Title
Some Results in Pointwise Ergodic Theory

Author
Krause, Benjamin

Publication Date
2015

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Some Results in Pointwise Ergodic Theory

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Mathematics

by

Benjamin Krause

2015

Abstract of the Dissertation

Some Results in Pointwise Ergodic Theory

by

Benjamin Krause

Doctor of Philosophy in Mathematics

University of California, Los Angeles, 2015

Professor Terence Chi-Shen Tao, Chair

In this dissertation we establish various pointwise ergodic theorems, of both quanitative and qualitative nature. Beyond an introductory chapter, we devote one chapter to establishing quantitative ergodic theorems for $\mathbb{Z}^d$ actions; one chapter to quantitative ergodic theorems for polynomial orbits; one chapter to uniform estimates for various modulated ergodic averages; and the final chapter to random modulated averages.

The dissertation of Benjamin Krause is approved.

William Duke

Rowan Killip

Amit Sahai

Terence Chi-Shen Tao, Committee Chair

University of California, Los Angeles

2015

To my parents and my sister

Table of Contents

Acknowledgments

I have been very fortunate to have had the opportunity to collaborate with many wonderful mathematicians and friends. Two of my collaborations have made it into this dissertation, and for that I feel indebted to Professors Tanja Eisner (body chapter 3) and Pavel Zorin-Kranich (body chapter 4). I would like to make special thanks to Professors Alex Iosevich, Michael Lacey, and Christoph Thiele for their great mentorship and friendship. Finally, the greatest thanks are due to my advisor, Professor Terence Tao, for his guidance, encouragement, and support.

VITA

1988	Born
2010	B.A. (Mathematics), Amherst College.
2010–2015	Teaching Assistant and Research Assistant, Mathematics Department, UCLA.

CHAPTER 1

Introduction

My current research interests to date lie at the interface of harmonic analysis and ergodic theory. This discipline has at its roots the classical theorem of Birkhoff [6], 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, $\int f \, d\mu$, 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].

Several questions immediately present themselves:

- How rapid is the convergence?
- What are the higher-dimensional analogues for $\mathbb{Z}^d$ actions of the Birkhoff's theorem? What sorts of shapes can one average over and still derive pointwise convergence results?

[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.

- Back in the one dimensional setting, for which other sequences does one obtain similar pointwise convergence results? Can one "sample" f at "fewer" points and still obtain convergence of the associated Cesaro means?

- In the previous context, what happens if one relaxes the condition of integrability to, say, square integrability?

- How "strongly" does convergence occur? Does pointwise conergence still hold when one considers the weighted averages $\{\frac{1}{N}\sum_{n\leq N} c_n f(\tau^n x)\}$, for a sequence of "oscillatory" complex numbers $|c_n| = 1$?

The purpose of this thesis will be to address the above questions above *simultaneously.*

One of the fundamental questions in ergodic theory is to address a shortcoming of the pointwise ergodic theorem: it doesnt tell us how fast the "time averages" converge (to the space average, in the ergodic case). *Variational estimates* offer a way to quantify this rate of convergence, and variational estimates for ergodic averages form the major theme of the first two body chapters of the thesis. In particular, we will prove variational estimates for ergodic averages in the higher-dimensional setting, and along polynomial orbits as well.

The second two body chapters of the thesis will focus on weighted, or *Wiener-Wintner*, ergodic theorems. The dichotomy between structure and randomness underlies the work in these chapters, as we seek to determine how "moral summation by parts" plays out in considering our weighted sums. We will prove a variety of results concerning uniform convergence of various weighted average, as well as providing the first known result combining weighted and *random* ergodic averages.

Our body chapters are designed to be stand-alone, though we will motivate many of our techniques in the introductory sections to follow.

1.0.1 Notation and Preliminaries

We let $e(t) := e^{2\pi i t}$ denote the exponential.

We will make use of the modified Vinogradov notation. We use $X \lesssim Y$, or $Y \gtrsim X$ to denote the estimate $X \leq CY$ for an absolute constant C. If we need C to depend on a parameter, we shall indicate this by subscripts, thus for instance $X \lesssim_p Y$ denotes the estimate $X \leq C_p Y$ for some C_p depending on p. We use $X \approx Y$ as shorthand for $Y \lesssim X \lesssim Y$.

We also make use of big-O notation: we let $O(Y)$ denote a quantity that is $\lesssim Y$, and similarly $O_p(Y)$ a quantity that is $\lesssim_p Y$. We let $\widehat{O}(Y)$ denote a quantity which is $\leq Y$. We let $o(Y)$ denote a quantity that is $\ll Y$, much smaller than Y.

For functions, f, g, we let $o_{f \to \infty}(g)$ denote a quantity whose quotient with g tends to zero as f tends to ∞, and for real numbers a, we define $o_{f \to a}(g)$ analogously.

Throughout our discussion, (X, μ, τ) will be a *measure-preserving system*, a non-atomic probability space, equipped with τ a measure preserving $\mathbb{Z}^d$-action, $\tau_y f(x) := f(\tau_{-y} x)$. For $\{E_i\} \subset \mathbb{Z}^d$, define the averaging operators $M_i f(x) := \frac{1}{|E_i|} \sum_{y \in E_i} (\tau_y f)(x)$. A major question in ergodic theory is to what extent one can make sense of the limiting behavior of the means $\{M_i f\}$ as $i \to \infty$. Because we will be interested in the pointwise regime, we introduce the following slightly nonstandard notion.

Definition 1.0.1. *A collection of subsets $\{E_i\} \subset \mathbb{Z}^d$ 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 $\{M_i f\}$ converge* pointwise *μ-a.e.*

With the above definition in mind, we are prepared to introduce and discuss the the classical pointwise ergodic theorem of Birkhoff [6].

1.1 Birkhoff's Theorem

Theorem 1.1.1 (Pointwise Ergodic Theorem)**.** *The collection of intervals $\{[0, i)\}$ is universally L^1-good.*

We first present the standard density-argument approach to this theorem:

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

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

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

$$|M_i f| = |M_i(h - \tau h)| = \frac{1}{i}\left|h - \tau^i h\right| \leq \frac{2\|h\|_{L^\infty(X)}}{i} \to 0.$$

The next claim is that this set is dense in $L^2(X)$, and hence in $L^1(X)$ as well (on a probability space, the L^2 norm is strictly stronger than the L^1 norm). To see this, one observes that the L^2-closure of the linear span of (bounded) coboundaries are precisely the invariant functions. Indeed, if

$$L^2(X) \ni f \perp \mathrm{Span}\{h - \tau h : h \in L^\infty(X)\},$$

then for M a large parameter to be sent to ∞, we estimate

$$\begin{aligned}
\|f - \tau f\|_{L^2(X)}^2 &= 2\|f\|_{L^2(X)}^2 - 2\Re\langle f, \tau f\rangle \\
&= 2\|f\|_{L^2(X)}^2 - 2\Re\langle f, f - (f - \tau f)\rangle \\
&= 2\Re\langle f, f1_{|f|\leq M} - \tau(f1_{|f|\leq M})\rangle + 2\Re\langle f, f1_{|f|>M} - \tau(f1_{|f|>M})\rangle \\
&= 2\Re\langle f, f1_{|f|>M} - \tau(f1_{|f|>M})\rangle = o_{M\to\infty}(1)
\end{aligned}$$

by Cauchy-Schwartz and dominated convergence.

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 Stein's maximal principle [32], the only way to do this is to prove a weak-type one-one estimate on the maximal function,

$$\mathcal{M}_\tau f := \sup_i |M_i f|.$$

This is an abstract problem; fortunately Calderón's transference principle [2] allows one to transfer this problem to one on the integer lattice.

Proposition 1.1.2. *For each $f \in L^1(X)$, we have $\|\mathcal{M}_\tau f\|_{L^{1,\infty}(X)} \leq \|f\|_{L^1(X)}$.*

Assuming this proposition, let us see how we may conclude the argument.

For $f \in L^1(X)$, and $1 \gg \epsilon > 0$ arbitrary but fixed, consider the oscillation set,

$$\Omega = \Omega_{f,\epsilon} := \left\{ \limsup_{i,j \to \infty} |M_i f - M_j f| > \epsilon \right\}.$$

Now, select g inside the dense subclass of functions for which $\{M_i g\}$ converges in such a way that $\|g - f\|_{L^1(X)} \ll \epsilon^2$ (say). By the pointwise convergence of $\{M_i g\}$ and the triangle inequality, we have the containment up to μ-null sets:

$$\Omega \subset \{\mathcal{M}_\tau (f - g) > \epsilon/2\};$$

The measure of this latter set is bounded by a constant multiple of

$$\frac{1}{\epsilon} \|f - g\|_{L^1(X)} \ll \epsilon,$$

by our approximation, which completes the proof, assuming the above proposition. □

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

$$\mathcal{M}_\tau f := \sup_{i \leq N} |M_i f|.$$

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 A_i denote the forward Lebesgue averaging operators,

$$A_i g(m) := \frac{1}{i} \sum_{n=0}^{i-1} g(m+n),$$

then for $n \leq R - N$, a moment's thought convinces one that

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

for all $i \leq N$. Let $\mathcal{M}_+ g := \sup_i |A_i g|$ denote the one-sided discrete Hardy-Littlewood maximal function, and let us see how the following lemma allows us to conclude the proof.

Lemma 1.1.3. *Let $\lambda > 0$ be an arbitrary altitude. Then for any $f \in L^1(\mathbb{Z})$,*

$$\lambda |E_\lambda| := \lambda |\{\mathcal{M}_+ f > \lambda\}| < \int_{E_\lambda} |f|.$$

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

$$\|\mathcal{M}_+ f\|_{L^{1,\infty}} \leq \|f\|_{L^1(\mathbb{Z})}.$$

Since we have defined

$$\mathcal{M}_\tau f(x) := \sup_{i \leq N} |M_i f|,$$

we have the pointwise inequality,

$$\mathcal{M}_\tau f(\tau^n x) \leq \mathcal{M}_+ F(x;n)$$

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 $\mathcal{M}_+$, we may estimate

$$\begin{aligned}(R-N)\mu(\{\mathcal{M}_\tau f > \lambda\}) &= \int_X \int_{[0,R-N)} 1_{\{(x,n):\mathcal{M}_\tau f(\tau^n x)>\lambda\}} \, d\#(n) d\mu(x) \\ &\leq \int_X \int_{[0,R-N)} 1_{\{(x,n):\mathcal{M}_+ F(x;n)>\lambda\}} \, d\#(n) d\mu(x) \\ &\leq \int_X |\{n : \mathcal{M}_+ F(x;n) > \lambda\}| d\mu(x) \\ &\leq \frac{1}{\lambda} \int_X \|F(x;n)\|_{L^1(\mathbb{Z})} d\mu(x).\end{aligned}$$

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

$$\begin{aligned}\frac{1}{\lambda} \int_X \|F(x;n)\|_{L^1(\mathbb{Z})} d\mu(x) &= \frac{1}{\lambda} \int_{\mathbb{Z}} \int_X |F(x;n)| \, d\mu(x) d\#(n) \\ &= R \frac{1}{\lambda} \|f\|_{L^1(X)}.\end{aligned}$$

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

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

$\square$

It remains only to prove Lemma 1.1.3, reproduced below for the reader's convenience:

Lemma 1.1.4. *Let $\lambda > 0$ be an arbitrary altitude. Then for any $f \in L^1(\mathbb{Z})$,*

$$\lambda |E_\lambda| := \lambda |\{\mathcal{M}_+ f > \lambda\}| < \int_{E_\lambda} |f|.$$

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

$$\|\mathcal{M}_+ f\|_{L^{1,\infty}} \leq \|f\|_{L^1(\mathbb{Z})}.$$

Proof. First, there is no loss of generality in assuming $f \geq 0$ is non-negative. For concreteness, let us suppose that $R = R(\lambda, f)$ is so large that

$$\|f \cdot 1_{\mathbb{Z} \smallsetminus [-R,R]}\|_{L^1(\mathbb{Z})} \leq \lambda/2,$$

which can be justified by (say) monotone convergence. With this in mind, we decompose $E_\lambda = \bigcup_j [a_j, b_j)$ as a union of intervals moving right. We quickly verify that there are at most finitely many intervals, and that all intervals are of finite length. To do so, we split

$$A_t f(x) \leq A_t(f \cdot 1_{[-R,R]})(x) + A_t(f \cdot 1_{\mathbb{Z} \smallsetminus [-R,R]})(x) \leq A_t(f \cdot 1_{[-R,R]})(x) + \lambda/2,$$

and observe that

$$A_t(f \cdot 1_{\mathbb{Z} \smallsetminus [-R,R]})(x) = 0$$

for $x > R$, while

$$A_t f(x) \leq \lambda/2$$

for $x < -R - \frac{2\|f\|_{L^1(\mathbb{Z})}}{\lambda}$: for x in the above range,

$$A_t(f \cdot 1_{[-R,R]})(x) \leq \frac{1}{t}\|f\|_{L^1(\mathbb{Z})} \cdot 1_{t \geq -R-x},$$

and if

$$x < -R - \frac{2\|f\|_{L^1(\mathbb{Z})}}{\lambda}$$

we necessarily have any such admissible $t > \frac{2\|f\|_{L^1(\mathbb{Z})}}{\lambda}$.

Now, suppose we knew that for each j,

$$\lambda |[a_j, b_j)| < \int_{[a_j,b_j)} f;$$

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

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

$$\int_{[a_k,b_k)} f \leq \lambda|[a_k,b_k)|.$$

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

$$l(x) := \max\left\{t : \int_{[x,t)} f > \lambda|[x,t)|\right\}.$$

We claim that $l(a) \leq b-1$. By assumption, we cannot have $l(a) = b$, and if $l(a) > b$, using the fact that $b \notin E_\lambda$, we would have

$$\lambda|[a,l(a))| < \int_{[a,l(a))} f = \int_{[a,b)} f + \int_{[b,l(a))} f \leq \lambda|[a,b)| + \lambda|[b,l(a))| = \lambda|[a,l(a))|,$$

for the desired contradiction. The upshot is that $l(a) < l(l(a))$ is non-trivial. But then

$$\lambda\left|\left[a,l(a)\right)\right| + \lambda\left|\left[l(a),l(l(a))\right)\right| < \int_{[a,l(a))} f + \int_{[l(a),l(l(a)))} f = \int_{[a,l(l(a)))} f \leq \lambda\left|\left[a,l(l(a))\right)\right|$$

by definition of our look-ahead function. This contradiction completes the proof. $\square$

Remark 1.1.5. *The proof of the above proposition is extendable to proving strong-type L^p bounds on $\mathcal{M}_\tau$ as well. Moreover, the argument relied on two – and only two – properties of the operator $\mathcal{M}_\tau$. Firstly, that $\mathcal{M}_\tau$ is sublinear; secondly, that $\mathcal{M}_\tau$ is* transferrable *to an operator on the integer lattice, $\mathcal{M}_+$, 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.*

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

Obtaining pointwise convergence results for $\{M_i f\}$ for rougher, exotic $\{E_i\} \subset \mathbb{Z}$ 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.

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

Indeed, to prove pointwise convergence of the ergodic averages of L^2-functions along the squares, Bourgain [10] attacked the issue of oscillation more directly, by showing that an appropriate *oscillation operator* was L^2 "controlled" ([10, §7].) In a redux of his argument [11], 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.

1.2 Introduction to Variation

Definition 1.2.1. *For a collection of functions $\{f_i\}$,*

$$\mathcal{V}^r(f_i)(x) := \sup \left(\sum_k |f_{i_k} - f_{i_{k+1}}|^r \right)^{1/r} (x)$$

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

These variation operators are more difficult to control than the maximal function $\sup_i |f_i|$: for any j, one may pointwise dominate

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

where $r < \infty$ is arbitrary. This difficulty is reflected in the fact that although having bounded r-variation, $r < \infty$, is enough to imply pointwise convergence, there are functions which converge, but which have unbounded r-variation for *any* $r < \infty$. (e.g. $\{(-1)^i \frac{1}{\log(i+1)}\}$)

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 [11].

1.2.1 Lépingle's Inequality

We pause here to introduce Lépingle's inequality in the special Euclidean/dyadic setting, which will be adequate for our purposes. We recall the dyadic filtration:

$$\mathcal{E}_n f(x) := \sum_Q f_Q 1_Q,$$

where the sum runs over all dyadic cubes of side-length 2^{-n}, and $f_Q := \frac{1}{|Q|}\int_Q f \ dx$. We let $\Delta_n := \mathcal{E}_{n+1} - \mathcal{E}_n$ denote the differencing operators.

We recall the (classical) the maximal function and square function, defined respectively as

$$\mathcal{M}f := \sup_n |\mathcal{E}_n f|, \quad \mathcal{S}f := \big(\sum_n |\Delta_n f|^2\big)^{1/2}.$$

These two operators, both of which roughly measure the "scales" where f oscillates, have long been thought of as being of "even strength;" this heuristic was beautifully quantified by Burkholder and Davis, who proved the following [8, Theorem 9], [11, Theorem 1]

Theorem 1.2.2 (Burkholder, $p > 1$, Davis, $p = 1$). *For each $p \geq 1$, there exist absolute constants $0 < c_p < C_p$ so that*

$$c_p \|\mathcal{S}f\|_p \leq \|\mathcal{M}f\|_p \leq C_p \|\mathcal{S}f\|_p.$$

With $\mathcal{V}^r f := \mathcal{V}^r(\mathcal{E}_n f)$, we have the following fundamental boundedness result concerning the r-variation operators on Euclidean space.

Theorem 1.2.3 (Lépingle's Inequality – Special Case). *For each $r > 2$, and $1 < p < \infty$, there exist absolute constants A_p so that*

$$\|\mathcal{V}^r f\|_{L^p(\mathbb{R}^d)} \leq A_p \|f\|_{L^p(\mathbb{R}^d)};$$

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

$$\|\mathcal{V}^r f\|_{L^{1,\infty}(\mathbb{R}^d)} \leq A_1 \|f\|_{L^1(\mathbb{R}^d)}$$

for some absolute constant A_1.

We remark that the range of $r > 2$ in the above theorem is sharp, since these estimates can fail for $r \leq 2$; see e.g. [13], [31].

By now, comparatively simple proofs of Lépingle's theorem can be found in Pisier and Xu [30] and Bourgain [11]. The idea was to leverage known estimates for jump inequalities for the $\{\mathcal{E}_n f\}$ to recover variational estimates:

Definition 1.2.4. *For each $\lambda > 0$, $N_\lambda(\mathcal{E}_n f)$ is the number of λ-jumps in $\{\mathcal{E}_n f\}$: the supremum over all $N \geq 1$ such that there exists an increasing sequence of indices $\{n_0 < n_1 < n_2 < \cdots < n_N\}$ with*

$$|\mathcal{E}_{n_i} f - \mathcal{E}_{n_{i-1}} f| > \lambda$$

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

Remark 1.2.5. *$N_\lambda(\mathcal{E}_n f)$ is comparable to λ-entropy of the $\{\mathcal{E}_n f\}$, i.e. the fewest number of λ-balls required to cover the collection $\{\mathcal{E}_n f\}$.*

The key result concerning λ-jumps is the following:

Theorem 1.2.6 ([9], [12]). *If N_λ is the number of λ-jumps in $\{\mathcal{E}_n f\}$ then for any $1 < p < \infty$*

there exist absolute constants B_p such that

$$\|\lambda N_\lambda^{1/2}\|_p \le B_p \|f\|_p \quad \textit{for all } \lambda > 0;$$

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

$$\|\lambda N_\lambda^{1/2}\|_{1,\infty} \le B_1 \|f\|_1 \quad \textit{for all } \lambda > 0$$

for some absolute constant B_1.

We refer the reader to [12] for a more comprehensive survey.

The significance of the argument provided below – inspired by the excellent exposition of [21, §2] – is that it sheds new insight into the relationship between maximal function, square function, and variation operator. Indeed, in light of Theorem 1.2.2, we recover Lépingle's result that the r-variation operators ($r > 2$) are themselves of "even strength" with the maximal functions.

Specifically, we prove the following

Theorem 1.2.7. *For $r > 2$, and all $\gamma > 0$ sufficiently small, we have the good-λ inequality*

$$\mu(\{\mathcal{V}^r f > C\lambda, \mathcal{M}f < \gamma\lambda, \mathcal{S}f < \gamma\lambda\}) \le A\gamma^2 \mu(\{\mathcal{V}^r f > \lambda\})$$

for appropriate (absolute) constants $A = A(r), C \gg 1$. Moreover, as $r \downarrow 2$, $A \approx (r-2)^{-2}$.

In particular, by integrating distribution functions we recover the optimal bound

$$\|\mathcal{V}^r f\|_2 \lesssim \frac{1}{r-2}\|f\|_2$$

as $r \downarrow 2$.

Remark 1.2.8. *The proof we provide translates readily to the integer lattice.*

This immediately proves that $\mathcal{V}^r f$ is bounded on L^p, $1 < p < \infty$ and moreover is integrable when the maximal function and the square function are (see above):

Corollary 1.2.9. *If $\mathcal{M}f \in L^1$ (and thus $\mathcal{S}f \in L^1$), $\mathcal{V}^r f \in L^1$ for each $r > 2$.*

1.2.2 The Proof of Lépingle's Inequality

We begin with a preliminary lemma.

Lemma 1.2.10. *For $\lambda > 0$, there exist absolute constants*

$$0 < c < 1 \ll C = C(r) \approx (r-2)^{-2}$$

(as $r \downarrow 2$) so that we have the following "weak-type" good-λ inequality

$$\lambda^2 |\{A : \mathcal{V}^r f > \lambda, \mathcal{M}f \leq c\lambda\}| \leq C \|f 1_A\|_{L^2(\mathbb{R}^d)}^2.$$

for any measurable $A \subset \mathbb{R}^d$ satisfying the localization criterion

$$N_\mu(f) 1_A = N_\mu(f 1_A)$$

for each μ. (The localization criterion is satisfied, for instance, if $A \in \mathcal{E}_t$, and $\mathcal{E}_n f = 0$ 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 [11, §3]

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

since $\mathcal{M}f < c < 1$, we may assume that the above sum runs over only $l \leq 0$, which leads to the containment

$$\{A : \mathcal{V}^r f > 1, \mathcal{M}f < c\} \subset \left\{A : \sum_{l\leq 0} 2^{rl} N_{2^l}(f) > 1/2\right\} \subset \bigcup_{l \leq 0} \left\{2^{sl} N_{2^l}(f 1_A) > c_s\right\},$$

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

In light of Lemma 1.2.6, this immediately leads to the majorization

$$\mu\left(\{A : \mathcal{V}^r f > 1, \mathcal{M}f < c\}\right) \leq \sum_{l \leq 0} 2^{(s-2)l} c_s^{-1} \|f 1_A\|_2^2 \lesssim (r-2)^{-2} \|f 1_A\|_2^2.$$

$\square$

Proof of Theorem 1.2.7. By homogeneity, it will suffice to prove our Theorem 1.2.7 with $\lambda = 1$.

We begin by collecting all maximal cubes, $\{I\}$, on which $\mathcal{V}^r f > C$, where C is a large absolute constant. We will show that for each selected I,

$$|\{I : \mathcal{V}^r f > 3C, \mathcal{M}f \leq \gamma, \mathcal{S}f \leq \gamma\}| \lesssim_r \gamma^2 |I| \tag{1.1}$$

which will yield the result. Freeze some such I, and let us assume for concreteness that I has side length 2^{-t}.

By the trivial observation that

$$\mathcal{V}^r f(x) \lesssim \mathcal{V}^r(\mathcal{E}_n f : n < i) + \mathcal{M}f + \mathcal{V}^r(\mathcal{E}_n f : n > i)$$

pointwise, we have

$$\{\mathcal{V}^r f > 3C\} \subset \{\mathcal{V}^r(\mathcal{E}_n f : n > t) > C\}.$$

Set now $g := f - f_I$ and observe that $\mathcal{M}f \leq \gamma$ implies $\mathcal{M}g \leq 2\gamma$, so that we have the containment

$$\{I : \mathcal{V}^r f > 3C, \mathcal{M}f \le \gamma, \mathcal{S}f \le \gamma\} \subset \{I : \mathcal{V}^r(\mathcal{E}_n g : n > t) > C, \mathcal{M}g \le 2\gamma, \mathcal{S}g \le \gamma\}$$
$$= \{I : \mathcal{V}^r g > C, \mathcal{M}g \le 2\gamma, \mathcal{S}g \le \gamma\}.$$

We next collect all maximal subcubes of I, $\{Q\}$, subject to the constraint that

$$\sum_{n=t}^{l(Q)} |\Delta_n g|^2 > \gamma^2;$$

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

$$h := \sum_Q g_Q 1_Q + g \cdot 1_{(\bigcup Q)^c},$$

and notice that for any cube P,

$$h_P \ne g_P$$

implies that P is properly contained in some selected cube Q. Consequently, $\mathcal{V}^r g > \mathcal{V}^r h$ inside $\bigcup Q \subset \{\mathcal{S}g > \gamma\}$.

Putting everything together, we may estimate

$$\begin{aligned}
|\{I : \mathcal{V}^r f > 3C, \mathcal{M}f \le \gamma, \mathcal{S}f \le \gamma\}| &\le |\{I : \mathcal{V}^r g > C, \mathcal{M}g \le 2\gamma, \mathcal{S}g \le \gamma\}| \\
&\le |\{I : \mathcal{V}^r h > C, \mathcal{M}h \le 2\gamma\}| \\
&\lesssim_r \|h 1_I\|_2^2 \\
&= \|\mathcal{S}(h 1_I)\|_2^2 \\
&\le \gamma^2 |I|,
\end{aligned}$$

as desired. We used Lemma 1.2.10 in the third line, so that the implicit constant grows like $(r-2)^{-2}$ as $r \downarrow 2$. $\square$

1.2.3 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. For the sake of simplicity, we will content ourselves with developing an L^2-theory; we will prove pointwise convergence of ergodic averages on $L^2(X)$, i.e. only that $\{[0,i)\}$ are L^2-good.

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

$$\|\mathcal{V}^r(M_i f)\|_{L^2(X)} \lesssim_r \|f\|_{L^2(X)};$$

this result first appeared in Bourgain's celebrated paper on polynomial ergodic averages, [11]. Since the variation operators are semi-local, we will derive this result from the corresponding one on the integer lattice:

Proposition 1.2.11. *For each $r > 2$*

$$\|\mathcal{V}^r(A_i f)\|_{L^2(\mathbb{Z})} \lesssim_r \|f\|_{L^2(\mathbb{Z})}.$$

The argument we provide here follows that of [10] – a simplification of [11] – and [8].

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:

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

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

Indeed, we have the following easy pointwise inequality:

Lemma 1.2.12 ([13], §3).

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

Sketch. Fix an increasing sequence of indices $\{i_k\}$. For each pair (i_k, i_{k+1}) so that there exists $n(k) < n(k+1)$ with

$$2^{n(k)-1} \leq i_k \leq 2^{n(k)} < 2^{n(k+1)} \leq i_{k+1} \leq 2^{n(k+1)+1},$$

replace

$$|f_{i_k} - f_{i_{k+1}}| \leq |f_{i_k} - f_{2^{n(k)}}| + |f_{2^{n(k)}} - f_{2^{n(k+1)}}| + |f_{i_{k+1}} - f_{2^{n(k+1)}}|,$$

so that

$$|f_{i_k} - f_{i_{k+1}}|^r \lesssim_r |f_{i_k} - f_{2^{n(k)}}|^r + |f_{2^{n(k)}} - f_{2^{n(k+1)}}|^r + |f_{i_{k+1}} - f_{2^{n(k+1)}}|^r.$$

Take l^r-norms in k, and make the above replacements when necessary; the first and third replaced terms become absorbed by the short variation, while each middle term becomes absorbed by the long variation. □

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

The Long Variation is $L^2(\mathbb{Z})$-Bounded. We begin with the long variation. Using Lépingle's inequality and the triangle inequality, it is enough to show that

$$\mathcal{V}^r(A_{2^n} f - \mathcal{E}_n f) \lesssim \left(\sum_{n \geq 0} |A_{2^n} f - \mathcal{E}_n f|^2 \right)^{1/2} =: \mathcal{S}_c f$$

is L^2-bounded (the subscript c stands for comparison); here we define

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

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

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

Indeed, expanding $f = \sum_{n \geq 0} \Delta_n f$ into increments, we majorize

$$\mathcal{S}_c f = \left(\sum_{n \geq 0} \left| (A_{2^n} - \mathcal{E}_n) \left(\sum_m \Delta_m f \right) \right|^2 \right)^{1/2} \leq \sum_t \left(\sum_{n \geq 0} |(A_{2^n} - \mathcal{E}_n)\Delta_{n+t} f|^2 \right)^{1/2}.$$

Estimating the tth term in L^2:

$$\begin{aligned}
\left\| \left(\sum_{n \geq 0} |(A_{2^n} - \mathcal{E}_n)\Delta_{n+t} f|^2 \right)^{1/2} \right\|_{L^2(\mathbb{Z})}^2 &= \sum_{n \geq 0} \|(A_{2^n} - \mathcal{E}_n)\Delta_{n+t} f\|_{L^2(\mathbb{Z})}^2 \\
&\lesssim 2^{-|t|} \sum_{n \geq 0} \|\Delta_{n+t} f\|_{L^2(\mathbb{Z})}^2 \\
&= 2^{-|t|} \left\| \left(\sum_{n \geq 0} |\Delta_{n+t} f|^2 \right)^{1/2} \right\|_{L^2(\mathbb{Z})}^2 \\
&\leq 2^{-|t|} \|f\|_{L^2(\mathbb{Z})}^2,
\end{aligned}$$

where we used the fact that the Δ_js are projections in the second line.

