\mathbb{F}_n\}$, so that for any (locally) integrable f (with respect to $\mathcal{B}$), $\mathbb{F}_n f$ is defined to be the unique function that is measurable with respect to $\mathcal{A}_n$ so that

$$\int_A \mathbb{F}_n f = \int_A f \tag{3.41}$$

for each $A \in \mathcal{A}_n$.

Exercise 3.42. Use the Radon-Nikodym theorem to prove that a unique such function exists.

A consequence of the nesting condition $\mathcal{A}_n \subset \mathcal{A}_{n+1}$ is contained in the following exercise.

Exercise 3.43. Prove that $\mathbb{F}_n \mathbb{F}_m = \mathbb{F}_{\min\{n,m\}}$.

A sequence of integrable functions $\{f_k\}$ form a martingale sequence for the sequence of σ-algebras $\{\mathcal{A}_n\}$ if

$$f_k = \mathbb{F}_k f_n$$

whenever $k \leq n$. A martingale is *complete* if there exists a function g so that $f_n = \mathbb{F}_n g$. The difference $d_k := f_k - f_{k-1}$ are known as *martingale increments.*

The most familiar example, which will be sufficient for our applications, is the *dyadic* martingale.

Exercise 3.44. Verify that inside $[0,1]$, if

$$\mathcal{A}_n := \{I \in \mathcal{D}(-n) : I \subset [0,1]\}$$

is the dyadic σ-algebra, then $\mathbb{F}_n = \mathbb{E}_{-n}$. Note that for any integrable function, $f \in L^1([0,1])$, $\{\mathbb{E}_{-n} f\}$ form a complete martingale.

Exercise 3.45. Verify that when $\mathcal{D}(n) := \{[m \cdot 2^n, (m+1) \cdot 2^n) \cap \mathbb{Z} : n\}$ are the dyadic σ-algebra on the lattice, then

$$\mathbb{E}_n' f(x) = \sum_{|I|=2^n \text{ dyadic}} \big(\frac{1}{I} \sum_{n \in I} f(n)\big) \cdot \mathbf{1}_I \tag{3.46}$$

form a *reverse martingale* in that the σ algebras satisfy a *reverse nesting* condition, and accordingly

$$\mathbb{E}_n' \mathbb{E}_m' = \mathbb{E}_m', \quad m \geq n.$$

3.1. Oscillation and Lépingle's Inequality. We pause here to introduce Lépingle's inequality in the special Euclidean/dyadic setting, which will be adequate for our purposes. To do so, however, we need some fundamentals of dyadic harmonic analysis. For ease of presentation, we will develop our theory on the real line, and will leave it as an exercise to translate these arguments to the integer setting.

The following section will concern functions and operators on Euclidean space.

3.1.1. *The Maximal Function and the Square Function.* We recall the dyadic martingale, see (2.1),

$$\mathbb{E}_n f(x) := \sum_{Q \in \mathcal{D}(n)} f_Q \cdot \mathbf{1}_Q, \tag{3.47}$$

where f_Q denotes the average value of f over Q:

$$f_Q := \frac{1}{|Q|} \int_Q f(x)\, dx. \tag{3.48}$$

We let $\Delta_n := \mathbb{E}_{n-1} - \mathbb{E}_n$ denote the differencing operators. It turns out that there is a convenient spatial representation of the operators Δ_n.

Exercise 3.49. For each dyadic $|I| = 2^{-n}$, define the *Haar function*

$$h_I(x) := \frac{1}{|I|^{1/2}} \cdot \big(\mathbf{1}_{I_l} - \mathbf{1}_{I_r}\big),$$

where I_l is the left half of the interval I, and I_r is the right half. Prove that the collection of Haar functions, $\{h_I : I \text{ dyadic}\}$ are an orthonormal subset of $L^2(\mathbb{R})$; we will upgrade this result in Exercise 3.79 below.

Exercise 3.50. Prove

$$\Delta_n f = \sum_{|I|=2^n} \langle f, h_I \rangle \cdot h_I,$$

Moreover, prove that $\{\Delta_n : n\}$ are *orthogonal projections*:

$$\Delta_n \Delta_n \equiv \Delta_n, \quad \text{and} \quad \Delta_n \Delta_m = 0, \ m \neq n.$$

Hint: Use Exercise 3.49*.*

With these preliminaries in mind, we recall the (one-dimensional) dyadic maximal functions, (1.37) and (1.43), and square function, defined respectively as

$$M_{\mathrm{D}} f := \sup_n \mathbb{E}_n |f|, \quad \text{and} \quad M_{\triangle} f := \sup_n |\mathbb{E}_n f|, \tag{3.51}$$

and

$$S_{\mathrm{D}} f := \left(\sum_n |\Delta_n f|^2 \right)^{1/2}; \tag{3.52}$$

recall the bound

$$\lambda \cdot |\{M_{\mathrm{D}} f \geq \lambda\}| \leq \int_{\{M_{\mathrm{D}} f \geq \lambda\}} |f| \tag{3.53}$$

for each $\lambda > 0$.

Remark 3.54. This argument abstracts to the setting of more general martingales, where it appears as *Doob's Maximal Inequality.*

Exercise 3.55. Prove that $S_{\mathrm{D}} f$ can be expressed as

$$\left(\sum_I |\langle f, h_I \rangle \cdot h_I|^2 \right)^{1/2} = \left(\sum_I |\langle f, h_I \rangle|^2 \cdot \frac{\mathbf{1}_I}{|I|} \right)^{1/2}.$$

Example 3.56. Set $f = \mathbf{1}_{[0,1)}$ and

$$h = \mathbf{1}_{[0,1/2)} - \mathbf{1}_{[1/2,1)} = h_{[0,1)},$$

as in Example 3.35.

We compute

$$S_{\mathrm{D}} f = \left(\sum_{|I| \geq 2} |\langle f, h_I \rangle \cdot h_I|^2 \right)^{1/2} \lesssim \sum_{I \supset [0,1) \text{ dyadic}} \frac{1}{|I|} \mathbf{1}_I,$$

see (3.34) and (1.42). On the other hand, by Exercise 3.55,

$$S_{\mathrm{D}} h = |h|,$$

as in Example 3.35.

These two operators, both of which roughly measure the "scales" where f oscillates, have long been thought of as being of "even strength," see Example 3.55 above. This heuristic was beautifully quantified by Burkholder and Davis, who proved the following [**32**, Theorem 9], [**50**, Theorem 1]

Theorem 3.6 (Special Case: Burkholder, $p > 1$, Davis, $p = 1$)**.** *For each* $1 \le p < \infty$,

$$\|S_D f\|_{L^p(\mathbb{R}^D)} \approx_p \|M_D f\|_{L^p(\mathbb{R}^D)} \approx_p \|M_{\triangle} f\|_{L^p(\mathbb{R}^D)}.$$

We leave the verification of this theorem in the special case where $p = 2$ to the following exercise.

Exercise 3.57. Prove Theorem 3.6 in the special case where $p = 2$.
Hint: Exercise 3.50 *will be useful.*

We begin our treatment of oscillation by proving Theorem 3.6 in the one-dimensional case, deferring the general set-up to Exercise 3.140 at the end of the chapter. Typically, these estimates are established for $1 < p < \infty$ by proving that S_{D} is bounded on $L^p(\mathbb{R})$ for $1 < p < \infty$, and then estimating, for an appropriate $\|g\|_{L^{p'}(\mathbb{R})} = 1$,

$$\begin{aligned} \|f\|_{L^p(\mathbb{R})} = \left| \int f \cdot g \right| &= \left| \int \sum_{n,m} \Delta_n f \cdot \Delta_m g \right| \\ &= \left| \int \sum_{n} \Delta_n f \cdot \Delta_n g \right| \le \int S_{\mathrm{D}} f \cdot S_{\mathrm{D}} g \\ &\lesssim \|S_{\mathrm{D}} f\|_{L^p(\mathbb{R})}, \end{aligned}$$

where we used Exercise 3.50 to pass to the third line, and the $L^{p'}(\mathbb{R})$ boundedness of the square function in the final estimate.

Establishing the L^p boundedness of $S_{\mathrm{D}} f$ typically goes through an endpoint estimate, which essentially dictates that it is challenging for $S_{\mathrm{D}} f$ to be large when $M_{\mathrm{D}} f$, or even $M_{\triangle} f$, is small. Our arguments, which essentially appear as the stopping time arguments used in [**186**, §2], provide a quick and direct proof of these estimates – but moreover provide a closer link between fluctuations in $S_{\mathrm{D}} f$ and $M_{\mathrm{D}} f$.

Our proof of Theorem 3.6 will follow from a *relative distributional*, or *good-λ*, inequality: for any $\lambda > 0$

$$|\{S_{\mathrm{D}} f > 3\lambda, M_{\mathrm{D}} f \leq \gamma\lambda\}| \lesssim \gamma^2 \cdot |\{S_{\mathrm{D}} f > \lambda\}| \tag{3.58}$$

and

$$|\{M_{\mathrm{D}} f > 3\lambda, S_{\mathrm{D}} f \leq \gamma\lambda\}| \lesssim \gamma^2 \cdot |\{M_{\mathrm{D}} f > \lambda\}|. \tag{3.59}$$

Although (3.58) and (3.59) appear complicated, the content is very simple: it is very challenging for either $S_{\mathrm{D}} f$ or $M_{\mathrm{D}} f$ to be large when the complementary operator is not. In fact, as our arguments below will show, this phenomenon holds *at all scales and locations.*

In particular, the strength of the good-λ approach is captured in the following lemma.

Lemma 3.7. *Suppose that for each $\lambda > 0$ and each $\gamma \ll 1$,*

$$\mu(\{|f| > \lambda, |g| \leq \gamma\lambda\}) = o_{\gamma\to 0}\big(\mu\{|f| > c\lambda\}\big), \quad c \leq 1.$$

Then for all $1 \leq p < \infty$,

$$\|f\|_{L^p(X,\mu)} \lesssim \|g\|_{L^p(X,\mu)} \tag{3.60}$$

and similarly

$$\|f\|_{L^{p,\infty}(X)} \lesssim \|g\|_{L^{p,\infty}(X)}. \tag{3.61}$$

Proof. We establish the strong type formulation (3.60), and leave the weak type statement (3.61) to Exercise 3.64 below. To do so, we use the wedding cake formulation of the L^r norms, see Lemma 1.5, which allows us to connect the size of super-level sets to L^r norms. In particular, for a sufficiently small parameter $\gamma > 0$, we bound

$$\begin{aligned} \mu(\{|f| > \lambda\}) &\leq \mu(\{|f| > \lambda, |g| \leq \gamma\lambda\}) + \mu(\{|g| > \gamma\lambda\}) \\ &= \epsilon \cdot \mu(\{|f| > c\lambda\}) + \mu(\{|g| > \gamma\lambda\}) \end{aligned}$$

for some sufficiently small $\epsilon = \epsilon(\gamma)$ *independent of c.*

By monotone convergence, it suffices to show that for each (large) $R < \infty$,

$$\|f \cdot \mathbf{1}_{|f| \leq R}\|_{L^r(X,\mu)} \lesssim \|g\|_{L^r(X,\mu)}.$$

Applying the wedding cake decomposition, we may express

$$\|f \cdot \mathbf{1}_{|f| \leq R}\|_{L^r(X,\mu)}^r \leq \int_0^R r t^{r-1} \cdot \mu(\{|f| > t\})\, dt \tag{3.62}$$

We estimate (3.62): for each $R < \infty$, we have the upper bound

$$\begin{aligned}
&\int_0^R rt^{r-1} \cdot \mu(\{|f| > t\}) \, dt \\
&\quad \leq \epsilon \cdot \int_0^R rt^{r-1} \cdot \mu(\{|f| > ct\}) \, dt + \int_0^R rt^{r-1} \cdot \mu(\{|g| > \gamma t\}) \, dt \\
&\quad \leq \frac{\epsilon}{c^r} \cdot \int_0^{cR} rt^{r-1} \cdot \mu(\{|f| > t\}) \, dt + \gamma^{-r} \int_0^{\gamma R} rt^{r-1} \cdot \mu(\{|g| > t\}) \, dt \\
&\quad \leq \frac{\epsilon}{c^r} \cdot \int_0^R rt^{r-1} \cdot \mu(\{|f| > t\}) \, dt + \gamma^{-r} \cdot \|g\|_{L^r(X,\mu)}^r.
\end{aligned}$$

If we choose ϵ so small that $\frac{\epsilon}{c^r} \leq 1/2$ (say), then we may re-arrange, to deduce that

$$(3.62) \leq \int_0^R rt^{r-1} \cdot \mu(\{|f| > t\}) \, dt \leq 2\gamma^{-r} \cdot \|g\|_{L^r(X,\mu)}^r,$$

which completes the proof. □

Exercise 3.63. Verify (3.62).

Exercise 3.64. Establish 3.61.

We immediately arrive at the following corollary.

Corollary 3.8. (3.58) *and* (3.59) *imply that*

$$\|M_D f\|_{L^r(\mathbb{R})} \approx_r \|S_D f\|_{L^r(\mathbb{R})}$$

for all $0 < r < \infty$ *– and Theorem* 3.6 *in particular.*

With Corollary 3.8 in hand, we begin the proof of Theorem 3.6.

Theorem 3.6, Part One. We begin by proving that the maximal function controls the square function: the first inequality in (3.58). In particular, this will already imply Theorem 3.6 in the regime $1 < p < \infty$.

Let Q be a maximal dyadic cube on which $S_{\mathrm{D}} f > \lambda$; we will prove that

$$|\{Q : S_{\mathrm{D}} f > 3\lambda, M_{\mathrm{D}} f \leq \gamma\lambda\}| \lesssim \gamma^2 \cdot |Q|. \tag{3.65}$$

By dilation and translation invariance, we may assume that $Q = [0,1]$ (why?), and by homogeneity we may assume that $\lambda = 1$; we will also assume that there exists some $z = z_Q \in Q$ so that $M_{\mathrm{D}} f(z) \leq \gamma$, otherwise there is nothing to show.

We claim that if we set

$$g = (f - f_Q) \cdot \mathbf{1}_Q = \big(f - \int_0^1 f\big) \cdot \mathbf{1}_{[0,1]},$$

it is enough to prove that

$$|\{S_{\mathrm{D}}g > 2, M_{\mathrm{D}}g \leq 2\gamma\}| \lesssim \gamma^2. \tag{3.66}$$

Exercise 3.67. Verify the reduction of (3.65) to (3.66).
Hint: There are two observations; first is that

$$\Big(\sum_{P \supsetneq Q} |\langle f, h_P\rangle|^2 \frac{\mathbf{1}_P}{|P|} \Big)^{1/2}$$

is constant on the parent *of Q; the second is that f_Q can be estimated using the existence of $z = z_Q$.*

We next observe that

$$\frac{|\langle g, h_J\rangle|}{|J|^{1/2}} > 2\gamma$$

only when $J \subset \{M_D g > 2\gamma\}$; this follows from the triangle inequality. In particular, if we collect all maximal intervals $\{K\} \subset [0,1]$ so that $M_{\mathrm{D}}g > 2\gamma$ and set

$$g_0 := \sum_K g_K \cdot \mathbf{1}_K + g \cdot \mathbf{1}_{[0,1] \smallsetminus \bigcup K},$$

then, as the following exercise will ask you to verify, we only have $S_{\mathrm{D}}g_0 < S_{\mathrm{D}}g$ inside

$$\{M_{\mathrm{D}}g > 2\gamma\}.$$

Exercise 3.68. Verify this pointwise inequality.
Hint: Try expanding out into Haar coefficients.

In particular, it remains only to show that $\|S_{\mathrm{D}}g_0\|_2^2 \lesssim \gamma^2$; but

$$\|S_{\mathrm{D}}g_0\|_{L^2(\mathbb{R})}^2 = \|g_0\|_{L^2(\mathbb{R})}^2 \leq \|M_{\mathrm{D}}g\|_{L^2(\mathbb{R})}^2 \lesssim \gamma^2$$

by pointwise considerations. □

We next show that the square function controls the maximal function; this argument is similar, and will allow us to conclude Theorem 3.6 in the case $p = 1$.

Proof of Theorem 3.6, Part Two. We now show that the square function controls the maximal function:

Let Q be a maximal dyadic cube on which $M_{\mathrm{D}}f > 1$; we will prove that

$$|\{Q : M_{\mathrm{D}}f > 3, S_{\mathrm{D}}f \leq \gamma\}| \lesssim \gamma^2 \cdot |Q|; \tag{3.69}$$

again, we may assume that $Q = [0,1]$.

Collect all maximal $\{P\} \subset [0,1]$ so that $S_{\mathrm{D}} g > \gamma$ on P, and set

$$g_0 := \sum_P g_P \cdot \mathbf{1}_P + g \cdot \mathbf{1}_{[0,1] \smallsetminus \bigcup P}.$$

The following exercise is the key to the argument.

Exercise 3.70. Prove that $S_{\mathrm{D}} g_0 \leq \gamma$, and that $M_{\mathrm{D}} g > M_{\mathrm{D}} g_0$ only inside $\bigcup P$.

Indeed, with this exercise in hand, the proof concludes immediately: we simply estimate the left-hand side of (3.69) by

$$|\{M_{\mathrm{D}} g_0 > 2\}| \lesssim \|g_0\|_{L^2(\mathbb{R})}^2 = \|S_{\mathrm{D}} g_0\|_{L^2(\mathbb{R})}^2 \lesssim \gamma^2,$$

again by pointwise considerations. □

With the maximal function and the square function in hand, we turn to Lépingle's inequality, in which we will investigate the boundedness of another, less classical, measurement of oscillation. First, though, we ask the reader to test themselves with the following exercise.

Exercise 3.71. Adapt the above argument to the integer lattice. On the integers, we set

$$\mathbb{E}_n' f(x) := \sum_{|Q|=2^n} f_Q \cdot \mathbf{1}_Q, \tag{3.72}$$

where $f_Q := \frac{1}{|Q|} \sum_{x \in Q} f(x)$.
Hint: One can either mimic the above arguments, or try rescaling the Euclidean arguments appropriately.

With these preliminaries in mind, we are prepared to study the variation of the $\{\mathbb{E}_n f\}$ via Lépingle's Inequality. We will focus on the one-dimensional formulation; the higher-dimensional formulation also presents, and will be addressed below, but the Haar function formulation of the one dimensional square function leads to formal simplification.

3.1.2. *Lépingle's Inequality.* The operator of consideration is the r-variation of the dyadic martingale,

$$\mathcal{V}^r f := \mathcal{V}^r(\mathbb{E}_n f).$$

We have the following fundamental boundedness result concerning the behavior of these operators.

Theorem 3.9 (Lépingle's Inequality – Special Case). *For each $r > 2$, and $1 < p < \infty$,*

$$\|\mathcal{V}^r f\|_{L^p(\mathbb{R}^D)} \lesssim_p \frac{r}{r-2} \cdot \|f\|_{L^p(\mathbb{R}^D)};$$

moreover, at the endpoint $p = 1$ we have the weak-type estimate

$$\|\mathcal{V}^r f\|_{L^{1,\infty}(\mathbb{R}^D)} \lesssim \frac{r}{r-2} \cdot \|f\|_{L^1(\mathbb{R}^D)}.$$

Remark 3.73. The range of $r > 2$ in the above theorem is sharp, since these estimates can fail for $r \leq 2$; see e.g. [**96**], [**150**].

By now, comparatively simple proofs of Lépingle's theorem can be found in Pisier and Xu [**149**] and Bourgain [**21**]. The idea was to leverage known estimates for jump inequalities for the $\{\mathbb{E}_n f\}$ to recover variational estimates:

Definition 3.74. For each $\lambda > 0$, $N_\lambda(\mathbb{E}_n f)(x)$ is the number of λ-jumps in $\{\mathbb{E}_n f(x)\}$: the supremum over all $N \geq 1$ such that there exists an increasing sequence of indices $\{n_1 < m_1 \leq n_2 < m_2 \leq \cdots < n_N \leq m_N\}$ with

$$|\mathbb{E}_{n_i} f(x) - \mathbb{E}_{m_i} f(x)| > \lambda$$

for each $1 \leq i \leq N$.

Exercise 3.75. Prove that $N_\lambda(\mathbb{E}_n f)(x)$ is comparable to the $\lambda/2$-entropy of the $\{\mathbb{E}_n f(x)\}$, i.e. the fewest number of $\lambda/2$-balls required to cover the collection $\{\mathbb{E}_n f(x)\}$.

The key result concerning λ-jumps is the following:

Theorem 3.10 ([**33**], [**95**]). *If N_λ is the number of λ-jumps in $\{\mathbb{E}_n f\}$ then for any $1 < p < \infty$*

$$\|\lambda N_\lambda^{1/2}\|_{L^p(\mathbb{R}^D)} \lesssim_p \|f\|_{L^p(\mathbb{R}^D)} \quad \textit{for all } \lambda > 0;$$

moreover, at the endpoint $p = 1$ we have the weak-type estimate

$$\|\lambda N_\lambda^{1/2}\|_{L^{1,\infty}(\mathbb{R}^D)} \lesssim \|f\|_{L^1(\mathbb{R}^D)} \quad \textit{for all } \lambda > 0.$$

Our proof of Theorem 3.10 above will follow from a good-λ inequality, which will relate the jump counting function to the Hardy-Littlewood maximal function. We will begin with the one-dimensional formulation, and will defer the D-dimensional theory to the end of this chapter; we first establish an $L^2(\mathbb{R})$-estimate.

A key technical point, whose proof is deferred to Exercise 3.78 below, is that $N_\lambda \leq N'_{\lambda/2}$where

$$N'_\lambda \tag{3.76}$$

is given by the supremum over all $k \geq 1$ such that there exists an increasing sequence of indices $\{n_1 < n_2 < \cdots < n_{k+1}\}$ with

$$|\mathbb{E}_{n_i} f(x) - \mathbb{E}_{n_{i+1}} f(x)| > \lambda$$

for each $1 \leq i \leq k$. Jump counting functions of the form $\{N'_\lambda : \lambda\}$ are known as *greedy* jump-counting functions, and are much simpler than the jump counting functions $\{N_\lambda : \lambda\}$, as they have a natural "stopping time" structure, which makes them very amenable to martingale-type arguments. We begin with the one-dimensional proof.

Proof of the $L^2(\mathbb{R})$-estimate. By Exercise 3.79 below, it suffices to prove the desired inequality for finite linear combinations of Haar functions

$$f = \sum_I a_I h_I.$$

Pick a maximal interval on which f is supported, call it J. It is enough to prove the desired inequality for $\sum_{I \subset J} a_I h_I$ (why?).

Let

$$F_1 := \bigcup_I \Big\{ I \subset J \text{ maximal} : \Big| \sum_{I \subsetneq I' \subset J} a_{I'} h_{I'}(x) \Big| > \lambda \Big\} \tag{3.77}$$

for some (any) $x \in I$. Let

$$T_1 := \{I \subset J : I \not\subset K \text{ for any } K \in F_1\}.$$

Next, if $x \in I \in F_1$, then

$$\{I' \in T_1 : x \in I'\} = \{I' : I \subsetneq I' \subset J\}.$$

But now:

$$\int \lambda^2 \sum_{I \in F_1} \mathbf{1}_I < \int \Big| \sum_{I \in T_1} a_I h_I \Big|^2 = \sum_{I \in T_1} |a_I|^2;$$

then iterate, considering the stopping regions for each $J_1 \in F_1$:

$$F_2(J_1) := \Big\{ I \subset J_1 \text{ maximal} : \Big| \sum_{I \subsetneq I' \subset J_1} a_{I'} h_{I'}(x) \Big| > \lambda \Big\}$$

for some (any) $x \in I$, and the trees

$$T_1(J_1) := \{I \subset J_1 : I \not\subset K \text{ for any } K \in F_2(J_1)\}.$$

□

We refer the reader to [**95**] for a more comprehensive survey.

Exercise 3.78. Verify the first claim concerning N_λ and $N'_{\lambda/2}$.
Hint: Suppose that $|a_{n_1} - a_{n_2}| > \lambda$. *Then there must exist some time* $n_1 \leq n \leq n_2$ *so that either*

$$|a_{n_1} - a_n| > \lambda/2$$

or

$$|a_n - a_{n_2}| > \lambda/2.$$

Exercise 3.79. Verify that $\{h_I : I \text{ dyadic}\}$ forms a basis for $L^2(\mathbb{R})$, and that finite linear combinations of Haar functions are dense in $L^p(\mathbb{R}),\ 1 < p < \infty$. Are they dense in $L^1(\mathbb{R})$?
Hint: Theorem 3.6 *will be useful for the density claims. For the final point, note that* $f \mapsto \int f$ *is continuous on* $L^1(\mathbb{R})$.