To prove our key estimate

$$\|(A_{2^i} - \mathcal{E}_i)\Delta_j\|_{L^2(\mathbb{Z}) \to L^2(\mathbb{Z})} \lesssim 2^{-\frac{|i-j|}{2}},$$

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

$$(A_{2^i} - \mathcal{E}_i)\Delta_j$$

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

Now, with $f \in L^2(\mathbb{Z})$ 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 $i \ll j$,

$$|(A_{2^i} - \mathcal{E}_i) h_J| = |A_i h_J - h_J| \leq \frac{1}{|J|^{1/2}} 1_{E_i(J)}$$

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,

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

and thus

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

On the other hand, when $i \gg j$,

$$(A_{2^i} - \mathcal{E}_i)\Delta_j \equiv A_{2^i} \Delta_j.$$

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

$$(A_{2^i} - \mathcal{E}_i)\Delta_j f(x_0) = A_{2^i} \Delta_j f(x_0) = \langle f, h_{J_0} \rangle A_{2^i}(h_{J_0})(x_0) + \langle f, h_{J_1} \rangle A_{2^i}(h_{J_1})(x_0).$$

Since

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

we may majorize

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

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

$$\|(A_{2^i} - \mathcal{E}_i)\Delta_j f\|_{L^2(\mathbb{Z})}^2 \lesssim 2^{2j-2i} \|\Delta_j f\|_{L^2(\mathbb{Z})}^2,$$

completing the proof of the claim. □

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 multipliers,

$$\widehat{A_t f}(\beta) := m_t(\beta) \widehat{f}(\beta) := \left(\frac{1}{t} \sum_{n \leq t} e(-n\beta) \right) \widehat{f}(\beta).$$

The Short Variation is $L^2(\mathbb{Z})$-Bounded, Take One. We will abbreviate

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

where

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

By arguing as previously, it will be enough to establish the estimate

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

We first notice that

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

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

The key observation here is that

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

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

$$\Delta_m f = \sum_Q a_Q 1_Q$$

into dyadic cubes Q which have side-lengths 2^m. Suppose $x \in P$, and is of distance greater than 2^{n+1} away from its boundary, so that

$$A_i \Delta_m f(x) = \sum_Q a_Q A_i 1_Q(x) = a_P A_i 1_P(x) \equiv a_P.$$

Since the short variation is formed by summing over *differences* of operators A_i, it follows that

$$\mathcal{V}_n^2(A_i(\Delta_m f))(x) = 0.$$

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

$$\mathcal{V}_n^2(A_i(\Delta_m f)) \lesssim \sum_Q \left(\sum_P |a_P| A_{2^{n+1}} 1_P \right) 1_{E_n(Q)} \leq \sum_Q (\sum_{i=0}^1 |a_{Q_i}|) 1_{E_n(Q)},$$

where we set $Q_0 := Q$ and Q_1 to be the immediate right dyadic neighbor of Q. Integrating,

$$\|\mathcal{V}_n^2(A_i(\Delta_m f))\|_{L^2(\mathbb{Z})}^2 \lesssim \sum_Q |a_Q|^2 |E_n(Q)| \lesssim 2^{n-m} \sum_Q |a_Q|^2 |Q| = 2^{n-m} \|\Delta_{n-m} f\|_{L^2(\mathbb{Z})}^2.$$

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

$$\mathcal{V}_n^2(A_i(\Delta_m f))^2 \lesssim 2^{m-n} A_{2^{n+1}}(|\Delta_m f|^2),$$

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

$$\sum_s |A_{i_s} - A_{i_{s+1}}(\Delta_m f)|^2(x) =$$
$$\sum_s \left|\frac{1}{i_s} - \frac{1}{i_{s+1}}\right|^2 \cdot \left|\int_{[0,i_s)} (\Delta_m f)(x+y)\ d\#(y)\right|^2 + \sum_s \left|\frac{1}{i_{s+1}}\right|^2 \cdot \left|\int_{[i_s,i_{s+1})} (\Delta_m f)(x+y)\ d\#(y)\right|^2.$$

Expand $\Delta_m f = \sum_Q \langle f, h_Q\rangle h_Q$ into Haar coefficients, and let $Q_0 = Q_0(x,s)$ and $Q_1 = Q_1(x,s)$ be appropriately chosen; one discovers that

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

This quantity is in turn majorized by

$$2^m \cdot \int_{[0,i_s)} |(\Delta_m f)(x+y)|^2 \leq 2^m \cdot \int_{[0,2^{n+1})} |\Delta_m f|(x+y)^2.$$

Inserting this upper bound into the sum

$$\sum_s \left|\frac{1}{i_s} - \frac{1}{i_{s+1}}\right|^2 \cdot \left|\int_{[0,i_s)} (\Delta_m f)(x+y)\ d\#(y)\right|^2$$

yields an upper estimate of

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

The contribution from the second sum,

$$\sum_s \left|\frac{1}{i_{s+1}}\right|^2 \cdot \left|\int_{[i_s,i_{s+1})} (\Delta_m f)(x+y)\ d\#(y)\right|^2,$$

is handled similarly; the key point is the "reverse-Hölder"-type estimate,

$$\left| \int_{[i_s, i_{s+1})} (\Delta_m f)(x+y) \ d\#(y) \right|^2 \lesssim 2^m \int_{[i_s, i_{s+1})} |\Delta_m f|^2(x+y) \ d\#(y),$$

which is established by expansion into Haar coefficients. The proof is complete. □

We now turn to a second proof that the short variation is $L^2(\mathbb{Z})$-bounded. Although the argument we provide is more involved, it relies on more general Fourier-analytic properties of the multipliers,

$$m_t(\beta) := \frac{1}{t} \sum_{n \leq t} e(-n\beta),$$

rather than concrete geometric considerations; indeed, the spirit of this argument will be present when treating short variation along polynomial orbits (below).

Recall that we are interested in showing that

$$\sum_{n \geq 0} \|\mathcal{V}^2_n(A_t f)\|^2_{L^2(\mathbb{Z})} \lesssim \|f\|^2_{L^2(\mathbb{Z})};$$

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

$$\sum_{n \geq 0} \|\mathcal{V}^2_n(C_t f)\|^2_{L^2(\mathbb{Z})} \lesssim \|f\|^2_{L^2(\mathbb{Z})},$$

where we define

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

We have seen that for each n, the operator A_{2^n} "annihilates" functions which oscillate at scales $\ll 2^n$, and "preserves" functions which oscillate only at scales $\gg 2^n$. The take-away from the above spatial argument was that there is great – and quantifiable – similarity between the operators A_{2^n} and A_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

$$v_t(\beta) := m_t(\beta) - m_{2^n}(\beta),$$

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

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

Lemma 1.2.13. *Suppose $2^n \le t < 2^{n+1}$. 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, A_t and A_{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 $\gtrsim 2^n$.

Proof. For the first point, one observes,

$$|v_t(\beta) - v_{t+1}(\beta)| \le \left(\frac{1}{t} - \frac{1}{t+1}\right)\left|\sum_{n=0}^{t-1} e(-n\beta)\right| + \frac{1}{t+1}|e(-t\beta)| \lesssim \frac{1}{t+1} \lesssim 2^{-n}.$$

For the second, a geometric sequence argument allows one to express

$$|m_t(\beta)| = \frac{1}{t}\left|\frac{1-e(-\beta t)}{1-e(-\beta)}\right| \lesssim \frac{1}{t|\beta|} \lesssim \frac{1}{2^n|\beta|},$$

from which the estimate on v_t follows from the triangle inequality.

Finally,

$$m_t(\beta) - 1 = \frac{1}{t}\sum_{n=0}^{t-1}(e(-n\beta) - 1) = O(t|\beta|) = O(2^n|\beta|),$$

and subtracting appropriately,

$$v_t(\beta) = (m_t(\beta) - 1) - (m_{2^n}(\beta) - 1),$$

yields the result. $\square$

The above lemma strongly motivates us to introduce a (rough) Littlewood-Paley decomposition of the torus $\mathbb{T} \cong [-1/2, 1/2)$ to study the behavior of the $\{v_t\}$. Since each multiplier $v_t(0) = 0$, it is in fact enough to decompose $\mathbb{T} \smallsetminus \{0\}$. We do so using the dyadic partitioning functions:

$$\chi_j := \begin{cases} 1_{[-2^{-j}, -2^{-j-1}) \cup [2^{-j-1}, 2^{-j})} & \text{if } j \geq 1 \\ 0 & \text{if } j \leq 0, \end{cases}$$

so that

$$\sum_j \chi_j = 1_{\mathbb{T} \smallsetminus \{0\}}.$$

Next, for $f \in L^2(\mathbb{Z})$, we let

$$f_j(n) := (\widehat{f}\chi_j)^\vee(n)$$

denote the inverse Fourier transform of the projection to frequency scales $\approx 2^{-j}$.

The idea will be to gain strong $L^2(\mathbb{Z})$-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 $2^n \leq t < 2^{n+1}$,

$$\|v_t \chi_{n+m}\|_{L^\infty(\mathbb{T})} \lesssim 2^{-|m|}.$$

To execute, we will use the following elementary lemma. This result is essentially due to Bourgain (cf. [11, Lemma 3.11], and [10, Proposition 2.12] as well), and we will make repeated use of it – here and below.

Lemma 1.2.14. *Suppose that $\{B_n\}_{n=1}^N$ are a family of operators which act on $L^2(\mathbb{Z})$ functions by multiplication on the Fourier side*

$$\widehat{B_n f}(\beta) := m_n(\beta)\widehat{f}(\beta).$$

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

$$\sup_{\beta \in \operatorname{supp} \widehat{f}} |m_n(\beta)| \leq A \textit{ and}$$

$$\sup_{\beta \in \operatorname{supp} \widehat{f}} |m_n(\beta) - m_{n+1}(\beta)| \leq a.$$

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

Proof. We may assume without loss of generality that $Na > A$, since otherwise we may use Cauchy-Schwartz to majorize

$$\mathcal{V}^2(B_n f) \leq \sum_{n \leq N} |B_n f - B_{n+1} f| \leq N^{1/2} \left(\sum_{n \leq N} |B_n f - B_{n+1} f|^2 \right)^{1/2};$$

taking $L^2(\mathbb{Z})$-norms yields the estimate

$$\begin{aligned}
\left\| \left(\sum_{n \leq N} |B_n f - B_{n+1} f|^2 \right)^{1/2} \right\|_{L^2(\mathbb{Z})} &= \left(\sum_{n \leq N} \|B_n f - B_{n+1} f\|_{L^2(\mathbb{Z})}^2 \right)^{1/2} \\
&= \left(\sum_{n \leq N} \|(m_n - m_{n+1})\widehat{f}\|_{L^2(\mathbb{T})}^2 \right)^{1/2} \\
&\leq \sqrt{N} a \|\widehat{f}\|_{L^2(\mathbb{T})} \\
&= \sqrt{N} a \|f\|_{L^2(\mathbb{Z})}
\end{aligned}$$

by repeated applications of Plancherel. Consequently, we have the desired estimate

$$\|\mathcal{V}^2(B_n f)\|_{L^2(\mathbb{Z})} \leq Na\|f\|_{L^2(\mathbb{Z})} \leq \sqrt{NAa}\|f\|_{L^2(\mathbb{Z})}.$$

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,

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

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

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

where we used Cauchy-Schwartz in the last inequality.

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

$$\begin{aligned}
\left\| \left(\sum_{i=1}^{L} |B_{a_i} f|^2 \right)^{1/2} \right\|_{L^2(\mathbb{Z})} &= \left(\sum_{i=1}^{L} \left\| m_{a_i}(\beta) \widehat{f}(\beta) \right\|_{L^2(\mathbb{T})}^2 \right)^{1/2} \\
&\leq \left(\sum_{i=1}^{L} A^2 \left\| \widehat{f} \right\|_{L^2(\mathbb{T})}^2 \right)^{1/2} \\
&\leq \sqrt{L} A \|f\|_{L^2(\mathbb{Z})}
\end{aligned}$$

and the second summand

$$\begin{aligned}
\left\| \left(\frac{N}{L} \right)^{1/2} \left(\sum_{n=1}^{N} |B_n f - B_{n+1} f|^2 \right)^{1/2} \right\|_{L^2(\mathbb{Z})} &= \left(\frac{N}{L} \right)^{1/2} \left(\sum_{n=1}^{N} \left\| (m_n - m_{n+1})(\beta) \widehat{f}(\beta) \right\|_{L^2(\mathbb{T})}^2 \right)^{1/2} \\
&\leq \left(\frac{N}{L} \right)^{1/2} \sqrt{N} a \|f\|_{L^2(\mathbb{Z})} \\
&= \frac{Na}{\sqrt{L}} \|f\|_{L^2(\mathbb{Z})}.
\end{aligned}$$

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

With this lemma in hand, the proof can be completed in a few short strokes:

The Short Variation is $L^2(\mathbb{Z})$-Bounded, Take Two. We begin with the pointwise majorizations

$$\mathcal{V}^{r,S}(A_t f) \leq \left(\sum_n \mathcal{V}_n^2(C_t f)^2 \right)^{1/2} \leq \sum_j \left(\sum_n \mathcal{V}_n^2(C_t f_{n+j})^2 \right)^{1/2}.$$

With j frozen, we consider

$$\left\| \left(\sum_n \mathcal{V}_n^2(C_t f_{n+j})^2 \right)^{1/2} \right\|_{L^2(\mathbb{Z})}^2 = \sum_n \|\mathcal{V}_n^2(C_t f_{n+j})\|_{L^2(\mathbb{Z})}^2.$$

In the language of the previous lemma, for any given n we have

$$N = 2^n, \; A = 2^{-|j|}, \; \text{ and } \; a = 2^{-n};$$

we estimate

$$\|\mathcal{V}_n^2(C_t f_{n+j})\|_{L^2(\mathbb{Z})}^2 \lesssim NAa\|f_{n+j}\|_{L^2(\mathbb{Z})}^2 = 2^{-|j|}\|f_{n+j}\|_{L^2(\mathbb{Z})}^2.$$

Summing over n, and using the L^2-orthogonality of the $\{f_m\}$, yields the upper estimate

$$\sum_n \|\mathcal{V}_n^2(C_t f_{n+j})\|_{L^2(\mathbb{Z})}^2 \lesssim 2^{-|j|}\|f\|_{L^2(\mathbb{Z})}^2.$$

Taking the appropriate square-root, and summing over j then completes the proof. □

So, despite the increased delicacy of the variation operators, we have shown that for $E_i = [0, i), r > 2$, the r-variation operators, $\mathcal{V}^r(M_i f)$, are of strong-type $(2, 2)$ [11, Corollary 3.26]. [2]

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

[2]It is shown in [10] that $\mathcal{V}^2(M_i f)$ is unbounded on L^2; the super-delicacy of the two-variation operator $\mathcal{V}^2$ will be addressed below.

1.2.4 Further (Smooth) Variations

Since Bourgain's celebrated 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. [10], where it is shown that the variation operators $\{\mathcal{V}^r\}_{r>2}$ are of of weak-type $(p,p), 1 \leq p < \infty$ for intervals $E_i = [0,i)$. These results were partially extended [8] to the higher-dimensional setting of $\mathbb{Z}^d$-actions, where E_i are taken e.g. to be cubes $[0,i)^d$.

A major point in proving $L^p \to L^p$ estimates on the variation operators is that, roughly speaking, it is challenging for the variation operators to be large where the governing maximal operator is not. Indeed, we have already seen an explicit example of this phenomenon in our good-λ approach to Lépingle's theorem.

The goal of [5] was to further strengthen this heuristic. In particular, provided that the pertaining sets in question, $\{E_i\}$, are geometrically "regular" (see below) we show that the $\mathcal{V}^r$ operators share the same $L^p(w) \to L^p(w)$ mapping behavior as do the maximal functions. Here, $w \geq 0$ are appropriate (discrete) $A_p(\mathbb{Z}^d)$ weights.

One example of my work in this direction is the following result:

Theorem 1.2.15 (Corollary 1.8 of [5] – Special Case)**.** *Suppose $r > 2$, and $E_i = [0,i)^d \subset \mathbb{Z}^d$ are cubes as above. Then for any A_p weight $w \in A_p(\mathbb{Z}^d), 1 < p < \infty$, there exists an absolute constant, $C_{r,p}$, depending only on r, p and the A_p-norm of w, so that*

$$\int_{\mathbb{Z}^d} \mathcal{V}^r(A_i f)^p w \leq C_{r,p} \int_{\mathbb{Z}^d} |f|^p w.$$

Moreover, at the endpoint $p = 1$, there exists an absolute constant $C_{r,1}$, depending only on r the A_1-norm of w so that for each $\lambda \geq 0$,

$$\lambda w\left(\{\mathcal{V}^r(A_i f) > \lambda\}\right) \leq C_{r,1} \int_{\mathbb{Z}^d} |f| w.$$

This theorem and related work on higher-dimensional ergodic averages [5] largely resolved the outstanding questions posed in [10] and [8] on variation of *smooth* averages (below). Little research, however, had been directed towards studying variations of averaging operators defined by "rough," polynomially-defined, sets.

1.2.5 Variation of Polynomial Ergodic Averages

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. Despite the utility of the variational approach, the following problem, posed and engaged in [22], and recently resolved by Mirek, Stein, and Trojan [29], remained almost untouched:

Problem 1.2.16. *Let $P(x) \in \mathbb{Z}[x]$ be an integer-valued polynomial, and set*

$$E_i := \{P(1), \ldots, P(i)\}.$$

For which $1 < p < \infty$, $2 < r < \infty$ do there exist a priori bounds

$$\|\mathcal{V}^r(M_N f)\|_p \leq C_{p,r,P} \|f\|_p?$$

My main result in this direction is that polynomial means converge rapidly in L^2:

Theorem 1.2.17 (Theorem 1.3 of [22])**.** *In the above setting, there exists an absolute constant $C_{r,P}$, depending only on $r > 2$ and the polynomial, P, so that for any measure preserving system (X, μ, τ), and any $f \in L^2(X)$,*

$$\|\mathcal{V}^r(M_N f)\|_{L^2(X)} \leq C_{r,P} \|f\|_{L^2(X)}.$$

Interpolating this result against Bourgain's celebrated result [11]

$$\|\mathcal{V}^\infty(M_N f)\|_{L^p(X)} \leq C_{p,P} \|f\|_{L^p(X)}, \; p > 1,$$

yields the following

Proposition 1.2.18 (Proposition 1.5 of [22]). *Suppose $r > \max\{p, p'\}$. Then for any dynamical system there exist absolute constants $C_{r,p,P}$ so that*

$$\|\mathcal{V}^r(M_N f)\|_{L^p(X)} \leq C_{r,p,P}\|f\|_{L^p(X)}.$$

As it so happens, the L^2 theory is, in general, sharp – as in the case of the standard Birkhoff means, the 2-variation operator is unbounded on L^2:

Theorem 1.2.19 (Theorem 1.7 of [22]). *For any $C > 0$, there exists an $f = f_C$ of L^2-norm one, but so that*

$$\|\mathcal{V}^2(M_N f)\|_{L^2(X)} \geq C,$$

where here we fix $M_N f(x) := \frac{1}{N}\sum_{n\leq N} \tau^{n^2} f(x)$, the (discrete) square means.

Remark 1.2.20. *Although this theorem is generalizable to actions associated to other integer-valued polynomials, for the sake of concreteness, we content ourselves with the case of the squares.*

These two papers more or less comprise my work on variations of ergodic averages, and questions concerning rates of convergence of ergodic averages. The remainder of my work in pointwise ergodic theory concerns understanding *modulated* ergodic averages, discussed in the following sections.

1.3 Wiener-Wintner Theorems: Convergence of Twisted Ergodic Averages

Prior to Bourgain's polynomial ergodic theorem, another generalization of Birkhoff's theorem was announced:

Theorem 1.3.1 (Wiener-Wintner Theorem [42], cf. [1]). *For every $f \in L^1(X)$ there exists* *a subset $X_f \subset X$ of full measure so that for* every $\theta \in [0,1]$ *the weighted averages,*

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

converge for all $x \in X_f$. In particular – X_f is entirely independent of θ.

Some remarks about this theorem are in order:

The density-argument proof of Birkhoff's theorem saw us decompose each $f \in L^1(X)$ function into a component fixed by the action, h, a co-boundary component, $g - \tau g$, and an arbitrarily small $L^1(X)$-error term, η, so that we could express

$$f = h + (g - \tau g) + \eta.$$

Heuristically, we can think of this decomposition as a reflection of the dichotomy between *structure* and *randomness*: roughly speaking, we begin with an L^1 function and separate it into a *structured* function, fixed by the action, and a *random* function, telescoping to zero in an average sense.

This same dichotomy informs the proof of the Wiener-Wintner theorem, but the "twist" is that we must be more careful with how we define our notion of randomness. Indeed, a naive application of our twisted means to a (bounded) coboundary yields:

$$\frac{1}{N} \sum_{n \leq N} e(n\theta) \tau^n (g - \tau g) = O\left(\frac{\|g\|_\infty}{N} \right) + \frac{1}{N} \sum_{n=2}^{N} (e(n\theta) - e((n-1)\theta)) T^n g.$$

Now, when $\theta = 0$, and we're back in the setting of Birkhoff's theorem, the latter sum vanishes; nothing can be said about the sum for general θ.

It turns out that the correct dichotomy between structure and randomness involves the so-called *Kronecker factor* of the system (X, μ, τ).

Definition 1.3.2. *For a measure preserving system* (X, μ, τ) *the* Kronecker factor[3] *of the system is*

$$\mathcal{K} := \overline{\bigoplus_{\lambda} \{\tau^n h_\lambda : h_\lambda \text{ is an eigenfunction of } \tau \text{ with eigenvalue } \lambda\}}.$$

One has the following convenient spectral characterization of $\mathcal{K}^\perp$:

Proposition 1.3.3. *The following three conditions are equivalent:*

1. $f \in \mathcal{K}^\perp$;
2. f *has a continuous spectral measure; and*
3. *The (bounded) correlation coefficients*

$$\langle \tau^n f, f \rangle$$

tend to zero outside an exceptional set $E \subset \mathbb{N}$ *of zero density.*

This proof will heavily exploit the Hilbert space properties of $L^2(X)$, so we will state here a powerful Hilbert space orthogonality lemma which will be used below.

Lemma 1.3.4 (Van der Corput)**.** *For a sequence of vectors in a Hilbert space,* $\{v_n\} \subset H$*, and any* $M \leq N$*,*

$$\left\| \sum_{n \leq N} v_n \right\|^2 \lesssim \frac{N}{M} \sum_{n \leq N} \|v_n\|^2 + \frac{N}{M} \sum_{m \leq M} \left| \sum_{n \leq N-m} \langle v_{n+m}, v_n \rangle \right|.$$

Remark 1.3.5. *Informally, the more orthogonal the* $\{v_n\}$ *are to their neighbors – the smaller their average* $\frac{1}{N}\sum_{n \leq N} v_n$ *becomes.*

[3] More can be said about $\mathcal{K}$:

$$\mathcal{K} := L^2(X, \mathcal{Z}, \mu),$$

where $\mathcal{Z}$ is the σ-algebra generated by the eigenfunctions of the action.

Proof of $1 \Rightarrow 2$. Let ν_f denote the spectral measure (on the torus, $\mathbb{T}$) associated to f; such a measure exists by the Bochner-Herglotz theorem since $\{\langle \tau^n f, f\rangle\}$ forms a positive semi-definite sequence, in that for each $\{z_n \in \mathbb{C} : n \leq K\}$,

$$\sum_{n,m \leq K} \langle \tau^{n-m} f, f\rangle z_n \overline{z_m} \geq 0.$$

Indeed, we just observe that

$$\sum_{n,m \leq K} \langle \tau^{n-m} f, f\rangle z_n \overline{z_m} = \left\| \sum_{n \leq K} z_n \tau^n f \right\|_{L^2(X)}^2 \geq 0,$$

since $\langle \tau^n f, \tau^m f\rangle = \langle \tau^{n-m} f, f\rangle$, by the unitarity of τ.

Now, if ν_f has an atom, $\nu_f(\{\alpha\}) > 0$, $g := 1_\alpha$ is a non-zero vector in $L^2(\nu_f)$, and corresponds via the spectral theorem to a non-zero function, $G \in \overline{\text{Span}\{\tau^n f\}}$. Note furthermore that the multiplication operator

$$Mg := e(\theta) g = e(\alpha) g,$$

and so by the spectral theorem G is an eigenfunction of τ with eigenvalue $e(\alpha)$.

If $f \in \mathcal{K}^\perp$, however, (finite) linear combinations of $\{\tau^n f\}$ also live in $\mathcal{K}^\perp$

$$\sum_j a_j \tau^j f \in \mathcal{K}^\perp,$$

since for any eigenfunction, h, with eigenvalue $e(\beta) = e(\beta(h))$,

$$\begin{aligned}\left\langle \sum_j a_j \tau^j f, h \right\rangle &= \sum_j a_j \langle \tau^j f, h\rangle \\ &= \sum_j a_j \langle f, \tau^{-j} h\rangle \\ &= \sum_j a_j e(j\beta) \langle f, h\rangle \\ &= 0.\end{aligned}$$

By continuity, if $f \in \mathcal{K}^\perp$, $\overline{\mathrm{Span}\{\tau^n f\}} \subset \mathcal{K}^\perp$. The upshot is that if ν_f is atomic, $f \notin \mathcal{K}^\perp$, or alternatively: each $f \in \mathcal{K}^\perp$ has a continuous spectral measure. $\square$

To prove that the second point of the above proposition implies the third, we will need the following lemma, which we will apply to the bounded, non-negative numbers $a_n = |\langle \tau^n f, f\rangle|$.

Lemma 1.3.6. *For a bounded sequence of non-negative numbers $\{a_n\}$ the following three conditions are equivalent:*

1. $\frac{1}{N} \sum_{n \le N} a_n \to 0$;
2. $\frac{1}{N} \sum_{n \le N} a_n^2 \to 0$; *and*
3. *There exists an exceptional set $E \subset \mathbb{N}$ of zero density, that is*

$$\lim_{N \to \infty} \frac{|E \cap \{1, \dots, N\}|}{N} = 0,$$

outside of which $\lim_{E \not\ni n \to \infty} a_n = 0$.

Proof. That the first point implies the second is clear: since the $\{a_n\}$ are bounded,

$$\frac{1}{N} \sum_{n \le N} a_n^2 \lesssim \frac{1}{N} \sum_{n \le N} a_n.$$

That the third implies the first is also not too hard: if $\epsilon > 0$ is arbitrary, and $N = N_\epsilon$ is so large that for all $n \geq N$, $n \notin E$, $a_n \ll \epsilon$ then for all $M \gg N$ we may split:

$$\begin{aligned}\frac{1}{M}\sum_{n\leq M} a_n &= \frac{1}{M}\sum_{E\ni n\leq M} a_n + \frac{1}{M}\sum_{E\not\ni n\leq N} a_n + \frac{1}{M}\sum_{E\not\ni n, N\leq n\leq M} a_n \\ &\lesssim \frac{|E\cap\{1,\ldots,M\}|}{M} + \frac{N}{M} + \epsilon \\ &= o_{M\to\infty}(1) + \epsilon,\end{aligned}$$

as desired.

So, it remains to prove that the second point implies the third. Define for each $m \geq 1$, the "obstruction sets"

$$O_m := \{n : a_n \geq 2^{-m}\};$$

by Markov's inequality,

$$2^{-2m}\frac{|O_m\cap\{1,\ldots,N\}|}{N} \leq \frac{1}{N}\sum_{O_m\ni n\leq N} a_n^2 \leq \frac{1}{N}\sum_{n\leq N} a_n^2 = o_{N\to\infty}(1).$$

Take now an increasing sequence of integers $\{M_m\}$ so that

$$\frac{|O_m\cap\{1,\ldots,N\}|}{N} \leq 2^{-m}$$

for all $N \geq M_m$. Define the exceptional set

$$E := \bigcup_{m\geq 1} O_m \cap [M_m, M_{m+1}),$$

and observe that $a_n \to 0$ outside E: if $k = k(n)$ is such that $M_k \leq n < M_{k+1}$, then for $a_n \notin E$, $a_n \leq 2^{-k}$. To see that E has zero density, we just observe that if $M_m \leq N < M_{m+1}$ is large, then

$$E\cap\{1,\ldots,N\} \subset O_m\cap\{1,\ldots,N\}$$

by the nesting property of the obstruction sets. But since $N \geq M_m$

$$\frac{|O_m \cap \{1, \ldots, N\}|}{N} \leq 2^{-m} \to 0,$$

as desired. □

Proof of $2 \Rightarrow 3$. By the previous lemma, it's enough to show that

$$\frac{1}{N} \sum_{n \leq N} |\langle \tau^n f, f \rangle|^2 \to 0;$$

by the spectral theorem, with ν_f as above, it's enough to show that

$$\frac{1}{N} \sum_{n \leq N} |\langle \tau^n f, f \rangle|^2 = \frac{1}{N} \sum_{n \leq N} \int_{\mathbb{T} \times \mathbb{T}} e(n(\beta - \theta)) \; d\nu_f(\theta) \times d\nu_f(\beta)$$

tends to zero. Taking a limit in N, and using the bounded convergence theorem to pass the limit inside the integral, we see that we need consider

$$\int_{\mathbb{T} \times \mathbb{T}} \lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} e(n(\beta - \theta)) \; d\nu_f(\theta) \times d\nu_f(\beta) = \int_{\mathbb{T} \times \mathbb{T}} 1_{\theta = \beta}(\theta, \beta) \; d\nu_f(\theta) \times d\nu_f(\beta),$$

which vanishes identically in the event that ν_f is continuous. □

It remains to show that if $\frac{1}{N} \sum_{n \leq N} |\langle \tau^n f, f \rangle|^2 \to 0$, then $f \in \mathcal{K}^\perp$.

Proof of $3 \Rightarrow 1$. We first prove the following lemma:

Lemma 1.3.7. *If* $\frac{1}{N} \sum_{n \leq N} |\langle \tau^n f, f \rangle|^2 \to 0$*, then for each* $g \in L^2(X)$*,*

$$\frac{1}{N} \sum_{n \leq N} |\langle \tau^n f, g \rangle|^2 = \frac{1}{N} \sum_{n \leq N} \langle \tau^n f, g \rangle \langle g, \tau^n f \rangle = \left\langle \frac{1}{N} \sum_{n \leq N} \langle g, \tau^n f \rangle \tau^n f, g \right\rangle \to 0.$$

Proof of the Lemma. We will prove, in fact, that

$$\frac{1}{N}\sum_{n\leq N}\langle g,\tau^n f\rangle\tau^n f \to 0.$$

To do this, we use Van der Corput's inequality with

$$v_n := \langle g,\tau^n f\rangle\tau^n f \text{ and } H := L^2(X);$$

for $M \leq N$, we bound

$$\begin{aligned}
&\left\|\frac{1}{N}\sum_{n\leq N}\langle g,\tau^n f\rangle\tau^n f\right\|_{L^2(X)}^2 \\
&\quad\lesssim \frac{1}{NM}\sum_{n\leq N}|\langle\tau^n f,g\rangle|^2\|f\|_{L^2(X)}^2 \\
&\qquad + \frac{1}{NM}\sum_{m\leq M}\left|\sum_{n\leq N}\langle g,\tau^{n+m}f\rangle\langle\tau^n f,g\rangle\langle\tau^m f,f\rangle\right| \\
&\qquad\quad + O\left(\frac{M}{N}\|g\|_{L^2(X)}^2\|f\|_{L^2(X)}^4\right).
\end{aligned}$$

Now, we majorize the first sum

$$\frac{1}{NM}\sum_{n\leq N}|\langle\tau^n f,g\rangle|^2\|f\|_{L^2(X)}^2 \leq \frac{1}{M}\cdot\|g\|_{L^2(X)}^2\|f\|_{L^2(X)}^4;$$

for the second sum we bound

$$\frac{1}{NM}\sum_{m\leq M}\left|\sum_{n\leq N}\langle g,\tau^{n+m}f\rangle\langle\tau^n f,g\rangle\langle\tau^m f,f\rangle\right| \leq \frac{1}{M}\sum_{m\leq M}|\langle\tau^m f,f\rangle|\cdot\|g\|_{L^2(X)}^2\|f\|_{L^2(X)}^2.$$

Sending $N\to\infty$, then $M\to\infty$ yields the conclusion. $\square$

With this stronger orthogonality property in mind, suppose that h is an eigenfunction of the action, with eigenvalue $e(\beta)$ for some $\beta = \beta(h)$. Then

$$|\langle f, h\rangle| = \frac{1}{N}\sum_{n\le N} |\langle f, h\rangle| = \frac{1}{N}\sum_{n\le N} |\langle f, \tau^{-n} h\rangle| = \frac{1}{N}\sum_{n\le N} |\langle \tau^n f, h\rangle| \to 0.$$

The triangle inequality yields the result for finite linear combinations of eigenfunctions, and continuity completes the proof. $\square$

We will prove the Wiener-Wintner theorem by establishing convergence for *finite* linear combinations of eigenfunctions in $\mathcal{K}$; by proving uniform-in-θ convergence to zero for functions in $\mathcal{K}^\perp$; and finally by showing that the set of functions for which the conclusion of the Wiener-Wintner theorem holds is closed in $L^1(X)$.

Proof of Wiener-Wintner Theorem for Finite Linear Combinations of Eigenfunctions. It is not too hard to see that if $f \in \mathcal{K}$ is a finite linear combination of eigenfunctions, there exists a set $X_f \subset X$ of full measure so that the Wiener-Wintner modulated averages converge for all $\theta \in [0,1]$:

If $\{f_j\} \subset L^2(X)$ are eigenfunctions of the action, with pertaining eigenvalues $e(\lambda_j)$, and $f \in \mathcal{K}$ is a finite linear combination

$$f = \sum a_j f_j,$$

then we have

$$\frac{1}{N}\sum_{n\le N} e(n\theta)\tau^n f(x) = \sum_j a_j \left(\frac{1}{N}\sum_{n\le N} e(n(\theta+\lambda_j)) \right) f_j(x),$$

which converges μ-a.e. for each $\theta \in [0,1]$. Indeed, for each j, the inner average converges to 0 if $\theta \neq -\lambda_j$ by a geometric series argument, while is identically one otherwise. $\square$

Informally, we can view the success of the above argument as a reflection that the eigenfunctions of τ share the same structure/random composition as do the sequences $\{e(n\theta)\}$.

We next turn to our random functions, where we might expect "moral summation by parts" to hold, and the weighted averages

$$\frac{1}{N}\sum_{n\leq N} e(n\theta)\tau^n f(x) \to 0.$$

Our method of proof will not directly proceed by summation by parts, but will still make heavy use of the idea that for $f \in \mathcal{K}^{\perp}$, successive iterates $\{\tau^n f\}$ have difficulty correlating.

Proof of Uniform Wiener-Wintner Theorem for $\mathcal{K}^{\perp}$. The goal here is to prove that for $f \in \mathcal{K}^{\perp}$,

$$\sup_{\theta\in[0,1]} \left|\frac{1}{N}\sum_{n\leq N} e(n\theta)\tau^n f(x)\right| \to 0 \ \mu - a.e.$$

To do so, we will use Van der Corput's lemma.

With $\theta \in [0,1]$ arbitrary, we will apply this lemma in the setting of

$$v_n := e(n\theta)\tau^n f(x), \ H := L^2(X);$$

We get

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{n\leq N} e(n\theta)\tau^n f(x)\right| \\
&\quad \lesssim \frac{1}{M}\cdot\frac{1}{N}\sum_{n\leq N}|\tau^n f(x)|^2 + \frac{1}{M}\cdot\frac{1}{N}\sum_{m\leq M}\left|e(m\theta)\sum_{n\leq N-m}\int_X \tau^{n+m} f\tau^n\overline{f}(x)\right| \\
&\quad = \frac{1}{M}\cdot\frac{1}{N}\sum_{n\leq N}|\tau^n f(x)|^2 \\
&\qquad + \frac{1}{M}\cdot\frac{1}{N}\sum_{m\leq M}\left|\sum_{n\leq N}\int_X \tau^m f\overline{f}(x)\right| \\
&\qquad\quad + O\left(\frac{M}{N}\|f\|_{L^2(X)}^2\right),
\end{aligned}$$

where we precomposed with τ^{-n} (using the measure-preserving nature of the action) and

Cauchy-Schwartz to majorize

$$\int_X \tau^m f\overline{f} = O(\|f\|_{L^2(X)}^2)$$

in passing to the final line above. Note that our above estimate is entirely independent of θ.

If we send $N \to \infty$ and take into account the pointwise ergodic theorem, we find that μ-a.e. we have

$$\frac{1}{M}\|f\|_{L^2(X)}^2 + \frac{1}{M} \cdot \sum_{m \leq M} \left| \int_X \tau^m f\overline{f}(x) \right|;$$

sending $M \to \infty$ and using the fact that $f \in \mathcal{K}^\perp$ completes the proof. □

As in our first proof of Birkhoff's theorem, we have proven our result for a dense subclass of $L^2(X)$. We will extend the result to all of $L^1(X)$ using the following (somewhat general) Wiener-Wintner-type Banach Principle.

Recall that a point $x \in X$ is called *generic* if the conclusion of the pointwise ergodic theorem holds for the Cesaro averages of the orbits $\{\tau x, \tau^2 x, \dots\}$, i.e. if

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

converges.

Proposition 1.3.8 (Wiener-Wintner-type Banach Principle)**.** *Let A be a set of uniformly bounded sequences $\{a_n\} \subset \mathbb{C}$. Then, for every measure-preserving system (X, μ, τ), the set of functions $f \in L^1(X)$ for which for μ-a.e. the weighted averages*

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

converge for every $\{a_n\} \in A$ is closed in the $L^1(X)$ norm.

Conclusion of the Wiener-Wintner Theorem. Using the uniform boundedness of $\{a_n\} \subset A$, one first observes that

$$\left|\frac{1}{N}\sum_{n\le N} a_n\tau^n f(x)\right| \lesssim \frac{1}{N}\sum_{n\le N}\tau^n|f|(x) \le \mathcal{M}_\tau f.$$

Assume now that there is a sequence of functions $\{f_j\} \subset L^1(X)$ so $\|f_j - f\|_{L^1(X)} \to 0$ so that for every f_j there is a set $X_j \subset X$ of full measure on which the averages

$$\frac{1}{N}\sum_{n\le N} a_n\tau^n f(x)$$

converge for every $x \in X_j$ for each $\{a_n\} \subset A$. Define the set

$$X' := \bigcap X_j \cap \{x : x \text{ is generic for each } |f - f_j|\};$$

note that X' has full measure. For each $x \in X$ we have the upper estimate – valid for any j –

$$\begin{aligned}
&\limsup_{N,M\to\infty}\sup_{\{a_n\}\in\mathcal{A}}\left|\frac{1}{N}\sum_{n\le N} a_n\tau^n f(x) - \frac{1}{M}\sum_{n\le M} a_n\tau^n f(x)\right| \\
&\quad\lesssim \limsup_{N,M\to\infty}\sup_{\{a_n\}\in\mathcal{A}}\left|\frac{1}{N}\sum_{n\le N} a_n\tau^n f_j(x) - \frac{1}{M}\sum_{n\le M} a_n\tau^n f_j(x)\right| + \mathcal{M}_\tau(f - f_j),
\end{aligned}$$

where the implicit constant can be taken as $2\sup_{\{a_n\}\in A}\|\{a_n\}\|_{l^\infty}$. The upshot is that for any $\epsilon > 0$, and any j, we have

$$\begin{aligned}
&\mu\left(\left\{\limsup_{N,M\to\infty}\sup_{\{a_n\}\in\mathcal{A}}|\frac{1}{N}\sum_{n\le N} a_n\tau^n f(x) - \frac{1}{M}\sum_{n\le M} a_n\tau^n f(x)| > \epsilon\right\}\right) \\
&\quad\le \mu\left(\{\mathcal{M}_\tau(f - f_j) \gtrsim \epsilon\}\right) \\
&\quad\lesssim \epsilon^{-1}\|f - f_j\|_{L^1(X)};
\end{aligned}$$

sending $j \to \infty$ concludes the argument. $\square$

1.3.1 Further Wiener-Wintner Results

A major generalization of the Wiener-Wintner theorem is due to Lesigne in [25]:

Theorem 1.3.9 (Special Case). *Let $P(x) \in \mathbb{Z}[x]$ be a polynomial with integer coefficients. For every $f \in L^1(X)$ there exists a subset $X_f \subset X$ of full measure so that for* every $\theta \in [0,1]$ *the weighted polynomial averages,*

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

converge for all $x \in X_f$. Again, X_f does not depend on the polynomial P.

Although this list is by no means comprehensive, we remark that Lesigne's result has been further generalized and "uniformized" by Frantzikinakis [18], Host-Kra [23], and Eisner and Zorin-Kranich [16].

In recent work with Tanja Eisner [12], we generalize Lesigne's result to *Hardy field* functions which are suitably "far" from polynomials. Our main result in this drection can be stated as follows:

Theorem 1.3.10. *For $0 < \delta < 1/2$, $m \geq 0$, and $M \geq 1$, consider the class of functions $\mathcal{M}_{\delta,M,m}$. For every $f \in L^1(X)$ there exists a subset $X_f \subset X$ of full measure so that the averages,*

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

converge uniformly *(in $p \in \mathcal{M}_{\delta,M,m}$) to zero for all $x \in X_f$, where X_f is independent of $p \in \mathcal{M}_{\delta,M,m}$.*

Moreover, if (X, μ, τ) is uniquely ergodic and $f \in C(X)$, one has

$$\sup_{p \in \mathcal{M}_{\delta,M,m}} \left\| \frac{1}{N} \sum_{n \leq N} e(p(n)) \tau^n f \right\|_{L^\infty(X)} \to 0.$$

See [12, §2] for precise definitions and references; informally, these smooth functions are uniformly "(δ, M, m)-far" from the class of polynomials.

Our second aim was to consider weighted averages when the polynomial weights "lie on τ." Specifically, we considered pointwise convergence of weighted averages of the form

$$\left\{ \frac{1}{N} \sum_{n \leq N} e(n\theta) \tau^{P(n)} f(x) \right\}$$

where $\theta \in [0,1]$ and P is a polynomial with integer coefficients, as above.

We state our main results in the case of the squares.

Theorem 1.3.11. *Let $0 < c < 1$ and $E = E_c \subset [0,1]$ be a set of c-badly approximable numbers with upper Minkowski dimension strictly less than $1/16$. For every $f \in L^2(X)$ there exists a subset $X_f \subset X$ of full measure so that the averages*

$$\left\{ \frac{1}{N} \sum_{n \leq N} e(n\theta) \tau^{n^2} f(x) \right\},$$

converge uniformly *(in $\theta \in E_c$) to zero for all $x \in X_f$, where X_f is independent of $\theta \in E_c$.*

We recall that a number, θ, is said to be *c-badly approximable* if for all $\frac{p}{q}$ reduced fractions, we have the lower bound $|\theta - \frac{p}{q}| \geq \frac{c}{q^2}$; note that by Dirichlet's principle we automatically have $0 < c < 1$.

A natural follow-up question concerns the behavior of the twisted square means in the case where θ is *not* badly-approximable. We provide a new, elementary proof of the following

Theorem 1.3.12. *Let P be as above, and $\theta \in [0,1]$ be arbitrary. For any $f \in L^p(X)$, $p > 1$, there exists a subset $X_f \subset X$ of full measure so that the averages,*

$$\left\{ \frac{1}{N} \sum_{n \leq N} e(n\theta) \tau^{P(n)} f(x) \right\},$$

converge pointwise on X_f.

This project segued to the following one:

1.3.2 Twisted (Sparse) Random Ergodic Averages in L^1

Our discussion of sparse ergodic averages must begin with

Theorem 1.3.13 (Bourgain's Polynomial Ergodic Theorem [11])**.** *For any polynomial $P(n)$ with integer coefficients, the sequence $\{P(1), \ldots, P(n)\}$ is universally L^p-good for each $p > 1$.*

Note that $\{P(1), \ldots, P(n)\}$ are zero-Banach-density subsequences of the integers; in fact, Bourgain used a probabilistic method to "construct" *extremely* sparse subsequences along which pointwise ergodic theorems held – in the range $p > 1$:

Theorem 1.3.14 (Bourgain's Random Ergodic Theorem, §8 of [10])**.** *Suppose $\{X_n\} : \Omega \to \{0,1\}$ are independent random variables with expectations $\{\sigma_n\}$, and let*

$$S^\omega = \{n : X_n(\omega) = 1\}.$$

If $\sigma_n = \frac{(\log \log n)^{B_p}}{n}$ with $B_p > \frac{1}{p-1}$, $1 < p \leq 2$, then almost surely, S^ω is universally L^p-good.

Bourgain's approach, which relied on Fourier analysis, left the endpoint $p = 1$ case unaddressed. Moreover, Bourgain's Polynomial Ergodic Theorem was recently shown to *fail* at the endpoint case $p = 1$ (see Buzcolich and Mauldin [7], or LaVictoire [27]); the distinction between pointwise ergodic theorems between $p > 1$ and $p = 1$ is fundamental.

Indeed, it was not until 2007 that an example of a Banach-density-zero universally L^1-good sequence was constructed by Buzcolich [6]; Urban and Zienkiewicz [33] subsequently proved that the sequence $\{\lfloor k^a \rfloor\}$ was universally L^1-good for $1 < a < 1 + \frac{1}{1000}$, and Mirek [28] recently extended their results to include a wider class of sequences; good representative examples of such sequences are $\{\lfloor k^a \rfloor\}$ for $1 < a < 1 + \frac{1}{29}$.

Before Mirek's result, LaVictoire succeeded in extending Bourgain's probabilistic construction to the L^1 setting; he proved the following

Theorem 1.3.15 (LaVictoire's Random Ergodic Theorem [26]). *In the setting of Bourgain's Random Ergodic Theorem, suppose $\sigma_n = n^{-a}$, with $0 < a < 1/2$. Then almost surely, S^ω is universally L^1-good.*

To compare this result with that of Urban and Zienkiewicz and Mirek, we note that by the strong law of large numbers, such sequences S^ω have $n_k \approx_\omega k^{\frac{1}{1-a}}$ with full probability (i.e. almost surely there exist absolute constants $c_\omega < C_\omega$ so that $c_\omega k^{\frac{1}{1-a}} \leq n_k \leq C_\omega k^{\frac{1}{1-a}}$).

This is as far as the story goes for *unweighted* ergodic averages in L^1; to the best of my knowledge, the question of weighted pointwise convergence for L^1 functions along (sparse) random sequences has remained unaddressed.

In joint work with Zorin-Kranich [23], we sought to bridge the gap between pointwise convergence for weighted averages, and pointwise convergence for sparse (random) averages:

Let $\{X_n\}$ be as in LaVictoire's Random Ergodic Theorem, i.e. independent $0-1$ random variables, with expectations $\sigma_n = n^{-a}$ for $0 < a < 1/2$. We define the *counting function* $a_n(\omega)$ to be the smallest integer subject to the constraint

$$X_1(\omega) + \cdots + X_{a_n(\omega)}(\omega) = n.$$

(Note that by the strong law of large numbers, almost surely $a_n(\omega)/n^{\frac{1}{1-a}}$ converges to a non-zero number.)

With $0 < \epsilon < 1$ arbitrary but fixed, suppose $p : \mathbb{R} \to \mathbb{R}$ is a *logarithmico-exponential* function satisfying

- the second-order difference relationship

$$p(x + y + z) - p(x + y) - p(x + z) + p(x) = O(x^{\epsilon-1}yz)$$

- for all $a(x) \in C \cdot \mathbb{Q}[x]$, the set of constant multiples of rational polynomials,

$$\frac{|a(x) - p(x)|}{\log x} \to \infty$$

For instance, $p(t) = t^{1+\epsilon}$ satisfies the above criterion. For a fuller discussion of logarithmico-exponential functions, see [3]; informally, these are all the functions one can get by combining real constants, the variable x, and the symbols exp, log, $\cdot$, and $+$. (e.g. $x^{1/2} = \exp(1/2 \cdot \log x)$ and $x^\pi / \log \log x$ are both logarithmico-exponential.)

Our main result is the following

Theorem 1.3.16. *Let $0 < \epsilon < 1$, $0 < a < 1/2$, $\{X_n\}$, a_n, and p be as above. There exists a set of full probability $\Omega_0 \subset \Omega$, so that for each $\omega \in \Omega_0$, the following L^1-phenomenon holds: for each (X, μ, τ) measure-preserving system, and each $f \in L^1(X)$, the averages*

$$\left\{ \frac{1}{N} \sum_{n=1}^{N} e(p(n)) \tau^{a_n(\omega)} f \right\}$$

converge pointwise to zero.

The novelty of this result is that there are, roughly speaking, two competing notations of "randomness" in the above modulated means: the honest randomness generated by the random sequence of times, $\{a_n\}$, and the moral randomness generated by the phase, $p(n)$. That the randomness generated by the phase "dominates" in the limit provides further evidence – beyond LaVictoire's maximal theory [26] – to latent smoothness (in an average sense) in the random averages along the $\{a_n\}$.

With these results in mind, we now turn to the main body of the dissertation.

References

[1] I. Assani. Wiener Wintner ergodic theorems, World Scientific Publishing Co. Inc., River Edge, NJ, 2003

[2] Birkhoff, G. "Proof of the ergodic theorem", Proc Natl Acad Sci USA 17 (12): 656-660 (1931)

[3] Boshernitzan, M.; Kolesnik, G.; Quas, A.; Wierdl, M. Ergodic averaging sequences. In: J. Anal. Math. 95 (2005), pp. 63-103. ISSN: 0021-7670.

[4] Bourgain, J. On the maximal ergodic theorem for certain subsets of the positive integers. Israel J. Math. 61 (1988), 39-72.

[5] Bourgain, J. Pointwise ergodic theorems for arithmetic sets. Inst. Hautes Études Sci. Publ. Math. (69):5-45, 1989. With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein.

Israel J. Math. 61 (1988), no. 1, 7384.

[6] Buczolich, Z. Universally L^1 good sequences with gaps tending to infinity. Acta Math. Hungar. 117 (2007), no. 1-2, 91-140.

[7] Buczolich, Z; Mauldin, D. Concepts behind divergent ergodic averages along the squares. Ergodic Theory and Related Fields, AMS, Contemporary Mathematics Vol. 430 (2007) 41-56

[8] Burkholder, D. L. *Martingale transforms.* Ann. Math. Statist. 37 1966 1494-1504.

[9] Burkholder, D.; David, B.; Gundy, R. *Integral inequalities for convex functions of operators on martingales.* Proc. Sixth Berkeley Semp. Math. Statist. Prob., 2 (1972), 223-240.

[10] Calderón, A. Ergodic theory and translation invariant operators, Proc. Nat. Acad. Sci., USA 59 (1968), 349-353

[11] Davis, B. On the integrability of the martingale square function. Israel J. Math. 8 1970 187-190.

[12] Eisner, T.; Krause, B. (Uniform) Convergence of Twisted Ergodic Averages. Preprint, http://arxiv.org/pdf/1407.4736.pdf.

[13] Eisner, T.; Zorin-Kranich, P. *On the uniformity in the Wiener-Wintner theorem for nilsequences*, Discrete Cont. Dyn. Syst., to appear.

[14] Frantzikinakis, N. *Uniformity in the polynomial Wiener-Wintner theorem*, Ergodic Theory Dynam. Systems **26** (2006), 1061–1071.

[15] Frantzikinakis, N.; Lesigne, E.; Wierdl, M. Random sequences and pointwise convergence of multiple ergodic averages. Indiana Univ. Math. J. 61 (2012), no. 2, 585617.

[16] Host, B.; Kra, B. *Uniformity seminorms on l^∞ and applications*, J. Anal. Math. **108** (2009), 219–276.

[17] Jones, R.; Kaufman, R.; Rosenblatt, J.; Wierdl, M. Oscillation in ergodic theory. Ergodic Theory Dynam. Systems 18 (1998), no. 4, 889-935.

[18] Jones, R.; Rosenblatt, J.; Wierdl, M. Oscillation in ergodic theory: higher dimensional results. Israel J. Math. 135 (2003), 1-27.

[19] Jones, R.; Seeger, A.; Wright, J. Strong variational and jump inequalities in harmonic analysis. Trans. Amer. Math. Soc. 360 (2008), no. 12, 67116742.

[20] Jones, R; Wang, G. Variation inequalities for the Féjer and Poisson kernels. Trans. Amer. Math. Soc., 356 (2004), 4493-4518.

[21] Krause, B. On Higher-Dimensional Oscillation in Ergodic Theory. Preprint, http://arxiv.org/pdf/1309.2336.pdf

[22] Krause, B. Polynomial Ergodic Averages Converge Quickly: Variations on a Theorem of Bourgain. Preprint, http://arxiv.org/pdf/1402.1803.pdf

[23] Krause, B.; Zorin-Kranich, P. A random pointwise ergodic theorem with Hardy field weights. Preprint, http://arxiv.org/pdf/1410.0806.pdf.

[24] Lépingle, D. La variation d'ordre p des semi-martingales. Z.F.W. 36, 1976, 295-316.

[25] E. Lesigne, Spectre quasi-discret et théorème ergodique de Wiener-Wintner pour les polynômes, Ergodic Theory Dynam. Systems **13** (1993), 767–784.

[26] LaVictoire, P. An L^1 ergodic theorem for sparse random subsequences. Math. Res. Lett. 16 (2009), no. 5, 849-859.

[27] LaVictoire, P. Universally L^1-bad arithmetic sequences. J. Anal. Math. 113 (2011), 241263.

[28] Mirek, M. Weak Type $(1,1)$ Inequalities for Discrete Rough Maximal Functions. Preprint, http://arxiv.org/pdf/1305.0575v2.pdf

[29] Mirek, M.; Stein, E.; Trojan, B. Variational estimates for discrete averaging operators and truncated singular integrals. Preprint.

[30] Pisier, G.; Xu, Q. The strong p-variation of martingales and orthogonal series. Prob. Theory, 77 (1988), 497-541.

[31] Qian, J. The p-variation of partial sum processes and the empirical process. Ann. of Prob. Theory, 77 (1988), 1370-1383.

[32] Stein, E. On Limits of Sequences of Operators. Annals of Mathematics, Second Series, Vol. 74, No. 1 (Jul., 1961), pp. 140-170

[33] Urban, R.; Zienkiewicz, J. Weak type $(1,1)$ estimates for a class of discrete rough maximal functions. Math. Res. Lett. 14 (2007), no. 2, 227237.

[34] Wiener, N.; Wintner, A. Harmonic analysis and ergodic theory, Amer. J. Math. 63 (1941), 415426.

[35] Wilson, M. Weighted Littlewood-Paley theory and exponential-square integrability. Lecture Notes in Mathematics, 1924. Springer, Berlin, 2008.