To complete the proof of Theorem 3.10, it suffices to prove that for any collection of trees $\mathcal{T} = \{T_i\}$ the square functions

$$S_{\mathcal{T}} f^2 := \sum_{i=1}^{\infty} \Big| \sum_{I \in T_i} \langle f, h_I \rangle \cdot h_I \Big|^2$$

satisfy the good-λ inequality, uniformly in $\mathcal{T}$:

$$|\{S_{\mathcal{T}} f > 3\lambda, M_{\mathrm{D}} f \leq \gamma\lambda\}| \lesssim \gamma^2 |\{S_{\mathcal{T}} f > \lambda\}|. \tag{3.80}$$

Exercise 3.81. Verify (3.80).
Hint: Mimic the argument used to establish (3.58)*: consider*

$$g_0 := \sum_{I \in T_1} g_I \cdot \mathbf{1}_I + g \cdot \mathbf{1}_{(\bigcup_{I \in T_1} I)^c}.$$

where g *is constructed as above.*

The significance of the argument we will use below to prove Lépingle's Inequality – which is similar in spirit to our proof of Theorem 3.6 – is that it sheds new insight into the relationship between maximal function, square function, and variation operator.[3] Indeed, in light of Theorem 3.6, we recover Lépingle's result that the r-variation operators ($r > 2$) are themselves of "even strength" with the maximal functions.

[3]Similar observations – valid in the more abstract setting of general martingales – have been made in [**91**] and [**134**].

Specifically, we have the following theorem, whose slightly involved proof will be presented at the conclusion of this chapter. Below, we abbreviate

$$\mathcal{V}^r f = \mathcal{V}^r(\mathbb{E}_k f) \tag{3.82}$$

Theorem 3.11. *For $r > 2$, and all $\gamma > 0$ sufficiently small, we have the good-λ inequality*
(3.83)

$$|\{\mathcal{V}^r f > C\lambda, M_D f < \gamma\lambda, S_D f < \gamma\lambda\}| \leq C \cdot \left(\frac{r}{r-2}\right)^2 \cdot \gamma^2 \cdot |\{\mathcal{V}^r f > \lambda\}|$$

where $C \gg 1$ is an appropriate (absolute) constant. In particular, by integrating distribution functions we recover the optimal bound

$$\|\mathcal{V}^r f\|_{L^2(\mathbb{R}^D)} \lesssim \frac{r}{r-2} \cdot \|f\|_{L^2(\mathbb{R}^D)}.$$

Above, we abuse notation and use $S_{\mathrm{D}} f$ to refer to the D-dimensional square function

$$\Big(\sum_k |\mathbb{E}_k f - \mathbb{E}_{k+1} f|^2 \Big)^{1/2},$$

where the expectation operators are D-dimensional; see (3.139) below.

Remark 3.84. The proof we provide translates readily to the integer lattice, see Exercise 3.71 above.

This immediately proves that $\mathcal{V}^r f$ is bounded on $L^p(\mathbb{R}^D)$, $1 < p < \infty$ and moreover is integrable when the maximal function and the square function are, see (3.139) for the D-dimensional formulation:

Corollary 3.12. *If $M_D f \in L^1(\mathbb{R}^D)$ (and thus $S_D f \in L^1(\mathbb{R}^D)$), $\mathcal{V}^r f \in L^1(\mathbb{R}^D)$ for each $r > 2$, see* (3.82).

To prepare for what follows, Exercise 3.85 will ask you to deduce Lépingle's Inequality in the integer setting from Theorem 3.11.

Exercise 3.85. Use a rescaling argument to deduce the integer version of (3.83), where the conditional expectation operators on $\mathbb{Z}$ are given by (3.72).

In the following section, we return to the sequence-space setting of $\mathbb{Z}$.

By an abuse of notation, in what follows for notational ease we will re-label the discrete martingale projections, $\mathbb{E}'_n$, see (3.72), $\mathbb{E}_n$.

3.2. A Quantitative Approach to Birkhoff's Theorem. With Lépingle's inequality in hand, we provide an alternative proof of Birkhoff's theorem – one which, we will see below, generalizes in a variety of ways. As we will see, the main issue is developing an appropriate L^2-theory, as the L^p-theory can be treated essentially as perturbations of the L^2-approach. Accordingly, we will spend the bulk of our attention proving pointwise convergence of ergodic averages on $L^2(X)$, i.e. that $\{[0,i)\}$ are L^2-good, before subsequently extending to L^p-goodness.

We will do so via a variational estimate; we will first show that for each $r > 2$,

$$\|\mathcal{V}^r(M_n^\tau f)\|_{L^2(X)} \lesssim_r \|f\|_{L^2(X)}; \tag{3.86}$$

this result first appeared in Bourgain's celebrated paper on polynomial ergodic averages, [**21**]. Since the variation operators are semi-local (why?), we will derive this result from the corresponding one on the integer lattice:

Proposition 3.13. *For each* $r > 2$

$$\|\mathcal{V}^r(M_n f)\|_{\ell^2(\mathbb{Z})} \lesssim_r \|f\|_{\ell^2(\mathbb{Z})},$$

where M_n *are defined in* (3.16).

Exercise 3.87. Rigorously derive the implication (3.86) from this theorem.

The argument we provide here follows that of [**94**] – a simplification of [**21**].

Before turning to the proof proper, we make some preliminary remarks. For a sequence of functions, $\{f_i\}$, we may divide our study of the *r-variation*, $\mathcal{V}^r(f_i)$, into the *long-* and *short-r variation*, respectively defined below:

$$\mathcal{V}^{r,L}(f_i)(x) := \sup_{(i_k) \text{ increasing, dyadic}} \Big(\sum_k |f_{i_k} - f_{i_{k+1}}|^r \Big)^{1/r}(x)$$

$$\mathcal{V}^{r,S}(f_i)(x) := \left(\sum_n \Big(\sup_{2^n \leq (i_k) \leq 2^{n+1} \text{ increasing}} \sum_k |f_{i_k} - f_{i_{k+1}}|^r \Big) \right)^{1/r} (x).$$

Indeed, we have the following pointwise inequality:

Lemma 3.14 ([**96**], §3)**.**

$$\mathcal{V}^r(f_i) \lesssim_r \mathcal{V}^{r,L}(f_i) + \mathcal{V}^{r,S}(f_i).$$

Exercise 3.88. Prove Lemma 3.14.

We may therefore focus on bounding each short and long variation of the ergodic averages independently on $\ell^2(\mathbb{Z})$.

Before proceeding, we remark that we also have the following majorization for the jump-counting function, N_λ.

Exercise 3.89 (Lemma 1.3 of [**95**])**.** Define $N_\lambda(f_i)$ in analogy with Definition 3.78, and $N_\lambda^D(f_i)$ as $N_\lambda(f_i)$, with the restriction that the only functions considered are indexed by the dyadics. Prove the following pointwise majorization:

$$\lambda N_\lambda(f_i)^{1/r} \lesssim \mathcal{V}^{r,S}(f_i) + \lambda N_{c\lambda}^D(f_i)^{1/r}$$

for some absolute $c > 0$.

The Long Variation is $\ell^p(\mathbb{Z})$-Bounded. We begin by developing an appropriate $\ell^2(\mathbb{Z})$-theory for the long variation. Using Lépingle's inequality and the triangle inequality, it is enough to show that

$$\mathcal{V}^r(M_{2^n}f - \mathbb{E}_n f) \lesssim \Big(\sum_{n\geq 0} |M_{2^n}f - \mathbb{E}_n f|^2\Big)^{1/2} =: S_c f \tag{3.90}$$

is $\ell^2(\mathbb{Z})$-bounded (the subscript c stands for comparison).

Exercise 3.91. By proving that $\mathcal{V}^r$ is sub-linear, verify the details of the above reduction.

Recalling that in the discrete setting

$$\mathbb{E}_n f := \sum_Q f_Q \mathbf{1}_Q,$$

see (3.72), with the sum running over dyadic cubes (intervals) of side-length 2^n, set $\Delta_n := \mathbb{E}_{n-1} - \mathbb{E}_n$ for $n \geq 1$ and $\Delta_n = 0$ otherwise. We claim that it is enough to show

$$\|(M_{2^i} - \mathbb{E}_i)\Delta_j\|_{\ell^2(\mathbb{Z})\to\ell^2(\mathbb{Z})} \lesssim 2^{-\frac{|i-j|}{2}}.$$

Indeed, expanding $f = \sum_{n\geq 0}\Delta_n f$ into increments (see Exercise 3.79), we majorize

$$\begin{aligned} S_c f &= \Big(\sum_{n\geq 0}\Big|(M_{2^n} - \mathbb{E}_n)\big(\sum_m \Delta_m f\big)\Big|^2\Big)^{1/2} \\ &\leq \sum_t \Big(\sum_{n\geq 0} |(M_{2^n} - \mathbb{E}_n)\Delta_{n+t} f|^2\Big)^{1/2}. \end{aligned} \tag{3.92}$$

Estimating the tth term in $\ell^2(\mathbb{Z})$:

$$\begin{aligned}\Big\|\Big(\sum_{n\geq 0}|(M_{2^n}-\mathbb{E}_n)\Delta_{n+t}f|^2\Big)^{1/2}\Big\|_{\ell^2(\mathbb{Z})}^2 &= \sum_{n\geq 0}\|(M_{2^n}-\mathbb{E}_n)\Delta_{n+t}f\|_{\ell^2(\mathbb{Z})}^2 \\ &\lesssim 2^{-|t|}\cdot\sum_{n\geq 0}\|\Delta_{n+t}f\|_{\ell^2(\mathbb{Z})}^2 \\ &= 2^{-|t|}\cdot\Big\|\Big(\sum_{n\geq 0}|\Delta_{n+t}f|^2\Big)^{1/2}\Big\|_{\ell^2(\mathbb{Z})}^2 \\ &\leq 2^{-|t|}\cdot\|f\|_{\ell^2(\mathbb{Z})}^2,\end{aligned}$$

where we used the fact that the Δ_js are projections in the second line, see Exercise 3.49.

To prove our key estimate

$$\|(M_{2^i}-\mathbb{E}_i)\Delta_j\|_{\ell^2(\mathbb{Z})\to\ell^2(\mathbb{Z})}\lesssim 2^{-\frac{|i-j|}{2}}, \tag{3.93}$$

we may without loss of generality assume that $|i-j|\gg 1$ is large, since for each i,j

$$(M_{2^i}-\mathbb{E}_i)\Delta_j$$

is a composition of two bounded operators on $\ell^2(\mathbb{Z})$ (check this).

Now, with $f\in\ell^2(\mathbb{Z})$ and $j\geq 0$ fixed, expand

$$\Delta_j f=\sum_J\langle f,h_J\rangle h_J$$

into Haar coefficients at scale $|J|=2^j$.

When $i\leq j-C$,

$$|(M_{2^i}-\mathbb{E}_i)h_J|=|M_{2^i}h_J-h_J|\leq\frac{1}{|J|^{1/2}}\cdot\mathbf{1}_{E_i(J)}$$

where $|E_i(J)|\lesssim 2^i$ is supported in a union of at most three intervals containg the points of discontinuity of h_J. Consequently,

$$|(M_{2^i}-\mathbb{E}_i)\Delta_j f|\leq\sum_J|\langle f,h_J\rangle|\cdot\frac{1}{|J|^{1/2}}\cdot\mathbf{1}_{E_i(J)}$$

and thus

$$\begin{aligned}\|(M_{2^i}-\mathbb{E}_i)\Delta_j f\|_{\ell^2(\mathbb{Z})}^2 &\leq\sum_J|\langle f,h_J\rangle|^2\cdot|J|^{-1}\cdot|E_i(J)| \\ &\lesssim 2^{i-j}\cdot\sum_J|\langle f,h_J\rangle|^2 \\ &= 2^{i-j}\cdot\|\Delta_j f\|_{\ell^2(\mathbb{Z})}^2.\end{aligned}$$

Exercise 3.94. Verify the final equality above.