CHAPTER 2

On Higher Dimensional Oscillation in Ergodic Theory

2.1 Introduction

Let (X, Σ, μ) be a non-atomic probability space, equipped with τ a measure preserving $\mathbb{Z}^d$-action

$$\tau_y f(x) := f(\tau_{-y} x).$$

For $\{E_i\} \subset \mathbb{Z}^d$, define the averaging operators

$$M_i f(x) := \frac{1}{|E_i|} \sum_{y \in E_i} (\tau_y f)(x);$$

the classical pointwise ergodic theorem of Birkhoff says that if $E_i = [0, i) \subset \mathbb{Z}^1$, then the one-dimensional averages $\{M_i f(x)\}$ converge pointwise μ-almost everywhere for $f \in L^p(X, \Sigma, \mu),\ 1 \leq p < \infty$. A standard proof proceeds by way of a density argument, where the key quantitative estimate is that the one-dimensional maximal function

$$f^{*,1} := \sup_i |M_i f|$$

is of weak-type $(1,1)$, and strong-type (p,p), $1 < p \leq \infty$. Explicitly, there exist absolute constants, $C_p,\ 1 \leq p \leq \infty$ so that

$$\lambda \mu \left(\left\{ f^{*,1} > \lambda \right\} \right) \leq C_1 \|f\|_1, \text{ for } \lambda \geq 0, \text{ and}$$
$$\left\| f^{*,1} \right\|_p \leq C_p \|f\|_p,\ 1 < p \leq \infty.$$

This result was generalized to higher dimensions by Wiener; a direct consequence of [22, Theorem II'] is that if $E_i \subset \mathbb{Z}^d$ are a nested, increasing sequence of cubes which contain the origin, then the d-dimensional averages $\{M_i f(x)\}$ converge pointwise μ-almost everywhere, and the d-dimensional maximal function

$$f^{*,d} := \sup_i |M_i f|$$

is similarly of weak-type $(1,1)$ and strong-type (p,p), $1 < p \leq \infty$.

A modern path to Wiener's result is through the transference principle of Calderón [2], which allows one to conduct the study of ergodic averages on the lattice, $\mathbb{Z}^d$: in particular, it is enough to consider the discrete convolution operators A_i, defined by

$$A_i f(n) := \frac{1}{|E_i|} \sum_{m \in E_i} f(n-m).$$

An advantage to this perspective is that real analytic methods – covering lemmas, Fourier analysis and further orthogonality techniques – can be brought to bear in studying more general regions $\{E_i\} \subset \mathbb{Z}^d$ and pertaining the averaging operators $\{A_i\}$. Indeed, provided the collection of sets $\{E_i\} \subset \mathbb{Z}^d$ being studied share some qualitative properties with an increasing collection of cubes [38, §§1-2]:

- A one-parameter structure; and
- Some geometric "smoothness"

the maximal functions on the lattice,

$$\sup_i |A_i f|$$

are of weak-type $(1,1)$ and strong-type $(p,p), 1 < p \leq \infty$.

Obtaining pointwise convergence results of $\{A_i f(n)\}$ for more exotic $\{E_i\} \subset \mathbb{Z}^d$ – those for which the above two properties are relaxed, or absent – 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.

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

Indeed, to prove pointwise convergence of the ergodic averages of L^2-functions along the squares, Bourgain [11] showed that for any lacunarily increasing sequence $\{i_k\}$, the *oscillation operator*,

$$\mathcal{O}f := \left(\sum_k \sup_{i_k < i \leq i_{k+1}} |A_{i_k} f - A_i f|^2 \right)^{1/2}$$

was of strong-type $(2,2)$. The L^2 result then anchored a density argument through which he was able to extend his result to all $p > 1$.

In proving this result, Bourgain made use of the *s-variation operators*, $s > 2$, classically used in probability theory to gain quantitative information on rates of convergence. [1] More precisely, Bourgain proved that, in the special case $E_i = [0, i) \subset \mathbb{Z}^1$, the s-variation operators

$$\mathcal{V}^s f = \mathcal{V}^s_{\{E_i\}} f := \sup_{i_1 < i_2 < \cdots < i_N} \left(\sum_{k=1}^N |A_{i_k} f - A_{i_{k+1}} f|^s \right)^{1/s}$$

were of strong-type $(2,2)$ [11, Corollary 3.26]. Here, the supremum is taken over all finite increasing sequences of integers $\{i_k\}$.

Now, these variation operators are more difficult to control than the maximal function $\sup_i |A_i f|$: for any j, one may pointwise dominate

$$\sup_i |A_i f| \leq \mathcal{V}^\infty f + |A_j f| \leq \mathcal{V}^s f + |A_j f|,$$

where $s < \infty$ is arbitrary. On the other hand, variational (or oscillation) estimates are a

[1] One representative example which in fact appears in Bourgain's argument is Lépingle's inequality for martingales [11].

powerful tool for proving pointwise convergence when a density argument seems unavailable.

Since Bourgain's celebrated result, establishing variational estimates for families of averaging operators has been the focus of much research in ergodic theory. [2]

A fundamental paper in this direction is due to Jones et. al. [10], where it is shown that the one-dimensional variation operators $\{\mathcal{V}^s\}$ are of weak-type $(1,1)$ and strong-type (p,p), $1 < p < \infty$. In other words, the variation operators enjoy the same boundedness properties (up to the endpoint $p = \infty$) as their associated maximal function. The argument in [10] proceeded by first controlling an oscillation operator adapted to $\{E_i\} = \{[0,i)\}$ and then using martingale-style techniques from probability theory; the approach to the oscillation operator itself, however, was driven by Fourier-based orthogonality arguments, initially found in [11].

A subsequent paper of Jones, Rosenblatt, and Wierdl, [8], refined the orthogonality methods used in [10] by eliminating the Fourier-analytic techniques, and more closely following (dyadic) martingale-style arguments. In so doing, the authors were able to establish analogous results in higher-dimensional settings for functions in the "low"-L^p regime, $1 < p \leq 2$.

To continue our discussion, we briefly introduce two representative operators studied in [8]: the "pointwise" square functions. In our current context, the significance of these operators is that establishing L^p bounds leads directly to bounds on the corresponding oscillation and variation operators [8, §1]:

Suppose that $\{E_i\}$ are cubes of side-length i, so that for $2^{k-1} \leq i < 2^k$,

$$E_i \subset H_k := \prod_{i=1}^{d} [-2^k, 2^k).$$

In other words, the $\{E_i\}$ are displaced from the origin by an amount comparable to, or less

[2] Variation estimates have also been studied from a harmonic analysis perspective in the context of truncations of singular integral operators. We refer the reader to [12] for further discussion.

than, their side lengths. Then, with

$$A^k f(x) := \frac{1}{|H_k|} \sum_{m \in H_k} f(n-m),$$

we define the *long* square function (on the integers) with respect the collection (E_i) as

$$\mathcal{S}^L_{\{E_i\}} f(x) := \left(\sum_k \left(\sup_{2^{k-1} \leq i < 2^k} |A_i f(x) - A^k f(x)|^2 \right) \right)^{1/2},$$

and the *short* square function as

$$\mathcal{S}^S_{\{E_i\}} f(x) := \left(\sum_k \left(\sup_{2^{k-1} \leq t_i < 2^k \text{ increasing}} |A_{t_i} f(x) - A_{t_{i+1}} f(x)|^2 \right) \right)^{1/2}.$$

A direct consequence of [8, Theorems A',B'] is the following:

Theorem 2.1.1. *There exist absolute constants C_1, C_p, $1 < p \leq 2$ so that*

$$\left\| \mathcal{S}^*_{\{E_i\}} f(x) \right\|_{L^p(\mathbb{Z}^d)} \leq C_p \|f\|_{L^p(\mathbb{Z}^d)}$$
$$\left\| \mathcal{S}^*_{\{E_i\}} f(x) \right\|_{L^{1,\infty}(\mathbb{Z}^d)} \leq C_1 \|f\|_{L^1(\mathbb{Z}^d)},$$

where $* = L, S$.

The two operators, $\mathcal{S}^L, \mathcal{S}^S$, both measure the scales and locations where f oscillates, i.e. differs from constant functions. The content of these results – and indeed the ideas that drive their proofs – is that there is sufficient orthogonality between the various scales to ensure that, in an average sense, even a *pointwise* measurement of oscillation is controllable by the size of the initial function. We refer the reader to §2, or to [8, §§3-4] for further discussion.

A key insight in [8] is that just as in the case of the maximal function, $\sup_i |A_i f|$, the $\{E_i\}$ do not actually need to be cubes for the pertaining square functions to enjoy the above control. Indeed, the above theorem holds provided the collection of sets $\{E_i\}$ being studied shared the same qualitative properties with a nested collection of cubes as in the case of the maximal function: [3]

- A one-parameter structure; and
- Some geometric "smoothness."

In light of the analogous approaches to studying maximal functions and square functions in our current context, the following informal question seems natural:

> To what (further) extent do the boundedness properties of the square functions under our consideration parallel those of the maximal functions?

More precisely, we organize our study of the operators introduced in [8] according to the following aims:

First, we seek to extend our control of the square functions under the assumptions outlined [8]. An immediate concern is the behavior of the square function in the "high"-$L^p(\mathbb{Z}^d)$ regime, which we investigate by using sharp-function techniques from harmonic analysis. We prove:

Theorem 2.1.2. *If $\mathcal{A} = \{E_i\}$ is* regular*, then the long and short square functions $\mathcal{S}_{\mathcal{A}}^L, \mathcal{S}_{\mathcal{A}}^S$, are $L^p(\mathbb{Z}^d)$-bounded, $2 < p < \infty$.*

We refer the reader to §2 for the precise definition of *regularity*; informally, regular sequences $\{E_i\}$ share the above-mentioned qualitative similarities to nested cubes.

This result leads directly to new control over jump and variation inequalities:

For $\lambda > 0$, we define, as in [8],

$$J((A_i), \lambda) = J((A_i(f, x), \lambda))$$

[3]For a more precise statement of these results, we refer the reader to [8, §1], or to §2 below; for an excellent treatment of continuous analogues of the square functions introduced above, we refer the reader to the discussion of the *intrinsic square function*, found in e.g. [21, §6].

as the largest N for which there are increasing indices (i_j) with

$$|(A_{i_j} - A_{i_{j+1}})f(x)| > \lambda, \ 1 \leq j < N.$$

Then, arguing as in [8], §1, under the assumption of regularity we have the following:

Corollary 2.1.3. *For $2 < p < \infty$ there exist absolute constants C_p so that for any $\lambda > 0$,*

$$\left\| \lambda \cdot J((A_i), \lambda)^{1/2} \right\|_{L^p(\mathbb{Z}^d)} \leq C_p \|f\|_{L^p(\mathbb{Z}^d)}.$$

Moreover, for any $s > 2$,

$$\|\mathcal{V}^s f\|_{L^p(\mathbb{Z}^d)} \leq C_p \|f\|_{L^p(\mathbb{Z}^d)}.$$

We remark that by transference this implies the corresponding result for dynamical systems.

Corollary 2.1.4. *For $2 < p < \infty$ there exist absolute constants C_p so that for any $\lambda > 0$,*

$$\left\| \lambda \cdot J((M_i), \lambda)^{1/2} \right\|_{L^p(X,\mu)} \leq C_p \|f\|_{L^p(X,\mu)}.$$

Moreover, for any $s > 2$,

$$\|\mathcal{V}^s f\|_{L^p(X,\mu)} \leq C_p \|f\|_{L^p(X,\mu)}.$$

To deepen the connection between square and maximal function, we also consider the behavior of square functions on the discrete $A_p(\mathbb{Z}^d)$-weighted classes. We are additionally motivated in this regard by the weighted estimates for one-parameter actions studied in [14], [6], and [35], and by the weighted theory of "rough" singular integrals which satisfy the so-called Hörmander conditions [13], [12].

We first recall a standard characterization of A_p weights; we refer the reader to [4, §7] or to [38, §5] for a more comprehensive treatment. We begin by recalling the discrete uncentered cubic Hardy-Littlewood maximal function.

Definition 2.1.5. *For $f : \mathbb{Z}^d \to \mathbb{C}$, we define the discrete uncentered cubic Hardy-Littlewood maximal function:*

$$M_{HL}f(x) := \sup_{Q \ni x} \frac{1}{|Q|} \int_Q |f|,$$

where the supremum is taken over all cubes Q which contain the point x.

With this operator in hand, we are free to define and introduce some properties of $A_p(\mathbb{Z}^d)$ weights.

Definition 2.1.6. *For $1 < p < \infty$, a positive function $w \in A_p(\mathbb{Z}^d)$ is a* discrete A_p weight *if M_{HL} is bounded on $L^p(\mathbb{Z}^d)$. Explicitly, $w \in A_p(\mathbb{Z}^d)$ if and only if there exists an absolute $C_p > 0$ so that*

$$\int_{\mathbb{Z}^d} |M_{HL}f|^p w \leq C_p^p \int_{\mathbb{Z}^d} |f|^p w \quad \text{or}$$

$$\|M_{HL}f\|_{L^p(\mathbb{Z}^d,w)} \leq C_p \|f\|_{L^p(\mathbb{Z}^d,w)}.$$

A positive function $w \in A_1(\mathbb{Z}^d)$ is an A_1 weight *if there exists a constant $C = C(w)$ so that $M_{HL}w \leq Cw$.*

We say $w \in A_\infty(\mathbb{Z}^d)$ is an A_∞ weight *if $w \in A_p(\mathbb{Z}^d)$ for some finite p.*

We isolate two important properties of $A_1(\mathbb{Z}^d)$ and $A_\infty(\mathbb{Z}^d)$ weights; the proofs of these facts can be found in e.g. [38, §5]:

- A_1 weights $w \in A_1$ satisfy
$$\frac{w(Q)}{|Q|} \leq B \inf_{x \in Q} w(x)$$
for any cube Q, and an absolute $B = B(w)$; and

- A_∞ weights are automatically doubling: $w \in A_\infty$ automatically satisfy
$$w(2Q) \leq Cw(Q)$$
for any cube Q, and an absolute $C = C(w)$. Here $2Q$ denotes the cube concentric with Q, but with twice the side length.

We say that a collection of sets $\{E_i\}$ is *cubic* if the maximal function $\sup_i |A_i f|$ is (up to constant factors) pointwise dominated by $M_{HL} f$. Under this natural assumption, we have the following theorem.

Theorem 2.1.7. *If $\mathcal{A} = \{E_i\}$ is* cubic*, then for both the long and short square functions $\mathcal{S}^*_{\mathcal{A}}$, $* = L, S$ there exist absolute constants $C_p = C_p([w]_{A_p(\mathbb{Z}^d)})$ dependent only on the $A_p(\mathbb{Z}^d)$ norm of w so that*

$$\int_{\mathbb{Z}^d} |\mathcal{S}^*_{\mathcal{A}} f|^p w \leq C_p \int_{\mathbb{Z}^d} |f|^p w$$

for any A_p-weight w, $1 < p < \infty$. Moreover, at the $p = 1$ endpoint, we have the weak-type bound: there exists an absolute constant $C_1 = C_1([w]_{A_1(\mathbb{Z}^d)})$ depending only on the $A_1(\mathbb{Z}^d)$ norm of w so that for each $\lambda \geq 0$

$$\lambda w\{\mathcal{S}^*_{\mathcal{A}} f \geq \lambda\} \leq C_1 \int_{\mathbb{Z}^d} |f| w.$$

Again arguing as in [8, §1], this implies the following corollary for cubic families:

Corollary 2.1.8. *For $1 < p < \infty$ there exist absolute constants C_p so that for any $\lambda > 0$,*

$$\left\| \lambda \cdot J((A_i), \lambda)^{1/2} \right\|_{L^p(\mathbb{Z}^d, w)} \leq C_p \|f\|_{L^p(\mathbb{Z}^d, w)}.$$

Moreover, for any $s > 2$,

$$\|\mathcal{V}^s f\|_{L^p(\mathbb{Z}^d, w)} \leq C_p \|f\|_{L^p(\mathbb{Z}^d, w)}.$$

Of course, by specializing to the weight $v \equiv 1$, we recover the primary results of [8].

Finally, we investigate the behavior of our square functions when the crucial smoothness assumption is relaxed. We focus our efforts in this regard on the following problem.

Problem 2.1.9 ([8], Problem 7.5)**.** *For each collection of nested rectangles $\{E_i\} \subset \mathbb{Z}^d$, is* *it true that for each* $\phi \in L^1(X, \Sigma, \mu)$,

$$\left(\sum_i |(M_i - M_{i+1})\phi|^2 \right)^{1/2} < \infty$$

μ-almost everywhere?

Certainly, this result would be implied by a weak-type bound

$$\left\| \left(\sum_i |(M_i - M_{i+1})\phi|^2 \right)^{1/2} \right\|_{L^{1,\infty}(X)} \lesssim \|\phi\|_{L^1(X)}.$$

In fact, as in shown in the Appendix §6 below, in many cases this weak-type bound is *necessary* for convergence to occur. We therefore focus our attention on the following slightly more general

Problem 2.1.10 ([8], Problem 7.5 – Working Version)**.** *For each collection of nested rectangles* $\{E_i\} \subset \mathbb{Z}^d$*, does there exist a bound*

$$\left\| \left(\sum_i |(A_i - A_{i+1})f|^2 \right)^{1/2} \right\|_{L^{1,\infty}(\mathbb{Z}^d)} \lesssim \|f\|_{L^1(\mathbb{Z}^d)}?$$

Though this problem remains out of reach in its fullest generality, we are able to answer the problem affirmatively under a lacunarity assumption on collection $\{E_i\} \subset \mathbb{Z}^d$.

Proposition 2.1.11. *For each collection of nested rectangles* $\{E_i\} \subset \mathbb{Z}^d$ *with* dyadic *side lengths,*

$$\left\| \left(\sum_i |(A_i - A_{i+1})f|^2 \right)^{1/2} \right\|_{L^{1,\infty}(\mathbb{Z}^d)} \lesssim \|f\|_{L^1(\mathbb{Z}^d)}.$$

The structure of the paper is as follows:

In §2 we introduce relevant definitions, and present a few reductions which will be used throughout;

In §3, we study our square functions' behavior in the high-$L^p(\mathbb{Z}^d)$ regime;

In §4, we prove weighted estimates; and

In §5 we relax the smoothness assumptions of [8], and discuss Problem 2.1.10.

Our appendix, §6, contains a weak-type principle for square functions in the spirit of [18].

2.1.1 Acknowledgements

The author would like to thank Benjamin Hayes and Michael Lacey for helpful conversations, and his advisor, Terence Tao, for his great patience and support.

2.1.2 Notation

We shall, whenever possible, maintain the notation introduced in [8].

For a set $E \subset \mathbb{Z}^d$, we use $|E|$ to denote the cardinality of the set E.

We let 1_E denote the indicator function of the set E, i.e.

$$1_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E. \end{cases}$$

We let $\chi_E := \frac{1}{|E|} \cdot 1_E$ denote the $L^1(\mathbb{Z}^d)$-normalized indicator function.

For a function f defined on $\mathbb{Z}^d$, our convention will be to let

$$\int f$$

denote the summation $\sum_{n \in \mathbb{Z}^d} f(n)$. Accordingly, we will use

$$\|f\|_{L^p(\mathbb{Z}^d)}$$

to denote the $L^p(\mathbb{Z}^d)$-norm, $(\sum_{n \in \mathbb{Z}^d} |f(n)|^p)^{1/p}$, with the obvious modification for $p = \infty$.

When integrating over other spaces, we will include the domain and measures.

We will make use of the modified Vinogradov notation. We use $X \lesssim Y$, or $Y \gtrsim X$ to denote the estimate $X \leq CY$ for an absolute constant C. If we need C to depend on a parameter, we shall indicate this by subscripts, thus for instance $X \lesssim_p Y$ denotes the estimate $X \leq C_p Y$ for some C_p depending on p. We use $X \approx Y$ as shorthand for $X \lesssim Y \lesssim X$.

2.2 Preliminaries

Let $\mathcal{R} = 0 \subset R_1 \subset R_2 \subset \dots$ denote a nested sequence of dyadic rectangles inside $\mathbb{Z}^d$, i.e.

$$R_k = \prod_{a=1}^{d} [0, 2^{a(k)});$$

here $\{a(k)\}$ are non-decreasing, and $a'(k) < a'(k+1)$ for at least one value $1 \leq a' \leq d$.

We also define the "symmetric" rectangles

$$H_k = \prod_{a=1}^{d} [-2^{a(k)}, 2^{a(k)}).$$

For $k \geq 0$, we let $\sigma_k = \sigma_k(\mathcal{R})$ denote the σ-algebra generated by R_k, i.e. the σ-algebra with atoms

$$\prod_{a=1}^{d} 2^{a(k)}[m, m+1).$$

Henceforth, we let $\mathcal{E}_k$ denote the expectation with respect to σ_k, $\mathcal{E}_0$ the identity operator,

$$\Delta_0 = \mathcal{E}_0 \text{ and } \Delta_k := \mathcal{E}_k - \mathcal{E}_{k-1}, \; k \geq 1$$

denote the martingale differences.

Closely connected to our family of rectangles, $\mathcal{R}$, are the collections of sets $\mathcal{A} = \{\mathcal{A}_k\}$, whose elements have controlled

1. Spatial location: for every $E \in \mathcal{A}_k$, $E \subset H_k$;

2. Size: for each $E \in \mathcal{A}_k$ there exists some $c > 0$ so that $c|H_k| \leq |E|$;

3. (Internal) Smoothness: for each $E \in \mathcal{A}_{k+l}$, for $l \geq 0$

$$\frac{|\{x : \partial E \cap (H_k - x) \neq \emptyset\}|}{|H_{k+l}|} =: \frac{B(E, H_k)}{|H_{k+l}|} \leq \iota(l)^2,$$

with $\sum_l \iota(l) < \infty$; and finally

4. Eccentricity:

$$\frac{B(H_{k+l}, H_k)}{|H_{k+l}|} \leq \varepsilon(l)^2,$$

where $\sum_l \varepsilon(l) < \infty$ as well.

We shall collectively refer to the above four criteria as the *JRW criteria*, and use the notation $\mathcal{A} = \{\mathcal{A}_k\}$ to denote the various collections.

The third of the above points forces regularity on the boundaries of the $E \in \mathcal{A}_k$, i.e. some smoothness on χ_E.

The fourth point is implicitly used in the proofs of the main theorems of [8], though not explicitly stated in the summary of [8, §5]. We shall replace it with the following equivalent formulation, which we isolate in the form of the following simple

Lemma 2.2.1. *Eccentricity control as above is equivalent to the existence of an L so that*

$$\min_{1 \leq a \leq d} \{a(k+L) - a(k)\} \geq 1$$

for each k, with $H_k = \prod_{a=1}^d [-2^{a(k)}), 2^{a(k)})$.

Moreover, the existence of such an L allows us to take

$$\varepsilon(l)^2 \lesssim 2^{-\frac{l}{L}},$$

and therefore

$$\sum_l \varepsilon(l)^\theta < \infty, \textit{ for any } \theta > 0.$$

We say that two cubes $Q = \prod_{i=1}^d I_i,\ R = \prod_{i=1}^d J_i, |I_i| \geq |J_i|$ with dyadic side-lengths are *s-separated* if

$$\min_{1 \leq i \leq d} \left\{ \frac{|J_i|}{|I_i|} \right\} = 2^{-s};$$

the content of the above theorem is that there exists an absolute L so that H_{k+L}, H_k are 1-separated for each k.

Proof. If $s = \min_{1 \leq i \leq d}\{a(k+l) - a(k)\}$, then

$$\frac{B(H_{k+l}, H_k)}{|H_{k+l}|} \approx 2^{-s};$$

if no such L were to exist, then we could find arbitrarily many $l, k = k(l)$ so that

$$\frac{B(H_{k+l}, H_k)}{|H_{k+l}|} \gtrsim 1,$$

which would force the sum $\sum_l \varepsilon(l)$ to diverge. □

For our purposes, the (alternate) eccentricity condition guarantees that for any dyadic $2^c, c \geq 1$,

$$\frac{B(H_{k+l}, 2^c H_k)}{|H_{k+l}|} \lesssim 2^{-\frac{l}{cL}}.$$

We shall be studying families of sets, $\{E_t\}$, which are *regular* with respect to our collections $\mathcal{A} = \{\mathcal{A}_k\}$: for each $2^k \leq t \leq t' < 2^{k+1}$,

$$E_t \subset E_{t'}, \text{ and } E_t, E_{t'} \in \mathcal{A}_k.$$

If in addition $\mathcal{A} = \{\mathcal{A}_k\}$ satisfy the above JRW criteria, we will say that $\mathcal{A}$ itself is *regular*.

With a regular collection $\mathcal{A} = \{\mathcal{A}_k\}$ and $\{E_t\}$ specified, we will let

$$\chi_t = \chi_t^{\mathcal{A}} := \frac{1}{|E_t|} 1_{E_t}$$

denote the $L^1(\mathbb{Z}^d)$-normalized indicator function.

We let

$$A_t g(x) = A_t^{\mathcal{A}} g(x) := \chi_t * g(x) = \frac{1}{|E_t|} \sum_{y \in E_t} g(x-y)$$

denote the convolution operator with kernel χ_t. We introduce the maximal function associated to $\mathcal{A}$

$$Mf = M_{\mathcal{A}} f := \sup_k \chi_{5H_k} * |f|(x),$$

which dominates $\sup_t A_t|f|$, and satisfies a weak-type $(1,1)$ inequality by the nesting properties of the $\{H_k\}$. (cf. e.g. [20, Lemma 5.3]).

With $\{E_t\}$ regular with respect to $\mathcal{A}$, we define the long square function

$$\mathcal{S}_{\mathcal{A}}^L f := \left(\sum_k |\mathcal{S}_{\mathcal{A},k}^L f|^2 \right)^{1/2} := \left(\sum_k \sup_{2^k \leq t < 2^{k+1}} |A_t f - \mathcal{E}_k f|^2 \right)^{1/2},$$

and short square function

$$\mathcal{S}_{\mathcal{A}}^S f := \left(\sum_k |\mathcal{S}_{\mathcal{A},k}^S f|^2 \right)^{1/2} := \left(\sum_k \sup_{t_i \text{ increasing}} \sum_{2^k \leq t_i < 2^{k+1}} |A_{t_i} f - A_{t_{i+1}} f|^2 \right)^{1/2}.$$

Often, we will suppress the subscript $\mathcal{A}$.