On the other hand, when $i \geq j + C$,

$$(M_{2^i} - \mathbb{E}_i)\Delta_j \equiv M_{2^i}\Delta_j.$$

Now, if $x_0 \in J_0$, and $J_1 =: J_0 - 2^i$, then – using the mean-zero nature of the Haar functions – one has the identity

$$\begin{aligned}(M_{2^i} - \mathbb{E}_i)\Delta_j f(x_0) &= M_{2^i}\Delta_j f(x_0) \\ &= \langle f, h_{J_0}\rangle M_{2^i}(h_{J_0})(x_0) + \langle f, h_{J_1}\rangle M_{2^i}(h_{J_1})(x_0).\end{aligned}$$

Since

$$|M_{2^i}(h_J)| \lesssim \frac{2^{j/2}}{2^i},$$

we may majorize

$$|(M_{2^i} - \mathbb{E}_i)\Delta_j f| \lesssim \sum_J \left(\sum_{i=0}^{1} |\langle f, h_{J_i}\rangle| \right) \frac{2^{j/2}}{2^i} \cdot \mathbf{1}_J,$$

where we let $J_0 := J$ and $J_1 := J - 2^i$. Square summing yields the (stronger) estimate

$$\|(M_{2^i} - \mathbb{E}_i)\Delta_j f\|_{\ell^2(\mathbb{Z})}^2 \lesssim 2^{2j-2i} \cdot \|\Delta_j f\|_{\ell^2(\mathbb{Z})}^2,$$

completing the proof of the claim. □

Exercise 3.95. Provide an alternative proof of (3.93) by showing that

$$\|(M_{2^i} - \mathbb{E}_i)\Delta_j\|_{\ell^1(\mathbb{Z}) \to \ell^1(\mathbb{Z})} \lesssim 2^{-|i-j|},$$

and interpolating against the $\ell^\infty(\mathbb{Z})$ estimate.
Hint: One can view

$$f \mapsto (M_{2^i} - \mathbb{E}_i)\Delta_j f$$

as an integral operator with kernel

(3.96)

$$K(x,z) = 2^{-i} \sum_{y \in \mathbb{Z}} \Big(\mathbf{1}_{[0,2^i)}(x-y) - \sum_{|I|=2^i} \mathbf{1}_I(x)\mathbf{1}_I(y) \sum_{|J|=2^j} h_J(y)h_J(z) \Big).$$

When $i \geq j$ one can bound

$$|K(x,z)| \lesssim 2^{-i} \cdot \mathbf{1}_{O(2^j) \cup 2^i + O(2^j)}(x-z),$$

while when $j \geq i$

$$|K(x,z)| \lesssim 2^{-j} \sum_{|J|=2^j} \mathbf{1}_{(\partial J_l \cup \partial J_r) + O(2^i)}(x) \cdot \mathbf{1}_J(z);$$

in both cases

$$\sup_z \sum_x |K(x,z)| \lesssim 2^{-|i-j|}.$$

Above, $J = J_l \cup J_r$ are the dyadic children of J, and we used the Minkowski sum.

We now quickly extend this result to the $\ell^p(\mathbb{Z})$-setting, via the following Corollary.

Corollary 3.15. *For each $1 < p < \infty$, $S_c f$ is $\ell^p(\mathbb{Z})$-bounded; in particular, the long variation is similarly bounded.*

We use an argument of [**59**], which will find repeated use throughout this section, and will feature later in Chapter 11 below. The following general lemma applies to any measure space and we will apply it both in the Euclidean and discrete settings.

Lemma 3.16. *Suppose that $\{K_j\}$ are a collection of $L^1(X)$-normalized functions, or measures that have uniformly bounded total variation, $\|K_j\|_{TV} \leq C$ for all j, and so that*

$$\| \sup_j |\tilde{K_j}| * |f| \|_{L^{p_0}(X)} \lesssim \|f\|_{L^{p_0}(X)}$$

for some p_0, where $\tilde{g}(x) := g(-x)$ is the antipodal reflection and is extended to measures by duality. Then for all $1 \leq p \leq \infty$ so that

$$|\frac{1}{2} - \frac{1}{p}| \leq \frac{1}{2p_0},$$

the following vector-valued inequality holds:

$$\Big\| \big(\sum_j |K_j * f_j|^2 \big)^{1/2} \Big\|_{L^p(X)} \lesssim \Big\| \big(\sum_j |f_j|^2 \big)^{1/2} \Big\|_{L^p(X)}.$$

Proof. By using mixed-norm duality, it suffices to prove the result for $p > 2$ (exercise!). So, with $r = (p/2)'$, for an appropriate $g \geq 0$ satisfying $\|g\|_{L^r(X)} = 1$, we estimate

$$\begin{aligned}
\Big\| \big(\sum_j |K_j * f_j|^2 \big)^{1/2} \Big\|_{L^p(X)}^2 &\lesssim \sum_{x \in \mathbb{Z}} \sum_j |K_j * f_j(x)|^2 \cdot g(x) \\
&\lesssim \sum_{x \in \mathbb{Z}} \sum_j |f_j(x)|^2 \cdot \sup_j |\tilde{K_j}| * g(x) \\
&\lesssim \| \sum_j |f_j|^2 \|_{L^{p/2}(X)} = \\
&\lesssim \Big\| \big(\sum_j |f_j|^2 \big)^{1/2} \Big\|_{L^p(X)}^2,
\end{aligned}$$

where $\tilde{h}(x) := h(-x)$ as above, and we have used convexity to replace

$$\sum_{x \in \mathbb{Z}} |K_j * f|^2(x) \cdot g(x) \leq \sum_{x \in \mathbb{Z}^2} |f|^2(x) \cdot \sup_j |\tilde{K_j}| * g(x).$$

□

Proof of Corollary 3.15; Sketch. By using mixed-norm duality, it suffices to prove that S_c is bounded on $\ell^p(\mathbb{Z})$ for $p > 2$ (exercise!); by interpolation against the $\ell^2(\mathbb{Z})$-estimate, it suffices to show that

$$\Big(\sum_{n \geq 0} |(M_{2^n} - \mathbb{E}_n)\Delta_{n+t} f|^2 \Big)^{1/2} \tag{3.97}$$

is bounded on $\ell^p(\mathbb{Z})$ for each $1 < p < \infty$. Although $M_{2^n} - \mathbb{E}_n$ are not convolution operators, one may apply the *proof* of Lemma 3.16 to conclude the result. □

Exercise 3.98. Fill in the details of the above proof.
Hint: The key point is to prove the bound

$$\sum_{\mathbb{Z}} |M_{2^n} f - \mathbb{E}_n f|^2 \cdot g \lesssim \sum_{\mathbb{Z}} |f|^2 \cdot M_{HL} g.$$

We next turn to the short variation. We provide two proofs; a spatial proof, reliant on the smoothness of the intervals $[0, i)$, and a Fourier-analytic proof which proceeds by analysis the *Fourier multipliers*,

$$\widehat{M_t f}(\beta) := \mathbf{m}_t(\beta) \cdot \hat{f}(\beta) := \left(\frac{1}{t} \sum_{n \leq t} e(-\beta n) \right) \hat{f}(\beta).$$

3.3. The Short Variation via Spatial Arguments.

The Short Variation is $\ell^p(\mathbb{Z})$-Bounded, Take One. We will abbreviate

$$\mathcal{V}^{r,S}(M_i f) \leq \mathcal{V}^{2,S}(M_i f) := \left(\sum_n \mathcal{V}_n^2(M_i f)^2 \right)^{1/2},$$

where

$$\mathcal{V}_n^2(M_i f) := \left(\sup_{2^n \leq (i_k) \leq 2^{n+1} \text{ increasing}} \sum_k |M_{i_k} f - M_{i_{k+1}} f|^2 \right)^{1/2}.$$

By arguing as previously, the $\ell^2(\mathbb{Z})$-theory will follow once we have established the estimate

$$\|\mathcal{V}_n^2(M_i(\Delta_m f))\|_{\ell^2(\mathbb{Z})} \lesssim 2^{-|n-m|/2} \cdot \|f\|_{\ell^2(\mathbb{Z})}.$$

We first notice that

$$\mathcal{V}_n^2(M_i g) \leq \sum_{i=2^n}^{2^{n+1}-1} |M_i f - M_{i+1} f| \lesssim M_{2^{n+1}} |f|;$$

consequently we may assume $|n-m| \gg 1$. We begin with the case where $n \leq m - C$.

The key observation here is that

$$\mathcal{V}_n^2(M_i(\Delta_m f))(x) \neq 0$$

implies that x is supported in the $\approx 2^n$-neighborhood of the 2^m-dyadic mesh.

Exercise 3.99. Verify this claim.

With this in mind, letting $E_n(Q) \subset Q$ denote the set (union of two intervals) of measure $\lesssim 2^n$ on which $\mathcal{V}_n^2(M_i(\Delta_m f)) \cdot \mathbf{1}_Q$ is supported, we majorize

$$\mathcal{V}_n^2(M_i(\Delta_m f)) \lesssim \sum_Q \big(\sum_P |a_P| \cdot M_{2^{n+1}} \mathbf{1}_P \big) \cdot \mathbf{1}_{E_n(Q)}$$

$$\leq \sum_Q (\sum_{i=0}^{1} |a_{Q_i}|) \cdot \mathbf{1}_{E_n(Q)}, \tag{3.100}$$

where we set $Q_0 := Q$ and Q_1 to be the immediate left dyadic neighbor of Q. Here, we define a_P via

$$\Delta_m f = \sum_{|Q|=2^{m-1}} a_Q \cdot \mathbf{1}_Q.$$

Summing,

$$\|\mathcal{V}_n^2(M_i(\Delta_m f))\|_{\ell^2(\mathbb{Z})}^2 \lesssim \sum_Q |a_Q|^2 \cdot |E_n(Q)|$$

$$\lesssim 2^{n-m} \cdot \sum_Q |a_Q|^2 \cdot |Q|$$

$$= 2^{n-m} \cdot \|\Delta_{n-m} f\|_{\ell^2(\mathbb{Z})}^2.$$

We next turn to the case where $n \geq m + C$; here, we will establish the pointwise majorization

$$\mathcal{V}_n^2(M_i(\Delta_m f))^2 \lesssim 2^{m-n} \cdot M_{2^{n+1}}(|\Delta_m f|^2), \tag{3.101}$$

which will yield the result.

To this end, let us freeze some x, and suppose that the nth increment of the short two-variation is realized as

$$\sum_s |(M_{i_s} - M_{i_{s+1}})\big(\Delta_m f\big)|^2(x)$$
$$= \sum_s \Big|\frac{1}{i_s} - \frac{1}{i_{s+1}}\Big|^2 \cdot \Big| \sum_{y\in[0,i_s)} (\Delta_m f)(x-y)\Big|^2$$
$$+ \sum_s \Big|\frac{1}{i_{s+1}}\Big|^2 \cdot \Big| \sum_{y\in[i_s,i_{s+1})} (\Delta_m f)(x-y)\Big|^2.$$

Expand

$$\Delta_m f = \sum_Q \langle f, h_Q\rangle \cdot h_Q$$

into Haar coefficients, and let $Q_0 = Q_0(x,s)$ and $Q_1 = Q_1(x,s)$ be appropriately chosen; one discovers that

$$\Big| \sum_{y\in[0,i_s)} (\Delta_m f)(x-y)\Big|^2 \lesssim \sum_{i=0}^{1} |\langle f, h_{Q_i}\rangle|^2 \cdot \Big| \sum_{[0,i_s)} h_{Q_i}\Big|^2$$
$$\lesssim \sum_{i=0}^{1} |\langle f, h_{Q_i}\rangle|^2 \cdot 2^m.$$

This quantity is in turn majorized by

$$2^m \cdot \sum_{y\in[0,i_s)} |(\Delta_m f)(x-y)|^2 \leq 2^m \cdot \sum_{y\in[0,2^{n+1})} |\Delta_m f(x-y)|^2, \tag{3.102}$$

which offers a power savings over the trivial estimate

$$\Big| \sum_{y\in[0,i_s)} (\Delta_m f)(x-y)\Big|^2 \leq i_s \cdot \sum_{y\in[0,i_s)} |\Delta_m f(x-y)|^2$$
$$\lesssim 2^n \sum_{y\in[0,2^{n+1})} |\Delta_m f(x-y)|^2.$$

Inserting the upper bound (3.102) into the sum

$$\sum_s \Big|\frac{1}{i_s} - \frac{1}{i_{s+1}}\Big|^2 \cdot \Big| \sum_{y\in[0,i_s)} (\Delta_m f)(x-y)\Big|^2$$

yields an upper estimate of

$$\lesssim 2^{m-n} \cdot M_{2^{n+1}}(|\Delta_m f|)^2(x).$$

The contribution from the second sum,

$$\sum_s \Big|\frac{1}{i_{s+1}}\Big|^2 \cdot \Big| \sum_{y\in[i_s,i_{s+1})} (\Delta_m f)(x-y)\Big|^2,$$

is handled similarly; the key point is the "improved Cauchy-Schwarz"-type estimate,

$$\Big|\sum_{y\in[i_s,i_{s+1})}(\Delta_m f)(x-y)\Big|^2 \lesssim 2^m\cdot\sum_{y\in[i_s,i_{s+1})}|\Delta_m f(x-y)|^2. \tag{3.103}$$

The proof is complete. □

Exercise 3.104. Establish (3.103).

By arguing as in the proof of Corollary 3.15, we may extend this result to all $\ell^p(\mathbb{Z})$.

Corollary 3.17. *For each* $1<p<\infty$, $\mathcal{V}^{2,S}(M_i f)$ *is* $\ell^p(\mathbb{Z})$*-bounded.*

Exercise 3.105. Prove Corollary 3.17.
Hint: The key point is the bound

$$\sum_{\mathbb{Z}}\mathcal{V}^2_n(M_i f)^2\cdot g\lesssim\sum_{\mathbb{Z}}|f|^2\cdot M_{HL}g.$$

Exercise 3.106. Suppose ϕ is a sufficiently smooth, rapidly decaying function, and for integers $r\geq 1$, consider

$$\Phi_r f(x):=\sum_{y\in\mathbb{Z}}r^{-1}\phi(y/r)f(x-y).$$

Show that the conclusion of the previous exercise continues to hold with

$$\sum_{\mathbb{Z}}\mathcal{V}^2_n(\Phi_i f)^2\cdot g\lesssim\sum_{\mathbb{Z}}|f|^2\cdot M_{\mathrm{HL}}g,$$

and conclude that

$$S_\psi f:=\Big(\sum_k|\psi_k*f|^2\Big)^{1/2} \tag{3.107}$$

is similarly $\ell^p(\mathbb{Z})$-bounded, where

$$\phi_k:=\phi_r,\quad r=2^k$$

and

$$\psi_k=\phi_k-\phi_{k+1}.$$

What is the minimal amount of decay/regularity needed?
Hint: For the final point, dominate

$$N_\lambda\Big(\sum_{j=1}^\infty f_k^j:k\Big)\leq\sum_{j=1}^\infty N_{c\lambda/j^2}\Big(f_k^j:k\Big)$$

for some absolute $c > 0$.

Exercise 3.108. The boundedness of the square function S'_ϕ is a special case of a broader phenomenon, presented in Euclidean space: for any collection of mean-zero $\{\psi_k\} \subset \mathcal{C}^1(\mathbb{R})$ so that

$$|\psi_k(x)| + 2^k \cdot |\partial \psi_k(x)| \leq 2^{-k} \cdot (1 + 2^{-k}|x|)^{-10}$$

(say), the square function

$$S_\psi f := \Big(\sum_k |\psi_k * f|^2 \Big)^{1/2} \tag{3.109}$$

is $L^p(\mathbb{R})$-bounded. Can you replace 10 with a smaller exponent?
Hint: For the $L^2(\mathbb{R})$-boundedness, prove that $|\widehat{\psi_k}(\xi)| \lesssim \min\{2^k|\xi|, \frac{1}{2^k|\xi|}\}$; to extend this to $L^p(\mathbb{R})$, argue as in Corollary 3.15, looking for orthogonality between $\psi_n \Delta_{n+t}$ as $|t|$ grows. One way to begin is via the convexity relationship

$$\psi_k(x) = - \int \frac{1}{t} \mathbf{1}_{[0,t]}(x) \cdot \ (t\partial_t \psi_k(t)) dt, \tag{3.110}$$

which expresses ψ_k as a convex-combination combination L^1-normalized indicator functions:

$$\sup_k \| t \partial_t \psi_k(t) \|_{L^1(\mathbb{R})} \lesssim 1.$$

Remark 3.111. The class of square functions identfied above are known as *Littlewood-Paley* square functions, and the collection $\{\psi_k\}$ are known as Littlewood-Paley projections, in the sense that

$$\widehat{f * \psi_k}$$

is highly concentrated on $\{|\xi| \approx 2^{-k}\}$. *Littlewood-Paley* theory is an approach to studying (typically) translation and dilation invariant operators by analyzing the effect of these operators on highly symmetrized Littlewood-Paley data, i.e. functions convolved with Littlewood-Paley projections, and then using various orthogonality techniques to reconstitute appropriately, see e.g. Lemma 3.16.

We finish this section by recording the following counterpart to Burkholder's Theorem 3.6 for the square functions considered above.

Lemma 3.18. *There exist (mean-zero) Schwartz functions* ψ *so that the square function*

$$Sf := S_\psi f := \Big(\sum_k |\psi_k * f|^2\Big)^{1/2} \tag{3.112}$$

satisfies

$$\|Sf\|_{L^p(\mathbb{R})} \approx_p \|f\|_{L^p(\mathbb{R})}$$

for each $1 < p < \infty$.

Proof. We use the duality argument given in our discussion of Theorem 3.6:

Choose a smooth even bump function

$$\mathbf{1}_{B(1/2)} \leq \eta \leq \mathbf{1}_{B(1)},$$

and define

$$\hat{\psi}(\xi) := \sqrt{\eta(\xi) - \eta(2\xi)},$$

so that ψ is a mean-zero Schwartz function with

$$\sum_j \widehat{\psi_j}(\xi)^2 := \sum_j \widehat{\psi}(2^j\xi)^2 = \mathbf{1}$$

for all $\xi \neq 0$; in particular, assuming the convergence

$$f = \sum_j \psi_j * \psi_j * f \tag{3.113}$$

for any $f \in L^p(\mathbb{R}),\ 1 < p < \infty$, we use Exercise 3.106 to simply express

$$\begin{aligned}\left|\int f \cdot g\right| &= \left|\int \sum_{|j| \leq N} \psi_j * f \cdot \psi_j * g\right| + o_{N\to\infty}(1) \\ &\leq \int Sf \cdot Sg + o_{N\to\infty}(1) \\ &\leq \|Sf\|_{L^p(\mathbb{R})} \cdot \|Sg\|_{L^{p'}(\mathbb{R})} + o_{N\to\infty}(1) \\ &\lesssim \|Sf\|_{L^p(\mathbb{R})} \cdot \|g\|_{L^{p'}(\mathbb{R})}\end{aligned}$$

upon sending $N \to \infty$ and using the $L^{p'}(\mathbb{R})$-boundedness of $Sg := \Big(\sum_j |\psi_j * g|^2\Big)^{1/2}$, see Exercises 3.106 and 3.108 and Exercise 3.115 below.

The result now follows by duality. □

Exercise 3.114. Establish the norm convergence (3.113).
Hint: By first showing that (3.113) *holds pointwise and in norm for the*

dense subset of $L^p(\mathbb{R})$ consisting of mean-zero Schwartz functions, use a density argument to reduce to proving that

$$\sup_N \| \sum_{|j|\leq N} \psi_j * \psi_j * f\|_{L^p(\mathbb{R})} \lesssim \|f\|_{L^p(\mathbb{R})}$$

for each $1 < p < \infty$. Apply Theorem 3.6, to reduce to bounding

$$\Big(\sum_k |\Delta_k(\sum_{|j|\leq N} \psi_j' * f)|^2\Big)^{1/2}, \quad \psi_j' := \psi_j * \psi_j$$

on $L^p(\mathbb{R})$, $1 < p < \infty$, independent on N. The foregoing is majorized by

$$\sum_l \Big(\sum_k |\Delta_k(\psi_{k+l}' * f)|^2\Big)^{1/2};$$

the square functions are uniformly bounded on $L^p(\mathbb{R})$ by arguing as in Lemma 3.16, while one can show that

$$\|\Delta_k(\psi_{k+l}' * f)\|_{L^2(\mathbb{R})} \lesssim 2^{-|l|/2} \cdot \|f\|_{L^2(\mathbb{R})}$$

by arguing as in (3.93) or Exercise 3.95, keeping in mind Examples 2.5 and 2.6; note that by duality

$$\sup_{\|f\|_{L^2(\mathbb{R})}=1} \|\Delta_k(\psi_{k+l}' * f)\|_{L^2(\mathbb{R})} = \sup_{\|f\|_{L^2(\mathbb{R})}=1} \|\tilde{\psi}_{k+l}' * (\Delta_k f)\|_{L^2(\mathbb{R})},$$

where $\tilde{g}(x) := g(-x)$ is given by relection.

Exercise 3.115. In higher dimensions, let $\psi_j \in \mathcal{C}^1(\mathbb{R}^D)$ be mean-zero functions satisfying the derivative conditions (say)

$$2^{jD} \cdot (1+2^{-j}|x|)^{D+1} \cdot |\psi_j(x)| + 2^{j(D+1)} \cdot (1+2^{-j}|x|)^{D+1} \cdot |\nabla \psi_j(x)| \leq C,$$

and establish the $L^p(\mathbb{R}^D)$-boundedness of

$$S_\psi g^2 := \sum_j |\psi_j * g|^2.$$

Hint: One can develop the theory of the higher-dimensional dyadic square function, $S_{D,D}$, see (3.139) in §4 below, and then compare to S_ψ spatially as in the one-dimensional case:

$$\begin{aligned} S_\psi g &= \Big(\sum_k |\psi_k * (\sum_l \Delta_{k+l} g)|^2\Big)^{1/2} \\ &\leq \sum_l \Big(\sum_k |\psi_k * (\Delta_{k+l} g)|^2\Big)^{1/2} \\ &= \sum_l \Big(\sum_k |\psi_k * \Delta_{k+l}(\Delta_{k+l} g)|^2\Big)^{1/2}. \end{aligned}$$

In particular, after applying the argument of Lemma 3.16 *to uniformly bound the lth square function on* $L^p(\mathbb{R}^D)$*, it suffices to exhibit gain in* $L^2(\mathbb{R}^D)$ *for the lth square function. To do so, prove that*

$$\|\psi_k * (\Delta_{k+l} f)\|_{L^2(\mathbb{R}^D)} \lesssim 2^{-|l|/2} \cdot \|f\|_{L^2(\mathbb{R}^D)}$$

as in Exercise 3.114*, see* (3.96) *for instance.*

We now turn to a second proof that the short variation is $\ell^p(\mathbb{Z})$-bounded. Although the argument we provide is more involved, it relies on more general Fourier-analytic properties of the multipliers,

$$\mathbf{m}_t(\beta) := \frac{1}{t} \sum_{n \le t} e(-\beta n),$$

rather than concrete geometric considerations.

Before presenting this argument, we pause to develop the properties of the Fourier transform that we will need.

3.4. The Short Variation via Fourier Analysis. Recall the definition of the discrete Fourier transform:

$$\hat{f}(\beta) := \sum_x f(x) e(-\beta x) \tag{3.116}$$

where we (initially) restrict attention to only finitely supported functions (in particular, there are no convergence issues in (3.116)).

We also recall the inverse Fourier transform

$$g^\vee(x) := \int_{\mathbb{T}} g(\beta) e(\beta x) \, d\beta \tag{3.117}$$

for $g \in L^1(\mathbb{T})$; see Chapter 1 §5 for a discussion.

We will repeatedly make use of the following heuristic, which is essentially a consequence of the "diagonalization" argument used to estimate S_c.