We briefly remark that that for $2^k \leq t < 2^{k+1}$, since

$$|A_t f|(x) \lesssim \chi_{H_k} * |f|(x) \text{ and } |\mathcal{E}_k f|(x) \lesssim \chi_{H_k} * |f|(x),$$

we may control $\mathcal{S}_k^L f \lesssim \chi_{H_k} * |f|$. We have similar control over $\mathcal{S}_k^S f$. Using the domination of the l^2 norm by the l^1 norm, and then the triangle inequality, we may bound

$$\begin{aligned}
\mathcal{S}_k^S f &:= \sup_{t_i \text{ increasing}} \left(\sum_{2^k \leq t_i < 2^{k+1}} |(A_{t_i} - A_{t_{i+1}}) f|^2 \right)^{1/2} (x) \\
&\leq \sum_{2^k \leq t < 2^{k+1}} |(A_t - A_{t+1}) f|(x) \\
&\leq \sum_{2^k \leq t < 2^{k+1}} \left(\left(\frac{1}{|E_t|} - \frac{1}{|E_{t+1}|} \right) 1_{E_t} * |f| \right) + \sum_{2^k \leq t < 2^{k+1}} \left(\frac{1}{|E_{t+1}|} 1_{E_{t+1} \smallsetminus E_t} * |f| \right) \\
&\lesssim \chi_{H_k} * |f|,
\end{aligned}$$

where we used the size control $c|H_k| \leq |E_t|$ in the final inequality. In particular, we may control each

$$\mathcal{S}_k^* f \lesssim Mf.$$

We also introduce the larger, shifted square functions:

$$\tilde{\mathcal{S}}_{\mathcal{A}}^* f(x) := \left(\sum_k |\tilde{\mathcal{S}}_{\mathcal{A},k}^* f(x)|^2 \right)^{1/2} := \left(\sum_k |\sup_{v \in H_k} \mathcal{S}_{\mathcal{A},k}^* f(x+v)|^2 \right)^{1/2},$$

$* = L, S$. The additional suprema affords the shifted square functions a useful degree of smoothness:

Lemma 2.2.2. *With x_Q denoting the center of each $Q \in \sigma_k$, $\tilde{\mathcal{S}}_k^* f$ is pointwise dominated, and $L^p(\mathbb{Z}^d)$-comparable $1 \leq p \leq \infty$, to a function which is constant on $Q \in \sigma_k$:*

$$\mathcal{S}_{D,k}^* f := \sum_{Q \in \sigma_k} \tilde{\mathcal{S}}_k^* f(x_Q) 1_{3Q}.$$

In particular,

$$\left\| \tilde{\mathcal{S}}_k^* f(x) \right\|_{L^p(\mathbb{Z}^d)} \approx_p \left\| \mathcal{S}_{D,k}^* f \right\|_{L^p(\mathbb{Z}^d)}.$$

Proof. If $Q_i \in \sigma_k$ lie in $3Q$, then for any $x \in Q$, we may bound

$$\tilde{\mathcal{S}}_k^* f(x) \leq \sum_i \tilde{\mathcal{S}}_k^* f(x_{Q_i}).$$

Indeed, if $v \in H_k$ is such that

$$\tilde{\mathcal{S}}_k^* f(x) = \mathcal{S}_k^* f(x+v),$$

then we may write $x+v = x_{Q_i} + v_{Q_i}$ for some center x_{Q_i} of a neighboring cube $Q_i \in 3Q$, and some (possibly different) $v_{Q_i} \in H_k$. Summing over all $Q \in \sigma_k$ therefore yields a pointwise majorization

$$\tilde{\mathcal{S}}_k^* f(x) \lesssim \sum_Q \tilde{\mathcal{S}}_k^* f(x_Q) 1_{3Q}.$$

On the other hand, if $v_Q \in H_k$ is such that

$$\tilde{\mathcal{S}}_k^* f(x_Q) = \mathcal{S}_k^* f(x_Q + v_Q),$$

then on the set

$$X(Q) = \{y \in Q : y + v = x_Q + v_Q \text{ for some } v \in H_k\} = Q \cap \{x_Q + v_Q - H_k\},$$

which has measure $|X(Q)| \gtrsim_d |Q|$, we may bound

$$\tilde{\mathcal{S}}_k^* f(x_Q) \leq \inf_{y \in X(Q)} \tilde{\mathcal{S}}_k^* f(y),$$

by definition of the set $X(Q)$. We therefore have a similar pointwise inequality for $p = \infty$, while for $1 \leq p < \infty$, using the finite overlap of $\{3Q\}$ to estimate pointwise

$$\left| \sum_Q \tilde{\mathcal{S}}_k^* f(x_Q) 1_{3Q} \right|^p \lesssim_p \sum_Q \tilde{\mathcal{S}}_k^* f(x_Q)^p 1_{3Q},$$

we estimate

$$\begin{aligned}
\left\|\sum_Q \tilde{\mathcal{S}}_k^* f(x_Q) 1_{3Q}\right\|_{L^p(\mathbb{Z}^d)}^p &\lesssim \sum_Q |\tilde{\mathcal{S}}_k^* f(x_Q)|^p |3Q| \\
&\lesssim \sum_Q |\tilde{\mathcal{S}}_k^* f(x_Q)|^p |X(Q)| \\
&\leq \sum_Q \int_{X(Q)} |\tilde{\mathcal{S}}_k^* f(y)|^p \\
&\leq \sum_Q \int_Q |\tilde{\mathcal{S}}_k^* f(y)|^p \\
&= \left\|\tilde{\mathcal{S}}_k^* f\right\|_{L^p(\mathbb{Z}^d)}^p .
\end{aligned}$$

□

For future use, we record the following additional

Lemma 2.2.3. *For any $v \in A_\infty$,*

$$\mathcal{S}_{D,k}^* f := \sum_{Q \in \sigma_k} \tilde{\mathcal{S}}_k^* f(x_Q) 1_{3Q}$$

is $L^p(v)$-comparable, $1 \leq p < \infty$ to

$$\mathcal{S}_{d,k}^* f := \sum_{Q \in \sigma_k} \tilde{\mathcal{S}}_k^* f(x_Q) 1_Q.$$

Proof. Using again the bounded overlap of $\{3Q\}$, one estimates

$$\begin{aligned}
\|\mathcal{S}_{D,k}^* f\|_{L^p(v)}^p &\lesssim \int \sum_Q |\tilde{\mathcal{S}}_k^* f(x_Q)|^p 1_{3Q} v \\
&\lesssim \sum_Q |\tilde{\mathcal{S}}_k^* f(x_Q)|^p v(3Q) \\
&\lesssim \sum_Q |\tilde{\mathcal{S}}_k^* f(x_Q)|^p v(Q) \\
&= \|\mathcal{S}_{d,k}^* f\|_{L^p(v)}^p,
\end{aligned}$$

where we used the doubling nature of $v \in A_\infty$ in passing to the third line. □

Though – as we shall see – more is true, for now we shall only need that our shifted square functions inherit L^2-boundedness from their centered associates:

Proposition 2.2.4. $\left\|\tilde{\mathcal{S}}^*_{\mathcal{A}} f\right\|_{L^2(\mathbb{Z}^d)} \lesssim \|f\|_{L^2(\mathbb{Z}^d)}$.

The argument here is very similar to the arguments of [8, §3]. The qualitative similarities between $\tilde{\mathcal{S}}^*_k$ and the projection operators $\Delta_k = \mathcal{E}_k - \mathcal{E}_{k-1}$ – informally, both measure the locations where f differs from being constant at $\approx \sigma_k$-scale – motivate the following orthogonality approach:

Proof. With

$$f = \sum_{k \geq 0} \Delta_k(f) =: \sum_{k \geq 0} d_k$$

we majorize

$$\tilde{\mathcal{S}}^* f = (\sum_n |\tilde{\mathcal{S}}^*_n(\sum_k d_k)|^2)^{1/2} \leq \sum_{j \in \mathbb{Z}} (\sum_n |\tilde{\mathcal{S}}^*_n(d_{n+j})|^2)^{1/2},$$

defining $d_t = 0$ for $t < 0$.

Since $\tilde{\mathcal{S}}^*_n(g) \lesssim Mg$ by previous discussion, we may ignore the $L^2(\mathbb{Z}^d)$-contribution of each of the $|j| \leq C = C(L)$ square functions:

$$\begin{aligned}
\sum_{|j|\leq C} \left\| \sum_n |\tilde{\mathcal{S}}_n^*(d_{n+j})|^2)^{1/2} \right\|_{L^2(\mathbb{Z}^d)} &\lesssim \sum_{|j|\leq C} \left\| (\sum_n |\tilde{\mathcal{S}}_n^*(d_{n+j})|^2)^{1/2} \right\|_{L^2(\mathbb{Z}^d)} \\
&\lesssim \sum_{|j|\leq C} \left\| (\sum_n |M d_{n+j}|^2)^{1/2} \right\|_{L^2(\mathbb{Z}^d)} \\
&= \sum_{|j|\leq C} \left(\sum_n \|M d_{n+j}\|^2_{L^2(\mathbb{Z}^d)} \right)^{1/2} \\
&\lesssim \sum_{|j|\leq C} \left(\sum_n \|d_{n+j}\|^2_{L^2(\mathbb{Z}^d)} \right)^{1/2} \\
&= \sum_{|j|\leq C} \left\| (\sum_n |d_{n+j}|^2)^{1/2} \right\|_{L^2(\mathbb{Z}^d)} \\
&= \sum_{|j|\leq C} \|f\|_{L^2(\mathbb{Z}^d)} \\
&\lesssim \|f\|_{L^2(\mathbb{Z}^d)},
\end{aligned}$$

where we used the $L^2(\mathbb{Z}^d)$-boundedness of M in the fourth line.

For $j > C$, let $Q \in \sigma_{n+j-1}$ be arbitrary, and consider $\tilde{\mathcal{S}}_n^* d_{n+j}(x)$ on Q. Since $j > C > 0$,

$$\tilde{\mathcal{S}}_n^* d_{n+j}(x) = \tilde{\mathcal{S}}_n^*(d_{n+j} 1_{5Q})(x),$$

we may bound

$$|\tilde{\mathcal{S}}_n^* d_{n+j}(x)| \lesssim \max_{x \in 5Q} |d_{n+j}|(x) \leq (\sum_{Q_i} |d_{n+j}|^2(x_{Q_i}))^{1/2},$$

where $Q_i \in \sigma_{n+j-1}$ lie in $5Q$, and x_{Q_i} is the center of each such Q_i; we note that d_{n+j} is constant-valued on $Q \in \sigma_{n+j-1}$. Let us call this constant value $c = c(Q)$. Then, for every $x \in Q \in \sigma_{n+j-1}$ such that $x + 2H_k \subset Q$,

$$A_t d_{n+j}(x) = c,$$

hence $\tilde{\mathcal{S}}_n^* d_{n+j}(x) = 0$ whenever $x + 2H_n \subset Q$. In particular, $\tilde{\mathcal{S}}_n^* d_{n+j}$ is supported inside

$$\{x : x + 2H_n \cap \partial Q \neq \emptyset\}.$$

We may consequently estimate

$$\begin{aligned}
\sum_{Q \in \sigma_{n+j-1}} \int_Q |\tilde{\mathcal{S}}_n^* d_{n+j}|^2(x) &\leq \sum_{Q \in \sigma_{n+j-1}} \int \sum_{Q_i} |d_{n+j}|^2(x_{Q_i}) \cdot 1_{B(Q,2H_n)} \\
&\lesssim \sum_{Q \in \sigma_{n+j-1}} \int |d_{n+j}|^2(x) 1_{B(Q,2H_n)} \\
&= \sum_{Q \in \sigma_{n+j-1}} |d_{n+j}|^2(x_Q)\ |B(Q,2H_n)| \\
&\leq \varepsilon(j-1-2L)^2 \sum_{Q \in \sigma_{n+j-1}} |d_{n+j}|^2(x_Q)\ |Q| \\
&= \varepsilon(j-1-2L)^2 \|d_{n+j}\|^2_{L^2(\mathbb{Z}^d)},
\end{aligned}$$

where L is the cost of separating scales, as in the alternative characterization of the eccentricity JRW-criterion.

Using the orthogonality of $\{d_{n+j}\}$ and summing over n exhibits

$$\left\| (\sum_n |\tilde{\mathcal{S}}_n^*(d_{n+j})|^2)^{1/2} \right\|^2_{L^2(\mathbb{Z}^d)} \leq \varepsilon(j-1-2L)^2 \|f\|_2^2,$$

and summing over $j > C$ shows

$$\sum_j \left\| (\sum_n |\tilde{\mathcal{S}}_n^*(d_{n+j})|^2)^{1/2} \right\|_{L^2(\mathbb{Z}^d)} \leq \sum_{j \geq C-2L} \varepsilon(j)^2 \|f\|_{L^2(\mathbb{Z}^d)} \lesssim \|f\|_{L^2(\mathbb{Z}^d)}.$$

Next, for $j < -C$ we establish the pointwise inequality

$$|\tilde{\mathcal{S}}_n^*(d_{n+j})|^2(x) \lesssim \iota^2(j) \cdot M d_{n+j}(x)^2,$$

where $\iota(j)$ appears as the quantitative measure of smoothness in the JRW-criterion.

Summing over n, j as above will then yield the desired bound.

To do so, for $Q \in \sigma_{n+j}$, we first observe that for any $E_i \in \mathcal{A}_n$, any $v \in H_n$, we may bound

$$\begin{aligned} |A_i d_{n+j}(x+v)|^2 &= \left| \frac{1}{|E_i|} \sum_{y \in x+v-E_i} d_{n+j}(y) \right|^2 \lesssim \left(\frac{1}{|H_n|} \right)^2 \left| \sum_{y \in x+v-E_i} d_{n+j}(y) \right|^2 \\ &= \left(\frac{1}{|H_n|} \right)^2 \left| \sum_{y \in x+v-E_i} \sum_{Q: Q \cap \partial(x+v-E_i)} d_{n+j} 1_Q(y) \right| \\ &\le \left(\frac{1}{|H_n|} \right)^2 |B(E_i, R_{n+j})| \cdot \sum_{y \in x+v-E_i} |d_{n+j}(y)|^2 \\ &\le \iota^2(j) \cdot \frac{1}{|H_n|} 1_{H_n} * |d_{n+j}|^2(x+v) \\ &\lesssim \iota^2(j) \cdot \frac{1}{|2H_n|} 1_{2H_n} * |d_{n+j}|^2(x) \\ &\le \iota^2(j) \cdot M d_{n+j}(x)^2. \end{aligned}$$

Since $\mathcal{E}_n d_{n+j} \equiv 0$, this immediately yields the result for the long variation.

For the short variation, the proof of [8, Theorem B] leads to the bound

$$|\mathcal{S}_n^S d_{n+j}(y)|^2 \lesssim \iota(j)^2 \cdot \chi_{H_k} * d_{n+j}^2(y).$$

Substituting $y = x + v$ where $v = v(x) \in H_k$ is such that

$$\tilde{\mathcal{S}}_n^S d_{n+j}(x) = \mathcal{S}_n^S d_{n+j}(x+v)$$

yields

$$|\tilde{\mathcal{S}}_n^S d_{n+j}(x)|^2 \lesssim \iota(j)^2 \cdot \chi_{H_k} * d_{n+j}^2(x+v) \lesssim \iota(j)^2 \cdot \chi_{4H_k} * d_{n+j}^2(x).$$

Integrating this estimate, then summing over n yields

$$\sum_n \int |\tilde{\mathcal{S}}_n^S d_{n+j}(x)|^2 \lesssim \iota(j)^2 \cdot \sum_n \|d_{n+j}\|_{L^2(\mathbb{Z}^d)}^2 = \iota(j)^2 \|f\|_{L^2(\mathbb{Z}^d)}^2.$$

A final sum over $j < -C$ shows

$$\sum_j \left\| (\sum_n |\tilde{\mathcal{S}}_n^*(d_{n+j})|^2)^{1/2} \right\|_{L^2(\mathbb{Z}^d)} \leq \sum_{j<-C} \iota(j) \cdot \|f\|_{L^2(\mathbb{Z}^d)} \lesssim \|f\|_{L^2(\mathbb{Z}^d)}.$$

□

The corollary below is a direct consequence of the previous propositions:

Corollary 2.2.5. *The discretized square functions*

$$\mathcal{S}_d^* f := \left(\sum_k |\mathcal{S}_{d,k}^* f|^2 \right)^{1/2},$$

$$\mathcal{S}_D^* f := \left(\sum_k |\mathcal{S}_{D,k}^* f|^2 \right)^{1/2}$$

$* = L, S$ *are* L^2 *bounded.*

2.3 High-$L^p(\mathbb{Z}^d)$ Estimates

In this section we prove Theorem 2.1.2 by way of the following

Proposition 2.3.1. *If* $\mathcal{A}$ *is regular, then* $\mathcal{S}_d^* f$, $* = L, S$ *is* $L^p(\mathbb{Z}^d)$*-bounded,* $2 < p < \infty$.

We refine our (reverse) filtration $\{\sigma_k\}_k$ to $\{\tau_j\}_j$, where $\tau_{j(k)} = \sigma_k$, and so successive atoms differ in size by a factor of 2: if

$$R_j = \prod_{a=1}^d [0, 2^{a(j)}) \in \tau_j$$

are generating atoms, with "symmetrized" rectangles $H_j' = \prod_{a=1}^d [-2^{a(j)}, 2^{a(j)})$, then we have

$$\sum_{a=1}^d |a(k+1) - a(k)| = 1.$$

We define the maximal operators

$$\mathcal{M}f(x) := \sup_j |\mathcal{E}'_j f|,$$

$$M'f(x) := \sup_j \sup_{x \in y+5H'_j} \frac{1}{|5H'_j|} \int_{y+5H'_j} |f|,$$

where $\mathcal{E}'_j$ is the expectation with respect to τ_j (so $\mathcal{E}'_{j(k)} = \mathcal{E}_k$). We also define the sharp function associated to our refined filtration

$$\mathcal{M}^{\#} f(x) := \sup_j \sup_{x \in R \in \tau_j} \inf_a \frac{1}{|R|} \int_R |f - a|.$$

We certainly have that $\mathcal{M}^{\#} f \leq \mathcal{M}f$; the familiar good-$\lambda$ inequality

$$|\{\mathcal{M}f > 2\lambda, \mathcal{M}^{\#} f \leq \gamma\lambda\}| \lesssim \gamma |\{\mathcal{M}f > \lambda\}|$$

holds (the implicit constant is in fact 2), and by integrating distribution functions we see that the sharp function controls the maximal function in $L^p(\mathbb{Z}^d)$.

The key result we need is the following:

Lemma 2.3.2.

$$\mathcal{M}^{\#}(\mathcal{S}_d^* f^2)(x) \lesssim M'(f^2)(x).$$

Remark 2.3.3. *Informally, $\mathcal{S}_d^* f$ is in "dyadic" BMO with respect to the filtration $\{\tau_j\}$ whenever $f \in L^\infty(\mathbb{Z}^d)$.*

Assuming this lemma, with $p = 2r > 2$, we will have

$$
\begin{aligned}
\|\mathcal{S}_d^* f\|_{L^p(\mathbb{Z}^d)}^2 &= \left\|\mathcal{S}_d^* f^2\right\|_{L^r(\mathbb{Z}^d)} \\
&\leq \left\|\mathcal{M}(\mathcal{S}_d^* f^2)\right\|_{L^r(\mathbb{Z}^d)} \\
&\lesssim \left\|\mathcal{M}^{\#}(\mathcal{S}_d^* f^2)\right\|_{L^r(\mathbb{Z}^d)} \\
&\lesssim \|M'(f^2)\|_{L^r(\mathbb{Z}^d)} \\
&\lesssim \|f^2\|_{L^r(\mathbb{Z}^d)} \\
&= \|f\|_{L^p(\mathbb{Z}^d)}^2,
\end{aligned}
$$

proving the proposition.

Proof of Lemma 2.3.2. Fix some $x \in \mathbb{Z}^d$, and let $x \in R \in \tau_j, j(k') < j \leq j(k'+1)$ be arbitrary. Express

$$
\begin{aligned}
\mathcal{S}_d^* f(x)^2 &= \sum_{k \leq k'} |\mathcal{S}_{d,k}^* f|^2 + \sum_{k > k'} |\mathcal{S}_{d,k}^* f|^2 \\
&= \sum_{k \leq k'} \sum_{Q \in \sigma_k} \tilde{\mathcal{S}}_k^* f(x_Q)^2 1_Q + \sum_{k > k'} \sum_{Q \in \sigma_k} \tilde{\mathcal{S}}_k^* f(x_Q)^2 1_Q \\
&= \sum_{k \leq k'} \sum_{Q \in \sigma_k} \tilde{\mathcal{S}}_k^* (f 1_{5R})(x_Q)^2 1_Q + c_R(f),
\end{aligned}
$$

by the nesting properties $R \supset Q \in \sigma_k$ for $k \leq k'$, $R \subset Q \in \sigma_k$ for $k > k'$ for $x \in Q$.

We bound

$$
\begin{aligned}
\frac{1}{|R|} \int_R |\mathcal{S}_d^* f(x)^2 - c_R(f)| &= \frac{1}{|R|} \int_R |\sum_{k \leq k'} \sum_{Q \in \sigma_k} \tilde{\mathcal{S}}_k^* (f 1_{5R})(x_Q)^2 1_Q| \\
&\leq \frac{1}{|R|} \int |\sum_{k \leq k'} \sum_{Q \in \sigma_k} \tilde{\mathcal{S}}_k^* (f 1_{5R})(x_Q)^2 1_Q| \\
&\leq \frac{1}{|R|} \|\mathcal{S}_d^* (f 1_{5R})\|_{L^2(\mathbb{Z}^d)}^2 \\
&\lesssim \frac{1}{|5R|} \int |f 1_{5R}|^2 \\
&\lesssim M'(f^2)(x),
\end{aligned}
$$

which leads to

$$\inf_a \frac{1}{|R|}\int_R |\mathcal{S}_d^* f(x)^2 - a| \leq M'(f^2)(x).$$

The lemma is proven by taking a final supremum over pertaining R. $\square$

We now transfer this result back to $\mathcal{S}_D^* f$:

Corollary 2.3.4. *For each $2 < p < \infty$, $\mathcal{S}_D^* f$ – the pointwise majorants of our shifted square functions – are bounded on $L^p(\mathbb{Z}^d)$.*

Proof. Let $p = 2r > 2$, and r' denote the dual exponent to $p/2 = r$. For an appropriate non-negative function $w \geq 0, \|w\|_{L^{r'}} = 1$ we estimate

$$\begin{aligned} \|\mathcal{S}_D^* f\|_p^2 &= \||\mathcal{S}_D^* f|^2\|_r \\ &= \sum_k \int |\mathcal{S}_{D,k}^* f|^2 w \\ &\leq \sum_k \int |\mathcal{S}_{D,k}^* f|^2 M_{HL,t} w, \end{aligned}$$

where

$$M_{HL,t} w := (M_{HL}(w^t))^{1/t}$$

for $1 < t < r'$. By e.g. [38, §V.6.15], we know that $M_{HL,t} w \in A_1 \subset A_\infty$, so that we have

$$\int |\mathcal{S}_{D,k}^* f|^2 M_{HL,t} w \lesssim \int |\mathcal{S}_{d,k}^* f|^2 M_{HL,t} w$$

for each k; summing appropriately yields

$$\begin{aligned} \|\mathcal{S}_D^* f\|_p^2 &\lesssim \int |\mathcal{S}_d^* f|^2 M_{HL,t} w \\ &\leq \||\mathcal{S}_d^* f|^2\|_r \|M_{HL,t} w\|_{r'} \\ &= \|\mathcal{S}_d^* f\|_p^2 \|M_{HL,t} w\|_{r'} \\ &\lesssim \|f\|_p^2, \end{aligned}$$

since $M_{HL,t}$ maps $L^{r'}$ to itself any $r' > t$. $\square$

As corollaries, we are now able to affirmatively answer the following problems.

Problem 2.3.5 ([8], Problem 7.1). *Let $D_1 \subset D_2 \subset \cdots \subset \mathbb{Z}^d$ be a nested sequence of (closed) disks (without loss of generality containing the origin) and let $p > 2$. If*

$$A_i f(n) := \frac{1}{|D_i|} \sum_{m \in D_i} f(n-m)$$

are convolution operators, is it true that the square function

$$\mathcal{S}_d f := (\sum_i |(A_i - A_{i+1})f|^2)^{1/2}$$

is bounded on L^p – and thus

$$\mathcal{S}_{abstract,d} := (\sum_i |(M_i - M_{i+1})f|^2)^{1/2}$$

is finite μ-a.e.?

Problem 2.3.6 ([10] Question 4.7). *Let $\mathcal{E}_k f$ be the usual dyadic martingale on $[0,1)$, and let $D_k f(x) = 2^k \int_{I_k(x)} f(t)\ dt$, where $I_k(x)$ is a measurably chosen interval of length 2^{-k} which contains the point x. Does the square function*

$$\mathcal{S}f(x) = (\sum_{k=0}^{\infty} |D_k f(x) - \mathcal{E}_k f(x)|^2)^{1/2}$$

map $L^p \to L^p$ for all $2 \le p < \infty$?

2.4 Weighted Estimates

2.4.1 Preliminaries

We continue to make use of the refined our (reverse) filtration $\{\tau_j\}_j$, where $\tau_{j(k)} = \sigma_k$, and the maximal operators

$$\mathcal{M}f(x) := \sup_j |\mathcal{E}'_j f|,$$
$$M'f(x) := \sup_j \sup_{x \in y+5H'_j} \frac{1}{|5H'_j|} \int_{y+5H'_j} |f|,$$

were introduced in the previous section.

Throughout this section, we will assume that $M'f \lesssim M_{HL}f$ pointwise, where we recall M_{HL} is the uncentered, cubic Hardy-Little maximal function. This condition forces additional smoothness on the *refined* collection $\{H'_j = \prod_{a=1}^d [-2^{a(j)}, 2^{a(j)})\}$:

$$\sup_j \max\{|a(j) - a'(j)| : 1 \leq a, a' \leq d\} \leq K$$

for some absolute $K = O(1)$. We shall call such families $\mathcal{A} = \{E_i\}$ *cubic.*

Remark 2.4.1. Cubicity *is a strictly stronger statement than the previous control over the eccentricities of the* $\{H_k\}$*, as seen for instance by considering*

$$\{H_k = [-2^{2^k}, 2^{2^k}) \times [-2^k, 2^k)\}$$

inside $\mathbb{Z}^2$*. Indeed, one may take* $\varepsilon(l)^2 \lesssim 2^{-l}$*, but the maximal function associated to the* $\{H_k\}$ *is pointwise incomparable to* M_{HLW}*.*

In this section, we will leverage especially strong weighted $L^2(\mathbb{Z}^d)$ bounds on our (sublinear) square functions to prove certain weighted weak-type $(1,1)$ estimates. The key tool used to achieve such lowering of exponents is the following classical decomposition technique due to Calderón and Zygmund.

Lemma 2.4.2 (Calderón-Zygmund Decomposition). *For any altitude $\lambda > 0$, any $f \in L^1(\mathbb{Z}^d)$ may be decomposed in the form*

$$f = g + b = g + \sum_P b_P,$$

where the sum runs over certain disjoint selected cubes $P \in \bigcup_j \tau_j$. g, the so-called "good" function, satisfies

1. $\|g\|_{L^\infty(\mathbb{Z}^d)} \lesssim \lambda$; *and*

2. $\|g\|_{L^1(\mathbb{Z}^d)} \leq \|f\|_{L^1(\mathbb{Z}^d)}$.

The "bad" function, $b = \sum_P b_P$, satisfies the following properties:

1. *Each b_P is supported inside P;*

2. $\int b_P = 0$;

3. $\|b_P\|_{L^1(\mathbb{Z}^d)} \lesssim \lambda |P|$;

4. $\sum_P |P| \leq \frac{1}{\lambda} \|f\|_{L^1(\mathbb{Z}^d)}$.

With this in hand, we are ready to turn to our main results.

2.4.2 Weighted Estimates

Following the approach of [21, §6], we prove the following weighted results:

Proposition 2.4.3 (Weighted L^2). *For any non-negative function $w \geq 0$, there exists an absolute constant $C_2 = C_2(\{\iota(l)\}, \{\varepsilon(l)\})$ so that*

$$\int |\mathcal{S}_D^* f|^2 w \leq C_2 \int |f|^2 M_{HL} w.$$

Consequently, for any A_1 weight $v \in A_1$, we have

$$\int |\mathcal{S}_D^* f|^2 v \leq C_2 \int |f|^2 v.$$

Proposition 2.4.4 (Weighted $L^{1,\infty}$). *For any $v \in A_1$, there exists an absolute constant $C_1 = C_1(\{\iota(l)\}, \{\varepsilon(l)\})$ so that*

$$\lambda v(\{\mathcal{S}_D^* f > \lambda\}) \leq C_1 \int |f| v$$

for all $\lambda \geq 0$.

By interpolation, we get that for each $v \in A_1$

$$\int |\mathcal{S}_D^* f|^p v \lesssim_p \int |f|^p v$$

for all cubic families, $1 < p \leq 2$. By Rubio de Francia extrapolation (cf. [21, §7, p.143] or [4, Theorem 7.8]), we arrive at Theorem 2.1.7:

Theorem 2.4.5. *For all cubic families, we have*

$$\int |\mathcal{S}_D^* f|^p v \lesssim_p \int |f|^p v$$

for any A_p-weight v, $1 < p < \infty$.

Proof of L^2 Estimate. The plan is to decompose our operator according to scale – and then the size of the weight w:

For $j \in \mathbb{Z}$, we collect

$$F(j,k) := \{P \in \sigma_k : 2^j < \frac{w(3P)}{|3P|} \leq 2^{j+1}\} \subset \{M_{HL} w \gtrsim 2^j\},$$

and let $F(j) = \cup_k F(j,k)$.

Using Tonelli's theorem to interchange the order of summation, we estimate

$$\begin{aligned}\int |\mathcal{S}_D^* f|^2 w &= \sum_k \int |\mathcal{S}_{D,k}^* f|^2 w \\ &\lesssim \sum_k \sum_{P \in \sigma_k} \mathcal{S}_k^* f(x_P)^2 w(3P) \\ &= \sum_j \left(\sum_k \sum_{P \in F(j,k)} \tilde{\mathcal{S}}_k^*(f)^2(x_P) w(3P) \right) \\ &\lesssim \sum_j 2^j \left(\sum_k \sum_{P \in F(j,k)} \tilde{\mathcal{S}}_k^*(f)^2(x_P) |P| \right).\end{aligned}$$

But now, for $P \in F(j)$ we know that $5P \subset \{M_{HL} w \gtrsim_d 2^j\}$, and we therefore have the equality:

$$\sum_k \sum_{P \in F(j,k)} \tilde{\mathcal{S}}_k^*(f)^2(x_P) = \sum_k \sum_{P \in F(j,k)} \tilde{\mathcal{S}}_k^*(f \cdot 1_{\{M_{HL} w \gtrsim_d 2^j\}})^2(x_P).$$

Consequently, using the L^2-boundedness of $\tilde{\mathcal{S}}_d$, we estimate

$$\begin{aligned}\sum_k \sum_{P \in F(j,k)} \tilde{\mathcal{S}}_k^*(f \cdot 1_{\{M_{HL} w \gtrsim_d 2^j\}})^2(x_P)|P| &\leq \left\| \mathcal{S}_d(f \cdot 1_{\{M_{HL} w \gtrsim_d 2^j\}}) \right\|_{L^2(\mathbb{Z}^d)}^2 \\ &\lesssim \int |f|^2 \cdot 1_{\{M_{HL} w \gtrsim_d 2^j\}},\end{aligned}$$

and summing over j shows

$$\begin{aligned}\int |\mathcal{S}_D^* f|^2 w &\lesssim \sum_j 2^j \left(\sum_k \sum_{P \in F(j,k)} \tilde{\mathcal{S}}_k^*(f \cdot 1_{\{M_{HL} w \gtrsim 2^j\}})^2(x_P)|P| \right) \\ &\lesssim \sum_j 2^j \int |f|^2 \cdot 1_{\{M_{HL} w \gtrsim 2^j\}} \\ &\lesssim \int |f|^2 M_{HL} w.\end{aligned}$$

□

Proof of $L^{1,\infty}$ Estimate. We maintain our convention of assuming $f \geq 0$. By multiplying f

by a suitable constant, it's enough to prove the result for $\lambda = 1$:

$$v(\{\mathcal{S}_D^* f > 1\}) \lesssim \int |f| v.$$

Using our Calderón-Zygmund decomposition Lemma 2.4.2, we perform a Calderón-Zygmund stopping-time decomposition at height $c = c(d, K) \lesssim 1$ depending on the dimension and the refined filtration using $\mathcal{M} f$.

With $E = \{\mathcal{M} f > c\}$, collect all maximal $R \in E$, $R \in \tau_j$ in E_j, so that we may decompose

$$E = \bigcup_k E(k) := \bigcup_k \left(\cup_{j(k-1) < j \leq j(k)} E_j\right).$$

We perform our Calderón-Zygmund decomposition and arrive at

$$f = g + b = g + \sum_k b^k = g + \sum_k \left(\sum_{P \in E(k)} b_P\right),$$

where g and b have the above-detailed properties. This may be achieved, for instance, by setting

$$b_P = (f - f_P) 1_P,$$

and

$$g = f 1_{E^c} + \sum_P f_P,$$

where $f_P = \frac{1}{|P|} \int_P f$ is the average of f on P.

Set $X = \bigcup 2^{2K} P$, where K is as above. We may choose c sufficiently large so that $X \subset \{M_{HL} f > 1\}$. We then have the standard estimate (cf. e.g. [38, §2.1.3])

$$|X| \lesssim \int |f| M_{HL} v \lesssim \int |f| v,$$

so our problem reduces to showing

- $v(\{X^c : \mathcal{S}_D^* g > 1/2\}) \lesssim \int |f|v$; and
- $v(\{X^c : \mathcal{S}_D^* b > 1/2\}) \lesssim \int |f|v$.

For the first point, we use Chebyshev's inequality and the established $L^2(\mathbb{Z}^d)$ bound to majorize

$$\begin{aligned} v(\{X^c : \mathcal{S}_D^* g > 1/2\}) &\lesssim \int \mathcal{S}_D^* g^2 \ (v 1_{X^c}) \\ &\lesssim \int g^2 \ M_{HL}(v 1_{X^c}) \\ &\lesssim \int g \ M_{HL}(v 1_{X^c}) \\ &\leq \int_{E^c} f M_{HL} v + \sum_P f_P \int_P M_{HL}(v 1_{X^c}). \end{aligned}$$

Note the use of the pointwise bound $|g| \lesssim c \lesssim 1$ in passing to the third line.

But

$$\sup_P M_{HL}(v 1_{X^c}) \lesssim \inf_P M_{HL} v,$$

since for any $x \in P$,

$$M_{HL}(v 1_{X^c})(x) \leq \sup_{x \in R} \frac{1}{|R|} \int_R v,$$

where the supremum is taken over R cubes, all of whose side lengths are at least as large as those of P. Consequently, we may estimate the above sum:

$$\sum_P f_P \int_P M_{HL}(v 1_{X^c}) \lesssim \sum_P (\int_P f) \cdot \inf_P M_{HL} v \leq \sum_P \int_P f M_{HL} v,$$

which yields the desired result, since $M_{HL} v \lesssim v \in A_1$.

Before beginning the second point we make the following observation: if P is a selected (bad) cube, then

$$\int_P |b| v \leq \int_P |f| v + \int_P |f_P| v \leq \int_P |f| v + |f_P| v(P).$$

But since $v \in A_1$,

$$v(P) \lesssim |P| \inf_{x \in P} v(x),$$

so that

$$|f_P| v(P) \lesssim \left| \int_P f \right| \cdot \inf_{x \in P} v(x) \leq \int_P |f| v.$$

Summing over all P, we have shown that

$$\int |b| v \lesssim \int |f| v,$$

and so we need only show that

$$v(\{X^c : \mathcal{S}_D^* b > 1/2\}) \lesssim \int |b| v.$$

With this in mind, we next note that for $R \in \sigma_k$,

$$|\tilde{\mathcal{S}}_k^* b^{k-l}|(x_R)|^2 \leq \iota(l)^2 \frac{1}{|R|} \int_{x_R + 4H_k} |b^{k-l}|.$$

Using this estimate, away from X, we majorize

$$\begin{aligned} |\mathcal{S}_{D,k}^* (\sum_{l \geq 1} b^{k-l})|^2 &\leq \left(\sum_{l \geq 1} |\mathcal{S}_{D,k}^* (b^{k-l})| \right)^2 \\ &\lesssim \left(\sum_{l \geq 1} \iota(l)^{1/2} \cdot \iota(l)^{1/2} \left(\sum_{R \in \sigma_k} \left(\int |b^{k-l}| 1_{x_R + 4H_k} \right) \frac{1_{3R}}{|R|} \right)^{1/2} \right)^2 . \end{aligned}$$

We now use Cauchy-Schwartz and the summability of $\sum_l \iota(l) < \infty$ to estimate the foregoing by a constant multiple of

$$\sum_{l \geq 1} \iota(l) \cdot \left(\sum_{R \in \sigma_k} \left(\int |b^{k-l}| 1_{x_R + 4H_k} \right) \frac{1_{3R}}{|R|} \right).$$

Immediately, we have

$$\int |\mathcal{S}^*_{D,k}(\sum_{l\geq 1} b^{k-l})|^2 v \lesssim \sum_{l\geq 1} \iota(l) \cdot \sum_{R\in\sigma_k} \left(\int |b^{k-l}| 1_{x_R+4H_k}\right) \frac{v(x_R+4H_k)}{|x_R+4H_k|}.$$

But for any $Q \subset E(k-l)$ which intersects $x_R + 4H_k$,

$$\frac{v(x_R+4H_k)}{|x_R+4H_k|} \lesssim \inf_{x\in Q\cap(x_R+4H_k)} M_{HL}v(x),$$

so we may replace

$$\left(\int |b^{k-l}| 1_{x_R+4H_k}\right) \frac{v(x_R+4H_k)}{|x_R+4H_k|} \lesssim \int_{x_R+4H_k} |b^{k-l}| M_{HL}v,$$

which leads directly to the bound

$$\begin{aligned}\int |\mathcal{S}^*_{D,k}(\sum_{l\geq 1} b^{k-l})|^2 v &\lesssim \sum_{l\geq 1} \iota(l) \sum_{R\in\sigma_k} \int_{x_R+4H_k} |b^{k-l}| M_{HL}v \\ &\lesssim \sum_{l\geq 1} \iota(l) \int |b^{k-l}| M_{HL}v,\end{aligned}$$

due to the bounded overlap of $\{x_R + 4H_k : R \in \sigma_k\}$.

Using Chebyshev and summing over k, l completes the second point,

$$\begin{aligned}v(\{X^c : \mathcal{S}^*_D b > 1/2\}) &\lesssim \int |\mathcal{S}^*_D b|^2 (v 1_{X^c}) \\ &\leq \sum_k \sum_{l\geq 1} \iota(l) \int |b^{k-l}| M_{HL}v \\ &= \sum_{l\geq 1} \iota(l) \int |b| M_{HL}v \\ &\lesssim \int |b| M_{HL}v,\end{aligned}$$

thereby concluding the proof. $\square$

2.5 The Rectangular Square Function

In this section we relax our smoothness/eccentricity control over our averaging families: suppose $\mathcal{A} = \{E_i\}$ is a nested sequence of rectangles in $\mathbb{Z}^d$, with sides parallel to the axes, but without any regularity assumptions.

We study the following

Problem 2.5.1 ([8] Problem 7.5). *For each collection rectangles $\mathcal{A} \subset \mathbb{Z}^d$ set*

$$\mathcal{S}f = (\sum_i |A_i - A_{i+1}f|^2)^{1/2}.$$

For each $\mathcal{A}$, does there exist a bound

$$\|\mathcal{S}f\|_{L^{1,\infty}(\mathbb{Z}^d)} \leq C(\mathcal{A})\|f\|_{L^1(\mathbb{Z}^d)}?$$

Following the approach of the previous sections, we study individually long/short square functions: with $\{H_k\}$ as above,

$$\mathcal{S}_{\mathcal{A}}^L f := (\sum_k |\chi_{H_k} * f - \chi_{H_{k+1}} f|^2)^{1/2}, \text{ and}$$
$$\mathcal{S}_{\mathcal{A}}^S f := (\sum_k |\mathcal{S}_{\mathcal{A}}^k f|^2)^{1/2} := (\sum_k \sum_{E_i \subset H_k, \text{ nested}} |A_i f - A_{i+1} f|^2)^{1/2}.$$

Establishing the weak type bound for the short square function $\mathcal{S}_{\mathcal{A}}^S$ remains currently out of reach; we are, however, able to establish the following partial result:

Proposition 2.5.2. *There exists an absolute constant independent of the collection $\mathcal{A}$ so that*

$$\left\|\mathcal{S}_{\mathcal{A}}^L f\right\|_{L^{1,\infty}(\mathbb{Z}^d)} \leq C\|f\|_{L^1(\mathbb{Z}^d)}.$$

This result is proven by combining the fibre-wise argument used in [7] with the Calderón-Zygmund technique used in establishing the $L^{1,\infty}(\mathbb{Z}^d)$ weighted estimate.

For the sake of exposition, we pause here to record the following lemmas which will be used in the main argument below.

Lemma 2.5.3. *Suppose that*

$$\Psi_n = \sum_J \psi_J,$$

where the $\{\psi_J\}$ *are a finite collection functions, disjointly supported in intervals* $J \subset \mathbb{Z}$, $|J| = 2^n$ *and satisfy* $\int \psi_J = 0$, $\|\psi_J\|_1 \lesssim 2^n$.

If $\chi_I := \frac{1}{|I|} 1_I$, $2^{n+s} \leq |I| < 2^{n+s+1}$, $s \geq 0$

$$\|\chi_I * \Psi_n\|_1 \lesssim 2^{-s} \|\Psi_n\|_1.$$

Proof. If $E := \{x : J \cap x - \partial I \neq \emptyset\}$, then $|E| \lesssim |J|$, and for each J,

$$|\chi_I * \psi_J(x)| \leq \frac{\|\psi_J\|_1}{|I|} 1_E \lesssim 2^{-s} 1_E.$$

Integrating, then summing over J, exhibits $\|\chi_I * \Psi_n\|_1 \lesssim 2^{-s} \|\Psi_n\|_1$, as desired. □

The following generalization of the Stein-Weiss Lemma [19, Lemma 2.3] on summing weak-type inequalities will also be of use.

Lemma 2.5.4. *If* $\|g_k\|_{1,\infty} \leq 1$, *and* c_k *are a collection of positive constants with* $\sum c_k = 1$. *Set* $\gamma := (\sum_k c_k^2)^{1/2}$. *Then* $G := (\sum |c_k g_k|^2)^{1/2}$ *has* $\|G\|_{1,\infty} \lesssim 1$. *By scaling:*

$$\left\| (\sum_k |g_k|^2)^{1/2} \right\|_{1,\infty} \lesssim \sum_k \|g_k\|_{1,\infty}.$$

Sketch. This proof is similar to Stein-Weiss. One splits each

$$g_k = l_k + m_k + u_k$$

according to size. The l_k are chosen so $(\sum |l_k|^2)^{1/2} \leq \lambda/2$, one uses a union bound to estimate the "high" component

$$|\{(\sum_k |u_k|^2)^{1/2} > \lambda/4\}| \leq |(\sum_k |u_k|^2)^{1/2} \neq 0|,$$

and Chebyshev to control

$$|\{(\sum_k |m_k|^2)^{1/2} > \lambda/4\}| \lesssim \lambda^{-2} \sum_k \|m_k\|_2^2.$$

□

Proof of Proposition 2.5.2. For notational ease, set $\mathcal{S}^L_{\mathcal{A}} = \mathcal{S}$. By separating into d families as in the proof of Theorem 2.1 in [7], we may sum over consecutive rectangles which differ in the first coordinate. We will express

$$x = (x_1, x') \in \mathbb{Z} \times \mathbb{Z}^{d-1},$$

and each

$$H_k = I_k \times J_k \subset \mathbb{Z} \times \mathbb{Z}^{d-1}$$

where J_k is a $d-1$-rectangle.

The set-up is quite similar to the proof of the $L^{1,\infty}(\mathbb{Z}^d)$ weighted estimate:

We continue to assume that $f \geq 0$, and seek to establish

$$|\{\mathcal{S}f > 1\}| \lesssim \|f\|_1.$$

We refine our (reverse) filtration $\{H_k\}$ as above, and consider once again the maximal operator

$$\mathcal{M}f(x) := \sup_{j \geq 0} |\mathcal{E}'_j f|.$$

We also define the (uncentered) fibred-maximal functions

$$M^{d-1}f(x) := \sup_k \sup_{x' \in y'+J_k} \frac{1}{|J_k|} \int_{y'+J_k} |f|(x_1, z') \, dz',$$

and

$$M^1 f(x) := \sup_k \sup_{x_1 \in y_1+I_k} \frac{1}{|I_k|} \int_{y_1+I_k} |f|(z_1, x') \, dz_1,$$

both of which enjoy the familiar one-parameter weak-type $(1,1)$ bounds.

With $E := \{\mathcal{M}f > c\}$, we collect all maximal $R \in E$, $R \in \tau_j$ in E_j, so that we may express

$$E = \bigcup_k E(k) := \bigcup_k \left(\cup_{j(k-1)<j\leq j(k)} E_j \right),$$

and decompose

$$f = g + b = g + \sum_k b^k = g + \sum_k \left(\sum_{Q \in E(k)} b_Q \right),$$

where $g := f \cdot 1_E + \sum_Q f_Q \cdot 1_Q$, and $b_Q = (f - f_Q) \cdot 1_Q$ as above.

With $X := \bigcup_Q 10Q$, so that $Q + H_k \subset X$ whenever $Q \in E(k)$, we have the estimate

$$|X| \lesssim \sum_Q |Q| \lesssim \sum_Q \int_Q f \leq \|f\|_1,$$

and using the $L^2(\mathbb{Z}^d)$-boundedness of $\mathcal{S}$ ([7], Theorem 2.1) estimate

$$|\{\mathcal{S}g > 1/2\}| \lesssim \|\mathcal{S}g\|_{L^2(\mathbb{Z}^d)}^2 \lesssim \|g\|_{L^2(\mathbb{Z}^d)}^2 \lesssim \|g\|_{L^1(\mathbb{Z}^d)} \leq \|f\|_{L^1(\mathbb{Z}^d)};$$

the argument reduces to showing

$$|\{X^c : \mathcal{S}b > 1/2\}| \lesssim \|f\|_{L^1(\mathbb{Z}^d)}.$$

To do this, we further decompose each

$$b_Q = b_Q^0 + b_Q^1$$

where each b_Q^i satisfies the size condition $\|b_Q^i\|_{L^1(\mathbb{Z}^d)} \sim |Q|$, but also the more specialized "fibred" moment conditions:

$$\int_{\mathbb{Z}} b_Q^0(x_1, x')dx_1 = 0 \text{ for each } x' \in \mathbb{Z}^{d-1},$$
$$\int_{\mathbb{Z}^{d-1}} b_Q^1(x_1, x')dx' = 0 \text{ for each } x_1 \in \mathbb{Z}.$$

This may be accomplished for instance by setting

$$\begin{aligned} b_Q &= b_Q^0 + b_Q^1 \\ &= \left(b_Q - \left(\int_{I_Q} b_Q(x_1, x') \, dx_1 \right) \cdot 1_{I_Q} \right) + \left(\left(\int_{I_Q} b_Q(x_1, x') \, dx_1 \right) \cdot 1_{I_Q} \right) \\ &:= \psi_{I_Q} \otimes 1_{J_Q} + 1_{I_Q} \otimes \psi_{J_Q}, \end{aligned}$$

where

$$\|\psi_{I_Q} \otimes 1_{J_Q}\|_{L^1(\mathbb{Z}^d)} + \|1_{I_Q} \otimes \psi_{J_Q}\|_{L^1(\mathbb{Z}^d)} \lesssim \|b_Q\|_{L^1(\mathbb{Z}^d)}.$$

We define

$$b^k = b_0^k + b_1^k$$

in the obvious way.

It suffices to separately estimate

$$|\{X^c : \mathcal{S}b^0 > 1/4\}|, \ |\{X^c : \mathcal{S}b^1 > 1/4\}|.$$

We begin by studying the square function's behavior on b^0 – i.e. we assume that each

$$b_Q^0 = \psi_{I_Q} \otimes 1_{J_Q}$$

has mean-zero when integrated in the x_1-direction.

We decompose the convolution operators

$$\begin{aligned}(\chi_{H_k} - \chi_{H_{k+1}}) * f &= \chi^2_{J_k} * ((\chi^1_{I_k} - \chi^1_{I_{k+1}}) * f) \\ &= (\chi^1_{I_k} - \chi^1_{I_{k+1}}) * (\chi^2_{J_k} * f) \\ &=: \eta^1_k * (\chi^2_{J_k} * f) \\ &= \chi^2_{J_k} * (\eta^1_k * f),\end{aligned}$$

with convolution involving $\chi^2_{J_k}$ taking place in the final $d-1$ coordinates, and $\chi^1_{I_k}, \chi^1_{I_{k+1}}$ in the first coordinate. This allows us to bound – on X^c –

$$\begin{aligned}\mathcal{S} b^0 &\leq \left(\sum_k \chi^2_{J_k} * |\eta^1_k * \left(\sum_{|I_Q| \leq |I_k|} b^0_Q \right) |^2 \right)^{1/2} \\ &\leq \sum_{s \geq 0} \left(\sum_k \chi^2_{J_k} * |\eta^1_k * \left(\sum_{|I_Q| = 2^{-s}|I_k|} b^0_Q \right) |^2 \right)^{1/2}.\end{aligned}$$

With $s \geq 0$ fixed, we use the Fefferman-Stein vector-valued maximal inequality (cf. e.g. [38], §2 Theorem 1) to estimate

$$\begin{aligned}&\left\| \left(\sum_k \chi^2_{J_k} * \left| \eta^1_k * \left(\sum_{|I_Q| = 2^{-s}|I_k|} b^0_Q \right) \right|^2 \right)^{1/2} \right\|_{L^{1,\infty}(\mathbb{Z}^d)} \\ &\qquad \leq \left\| \left(\sum_k M^{d-1} \left| \eta^1_k * \left(\sum_{|I_Q| = 2^{-s}|I_k|} b^0_Q \right) \right|^2 \right)^{1/2} \right\|_{L^{1,\infty}(\mathbb{Z}^d)} \\ &\qquad \lesssim \left\| \left(\sum_k \left| \eta^1_k * \left(\sum_{|I_Q| = 2^{-s}|I_k|} b^0_Q \right) \right|^2 \right)^{1/2} \right\|_{L^1(\mathbb{Z}^d)} \\ &\qquad \leq \sum_k \left\| \eta^1_k * \left(\sum_{|I_Q| = 2^{-s}|I_k|} b^0_Q \right) \right\|_{L^1(\mathbb{Z}^d)}.\end{aligned}$$

By applying Lemma 2.5.3 to the functions

$$x_1 \mapsto \chi_{I_k} * \left(\sum_{|I_Q|=2^{-s}|I_k|} b_Q^0(x_1, x') \right),$$

integrating first in $x_1 \in \mathbb{Z}$, then over $x' \in \mathbb{Z}^{d-1}$, we estimate this final sum by a constant multiple of

$$\sum_k 2^{-s} \left\| \sum_{|I_Q|=2^{-s}|I_k|} b_Q^0 \right\|_1 \leq 2^{-s} \|b^0\|_{L^1(\mathbb{Z}^d)} \lesssim 2^{-s} \|f\|_{L^1(\mathbb{Z}^d)}.$$

Combining our estimates in $s \geq 0$ and choosing $c > 0$ appropriately small leads to the desired bound:

$$\begin{aligned} |\{\mathcal{S}b^0 > 1/4\}| &\leq \sum_{s \geq 0} \left| \left\{ \left(\sum_k \chi_{J_k}^2 * \left| \eta_k^1 * \left(\sum_{|I_Q|=2^{-s}|I_k|} b_Q^0 \right) \right|^2 \right)^{1/2} > c2^{-s/2} \right\} \right| \\ &\lesssim \sum_{s \geq 0} 2^{s/2} \left\| \left(\sum_k \left| \eta_k^1 * \left(\sum_{|I_Q|=2^{-s}|I_k|} b_Q^0 \right) \right|^2 \right)^{1/2} \right\|_{L^1(\mathbb{Z}^d)} \\ &\lesssim \sum_{s \geq 0} 2^{s/2} \cdot 2^{-s} \|f\|_{L^1(\mathbb{Z}^d)} \\ &\lesssim \|f\|_{L^1(\mathbb{Z}^d)}. \end{aligned}$$

In passing to the second case, where we must estimate

$$|\{X^c : \mathcal{S}(b^1) > 1/4\}|,$$

we collect cubes according to separation of scales of the final $(d-1)$-coordinates.

For

$$Q = I_Q \times J_Q \in \tau_j, j < j(k),$$

we use

$$\triangle(Q, H_k) = \triangle_1(Q, H_k)$$

to denote the degree of separation between J_Q, J_k.

Now, away from $X = \bigcup_Q 10Q$, we use the triangle inequality to obtain the pointwise bound

$$\begin{aligned}
\mathcal{S}b^1 &\leq \left(\sum_k \left| (\chi_{H_k} - \chi_{H_{k+1}}) * \left(\sum_{s\geq 0} \sum_{\triangle(Q,H_k)=s} b_Q^1 \right) \right|^2 \right)^{1/2} \\
&\leq \sum_{s\geq 0} \left(\sum_k \left| (\chi_{H_k} - \chi_{H_{k+1}}) * \left(\sum_{\triangle(Q,H_k)=s} b_Q^1 \right) \right|^2 \right)^{1/2} \\
&= \sum_{s\geq 0} (\sum_i |h_{J_{k_i},s}|^2)^{1/2},
\end{aligned}$$

where

$$h_{J_{k_i},s} := \left(\sum_{H_k : J_k = J_{k_i}} \left| (\chi_{H_k} - \chi_{H_{k+1}}) * \left(\sum_{\triangle(Q,H_{k_i})=s} b_Q^1 \right) \right| \right)^{1/2},$$

with the sum taken over all rectangles $\{H_k\}$ which share the same final $d-1$-dimensions.

We will show that for each s,

$$\|h_{J_{k_i},s}\|_{L^{1,\infty}(\mathbb{Z}^d)} \lesssim 2^{-s} \left\| \sum_{\triangle(Q,H_{k_i})=s} b_Q^1 \right\|_{L^1(\mathbb{Z}^d)},$$

then will sum on J_i and apply Lemma 2.5.4 to conclude that

$$\begin{aligned}
\left\| (\sum_i |h_{J_{k_i},s}|^2)^{1/2} \right\|_{L^{1,\infty}(\mathbb{Z}^d)} &\lesssim \sum_i \|h_{J_{k_i},s}\|_{L^{1,\infty}(\mathbb{Z}^d)} \\
&\lesssim 2^{-s} \sum_i \left\| \sum_{\triangle(Q,H_{k_i})=s} b_Q^1 \right\|_{L^1(\mathbb{Z}^d)} \\
&\leq 2^{-s} \|b^1\|_{L^1(\mathbb{Z}^d)} \\
&\lesssim 2^{-s} \|f\|_{L^1(\mathbb{Z}^d)},
\end{aligned}$$

so that, once again, with $c > 0$ an absolute constant, we estimate

$$\begin{aligned}
|\{X^c : \mathcal{S}b^1 > 1/4\}| &\leq \sum_{s \geq 0} |\{(\sum_i |h_{J_{k_i},s}|^2)^{1/2} > c2^{-s/2}\}| \\
&\lesssim \sum_{s \geq 0} 2^{s/2} \cdot 2^{-s} \|f\|_{L^1(\mathbb{Z}^d)} \\
&\lesssim \|f\|_{L^1(\mathbb{Z}^d)}.
\end{aligned}$$

To do this, we express

$$h_{J_{k_i},s} = \left(\sum_{H_k : J_k = J_{k_i}} \left| \eta_k^1 * \left(\chi_{J_{k_i}} * \left(\sum_{\triangle(Q, H_{k_i}) = s} b_Q^1 \right) \right) \right|^2 \right)^{1/2}$$

as a one-dimensional square function applied to the function

$$\chi_{J_{k_i}} * \left(\sum_{\triangle(Q, H_{k_i}) = s} b_Q^1 \right),$$

which has small $L^1(\mathbb{Z}^d)$ norm by the separation of scales Lemma 2.5.3. In particular, by considering the functions

$$x' \mapsto \chi_{J_{k_i}} * \left(\sum_{\triangle(Q, H_{k_i}) = s} b_Q^1(x_1, x') \right)$$

we estimate

$$\begin{aligned}
\left\| \chi_{J_{k_i}} * \left(\sum_{\triangle(Q, H_{k_i}) = s} b_Q^1(x_1, x') \right) \right\|_{L^1(\mathbb{Z}^d)} &= \int_{\mathbb{Z}} \left\| \chi_{J_{k_i}} * \left(\sum_{\triangle(Q, H_{k_i}) = s} b_Q^1(x_1, -) \right) \right\|_{L^1(\mathbb{Z}^{d-1})} dx_1 \\
&\lesssim \int_{\mathbb{Z}} 2^{-s} \left\| \sum_{\triangle(Q, H_{k_i}) = s} b_Q^1(x_1, -) \right\|_{l^1(\mathbb{Z}^{d-1})} dx_1 \\
&= 2^{-s} \left\| \sum_{\triangle(Q, H_{k_i}) = s} b_Q^1 \right\|_{L^1(\mathbb{Z}^d)}.
\end{aligned}$$

But now, using the one-dimensional square function result of [10], for any $\lambda \geq 0$ we may bound

$$\begin{aligned}
&\lambda|\{(x_1,x') \in \mathbb{Z} \times \mathbb{Z}^{d-1} : h_{J_{k_i},s}(x_1,x') > \lambda\}| \\
&\qquad = \int_{\mathbb{Z}^{d-1}} \lambda|\{x_1 \in \mathbb{Z} : h_{J_i}(-,x') > \lambda\}| \, dx' \\
&\qquad \lesssim \int_{\mathbb{Z}^{d-1}} \left(\int_{\mathbb{Z}} \left| \chi_{J_{k_i}} * \left(\sum_{\triangle(Q,H_{k_i})=s} b_Q^1 \right) (x_1,x') \right| \, dx_1 \right) \, dx' \\
&\qquad = \left\| \chi_{J_{k_i}} * \left(\sum_{\triangle(Q,H_{k_i})=s} b_Q^1 \right) \right\|_{l^1(\mathbb{Z}^d)} \\
&\qquad \leq 2^{-s} \left\| \sum_{\triangle(Q,H_{k_i})=s} b_Q^1 \right\|_{l^1(\mathbb{Z}^d)},
\end{aligned}$$

as desired. $\square$

2.6 Appendix: A Weak-Type Principle for Square Functions

With T a free $\mathbb{Z}^d$ action,

$$T_y f(x) := f(T_{-y}x),$$

on a non-atomic probability space (X, Σ, μ), we consider the square function:

$$Sf(x) = S^{\tau}_{(E_m)} f(x) := (\sum_m |(M_m - M_{m+1})f|^2)^{1/2}(x) := (\sum_m |K_m f|^2)^{1/2}(x).$$

Note that by the measure-preserving action of T and the triangle inequality

$$\|K_m f\|_{L^1(X)} \leq 2\|f\|_{L^1(X)},$$

i.e. the $\{K_m\}$ are uniformly bounded in operator norm.