> **If $\hat{f}$ has large Fourier coefficients in the range $\{|\beta| \approx 2^{-n}\}$, then $\Delta_n f$ has large ℓ^2 norm.**

This heuristic is an instance of the *uncertainty principle.*

With the above discussion in mind, we are prepared for our second short-variation argument:

Recall that we are interested in showing that

$$\Big\| \big(\sum_{n \geq 0} |\mathcal{V}^2_n (M_t f)|^2 \big)^{1/2} \Big\|_{\ell^p(\mathbb{Z})} \lesssim \|f\|_{\ell^p(\mathbb{Z})};$$

since we are summing over *differences* of operators, this is equivalent to showing

$$\Big\| \big(\sum_{n \geq 0} |\mathcal{V}^2_n (C_t f)|^2 \big)^{1/2} \Big\|_{\ell^p(\mathbb{Z})} \lesssim \|f\|_{\ell^p(\mathbb{Z})}$$

where we define

$$C_t := M_t - M_{2^n} \text{ for } 2^n \leq t < 2^{n+1}.$$

We have seen that for each n, the operator M_{2^n} "annihilates" functions which oscillate at scales $\ll 2^n$, and "preserves" functions which oscillate only at scales $\gg 2^n$, see Examples 2.5 and 2.7 for the analogous continuous formulation. The take-away from the above spatial argument was that there is great – and quantifiable – similarity between the operators M_{2^n} and M_t for $2^n \leq t < 2^{n+1}$. We might expect, therefore, the operators $\{C_t : 2^n \leq t < 2^{n+1}\}$ to be most sensitive to functions which oscillate at scales $\approx 2^n$. These heuristics can be made rather precise using Fourier analysis. In what follows, we will let

$$v_t(\beta) := \mathbf{m}_t(\beta) - \mathbf{m}_{2^n}(\beta),$$

for $2^n \leq t < 2^{n+1}$, so that

$$\widehat{C_t f}(\beta) = v_t(\beta) \cdot \hat{f}(\beta).$$

Lemma 3.19. *Suppose $2^n \leq t < 2^{n+1}$. For any $\beta \in \mathbb{T}$, we have the following estimates:*

(1) $|v_t(\beta) - v_{t+1}(\beta)| \lesssim 2^{-n}$;
(2) $|v_t(\beta)| \lesssim \min\{2^n|\beta|, \frac{1}{2^n|\beta|}\}$.

This lemma spectrally quantifies the above discussion: the qualitative similarity between averaging operators C_t and C_{t+1} (or, M_t and M_{t+1}) is felt via strong Fourier-dampening; the multiplier $v_t(\beta)$ is morally felt only by functions built up of frequencies on the order of 2^{-n}, i.e. functions which are morally constant at scales $\ll 2^n$, and morally mean-zero at scales $\gtrsim 2^n$.

Exercise 3.118. Verify Lemma 3.19.

The above lemma strongly motivates us to introduce a Littlewood-Paley decomposition of the torus $\mathbb{T} \cong [-1/2, 1/2)$ to study the behavior of the $\{v_t\}$. Accordingly, let $\mathbf{1}_{[-1/10,1/10]} \leq \hat{\eta} \leq \mathbf{1}_{[-1/5,1/5]}$ be a smooth bump function, let

$$\widehat{\psi_j}(\beta) := \hat{\eta}(2^j \beta) - \hat{\eta}(2^{j+1}\beta) \tag{3.119}$$

for $j \geq 1$ and let

$$\psi_0(x) := \delta_0(x) - \eta(x/2)/2, \tag{3.120}$$

where δ_0 is the point mass at the origin. Notice that

$$\sum_{j \geq 0} \widehat{\psi_j}(\beta) = \mathbf{1}_{\mathbb{T} \smallsetminus \{0\}},$$

so that

$$\sum_{j \geq 0} \psi_j * f = f$$

with convergence in $\ell^p(\mathbb{Z})$ for any $1 < p < \infty$ (what fails at the endpoints?) by Exercise 3.121 below.

Exercise 3.121. Prove the convergence in norm,

$$\sum_{j \geq 0} \psi_j * f = f.$$

Hint: Combine the approach of Exercise 3.114 *with an appropriate analogue of* (3.96).

For $f \in \ell^p(\mathbb{Z})$, we let

$$f_j(n) := f * \psi_j \tag{3.122}$$

denote the inverse Fourier transform of the projection to frequency scales $\approx 2^{-j}$; these form rough analogues for the projection operators $\Delta_j f$.

Accordingly, our arguments will follow a similar strategy to that appearing in the previous section:

The idea will be to first gain strong $\ell^2(\mathbb{Z})$-decay in $|m|$ as we consider

$$\mathcal{V}^2_n(C_t f_{n+m});$$

we are lead to this approach by the fact that for $2^n \leq t < 2^{n+1}$,

$$\|v_t \widehat{\psi_j}\|_{L^\infty(\mathbb{T})} \lesssim 2^{-|m|}.$$

We will then seek to show that $\mathcal{V}^2_n(C_t f_{n+m})$ are uniformly bounded on $\ell^p(\mathbb{Z})$, and will ultimately recombine scales using variants of the square function arguments used above.

To execute, we begin with the following elementary lemma. This result is essentially due to Bourgain (cf. [**21**, Lemma 3.11], and [**94**, Proposition 2.12] as well).

Lemma 3.20. *Suppose that* $\{T_n\}_{n=1}^N$ *are a family of operators which act on* $\ell^2(\mathbb{Z})$*-functions by multiplication on the Fourier side*

$$\widehat{T_n f}(\beta) := \mathbf{m}_n(\beta) \cdot \hat{f}(\beta).$$

Suppose that for each $1 \le n \le N$

$$\sup_{\beta \in \operatorname{supp} \hat{f}} |\mathbf{m}_n(\beta)| \le A \text{ and}$$
$$\sup_{\beta \in \operatorname{supp} \hat{f}} |\mathbf{m}_n(\beta) - \mathbf{m}_{n+1}(\beta)| \le a.$$

Then $\|\mathcal{V}^2(T_n f)\|_{\ell^2(\mathbb{Z})} \lesssim \sqrt{NAa} \cdot \|f\|_{\ell^2(\mathbb{Z})}$.

Proof. We may assume without loss of generality that $Na > A$.

Exercise 3.123. Establish this reduction.
Hint: Begin by majorizing $\mathcal{V}^2$ *by* $\mathcal{V}^1$.

Proceeding therefore under the assumption that $Na > A$, we let L be a positive integer to be determined, and choose an $\approx \frac{N}{L}$-net,

$$1 = a_1 < a_2 < \dots a_L = N$$

of (almost) equally spaced indices (so $a_{i+1} - a_i \approx \frac{N}{L}$). We use this net to pointwise dominate

$$\begin{aligned}
&\mathcal{V}^2(T_n f) \\
&\le \Big(\sum_{i=1}^{L} |T_{a_i} f|^2\Big)^{1/2} + \Big(\sum_{i=1}^{L} \mathcal{V}^2(T_n f : a_i \le n \le a_{i+1})^2\Big)^{1/2} \\
&\le \Big(\sum_{i=1}^{L} |T_{a_i} f|^2\Big)^{1/2} + \Big(\sum_{i=1}^{L} \Big(\sum_{n=a_i}^{a_{i+1}} |T_n f - T_{n+1} f|\Big)^2\Big)^{1/2} \\
&\lesssim \Big(\sum_{i=1}^{L} |T_{a_i} f|^2\Big)^{1/2} + \left(\frac{N}{L}\right)^{1/2} \cdot \Big(\sum_{n=1}^{N} |T_n f - T_{n+1} f|^2\Big)^{1/2},
\end{aligned}$$

where we used Cauchy-Schwartz in the last inequality.

We take $\ell^2(\mathbb{Z})$-norms, and use Plancherel's theorem to majorize the first summand

$$\begin{aligned}
\Big\| \Big(\sum_{i=1}^{L} |T_{a_i} f|^2\Big)^{1/2} \Big\|_{\ell^2(\mathbb{Z})} &= \Big(\sum_{i=1}^{L} \|\mathbf{m}_{a_i} \cdot \hat{f}\|_{L^2(\mathbb{T})}^2\Big)^{1/2} \\
&\le \Big(\sum_{i=1}^{L} A^2 \cdot \|\hat{f}\|_{L^2(\mathbb{T})}^2\Big)^{1/2} \\
&\le \sqrt{L} A \cdot \|f\|_{\ell^2(\mathbb{Z})}
\end{aligned}$$

and the second summand

$$\begin{aligned}
\left(\frac{N}{L}\right)^{1/2} &\cdot \Big\| \big(\sum_{n=1}^{N} |B_n f - B_{n+1} f|^2 \big)^{1/2} \Big\|_{\ell^2(\mathbb{Z})} \\
&= \left(\frac{N}{L}\right)^{1/2} \cdot \big(\sum_{n=1}^{N} \| (\mathbf{m}_n - \mathbf{m}_{n+1}) \cdot \hat{f} \|_{L^2(\mathbb{T})}^2 \big)^{1/2} \\
&\leq \left(\frac{N}{L}\right)^{1/2} \cdot \sqrt{N} a \cdot \|f\|_{\ell^2(\mathbb{Z})} \\
&= \frac{Na}{\sqrt{L}} \cdot \|f\|_{\ell^2(\mathbb{Z})}.
\end{aligned}$$

Setting $L \approx \frac{Na}{A}$ yields the result. □

With this lemma in hand, the $\ell^2(\mathbb{Z})$-proof can be completed in a few short strokes.

Exercise 3.124. Use Lemma 3.19, Lemma 3.20, and the orthogonality approach of (3.92) to prove the $\ell^2(\mathbb{Z})$-result:

$$\left(\sum_n \mathcal{V}_n^2 (C_t f_{n+m})^2 \right)^{1/2} \tag{3.125}$$

has $\ell^2(\mathbb{Z})$ norm $\lesssim 2^{-|m|/2} \cdot \|f\|_{\ell^2(\mathbb{Z})}$.

The passage to $\ell^p(\mathbb{Z})$-theory is similar, and one begins with the following refinement of Lemma 3.20.

Exercise 3.126. Prove that for any $1 \leq p \leq \infty$, one may estimate

$$\|\mathcal{V}^2(T_n f)\|_{\ell^p(\mathbb{Z})} \lesssim (Na)^{1/p} \cdot A^{1-1/p} \cdot \|f\|_{\ell^p(\mathbb{Z})} \tag{3.127}$$

where here

$$\begin{aligned}
a &:= \sup_{1 \leq n < N} \|T_n - T_{n+1}\|_{\ell^p(\mathbb{Z}) \to \ell^p(\mathbb{Z})} \text{ and} \\
A &:= \sup_{1 \leq n \leq N} \|T_n\|_{\ell^p(\mathbb{Z}) \to \ell^p(\mathbb{Z})}.
\end{aligned}$$

Conclude that $\|\mathcal{V}_n^2(f_{n+m})\|_{\ell^p(\mathbb{Z})} \lesssim \|f_{n+m}\|_{\ell^p(\mathbb{Z})}$.

By interpolation, Exercise 3.126 allows us to estimate

$$\|\mathcal{V}_n^2(f_{n+m})\|_{\ell^p(\mathbb{Z})} \lesssim 2^{-\epsilon_p |m|} \cdot \|f_{n+m}\|_{\ell^p(\mathbb{Z})}, \ \epsilon_p > 0 \tag{3.128}$$

for each $1 < p < \infty$. The problem will ultimately amount to combining scales, which we do via square function arguments; the following exercise contains the details of this reduction.

Exercise 3.129. Prove that

$$\sum_{x\in\mathbb{Z}} \mathcal{V}_n^2(C_t f)^2 \cdot g \lesssim \sum_{x\in\mathbb{Z}} |f|^2 \cdot M_{HL} g.$$

Temporarily assuming the $\ell^p(\mathbb{Z})$-boundedness, $1 < p < \infty$, of the square function

$$S' f := \left(\sum_n |f_n|^2 \right)^{1/2}, \tag{3.130}$$

see Lemma 3.21 below, conclude that

$$\|(3.125)\|_{\ell^p(\mathbb{Z})} \lesssim 2^{-\epsilon_p |m|} \cdot \|f\|_{\ell^p(\mathbb{Z})}, \ \epsilon_p > 0$$

provided that $1 < p < \infty$.
Hint: For the first point, see (3.100) *and* (3.101)*; Lemma* 3.16 *will be helpful.*

By the previous exercise, the argument will be complete once we prove the following lemma.

Lemma 3.21. *For each* $1 < p < \infty$,

$$\|S' f\|_{\ell^p(\mathbb{Z})} \lesssim \|f\|_{\ell^p(\mathbb{Z})},$$

see (3.130) *and* (3.122)*.*

The Euclidean analogue of these estimates are classical, and Lemma 3.21 can be established by mimicking these arguments; in particular, endpoint estimates are also available. We avoid these complications by making use of our martingale technology.