Theorem 2.6.1. *We have the following equivalence*

- *For each $f \in L^1(X, \Sigma, \mu)$, $Sf < \infty$ μ-a.e.; and*

- $\|Sf\|_{L^{1,\infty}(X)} \lesssim \|f\|_{L^1(X)}$.

Remark 2.6.2. *The only key property about the integrability class $L^1(X)$ used in the below proof is that $1 \leq 2$; the above equivalence persists for all $1 \leq p \leq 2$.*

That the second point implies the first is clear, so we concern ourselves with the remaining implication. We actually will prove (a strengthened version of) the contrapositive, namely that if no bound on the operator norm $S : L^1(X) \to L^{1,\infty}(X)$ exists, then there exists an $f \in L^1(X)$ so that $Sf = +\infty$ a.e.

The argument below is a similar to that of [18], though to the best of our knowledge has yet to appear in print. In the spirit of the Conze principle [3], we reduce to the ergodic case:

Proposition 2.6.3. *Suppose that either (and hence both) of the conditions in Theorem 2.6.1 hold for some (free) system. Then the same is true of every such system.*

In particular: $\|S\phi\|_{L^{1,\infty}(X)} \leq C\|\phi\|_{L^1(X)}$ *if and only if* $\|\mathcal{S}f\|_{L^{1,\infty}(\mathbb{Z}^d)} \leq C\|f\|_{L^1(\mathbb{Z}^d)}$.

The forward implication is established by the Rokhlin Lemma [16]; it is here that we make use of the freeness of our action T. The reverse implication is a direct application of Calderón's transference principle [2].

For our purposes, we will henceforth assume that our action T is ergodic.

2.6.1 Preliminaries

The following results will be of use.

Lemma 2.6.4 (Randomization Lemma)**.** *If $\{E_n\} \subset X$ have $\sum_n \mu(E_n) = +\infty$, then there exist a collection of vectors $y(n) \subset \mathbb{Z}^d$ so that*

$$\limsup T_{-y(n)} E_n = X \ \mu - a.e.$$

In the probabilistic setting, the Borel-Cantelli lemma says that if the $\{E_n\}$ are independent events with $\sum_n \mu(E_n) = \infty$, then $\limsup E_n = X$ almost surely. The content of this lemma is that the ergodicity of the T-action is sufficiently randomizing to force similar independence.

Proof. If $Q(N) := \{y \in \mathbb{Z}^d : |y(i)| \leq N, 1 \leq i \leq d\}$, by the ergodicity of T, we know that for any measurable $A, B \subset X$

$$\frac{1}{(2N+1)^d} \sum_{y \in Q(N)} \mu(T_{-y}A \cap B) \to \mu(A)\mu(B).$$

Consequently, we may find a $y = y(A, B)$ so that

$$\mu(T_{-y}A \cap B) \geq 1/2\mu(A)\mu(B);$$

it is this point which anchors the volume-packing argument found e.g. in [38, §10]. $\square$

We also recall the following orthogonality lemma concerning the well-known Rademacher functions, $\{r_m(s)\}$:

Lemma 2.6.5 ([23], §V, Theorem 8.2)**.** *If $\sum_m |a_m|^2 < \infty$, then for (Lebesgue) almost every $s \in [0,1]$, $|\sum_m r_m(s)a_m|$ converges. Conversely, if $\sum_m |a_m|^2 = +\infty$, then $|\sum_m r_m(s)a_m|$ diverges a.e.*

For what is to follow, we shall need the following two-dimensional variant.

Lemma 2.6.6 (Two-Dimensional Orthogonality Lemma)**.** *Suppose that $E \subset [0,1]^2$ is non-null, and assume that*

$$G(t)^2 := \lim_M \lim_N |\sum_{m \leq M} \sum_{n \leq N} r_m(s)r_n(t)a_{mn}|^2$$

satisfies $|G(t)| < A$ on E.

Then, there exists a M so that

$$\sum_{m>M} \sum_n |a_{mn}|^2 \leq CA^2$$

for some absolute constant C.

Proof. Set $\gamma_{mn}(s,t) := r_m(s)r_n(t)$, and note that since

$$\{\gamma_{mn}\gamma_{m'n'} : (m,n) \neq (m',n')\}$$

are orthonormal over $[0,1]^2$,

$$\sum_{(m,n)\neq(m',n')} |\langle \gamma_{mn}\gamma_{m'n'} | 1_E \rangle|^2 \leq |E|$$

converges absolutely. Consequently, we may pick an M_0 sufficiently large so that for all $M \geq M_0$,

$$\sum_{(m,n)\neq(m',n'),m,m'>M} |\langle \gamma_{mn}\gamma_{m'n'}, 1_E \rangle|^2 < (|E|/2)^2.$$

Possibly after increasing M_0, we may also assume that for all $M \geq M_0$

$$\limsup_N |\sum_{m\leq M}\sum_{n\leq N} r_m(s)r_n(t)a_{mn}| < A.$$

Now, for any $M > M' \geq M_0$

$$\begin{aligned}
&\limsup_{N\to\infty} |\sum_{M'<m\leq M}\sum_{n\leq N} r_m(s)r_n(t)a_{mn}|^2 \\
&\qquad \lesssim \limsup_{N\to\infty} |\sum_{m\leq M'}\sum_{n\leq N} r_m(s)r_n(t)a_{mn}|^2 + \limsup_{N\to\infty} |\sum_{m\leq M}\sum_{n\leq N} r_m(s)r_n(t)a_{mn}|^2 \\
&\qquad \lesssim A^2 + A^2,
\end{aligned}$$

so we may use dominated convergence to estimate

$$
\begin{aligned}
\lim_{M\to\infty}\lim_{N\to\infty}\int_E |\sum_{M'<m\le M}\sum_{n\le N} r_m(s)r_n(t)a_{mn}|^2 \\
&= \lim_{M\to\infty}\int_E \lim_{N\to\infty} |\sum_{M'<m\le M}\sum_{n\le N} r_m(s)r_n(t)a_{mn}|^2 \\
&= \int_E \lim_{M\to\infty}\lim_{N\to\infty} |\sum_{M'<m\le M}\sum_{n\le N} r_m(s)r_n(t)a_{mn}|^2 \\
&\lesssim \int_E |G(t)|^2 + \int \limsup_{N\to\infty} |\sum_{m\le M'}\sum_{n\le N} r_m(s)r_n(t)a_{mn}|^2 \\
&\lesssim A^2|E| + A^2|E| \\
&\lesssim A^2|E|.
\end{aligned}
$$

We now seek a lower bound on the $\lim_M \lim_N \int_E |\sum_{M'<m\le M}\sum_{n\le N} r_m(s)r_n(t)a_{mn}|^2$. To do this, we expand the square

$$
\int_E |\sum_{M'<m\le M}\sum_{n\le N} r_m(s)r_n(t)a_{mn}|^2
$$

to get

$$
\begin{aligned}
\int_E \sum_{M'<m\le M}\sum_{n\le N} |a_{mn}|^2 + \int_E \sum_{M'<m\ne m'\le M}\sum_{n\ne n'\le N} \gamma_{mn}\gamma_{m'n'}a_{mn}\bar{a}_{m'n'} \\
=: I(M,N) + II(M,N).
\end{aligned}
$$

We will show that

$$
|II(M,N)| < \frac{1}{2}I(M,N) = \frac{|E|}{2}\sum_{M'<m\le M}\sum_{n\le N}|a_{mn}|^2.
$$

This will allow us to bound from below:

$$\lim_{M\to\infty}\lim_{N\to\infty}\int_E |\sum_{M'<m\leq M}\sum_{n\leq N} r_m(s)r_n(t)a_{mn}|^2$$
$$\geq \lim_{M\to\infty}\lim_{N\to\infty}\frac{|E|}{2}\sum_{M'<m\leq M}\sum_{n\leq N}|a_{mn}|^2$$
$$= \frac{|E|}{2}\sum_{M'<m}\sum_{n}|a_{mn}|^2.$$

Combining this with our upper bound will allow us to conclude:

$$\sum_{M'<m}\sum_{n}|a_{mn}|^2 \lesssim A^2.$$

But we may use Cauchy-Schwarz to majorize:

$$\left|\int_E \sum_{M'<m\neq m'\leq M}\sum_{n\neq n'\leq N}\gamma_{mn}\gamma_{m'n'}a_{mn}\bar{a}_{m'n'}\right|$$
$$\leq \sum_{M'<m\neq m'\leq M}\sum_{n\neq n'\leq N}|\langle\gamma_{mn}\gamma_{m'n'}|1_E\rangle||a_{mn}a_{m'n'}|$$
$$\leq (\sum_{M'<m\neq m'\leq M}\sum_{n\neq n'\leq N}|\langle\gamma_{mn}\gamma_{m'n'}|1_E\rangle|^2)^{1/2}\cdot(\sum_{M'<m\neq m'\leq M}\sum_{n\neq n'\leq N}|a_{mn}a_{m'n'}|^2)^{1/2}$$
$$\leq \frac{|E|}{2}\cdot\sum_{M'<m\leq M}\sum_{n\leq N}|a_{mn}|^2.$$

□

We now turn to the proof proper.

Proof of the Theorem 2.6.1. We proceed as in [18]:

Suppose that no bound $\|Sf\|_{L^{1,\infty}(X)} \leq N\|f\|_{L^1(X)}$ existed. In this case we could find a sequence of functions $\{g_n\} \in L^1(X)$, and a monotonically increasing sequence $\{R_n\}$ so

$$\sum_n \mu(Sg_n > R_n) = \infty$$

diverges, while

$$\sum_n \|g_n\|_1 \lesssim 1$$

converges.

We apply our Randomization Lemma 2.6.4 to the collection of sets

$$E_n := \{Sg_n > R_n\}$$

so that

$$X_0 := \limsup T_{-y(n)} E_n \subset X$$

has full measure, and set $f_n := T_{y(n)} g_n$.

Then $Sf_n > R_n \to \infty$ on X_0, while $\sum_n \|f_n\|_1 = \sum_n \|T_{y(n)} g_n\|_1 < \infty$.

We consider the formal sum

$$\sum_n r_n(t) f_n(x);$$

as in [18] we may find a subsequence $\{N_k\}$ along which $F(x,t) := \lim_k \sum_{n \leq N_k} r_n(t) f_n(x)$ satisfies

1. For almost every $t \in [0,1]$, $F(x,t) \in L^1(X)$;

2. For μ-a.e. $x \in X$, for each m

   $$K_m F(x,t) \equiv \lim_k \sum_{n \leq N_k} r_n(t) K_m f_n(x)$$

   as functions of t

We will prove:

$$\text{for almost every } t,\ SF(x,t) = +\infty \ \mu\text{-a.e.}$$

To do this, we proceed by contradiction, and assume that the sum

$$\sum_m |K_m F(x,t)|^2 < \infty$$

converges on a set $D \subset X \times [0,1]$ of positive product measure.

By Lemma 2.6.5, for each $(x,t) \in D$

$$|\sum_m r_m(s) K_m F(x,t)| < \infty$$

s-a.e. so we may extract a subset $E \subset X \times [0,1]^2$, $A > 0$, so that for each $M \geq M'$ sufficiently large partial summand

$$|\sum_{m \leq M} r_m(s) K_m F(x,t)| < A$$

on E.

For each measurable section (μ-almost all sections are measurable), we set

$$E_x := \{(s,t) \in [0,1]^2 : (x,s,t) \in E\} \subset [0,1].$$

Since E has positive product measure, we may extract some $\delta > 0$ and a set X_δ of positive μ-measure so that for each $x \in X_\delta$, $|E_x| \geq \delta$.

For each $x \in X_\delta$, we apply our Two-Dimensional Orthogonality Lemma 2.6.6 to

$$\lim_{M\to\infty} |\sum_{m \leq M} r_m(s) K_m F(x,t)| = \lim_{M\to\infty} \lim_{N_k \to\infty} |\sum_{m \leq M} \sum_{n \leq N_k} r_m(s) r_n(t) K_m f_n(x)|,$$

with $K_m f_n(x)$ in the role of a_{mn}.

In particular, we extract $M(x)$ so that

$$\sum_{M(x) < m} \sum_n |K_m f_n(x)|^2 \leq CA^2.$$

Assume that $M(x)$ is minimal subject to the condition that

$$\sum_{M(x)<m} \sum_n |K_m f_n(x)|^2 \leq 2CA^2,$$

and collect

$$A_k := \{x \in X_\delta : M(x) = k\},$$

so that we may express $X_\delta = \bigcup_{k\geq 1} A_k$ as a disjoint union. Choose k' as small as possible subject to the constraint that $\mu(A_{k'}) > 0$. We have:

$$\sum_{k'<m} \sum_n |K_m f_n(x)|^2 \lesssim A^2$$

on $A_{k'}$.

We now wish to show that

$$\left\{ y : \sum_{m\leq k'} \sum_n |K_m f_n(y)| \geq \frac{R_n}{2} \text{ for all but finitely many } n \right\}$$
$$= \liminf_n \left\{ y : \sum_{m\leq k'} \sum_n |K_m f_n(y)| \geq \frac{R_n}{2} \right\}$$

is μ-null. This will allow us to conclude that for almost every $x \in A_{k'} \cap X_0$ – and thus almost every $x \in A_{k'}$, since $\mu(X_0) = 1$ – there exist infinitely many n with

$$\sum_{k'<m} |K_m f_n(x)|^2 \geq |Sf_n(x)|^2 - \sum_{m\leq k'} |K_m f_n(x)|^2 \geq \frac{3}{4} R_n^2.$$

Summing in n would then force

$$\sum_{k'<m} \sum_n |K_m f_n(x)|^2$$

to diverge, contradicting the upper bound of $\lesssim A^2$, and concluding the argument.

To this end, for each n we estimate

$$\sum_{m \le k'} \|K_m f_n(x)\|_1 \lesssim k' \|f_n\|_1,$$

from which it follows that

$$\mu \left(\left\{ y \in X : \sum_{m \le k'} |K_m f_n|(y) > R_n/2 \right\} \right) \lesssim k' \cdot \frac{\|f_n\|_1}{R_n}.$$

Consequently, for each l, we may estimate

$$\mu \left(\bigcap_{n \ge l} \left\{ y \in X : \sum_{m \le k'} |K_m f_n|(y) > R_n/2 \right\} \right) \lesssim \lim_{n \to \infty} k' \cdot \frac{\|f_n\|_1}{R_n} = 0,$$

which yields

$$\mu \left(\liminf_n \left\{ y \in X : \sum_{m \le k'} |K_m f_n|(y) > R_n/2 \right\} \right) = 0,$$

as desired. $\square$

References

[1] Bourgain, J. Pointwise ergodic theorems for arithmetic sets. Inst. Hautes Études Sci. Publ. Math. (69):545, 1989. With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein.

[2] Calderón, A. Ergodic theory and translation invariant operators, Proc. Nat. Acad. Sci., USA 59 (1968), 349-353

[3] Conze, J. Convergences des moyennes ergodiques pour sous-suites. Bull. Soc. Math. France 35, 7-15.

[4] Duoandikoetxea, J. Fourier Analysis. Graduate Studies in Mathematics, vol. 29, 2001.

[5] Duoandikoetxea, J; Rubio de Francia, J. Maximal and singular integral operators via Fourier transform estimates. Invent. Math. 84 (1986), no. 3, 541561.

[6] Gillespie, T. A.; Torrea, J. Weighted ergodic theory and dimension free estimates. Q. J. Math. 54 (2003), no. 3, 257280.1464-3847

[7] Jones, R.; Rosenblatt, J.; Wierdl, M. Oscillation inequalities for rectangles. Proc. Amer. Math. Soc. 129 (2001), no. 5, 13491358 (electronic).

[8] Jones, R.; Rosenblatt, J.; Wierdl, M. Oscillation in ergodic theory: higher dimensional results. Israel J. Math. 135 (2003), 127.

[9] Jones, R.; Kaufman, R.; Rosenblatt, J.; Wierdl, M. Oscillation in ergodic theory. Ergodic Theory Dynam. Systems 18 (1998), no. 4, 889-935.

[10] Jones, R.; Seeger, A.; Wright, J. Strong variational and jump inequalities in harmonic analysis. Trans. Amer. Math. Soc. 360 (2008), no. 12, 67116742.

[11] Lépingle, D. La variation d'ordre p des semi-martingales. Z.F.W. 36, 1976, 295-316.

[12] Lorente, M.; Martell, J.; Pérez, C.; Riveros, M. "Generalized Hormander conditions and weighted endpoint estimates." Studia Math. 195 (2009), no. 2, 157192.

[13] Lorente, M; Riveros, M. S. ; de la Torre, A. "On Hormanders condition for singular integrals." Rev. Un. Mat. Argentina 45 (2004), no. 1, 714 (2005).

[14] Lorente, M.; de la Torre, A. "Weighted and ergodic theorems for series of differences of convolutions." J. Fourier Anal. Appl. 14 (2008), no. 1, 3959.

[15] Martín-Reyes F.; de la Torre, A. "The dominated ergodic theorem for invertible, positive operators." Semesterbericht Funktionalysis Tubingen (1985), 143150.

[16] Ornstein, D.; Weiss, B. Ergodic theory of amenable group actions. I: The Rohlin lemma. Bull. Amer. Math. Soc. (N.S.) Volume 2, Number 1 (1980), 161-164.

[17] Stein, E. Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals. Princeton Mathematical Series, 43. Monographs in Harmonic Analysis, III. Princeton University Press, Princeton, NJ, 1993.

[18] Stein, E. On Limits of Sequences of Operators. Annals of Mathematics, Second Series, Vol. 74, No. 1 (Jul., 1961), pp. 140-170

[19] Stein, E.; Weiss, N. On the Convergence of Poisson Integrals. Transactions of the American Mathematical Society, Vol. 140, (Jun., 1969), pp. 35-54

[20] Stein, E.; Weiss, N. Introduction to Fourier Analysis on Euclidean Spaces. Princeton Mathematical Series, No. 32. Princeton University Press, Princeton, N.J., 1971.

[21] Wilson, M. Littlewood-Paley Theory and Exponential-Square Integrability. Lecture Notes in Mathematics, Volume 1924 2008.

[22] Wiener, N. The ergodic theorem. Duke Math. J. Volume 5, Number 1 (1939), 1-18.

[23] Zygmund, A. Trigonometric Series, Third Edition, Volume I & II Combined (Cambridge Mathematical Library)

CHAPTER 3

Polynomial Ergodic Averages Converge Rapidly: Variations on a Theorem of Bourgain

3.1 Introduction

Let (X, Σ, μ) be a non-atomic probability space, equipped with τ a measure-preserving $\mathbb{Z}$-action

$$\tau_y f(x) := f(\tau_{-y} x).$$

For $(E_i) \subset \mathbb{Z}$, define the averaging operators

$$M_i f(x) := \frac{1}{|E_i|} \sum_{y \in E_i} (\tau_y f)(x);$$

the classical (L^2-) pointwise ergodic theorem of Birkhoff [6] says that if

$$E_i = [0, i) \subset \mathbb{Z},$$

then the one-dimensional averages $\{M_i f(x)\}$ converge pointwise μ-almost everywhere for $f \in L^2(X, \Sigma, \mu)$.

A standard proof proceeds by way of a density argument: one begins with the the dense subset

$$\{\phi \in L^2 \cap L^\infty : \tau\phi = \phi\} \oplus \operatorname{Span}\{h - \tau h : h \in L^\infty\} \subset L^2$$

on which convergence holds, and absorbs small errors using the L^2-boundedness of the maximal function

$$f \mapsto \sup_i |M_i f|.$$

This density argument relies crucially on the *smoothness* of the intervals $[0, i)$.

Obtaining pointwise convergence results of $\{M_i f\}$ for rougher, exotic $\{E_i\} \subset \mathbb{Z}$ 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.

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

Indeed, to prove pointwise convergence of the ergodic averages of L^2-functions along the squares, Bourgain [10] attacked the issue of oscillation more directly, by showing that an appropriate *oscillation operator* was L^2 "controlled" ([10, §7].) [1]

In a redux of his argument [11], 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.

Definition 3.1.1. *For a collection of functions $\{f_i\}$*

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

*is the r-*variation *of the $\{f_i\}$.*

These variation operators are more difficult to control than the maximal function $\sup_i |f_i|$: for any j, one may pointwise dominate

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

[1] This result was later extended by Quas-Wierdl-Lacey [18], who showed that the *oscillation operators* were actually L^2-bounded.

where $r < \infty$ is arbitrary. This difficulty is reflected in the fact that although having bounded r-variation, $r < \infty$, is enough to imply pointwise convergence, there are functions which converge, but which have unbounded r variation for *any* $r < \infty$. (e.g. $\{(-1)^i \frac{1}{\log i+1}\}$)

Despite the increased delicacy of the variation operators, Bourgain proved that for $E_i = [0, i)$, the r-variation operators

$$\mathcal{V}^r(M_i f)$$

were of strong-type $(2, 2)$ [11, Corollary 3.26] for $r > 2$,

$$\|\mathcal{V}^r(M_i f)\|_{L^2(X)} \leq C\frac{r}{r-2}\|f\|_{L^2(X)}$$

for some absolute C. [2]

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

Since Bourgain's celebrated result, establishing variational estimates for families of averaging operators has been the focus of much research in ergodic theory and harmonic analysis (cf. e.g. [10], [11], or [12]). Nevertheless, little research has been directed towards studying variations of averaging operators defined by "rough," polynomially-defined, sets. This is a natural object of consideration: variational estimates are a strong tool for proving pointwise convergence of averages when a density argument is unavailable.

Indeed, using an easy modification of Bourgain's earliest – and most straightforward – proof of the boundedness of maximal function along the squares (i.e. $E_i = \{1, 2^2, \ldots, i^2\}$) [10],

$$\text{there exists an absolute } C \text{ so that } \|\sup_N |M_N f|\|_{L^2(X)} \leq C\|f\|_{L^2(X)},$$

one can prove (discussed in §3) that for each $r > 2,\ \epsilon > 0$ there exists an absolute $C_{r,\epsilon}$ so

[2] It was later shown in [10] that $\mathcal{V}^2(M_i f)$ was unbounded on L^2; the super-delicacy of the two-variation operator $\mathcal{V}^2$ will be addressed in our context in §7.

that

$$\|\mathcal{V}^r(M_N f : N \in \lfloor (1+\epsilon)^n \rfloor)\|_{L^2(X)} \leq C_{r,\epsilon}\|f\|_{L^2(X)},$$

which shows that the means $\{M_N f(x) : N \in \lfloor (1+\epsilon)^n \rfloor\}$ converge pointwise almost everywhere. Since ϵ can be taken arbitrarily small, and general means differ from the $(1+\epsilon)$-lacunary means by a multiplicative factor of at most $1+\epsilon$, this result is enough to recover the full pointwise convergence result – without recourse to Bourgain's difficult "oscillation" argument [10, §7] or his metric-entropy approach [11, §6].

Despite the utility of the variational approach, the following problem remains almost untouched:

Problem 3.1.2. *With (X, μ, τ) as above, let*

$$P(n) := b_d n^d + \cdots + b_1 n + b_0, \ b_j \in \mathbb{Z}, \ b_d > 0$$

be an integer-valued polynomial, and set $E_i := \{P(1), \ldots, P(i)\}$.

For which $1 < p < \infty$, $2 < r < \infty$ do there exist a priori bounds

$$\|\mathcal{V}^r(M_N f)\|_{L^p(X)} \leq C_{p,r,P}\|f\|_{L^p(X)}?$$

The main result of this note is a first step towards resolving the above Problem 3.1.2. We prove

Theorem 3.1.3 (Polynomial Means Converge Rapidly in L^2)**.** *Let $r > 2$ be arbitrary, and let*

$$P(n) := b_d n^d + \cdots + b_1 n + b_0, \ b_j \in \mathbb{Z}, \ b_d > 0$$

be an integer-valued polynomial. Then there exists an absolute constant $C_{r,P}$, depending only on r and the polynomial, P, so that for any measure-preserving system (X, μ, τ), and any $f \in L^2(X)$,

$$\|\mathcal{V}^r(M_N f)\|_{L^2(X)} \leq C_{r,P}\|f\|_{L^2(X)}.$$

Since variation operators are *semi-local* in the sense of [13], by Calderón's transference principle [13], Theorem 3.1.3 will follow from the result below:

Proposition 3.1.4. *Let $r > 2$ be arbitrary, and let*

$$P(n) := b_d n^d + \cdots + b_1 n + b_0,\ b_j \in \mathbb{Z},\ b_d > 0$$

be an integer-valued polynomial. Then, with $E_i := \{P(1), P(2), \ldots, P(i)\}$ there exists an absolute constant $C_{r,P}$ so that for any $f \in l^2(\mathbb{Z})$,

$$\|\mathcal{V}^r(K_N * f)\|_{l^2(\mathbb{Z})} \leq C_{r,P} \|f\|_{l^2(\mathbb{Z})}.$$

where

$$K_N * f(x) := \frac{1}{N} \sum_{n \leq N} f(x + P(n))$$

is the discrete convolution operator.

Interpolating this result against Bourgain's celebrated result [11]

$$\|\mathcal{V}^\infty(M_N f)\|_{L^p(X)} \leq C_{p,P} \|f\|_{L^p(X)},\ p > 1,$$

yields the following

Proposition 3.1.5. *Suppose $r > \max\{p, p'\}$. Then for any dynamical system there exist absolute constants $C_{r,p,P}$ so that*

$$\|\mathcal{V}^r(M_N f)\|_{L^p(X)} \leq C_{r,p,P} \|f\|_{L^p(X)}.$$

Again, this result follows from the analogous one on the integer lattice (see §7):

Proposition 3.1.6. *For $r > \max\{p, p'\}$,*

$$\|\mathcal{V}^r(K_N * f)\|_{l^p} \leq C_{r,p,P} \|f\|_{l^p}.$$

Finally, we prove that our L^2 theory is, in general, sharp. We do so by studying the more delicate 2-variation operator. Specifically, we prove:

Theorem 3.1.7 (The 2-Variation Operator is Unbounded on L^2)**.** *For any $C > 0$, there exists an $f = f_C$ of L^2-norm one, but so that*

$$\|\mathcal{V}^2(M_N * f)\|_{L^2(X)} \geq C,$$

where here we fix

$$M_N * f(x) := \frac{1}{N} \sum_{n \leq N} \tau^{n^2} f(x),$$

the (discrete) square means.

Remark 3.1.8. *Although this theorem is generalizable to actions associated to other integer-valued polynomials, for the sake of clarity, we have contented ourselves with the case of the squares.*

This result will follow, by Calderón's transference principle [13], from an analogous version on the torus system

$$(\mathbb{T}, dx, T : x \mapsto x + \alpha)$$

for $0 \leq \alpha < 1$; the Kakutani-Rokhlin Lemma [17, Lemma 4.7] may be then used to transfer the unboundedness of the variation operator to any aperiodic measure-preserving system.

The structure of the paper is as follows:

In §2 we introduce relevant definitions, and present a few reductions which will be used throughout;

In §3, we review long variation arguments, and sketch a proof of pointwise convergence along polynomial means;

In §4, we collect preliminary definitions and lemmas;

In §5, assuming the number-theoretic Proposition 3.4.1, we prove full variational estimates along the polynomial sequences (i.e. Proposition 3.1.4) by studying the short variation;

In §6, we prove Proposition 3.4.1;

In §7, we interpolate our result against Bourgain's theorem to obtain (partial) variational estimates for other L^p spaces; and

In §8 we prove that our L^2-result is sharp – that in general the 2-variation associated to our polynomial averages is an unbounded operator.

3.1.1 Acknowledgements

The author would like to thank Lewis Bowen and Akos Magyar for helpful conversations, Michael Lacey for early encouragement, and his advisor, Terence Tao, for his great patience and support.