Sketch. By expanding $f = \sum_m \Delta_m f$ into martingale increments and arguing as previously, it suffices to show that

$$\left(\sum_n |\psi_n * \Delta_{n+m} f|^2 \right)^{1/2}$$

has $\ell^p(\mathbb{Z})$-norms which decay geometrically in $|m|$. An $\ell^p(\mathbb{Z})$-estimate without any gain in $|m|$ follows from Corollary 3.15, so it suffices to establish the relevant gain in $\ell^2(\mathbb{Z})$; see Exercise 3.108. □

Exercise 3.131. Make the proof of Lemma 3.21 rigorous.
Hint: View

$$\psi_n * (\Delta_{n+m} f) = \psi_n * \Delta_{n+m}\big((\Delta_{n+m} f)\big)$$

as an integral operator as in (3.96) *and estimate the kernel as in Exercise* 3.95*.*

So, despite the increased delicacy of the variation operators, we have shown that for $E_n = [0, n), r > 2$, the r-variation operators, $\mathcal{V}^r(M_n f)$, are of strong-type (p, p) [**21**, Corollary 3.26].[4]

In other words, not only do the classical Birkhoff ergodic means $\{M_n^\tau f\}$ converge for $f \in L^2(X)$, but they do so *rapidly.*

Exercise 3.132. Prove that for $1 < p < \infty$

$$\sup_\lambda \|\lambda \cdot N_\lambda(M_t f)^{1/2}\|_{\ell^p(\mathbb{Z})} \lesssim_p \|f\|_{\ell^p(\mathbb{Z})}.$$

Exercise 3.133. Suppose $\{\nu_k\}$ are a sequence of non-negative finitely supported functions on the integers with

$$\|\nu_k\|_{\ell^1(\mathbb{Z})} = 1$$

for all k, and $\operatorname{supp} \nu_k \subset \{|x| \lesssim 2^k\}$, and

$$|\widehat{\nu_k}(\beta)| \lesssim |2^k \beta|^{-c},$$

for some $c > 0$. Prove that for each $r > 2$

$$\|\mathcal{V}^r(\nu_k * f : k)\|_{\ell^2(\mathbb{Z})} \lesssim \frac{r}{r-2} \cdot \|f\|_{\ell^2(\mathbb{Z})}$$

and

$$\sup_\lambda \|\lambda \cdot N_\lambda(\nu_k * f)^{1/2}\|_{\ell^2(\mathbb{Z})} \lesssim \|f\|_{\ell^2(\mathbb{Z})}.$$

Hint: Compute the Fourier behaviour of $\widehat{\nu_k}(\beta)$ for $|\beta| \lesssim 2^k$, and then consider the square function

$$\left(\sum_k |\nu_k * f - M_{2^k} f|^2\right)^{1/2};$$

this argument also applies to $\lambda \cdot N_\lambda(\nu_k * f : k)^{1/2}$.

Remark 3.134. The conclusion of this lemma holds on $\ell^p(\mathbb{Z}),\ 1 < p < \infty$ as well provided that the maximal function

$$\sup_k \nu_k * |f|$$

is bounded on $\ell^{p_0}(\mathbb{Z})$ for some $p_0 < \infty$, see Chapter 11 below.

[4]It is shown in [**94**] that $\mathcal{V}^2(M_n f)$ is unbounded on L^2.

Exercise 3.135. Prove the continuous analogue of the previous results for the Lebesgue averaging operators: if we define

$$M_t^{\mathbb{R}} f(x) := \frac{1}{t} \int_0^t f(x-y)\, dy \tag{3.136}$$

for functions on the line, prove that for $1 < p < \infty$

$$\sup_\lambda \|\lambda \cdot N_\lambda(M_t^{\mathbb{R}} f)^{1/2}\|_{L^p(\mathbb{R})} + \|\mathcal{V}^r(M_t^{\mathbb{R}} f)\|_{L^p(\mathbb{R})} \lesssim_p \frac{r}{r-2} \cdot \|f\|_{L^p(\mathbb{R})}.$$

So concludes our treatment of Birkhoff's theorem. In the next chapter we will begin our study of the maximal functions,

$$\sup_N \left| \frac{1}{N} \sum_{n \leq N} f(x - P(n)) \right|,$$

where $P \in \mathbb{Z}[\cdot]$. Before proceeding, we conclude with a few further remarks about variation in discrete harmonic analysis.

3.5. Further Variations. Since Bourgain's variational result, establishing variational estimates for families of averaging operators has been the focus of much research in ergodic theory. A fundamental paper in this direction is due to Jones et. al. [**94**], where it is shown that the variation operators $\{\mathcal{V}^r(M_i f)\}_{r>2}$ are bounded on $\ell^p(\mathbb{Z})$, $1 < p < \infty$, of weak-type $(1,1)$, and bounded from $\ell^\infty(\mathbb{Z}) \to \text{BMO}$, for intervals $E_n = [0, n)$.

A major point in proving $L^p \to L^p$ estimates on the variation operators taken over *linear* averaging operators is that, roughly speaking, it is challenging for the variation operators to be large where the governing maximal operator is not. Indeed, we will see an explicit example of this phenomenon in our good-λ approach to Lépingle's theorem, Theorem 3.11. (This heuristic was further strengthened in [**102**].)

In the case when E_n are *polynomially defined*, the situation becomes much more complicated:

Variation operators "along" polynomially defined sets are a natural object of consideration: variational estimates are a strong tool for proving pointwise convergence of averages when a density argument is unavailable. This theme has lead to several papers in recent years (see e.g. [**103**], [**139**], [**189**] and the references therein); full variational estimates for averaging operators (and maximally truncated discrete singular integrals) were proven by Mirek, Stein, and Trojan in 2015 [**134**], with the end-point jump inequality addressed in 2018 by Mirek, Stein, and Zorin-Kranich, [**137**]. Using more modern machinery, we will be able to give a brief proof of these results in Chapter 9.

We conclude these notes with the stopping-time proof of Theorem 3.11, and an extension to the related convolution operators.

4. The Proof of Lépingle's Inequality

Throughout this section we will abbreviate

$$\mathcal{V}^r f := \mathcal{V}^r(\mathbb{E}_k f).$$

Since we will be interested in establishing Lépingle's Inequality in the general D-dimensional setting, we begin by establishing the D-dimensional formulation of Theorem 3.10, which we do by way of an appropriate good-λ inequality.

First, we begin by localizing the martingale increments $\{\Delta_k f : k\}$:

For each $Q \in \mathcal{D}(k)$, define

$$a_Q(f) := \Delta_{k+1} f \cdot \mathbf{1}_Q,$$

so that $a_Q(f)$ is constant on Q.

Exercise 3.137. Verify that

$$\bigcup_k \{a_Q(f) : Q \in \mathcal{D}(k)\}$$

are orthogonal.

Similar to the one-dimensional setting, we may express each increment

$$a_Q(f)$$

using an orthonormal basis for mean-zero functions on Q

$$\{h_{Q,1}, \ldots, h_{Q;D'}\}, \quad D' := 2^D - 1,$$

which we now proceed to define inductively, using an elegant argument of [**186**]; by exploiting the symmetries of the dyadic grid, we can assume that $Q_D = [0,1]^D$.

So, suppose that

$$\big\{(E_j, F_j) : 1 \leq j \leq D'\big\} \subset Q_D := [0,1]^D$$

are a collection of pairwise disjoint sets (up to measure zero) satisfing

$$|E_j| = |F_j|$$

and the nesting property that whenever $j \neq k$, either

$$E_j \cup F_j \subset E_k \text{ or } F_k,$$

or the reverse containment with j, k interchanged. Define

$$h_j := h_{Q_D,j} := \frac{\left(\mathbf{1}_{E_j} - \mathbf{1}_{F_j}\right)}{|E_j \cup F_j|^{1/2}} = \frac{\left(\mathbf{1}_{E_j} - \mathbf{1}_{F_j}\right)}{2^{1/2}|E_j|^{1/2}}$$

When $D = 1$ just take $E_1 = [0, 1/2),\ F_1 = [1/2, 1)$; assuming inductively that

$$\left\{(E_j, F_j) : 1 \leq j \leq D'\right\}$$

have been chosen, we define

$$\left\{(E'_j, F'_j) : 1 \leq j \leq (D+1)'\right\} \subset [0,1]^{D+1} = Q_D \times [0,1]$$

by setting

$$E_j^1 := E_j \times [0, 1/2), \quad E_j^2 := E_j \times [1/2, 1)$$

and

$$F_j^1 := F_j \times [0, 1/2), \quad F_j^2 := E_j \times [1/2, 1),$$

and including

$$G_1 := Q_D \times [0, 1/2), \quad G_2 := Q_D \times [1/2, 1),$$

and collecting

$$\bigcup_{i=1}^{2} \left\{(E_j^i, F_j^i) : 1 \leq j \leq D'\right\} \cup (G_1, G_2).$$

Exercise 3.138. Prove that

$$\{h_{Q_D,j}\}$$

are an orthonormal basis for mean-zero functions on Q_D, and use scaling to prove that

$$a_Q(f) := \sum_{j \leq D'} \langle f, h_{Q,j} \rangle \cdot h_{Q,j}.$$

With this in mind, we can introduce the D-dimensional square function,

$$S_{\mathrm{D},D} f^2 := S_{\mathrm{D}} f^2 := \sum_k |\Delta_k f|^2 \tag{3.139}$$

$$= \sum_Q \sum_j |\langle f, h_{Q,j} \rangle|^2 \cdot \frac{\mathbf{1}_Q}{|Q|},$$

$$= \sum_Q \|a_Q(f)\|^2_{L^2(\mathbb{R}^D)} \cdot \frac{\mathbf{1}_Q}{|Q|},$$

and prove the D-dimensional analgoue of Theorem 3.6 by establishing analogues of the good-λ inequalities (3.58) and (3.59).

Exercise 3.140. Prove Theorem 3.6 in $\mathbb{R}^D$ and more generally

$$\|M_{\mathrm{D}} f\|_{L^r(\mathbb{R}^D)} \approx_{r,D} \|S_{\mathrm{D}} f\|_{L^r(\mathbb{R}^D)}$$

for all $0 < r < \infty$.
Hint: With C_D an appropriately large dimensional constant, prove that if

$$|\{S_{D,D}f > C_D\lambda, M_D f \leq \gamma\lambda\}| \lesssim_D \gamma^2 \cdot |\{S_{D,D}f > \lambda\}| \tag{3.141}$$

and

$$|\{M_D f > C_D\lambda, S_{D,D}f \leq \gamma\lambda\}| \lesssim_D \gamma^2 \cdot |\{M_D f > \lambda\}|. \tag{3.142}$$

For the first inequality, note that

$$\frac{|\langle f, h_{Q;i}\rangle|}{|Q|^{1/2}} \leq \inf_Q M_D f,$$

so that

$$\frac{\|a_Q(f)\|^2_{L^2(\mathbb{R}^D)}}{|Q|} \cdot \mathbf{1}_Q \lesssim_D \inf_Q M_D f^2.$$

With these preliminaries in mind, we present the D-dimensional version of Theorem 3.10.

Proof of Theorem 3.10, *D*-Dimensional Case; Sketch. The proof is similar to the 1-dimensional case.

Using the higher-dimensional version of Theorem 3.6, namely Exercise 3.140, it suffices to prove the desired inequality for finite linear combinations of Haar functions

$$f = \sum_Q \sum_j a_{Q;j} \cdot h_{Q;j}.$$

Pick a maximal interval on which f is supported, call it J. It is enough to prove the desired inequality for $\sum_{Q\subset J} a_Q(f)$.

Let

$$F_1 := \bigcup_I \Big\{ I \subset J \text{ maximal} : \Big| \sum_{I \subsetneq I' \subset J} a_{I'}(f)(x) \Big| > \lambda \Big\} \tag{3.143}$$

for some (any) $x \in I$. Let

$$T_1 := \{I \subset J : I \not\subset K \text{ for any } K \in F_1\},$$

Next, if $x \in I \in F_1$, then

$$\{I' \in T_1 : x \in I'\} = \{I' : I \subsetneq I' \subset J\}.$$

But now:

$$\int \lambda^2 \sum_{I \in F_1} \mathbf{1}_I < \int \Big| \sum_{I \in T_1} a_I(f) \Big|^2$$
$$= \sum_{I \in T_1} \|a_I(f)\|_{L^2(\mathbb{R}^D)}^2$$
$$= \sum_{I \in T_1} \sum_{j=1}^{D'} |\langle f, h_{I,j} \rangle|^2,$$

then iterate, considering the stopping regions for each $J_1 \in F_1$:

$$F_2(J_1) := \Big\{ I \subset J_1 \text{ maximal} : \Big| \sum_{I \subsetneq I' \subset J_1} a_Q(f)(x) \Big| > \lambda \Big\}$$

for some (any) $x \in I$, and the trees

$$T_1(J_1) := \{ I \subset J_1 : I \not\subset K \text{ for any } K \in F_2(J_1) \}.$$

□

Exercise 3.144. Complete the higher-dimensional proof of Theorem 3.10, by proving that for any collection of trees $\mathcal{T} = \{T_i\}$ the square functions

$$S_{\mathcal{T}} f^2 := \sum_{i=1}^{\infty} \sum_{I \in T_i} \|a_I(f)\|_{L^2(\mathbb{R}^D)}^2 \cdot \frac{\mathbf{1}_I}{|I|}$$

satisfy the good-λ inequality, uniformly in $\mathcal{T}$:

$$|\{S_{\mathcal{T}} f > C_D \lambda, M_{\mathrm{D}} f \leq \gamma \lambda\}| \lesssim \gamma^2 |\{S_{\mathcal{T}} f > \lambda\}|. \tag{3.145}$$

Hint: See Exercise 3.81.

With this machinery in hand, we turn to the proof proper of Lépingle's Inequality.

We begin with a preliminary lemma.

Lemma 3.22. *For $\lambda > 0$, there exists an absolute $C \gg 1$ so that we have the following "weak-type" good-λ inequality*

$$\lambda^2 \cdot |\{A : \mathcal{V}^r f > \lambda, M_D f \leq \lambda/2\}| \leq C \cdot \left(\frac{r}{r-2} \right)^2 \cdot \|f \cdot \mathbf{1}_A\|_{L^2(\mathbb{R}^D)}^2$$

whenever the following situation holds: $A \subset \mathbb{R}^D$ is a union of dyadic intervals of side length 2^{-t}, and $\mathbb{E}_{-n} f = 0$ for all $n \leq t$.

Proof. By homogeneity, it suffices to prove the result with $\lambda = 1$. Fix

$$2 < s = s(r) := \frac{r+2}{2} < r,$$

and pointwise dominate the variation as in [**21**, §3]

$$(\mathcal{V}^r f)^r \le \sum_{l\in\mathbb{Z}} 2^{rl} \cdot N_{2^l}(f);$$

since $M_{\mathrm{D}} f \le 1/2$, we may assume that the above sum runs over only $l \le 0$, which leads to the containment

$$\begin{aligned}\{A : \mathcal{V}^r f > 1, M_{\mathrm{D}} f \le 1/2\} &\subset \Big\{A : \sum_{l\le 0} 2^{rl} \cdot N_{2^l}(f) > 1/2\Big\} \\ &\subset \bigcup_{l\le 0} \Big\{2^{sl} \cdot N_{2^l}(f \cdot \mathbf{1}_A) > c_s\Big\},\end{aligned}$$

for an appropriate absolute constant $c_s \approx r-2$, with the corresponding inequality for measures.

In light of Lemma 3.10, this immediately leads to the majorization

$$\begin{aligned}|\{A : \mathcal{V}^r f > 1, M_{\mathrm{D}} f \le 1/2\}| &\le \sum_{l\le 0} 2^{(s-2)l} \cdot c_s^{-1} \cdot \|f \cdot \mathbf{1}_A\|_{L^2(\mathbb{R}^D)}^2 \\ &\lesssim \left(\frac{r}{r-2}\right)^2 \cdot \|f \cdot \mathbf{1}_A\|_{L^2(\mathbb{R}^D)}^2.\end{aligned}$$

□

Exercise 3.146. Verify that under the hypotheses of the theorem, we have the equality

$$N_\mu(f) \cdot \mathbf{1}_A = N_\mu(f \cdot \mathbf{1}_A)$$

for each $\mu > 0$.

Proof of Theorem 3.11. By homogeneity, it will suffice to prove our Theorem 3.11 with $\lambda = 1$.

We begin by collecting all maximal intervals, $\{I\}$, on which $\mathcal{V}^r f > C$, where C is a large absolute constant. We will show that for each selected I,

$$|\{I : \mathcal{V}^r f > 3C, M_{\mathrm{D}} f \le \gamma, S_{\mathrm{D}} f \le \gamma\}| \lesssim_r \gamma^2 \cdot |I|$$

which will yield the result. Freeze some such I, and let us assume for concreteness that I has side length 2^t.

The argument begins by establishing the following pointwise bound:

$$\mathcal{V}^r f(x) \lesssim \mathcal{V}^r(\mathbb{E}_n f : n < t) + M_{\mathrm{D}} f + \mathcal{V}^r(\mathbb{E}_n f : n > t). \tag{3.147}$$

Exercise 3.148. Verify the inequality (3.147).

The significance of (3.147) is that it allows us to "localize" our operators:

$$\{\mathcal{V}^r f > 3C\} \subset \{\mathcal{V}^r(\mathbb{E}_n f : n < t) > C\}.$$

Set now $g := f - f_I$ and observe that $M_{\mathrm{D}} f \leq \gamma$ implies $M_{\mathrm{D}} g \leq 2\gamma$, so that we have the containment

$$\begin{aligned} &\{I : \mathcal{V}^r f > 3C, M_{\mathrm{D}} f \leq \gamma, S_{\mathrm{D}} f \leq \gamma\} \\ &\quad \subset \{I : \mathcal{V}^r(\mathbb{E}_n f : n < t) > C, M_{\mathrm{D}} g \leq 2\gamma, S_{\mathrm{D}} g \leq \gamma\} \\ &\quad = \{I : \mathcal{V}^r g > C, M_{\mathrm{D}} g \leq 2\gamma, S_{\mathrm{D}} g \leq \gamma\}. \end{aligned}$$

We next collect all maximal subcubes of I, $\{Q\}$, subject to the constraint that

$$\sum_{n=l(Q)}^{t} |\Delta_n g|^2 > \gamma^2;$$

here, we define $l(Q)$ to be the integer so that the side length of Q is $2^{l(Q)}$.

Define the second auxiliary function

$$h := \sum_Q g_Q \cdot \mathbf{1}_Q + g \cdot \mathbf{1}_{(\bigcup Q)^c},$$

and notice that for any cube P,

$$h_P \neq g_P$$

implies that P is properly contained in some selected cube Q. The significance of this implication is contained in the following exercise.

Exercise 3.149. Prove that $\mathcal{V}^r g > \mathcal{V}^r h$ only inside $\bigcup Q \subset \{S_{\mathrm{D}} g > \gamma\}$.
Hint: If g is constant on Q, then

$$\sum_{n=l(Q)}^{t} |\Delta_n g|^2 = \sum_{n=l(Q)+1}^{t} |\Delta_n g|^2.$$

In light of the previous exercise, we may estimate

$$\begin{aligned} &|\{I : \mathcal{V}^r f > 3C, M_{\mathrm{D}} f \leq \gamma, S_{\mathrm{D}} f \leq \gamma\}| \\ &\quad \leq |\{I : \mathcal{V}^r g > C, M_{\mathrm{D}} g \leq 2\gamma, S_{\mathrm{D}} g \leq \gamma\}| \\ &\quad \leq |\{I : \mathcal{V}^r h > C, M_{\mathrm{D}} h \leq 2\gamma\}| \\ &\quad \lesssim \left(\frac{r}{r-2}\right)^2 \cdot \|h \cdot \mathbf{1}_I\|_{L^2(\mathbb{R})}^2 \\ &\quad = \left(\frac{r}{r-2}\right)^2 \cdot \|S_{\mathrm{D}}(h \cdot \mathbf{1}_I)\|_{L^2(\mathbb{R})}^2 \\ &\quad \leq \left(\frac{r}{r-2}\right)^2 \cdot \gamma^2 \cdot |I|, \end{aligned}$$

as desired. □

A similar inequality holds for the $\mathcal{V}^r(M_i f)$ operators. We present it in the Euclidean context, see (3.136). To do so, we introduce the *shifted* square functions.

Definition 3.150. For each $k \in \mathbb{Z}$, let $Q_k(x)$ denote the dyadic interval of length 2^k which contains x, and $3Q_k(x)$ to be the three-fold dilate. Set

$$\overline{S}_k f(x) := \sup_{y \in 3Q_k(x)} \mathcal{V}^2(M_i^{\mathbb{R}} f - \mathbb{E}_k f : 2^{k-1} \leq i < 2^k)(y) \cdot \mathbf{1}_{Q_k(x)}(x)$$

and define the *shifted square function*:

$$\overline{S} f^2 := \sum_k \overline{S}_k f^2$$

Exercise 3.151. Prove that $\mathcal{V}^r(M_t^{\mathbb{R}} f) \leq \overline{S} f + \mathcal{V}^r f$. By arguing spatially, prove that

$$\left\| \Big(\sum_k \overline{S}_k \Delta_{j+k} f^2 \Big)^{1/2} \right\|_{L^2(\mathbb{R})} \lesssim 2^{-c|j|} \cdot \|f\|_{L^2(\mathbb{R})} \tag{3.152}$$

for some $c > 0$, and conclude that $\overline{S} f$ is $L^2(\mathbb{R})$-bounded.
Hint: Follow the first proof that the Short Variation is $\ell^2(\mathbb{Z})$-Bounded.

Conveniently, the shifted square functions, and thus the r-variation operators, share a good-λ estimate with the Hardy-Littlewood maximal function; see [**109**] for a slightly stronger statement.

Lemma 3.23. *For each $\lambda > 0$, the following estimate holds:*

$$|\{\overline{S} f > 3\lambda, \ M_{HL} f \leq \gamma\lambda\}| \lesssim \gamma^2 \cdot |\{\overline{S} f > \lambda\}| \tag{3.153}$$

Proof. Let Q denote a maximal dyadic interval on which $\overline{S} f > \lambda$; we will prove that

$$|\{Q : \overline{S} f > 3\lambda, \ M_{\mathrm{HL}} f \leq \gamma\lambda\}| \lesssim \gamma^2 \cdot |Q|.$$

Without loss of generality, we may choose $z_Q \in Q$ so that $M_{\mathrm{HL}} f(z_Q) \leq \gamma\lambda$, otherwise there is nothing to show, and for concreteness, assume that Q has side length 2^{k_0}.

Since there exists some point in $p \in Q'$, the parent of Q, so that $\overline{S} f(p) \leq \lambda$,

$$\left(\sum_{j > k_0} \overline{S}_j f(x)^2 \right)^{1/2} \leq \lambda,$$

and thus

$$\{Q : \overline{S} f > 3\lambda, \ M_{\mathrm{HL}} f \leq \gamma\lambda\} \subset \{\overline{S}_0(f \cdot \mathbf{1}_{5Q}) > 2\lambda, \ M_{HL} f \leq \gamma\lambda\},$$

where

$$\overline{S}_0 f^2 := \sum_{j \le k_0} \overline{S}_j f^2.$$

By Exercise 3.151,

$$\begin{aligned}|\{\overline{S}_0(f \cdot \mathbf{1}_{5Q}) > 2\lambda,\ M_{\mathrm{HL}} f \le \gamma\lambda\}| &\lesssim \lambda^{-2} \cdot \|f \cdot \mathbf{1}_{5Q}\|_{L^2(\mathbb{R})} \\ &\lesssim \lambda^{-2} \cdot |Q| \inf_{y \in 5Q} M_{\mathrm{HL}} f(y)^2 \\ &\le \gamma^2 \cdot |Q|,\end{aligned}$$

as desired. □

Exercise 3.154. With

$$\overline{S}_k^D f(x) := \sup_{y \in 3Q_k(x)} \mathcal{V}^2(M_i^{\mathbb{R}^D} f - \mathbb{E}_k f : 2^{k-1} \le i < 2^k)(y) \cdot \mathbf{1}_{Q_k(x)}(x)$$

for

$$M_i^{\mathbb{R}^D} f := \frac{1}{i^D} \int_{[0,i]^D} f(x-t)\, dt,$$

establish the higher-dimensional version of (3.153).
Hint: The key inequality is the higher dimensional formulation of (3.152)*. To establish this, the key insight is that whenever* $2^k \le i < 2^{k+1}$*,*

$$M_i^{\mathbb{R}^D} \Delta_{k-n} f(x) = \sum_{|Q| = 2^{k-n}:\ (*)} M_i^{\mathbb{R}^D} a_Q(f)(x)$$

where the sum runs only over Q *that intersect the boundary of* $x - [0,i]^D$*:*

$$Q \cap \partial\big(x - [0,i]^D\big) \ne \emptyset;$$

since

$$\sup_x \Big| \bigcup_{|Q| = 2^{k-n}:\ (*)} Q \Big| \lesssim 2^{kD-n},$$

one can establish a higher-dimensional "improved Cauchy-Schwarz"-inequality, see (3.103)*.*