3.1.2 Notation

For a set $E \subset \mathbb{Z}$, we use $|E|$ to denote $\#E$ the counting measure (cardinality) of the set E. We also let $e(t) := e^{2\pi i t}$. For subsets of the torus, $D \subset \mathbb{T}$, we denote the frequency projection onto D of (finitely supported) functions on the integers by

$$f_D(n) := \left(\widehat{f} \cdot 1_D\right)^{\vee}(n).$$

We will make use of the modified Vinogradov notation. We use $X \lesssim Y$, or $Y \gtrsim X$ to denote the estimate $X \leq CY$ for an absolute constant C. If we need C to depend on a parameter, we shall indicate this by subscripts, thus for instance $X \lesssim_p Y$ denotes the estimate $X \leq C_p Y$ for some C_p depending on p. We use $X \approx Y$ as shorthand for $Y \leq X < 2Y$.

We also make use of big-O notation: we let $O(Y)$ denote a quantity that is $\lesssim Y$, and similarly $O_p(Y)$ a quantity that is $\lesssim_p Y$.

3.2 Preliminaries

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:

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

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

Indeed, we have the following easy pointwise inequality:

Lemma 3.2.1 ([13], §3)**.**

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

Sketch. Fix an increasing sequence of indices $\{i_k\}$. For each pair (i_k, i_{k+1}) so that there exists $n(k) < n(k+1)$ with

$$2^{n(k)-1} \le i_k \le 2^{n(k)} < 2^{n(k+1)} \le i_{k+1} \le 2^{n(k+1)+1},$$

replace

$$|f_{i_k} - f_{i_{k+1}}| \le |f_{i_k} - f_{2^{n(k)}}| + |f_{2^{n(k)}} - f_{2^{n(k+1)}}| + |f_{i_{k+1}} - f_{2^{n(k+1)}}|,$$

so that

$$|f_{i_k} - f_{i_{k+1}}|^r \lesssim_r |f_{i_k} - f_{2^{n(k)}}|^r + |f_{2^{n(k)}} - f_{2^{n(k+1)}}|^r + |f_{i_{k+1}} - f_{2^{n(k+1)}}|^r.$$

Take l^r-norms in k, and make the above replacements when necessary; the first and third replaced terms become absorbed by the short variation, while each middle term becomes absorbed by the long variation. $\square$

3.3 A Sketch of Bourgain's Argument, and the Long Variation

In this section, we abbreviate the long variation $\mathcal{V}^{r,L}$ by $\mathcal{V}^r$.

We use the argument of [11], combined with (an appropriately scaled version of) the variational result of [6, Proposition 1.2] (cf. also [16, Proposition 4.1]). The square-function argument of [11, §6] then is robust enough that one may replace the maximal function along dyadic scales $\sup_{N\in 2^{\mathbb{N}}} |K_N * f|$ with the long variation $\mathcal{V}^r(K_N * f : N \in 2^{\mathbb{N}})$.

For the sake of clarity, we provide some details of the argument in the case of the squares [3]

$$K_N * f(x) := \frac{1}{N}\sum_{n\leq N} f(x+n^2).$$

The departure point is that by using techniques from the Hardy-Littlewood circle method, the multipliers

$$\widehat{K_t}(\alpha) = \frac{1}{t}\sum_{n\leq t} e(-\alpha n^2)$$

can be well-approximated in the L^2 sense by a more tractable family of multipliers:

$$\begin{aligned}\widehat{L}_t(\alpha) &:= \sum_{s\geq 0}\widehat{L_{s,t}}(\alpha)\\ &:= \sum_{s\geq 0}\sum_{a/q\in\mathcal{R}_s} S(a/q)v_t(\alpha-a/q)\phi(10^s(\alpha-a/q)),\end{aligned}$$

where

[3] In fact, in the case of the squares, the argument of [10, §§3-5] goes through essentially unchanged. Indeed, the only new observation is that, with $k(x) := \frac{1}{2\sqrt{x}}1_{[0,1]}$ the quadratic density, not only does the family $\{k_t : t > 0\}$ satisfy the maximal inequality

$$\|\sup_t |k_t * f|\|_{L^2(\mathbb{R})} \lesssim \|f\|_{L^2(\mathbb{R})}$$

but also the stronger variational inequality

$$\|\mathcal{V}^r(k_t * f)\|_{L^2(\mathbb{R})} \lesssim \|f\|_{L^2(\mathbb{R})},$$

$r > 2$. Working on the spatial side, this follows from Bourgain's [11, Lemma 3.11] and convexity; alternatively, one can work on the Fourier side and reduce matters to the more general [12, Theorem 1.5] using Stein's "universal" lifting argument [38, Lemma 11.2.4].

- $1_{[-0.1,0.1]} \leq \phi \leq 1_{[-0.2,0.2]}$ is a smooth cut-off;

- the sets $\{\mathcal{R}_s\}$ form an exhaustion of the rationals inside $\mathbb{T}$:

$$\mathcal{R}_s := \{a/q : (a,q) = 1, q \approx 2^s\}$$

  (we identify $\mathcal{R}_0 = \{0/1 \equiv 1/1\}$, and recall our convention, $X \approx Y$ means $Y \leq X < 2Y$);

- the weights

$$S(a/q) := \frac{1}{q} \sum_{r=1}^{q} e\left(-\frac{r^2 a}{q}\right),$$

  satisfy $|S(a/q)| \lesssim q^{-\nu}$ for some $\nu > 0$ by Hua's [25, §7, Theorem 10.1]; and

- the v_t are oscillatory "pseudo-projections"

$$v_t(\beta) := \int_0^1 e(-\beta t^2 s^2) \, ds,$$

  which satisfy the estimates

$$|v_t(\beta) - 1| \lesssim t^2|\beta|, \quad \text{and } |v_t(\beta)| \lesssim \frac{1}{t|\beta|^{1/2}}$$

  by the mean value theorem and van der Corput's estimate on oscillatory integrals.

These approximation techniques will reappear in our present context §§4-5; we include a heuristic discussion of Bourgain's (and our) approach.

For each parameter t, Bourgain divided the torus into two distinct regions:

- The t-major arcs, which consist of points "t-near" rationals a/q with "t-small" denominators; and and their complements

- the t-minor arcs.

On each t-major arc, when $\alpha \sim a/q$ lies near a rational with small denominator, Bourgain showed that

$$\widehat{K_t}(\alpha) \sim S(a/q)v_t(\alpha - a/q),$$

where v_t is a continuous (integral) analogue of the discrete exponential sum, and the weight $S(a/q)$ measures the lack of uniform distribution of the squares in the residue classes $\mod q$. On each t-minor arc, where α is "t-far" from rational numbers with small denominator, the exponential sum $\widehat{K_t}(\alpha)$ is "t-negligible" – α lives too far from any rational a/q which correlates sufficiently quickly to prevent $\widehat{K_t}(\alpha)$ from oscillating itself out of (moral) consideration.

Bourgain was able to quantify these heuristic ideas in his

Lemma 3.3.1 (Lemma 6.14 of [11])**.** *There exists some $\nu > 0$ sufficiently small so that*

$$|\widehat{K_t}(\alpha) - \sum_{s \geq 0} \widehat{L_{s,t}}(\alpha)| \lesssim t^{-\nu}.$$

For our purposes, for any lacunary constant $\sigma > 1$, and $\|f\|_2 = 1$, we may estimate

$$\begin{aligned}
\|\mathcal{V}^r(K_t * f : t \in \sigma^{\mathbb{N}})\|_2 &\leq \|\mathcal{V}^r(L_t * f : t \in \sigma^{\mathbb{N}})\|_2 + \left\| \left(\sum_k |K_{\sigma^k} * f - L_{\sigma^k} * f|^2 \right)^{1/2} \right\|_2 \\
&\leq \sum_{s \geq 0} \|\mathcal{V}^r(L_{s,t} * f : t \in \sigma^{\mathbb{N}})\|_2 + \sum_{k \geq 0} (\sigma^{-\nu})^k \\
&\leq \sum_{s \geq 0} \|\mathcal{V}^r(L_{s,t} * f : t \in \sigma^{\mathbb{N}})\|_2 + \frac{1}{1 - \sigma^{-\nu}}.
\end{aligned}$$

Bourgain's next idea was to replace the "pseudo-projective" multipliers

$$\widehat{L_{s,t}}(\alpha) := \sum_{a/q \in \mathcal{R}_s} S(a/q)v_t(\alpha - a/q)\phi(10^s(\alpha - a/q))$$

with more honestly projective ones

$$\widehat{L'_{s,t}}(\alpha) := \sum_{a/q \in \mathcal{R}_s} S(a/q)1_{[-1,1]}(t^2(\alpha - a/q))\phi(10^s(\alpha - a/q)).$$

Indeed, using our "pseudo-projective" estimates on v_t, and our bound on $S(a/q)$, we may upper bound each

$$\|\mathcal{V}^r(L_{s,t} * f : t \in \sigma^{\mathbb{N}})\|_2 \leq \|\mathcal{V}^r(L'_{s,t} * f : t \in \sigma^{\mathbb{N}})\|_2 + \left\| \left(\sum_k |L_{\sigma^k} * f - L'_{\sigma^k} * f|^2 \right)^{1/2} \right\|_2$$
$$\lesssim \|\mathcal{V}^r(L'_{s,t} * f : t \in \sigma^{\mathbb{N}})\|_2 + 2^{-s\nu} \cdot \frac{1}{\sigma - 1},$$

and thus

$$\sum_{s \geq 0} \|\mathcal{V}^r(L_{s,t} * f : t \in \sigma^{\mathbb{N}})\|_2 \lesssim \sum_{s \geq 0} \|\mathcal{V}^r(L'_{s,t} * f : t \in \sigma^{\mathbb{N}})\|_2 + \frac{1}{\sigma - 1}.$$

Define now the multiplier

$$\widehat{B_{s,t}g} := \sum_{a/q \in \mathcal{R}_s} 1_{[-1,1]}(t^2(\alpha - a/q))\widehat{g}(\alpha).$$

With

$$\widehat{g_s}(\alpha) := \widehat{f}(\alpha) \cdot \sum_{a/q \in \mathcal{R}_s} S(a/q)\phi(10^s(\alpha - a/q))$$

defined via the Fourier transform, so

$$\|g_s\|_2 \leq 2^{-s\nu},$$

we have

$$\widehat{B_{s,t}g_s} \equiv \widehat{L'_{s,t}f}.$$

Now, suppose we knew that for each g,

$$(*) \qquad \|\mathcal{V}^r(\widehat{B_{s,t}g} : t \in \sigma^{\mathbb{N}})\|_2 \lesssim \left(\frac{r}{r-2}\right)^2 s^2 \cdot \frac{1}{\sigma - 1}\|g\|_2;$$

then we could conclude

$$\begin{aligned}\sum_{s\geq 0}\|\mathcal{V}^r(L'_{s,t}*f:t\in\sigma^{\mathbb{N}})\|_2 &\equiv \sum_{s\geq 0}\|\mathcal{V}^r(B_{s,t}*g_s:t\in\sigma^{\mathbb{N}})\|_2\\ &\lesssim \frac{1}{\sigma-1}\cdot\left(\frac{r}{r-2}\right)^2\cdot\sum_{s\geq 0}s^2\|g_s\|_2\\ &\lesssim \frac{1}{\sigma-1}\cdot\left(\frac{r}{r-2}\right)^2\sum_{s\geq 0}s^2 2^{-s\nu}\\ &\lesssim \frac{1}{\sigma-1}\cdot\left(\frac{r}{r-2}\right)^2,\end{aligned}$$

from which it follows that

$$\|\mathcal{V}^r(K_t*f:t\in\sigma^{\mathbb{N}})\|_2\lesssim\left(\left(\frac{r}{r-2}\right)^2\cdot\frac{1}{\sigma-1}\cdot\frac{1}{1-\sigma^{-\nu}}\right).$$

It remains to prove the following (slightly more general) proposition, in the spirit of [11, §4]:

Proposition 3.3.2. *Suppose $\lambda_1<\cdots<\lambda_N\subset\mathbb{T}$ are τ-separated frequencies: $|\lambda_i-\lambda_j|>\tau$ for $i\neq j$. For k so that $\sigma^{-k}<\frac{1}{100}\tau$ (say) let*

$$R_k:=\{\alpha\in\mathbb{T}:|\alpha-\lambda_j|\leq\sigma^{-k}\},$$

and define

$$\Delta_k^{\mathbb{T}}f(n):=\left(1_{R_k}\widehat{f}\right)^{\vee}(n).$$

Then

$$\|\mathcal{V}^r(\Delta_k^{\mathbb{T}}f)\|_{l^2}\lesssim\left(\frac{r}{r-2}\log N\right)^2\cdot\frac{1}{\sigma-1}\|f\|_{l^2}.$$

In particular, taking $\{\lambda_1,\ldots,\lambda_N\}$ to be the $\lesssim 4^s$ elements of $\mathcal{R}_s$, and R_k the pertaining

σ^{-k}-neighborhood, we find that

$$\|\mathcal{V}^r(B_{s,t} * g_s : t \in \sigma^{\mathbb{N}})\|_2 \lesssim \frac{1}{\sigma - 1} \cdot \left(\frac{r}{r-2}\right)^2 (\log |\mathcal{R}_s|)^2 \|g_s\|_2$$
$$\lesssim \frac{1}{\sigma - 1} \cdot \left(\frac{r}{r-2}\right)^2 s^2 \|g_s\|_2,$$

proving $(*)$. Indeed, upon establishing Proposition 3.3.2, we will have proven the following:

Proposition 3.3.3. *With $K_N * f(x) := \frac{1}{N}\sum_{n \leq N} f(x + P(n))$ as above, for any $r > 2$*

$$\|\mathcal{V}^r(K_N * f : N \in 2^{\mathbb{N}})\|_{l^2} \lesssim_{r,P} \|f\|_{l^2}.$$

And moreover

Corollary 3.3.4. *For $r > 2$, $\sigma > 1$, maintaining the above notation,*

$$\|\mathcal{V}^r(K_N * f : N \in \sigma^{\mathbb{N}})\|_{L^2(X)} \lesssim_{r,P} \left(\frac{1}{\sigma - 1} \cdot \frac{1}{1 - \sigma^{-\nu}}\right) \|f\|_{L^2(X)}.$$

In particular, the means $\{K_N * f(x) : N \in \lfloor \sigma^n \rfloor\}$ converge pointwise almost everywhere.

Since σ can be taken arbitrarily close to one, and general means differ from the σ-lacunary means by a multiplicative factor of at most σ, this result is enough to recover the L^2-version of the following result:

Theorem 3.3.5 ([11], Theorem 5)**.** *For any measure-preserving system (X, Σ, μ, τ) and any function $f \in L^2(X)$, the means*

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

converge pointwise μ-a.e.

Remark 3.3.6. *The following theorem in fact holds for $f \in L^p(X)$, $p > 1$: provided that f has "mild enough" singularities, polynomial means converge μ-a.e. This result was later proven to be sharp [7].*

We establish Proposition 3.3.2 in the following subsection.

3.3.1 Proof of Proposition 3.3.2

By arguing as in [11, Lemma 4.4] (see also [40, Lemma 5]), it is enough to prove the analogous result on $\mathbb{R}$, where we may take advantage of the dilation structure, and prove the analogous result under the hypothesis that the frequencies are 1-separated.

Lemma 3.3.7. *Suppose $\xi_1 < \cdots < \xi_N \subset \mathbb{R}$ are 1-separated, and similarly define for k so large that $\sigma^{-k} \leq \frac{1}{100}$*

$$\Delta_k f(x) := \left(1_{R_k} \widehat{f}\right)^{\vee}(x),$$

where now $R_k := \{\xi : |\xi - \xi_j| \leq \sigma^{-k} \text{ for some } j\}$. Then

$$\|\mathcal{V}^r(\Delta_k f)\|_2 \lesssim \left(\frac{r}{r-2}\log N\right)^2 \cdot \frac{1}{\sigma - 1}\|f\|_2.$$

Proof of Lemma 3.3.7. We begin with a reduction:

Let $1_{[-1/2,1/2]} \leq \widehat{\phi} \leq 1_{[-1,1]}$ be a smooth function,

$$\phi_k := \frac{1}{\sigma^k}\phi(\frac{x}{\sigma^k}),$$

and define the operators

$$\begin{aligned} D_k f &:= \sum_{j=1}^N \int \widehat{f}(\xi)\widehat{\phi_k}(\xi - \xi_j)e(\xi x)\, d\xi \\ &= \sum_{j=1}^N e(\xi_j x) \int \widehat{f}(\xi + \xi_j)\widehat{\phi_k}(\xi)e(\xi x)\, d\xi. \end{aligned}$$

We may majorize

$$\mathcal{V}^r(\Delta_k f) \leq \mathcal{V}^r(D_k f) + \left(\sum_k |\Delta_k f - D_k f|^2\right)^{1/2},$$

and using the separation hypothesis, we see that

$$\| \left(\sum_k |\Delta_k f - D_k f|^2 \right)^{1/2} \|_2 \lesssim \frac{1}{\sigma - 1},$$

since

$$\sup_\xi \sum_k |1_{R_k}(\xi) - \sum_{j=1}^N \widehat{\phi}_k(\xi - \xi_j)| \lesssim \frac{1}{\sigma - 1}$$

by sparsification.

In particular, we have

$$\|\mathcal{V}^r(\Delta_k f)\|_2 \leq \|\mathcal{V}^r(D_k f)\|_2 + \frac{1}{\sigma - 1}\|f\|_2.$$

But the arguments of [6, §2] show that

$$\|\mathcal{V}^r(D_k f)\|_2 \lesssim \left(\frac{r}{r-2} \log N \right)^2 \|f\|_2,$$

which proves the lemma, and with it, Proposition 3.3.2. □

This concludes our treatment of the Long Variation result in the case of the squares.

The case of general polynomial averages follows a similar argument; the only modifications are to the definitions of the weights and the "pseudo-projections." See [11, §§5-6] or §4 below for the appropriate generalizations.

3.4 The Short Variation

In this section, we set the stage for a proof that $\mathcal{V}^{r,S}$, and therefore $\mathcal{V}^r$ itself (see above), is L^2-bounded. Henceforth, we abbreviate the short variation $\mathcal{V}^{r,S}$ by $\mathcal{V}^r$;

3.4.1 Fourier Preliminaries

This is an L^2-problem, so we will make use of the Fourier transform. We begin with some notation; wherever possible, we will maintain that of Bourgain.

Throughout, we shall regard our polynomial

$$P(n) := b_d n^d + \dots b_1 n + b_0$$

as fixed, and $1 \gg \delta > 0$ will be a small but fixed constant.

We recall the exhaustion of the rationals inside $\mathbb{T}$

$$\bigcup_s \mathcal{R}_s := \bigcup_s \left\{ \frac{a}{q} : (a,q) = 1, q \approx 2^s \right\},$$

where we use $X \approx Y$ to mean $Y \leq X \leq 2Y$.

For $t \approx 2^n$, we define the *n-major arc*

$$\begin{aligned} \mathfrak{M}_n &:= \bigcup_{s \leq n\delta} \bigcup_{a/q \in \mathcal{R}_s} \mathfrak{M}_n(a/q) \\ &:= \bigcup_{a/q \in \mathcal{R}_s, s \leq n\delta} \left\{ \alpha : |\{b_d \alpha\} - a/q| < 2^{-n(d-\delta)} \right\}, \end{aligned}$$

where we use $\{x\} := x - \lfloor x \rfloor$ to denote the fractional part. We continue to identify $0 \equiv 1$, so $\mathcal{R}_0 := \{\frac{0}{1}\}$ and

$$\mathfrak{M}_n(0/1) := \{\alpha : \|\{b_d \alpha\}\|_{\mathbb{T}} < 2^{-n(d-\delta)}\},$$

where we use $\|x\|_{\mathbb{T}}$ to denote the distance to the nearest integer.

For $0 \leq i < b_d$, [4] we further define

$$\mathfrak{M}_n^i(a/q) := \left\{ \alpha \in \left[\frac{i}{b_d}, \frac{i+1}{b_d} \right) \right\} \cap \mathfrak{M}_n(a/q)$$

[4]For technical ease, on first reading we recommend assuming that $b_d = 1$, i.e. that P is a monic polynomial.

to be intersection of each major arc with each of the b_d distinct "pre-intervals"

$$\left\{|\beta - a/q| \leq 2^{-n(d-\delta)}\right\}$$

under the map $\alpha \mapsto b_d\alpha$. (Throughout, we assume that $n \gg_P 1$ is sufficiently large that there are b_d distinct such pre-intervals.)

We simply define the minor arcs

$$\mathfrak{m}_n := \mathbb{T} \smallsetminus \mathfrak{M}_n.$$

With

$$s, s' \leq n\delta, \ a/q \in \mathcal{R}_s, \ b/r \in \mathcal{R}_{s'},$$

we remark that any two pre-intervals corresponding to a/q, b/r are distinct. For otherwise we would have α such that

$$|\{b_d\alpha\} - a/q| + |\{b_d\alpha\} - b/r| \lesssim 2^{-n(d-\delta)},$$

while

$$|a/q - b/r| \geq 1/qr \gtrsim 2^{-2n\delta},$$

for the desired contradiction, since δ is sufficiently small.

Now, for each $a/q \in \mathcal{R}_s, \ s \leq n\delta$, we define

$$q_i = q_i(a/q, P, i), \ 0 \leq i < b_d$$

to be the least common denominator of

$$a/q, \ b_{d-1}/b_d \cdot (a/q + i), \ \ldots, \ b_1/b_d \cdot (a/q + i)$$

i.e. q_i is as small as possible such that there exist integers

$$a_d^i = a_d^i(a/q, P, i),\ a_{d-1}^i = a_{d-1}^i(a/q, P, i),\ \ldots,\ a_1^i = a_1^i(a/q, P, i)$$

satisfying

$$\big(a/q, b_{d-1}/b_d \cdot (a/q + i), \ldots, b_1/b_d \cdot (a/q + i)\big) = \big(a_d^i/q_i, a_{d-1}^i/q_i, \ldots, a_1^i/q_i\big) \text{ and}$$
$$(a_d^i, a_{d-1}^i, \ldots, a_1^i, q_i) = 1.$$

We remark that for each i, $q|q_i$, and that $q_i \leq b_d q \lesssim_P q$.

With the above notation in mind, we define the weight

$$S_P^i(a/q) := \frac{1}{q_i} \sum_{r=1}^{q_i} e\left(\frac{-a_d^i r^d - \cdots - a_1^i r}{q_i} \right).$$

By [25, §7, Theorem 10.1] for any $\epsilon > 0$ we have the estimate

$$|S_P^i(a/q)| \lesssim_\epsilon (q_i)^{\epsilon - 1/d} \lesssim q^{\epsilon - 1/d};$$

all we need is that there exist small $\nu > 0$ (determined below) so that

$$|S_P^i(a/q)| \lesssim q^{-\nu}.$$

We also define the oscillatory "pseudo-projections"

$$v_t(\beta) := \int_0^1 e(-\beta t^d s^d)\ ds;$$

for $t \approx 2^n$ we have

$$|v_t(\beta) - 1| \lesssim 2^{nd}|\beta|,$$
$$|v_t(\beta)| \lesssim \frac{1}{2^n|\beta|^{1/d}},$$

by the mean-value theorem and van der Corput's lemma on oscillatory integrals.

For $t \approx 2^n$, we define the multipliers

$$\widehat{C_t}(\alpha) = \widehat{K_t}(\alpha) - \widehat{K_{2^n}}(\alpha).$$

The L^2-smoothness of the family of multipliers

$$t \mapsto \widehat{K_t}(\alpha)$$

is captured by the following proposition, whose proof we defer to §6 below.

Proposition 3.4.1. *Suppose* $\frac{a}{q} \in \mathcal{R}_s, s \leq n\delta$, *and that* $t \approx 2^n$. *Then there exists* $\nu > 0$ *so that*

1. *For any* $\alpha \in \mathbb{T}$, $|\widehat{C_t - C_{t+1}}(\alpha)| \equiv |\widehat{K_t - K_{t+1}}(\alpha)| \lesssim 2^{-n}$;
2. *For* $\alpha \in \mathfrak{m}_n$ *a minor arc,* $|\widehat{C_t}(\alpha)| \lesssim 2^{-n\nu}$; *and*
3. *For* $\alpha \in \mathfrak{M}_n(\frac{a}{q})$ *a major arc*

$$|\widehat{C_t}(\alpha)| \lesssim 2^{-\nu s} \left(\min \left\{ 2^n |\{b_d \alpha\} - a/q|^{1/d}, \frac{1}{2^n |\{b_d \alpha\} - a/q|^{1/d}} \right\} + 2^{-n/2} \right).$$

We also include the following elementary lemma. This result is essentially due to Bourgain (cf. [11, Lemma 3.11], and [10, Proposition 2.12] as well), and we will make repeated use of it.

Lemma 3.4.2. *Suppose that* $\{B_n\}_{n=1}^N$ *are a family of operators which act on (finitely supported)* l^2 *functions by multiplication on the fourier side*

$$\widehat{B_n f}(\beta) := m_n(\beta) \widehat{f}(\beta).$$

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

$$\sup_{\beta \in \operatorname{supp} \widehat{f}} |m_n(\beta)| \le A \text{ and}$$

$$\sup_{\beta \in \operatorname{supp} \widehat{f}} |m_n(\beta) - m_{n+1}(\beta)| \le a.$$

Then $\|\mathcal{V}^2(B_n f)\|_{l^2} \lesssim \sqrt{NAa}\|f\|_{l^2}$.

Proof. With L a positive integer to be determined, and $1 = a_1 < a_2 < \dots a_L = N$ (almost) equally spaced indices (so $a_{i+1} - a_i \approx \frac{N}{L}$), we may pointwise dominate

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

where we used Cauchy-Schwartz in the last inequality.

We take l^2-norms, and use Plancherel's theorem to majorize the first summand

$$\begin{aligned}
\left\| \left(\sum_{i=1}^{L} |B_{a_i} f|^2\right)^{1/2} \right\|_{l^2} &= \left(\sum_{i=1}^{L} \left\| m_{a_i}(\beta) \widehat{f}(\beta) \right\|_{L^2}^2\right)^{1/2} \\
&\le \left(\sum_{i=1}^{L} A^2 \left\| \widehat{f} \right\|_{L^2}^2\right)^{1/2} \\
&\le \sqrt{L} A \|f\|_{l^2}
\end{aligned}$$

and the second summand

$$\begin{aligned}\left\|\left(\frac{N}{L}\right)^{1/2}\left(\sum_{n=1}^{N}|B_n f - B_{n+1} f|^2\right)^{1/2}\right\|_{l^2} &= \left(\frac{N}{L}\right)^{1/2}\left(\sum_{n=1}^{N}\left\|(m_n - m_{n+1})(\beta)\widehat{f}(\beta)\right\|_{L^2}^2\right)^{1/2}\\ &\leq \left(\frac{N}{L}\right)^{1/2}\sqrt{N}a\|f\|_{l^2}\\ &= \frac{Na}{\sqrt{L}}\|f\|_{l^2}.\end{aligned}$$

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

3.5 The Proof of Proposition 3.1.4

By our long variation result, it suffices to prove

$$\left(\sum_n \|\mathcal{V}_n^2(K_t * f)\|_2^2\right)^{1/2} \lesssim \|f\|_2,$$

where we abbreviate

$$\mathcal{V}_n^2(K_t * f) := \mathcal{V}^2(K_t * f : t \approx 2^n).$$

We split $f = f_{\mathfrak{m}_n} + f_{\mathfrak{M}_n}$ as a projection onto n-minor and n-major arcs, and majorize

$$\left(\sum_n \|\mathcal{V}_n^2(K_t * f)\|_2^2\right)^{1/2} \leq \left(\sum_n \|\mathcal{V}_n^2(K_t * f_{\mathfrak{m}_n})\|_2^2\right)^{1/2} + \left(\sum_n \|\mathcal{V}_n^2(K_t * f_{\mathfrak{M}_n})\|_2^2\right)^{1/2}.$$

We use Proposition 3.4.1 and Lemma 3.4.2 to control the first summand. Specifically, in the notation of Lemma 3.4.2 we may take

$$N = 2^n, A = 2^{-n\nu}, \text{ and } a = 2^{-n},$$

so that we may estimate

$$\left(\sum_n 2^{-n\nu}\|f_{\mathfrak{m}_n}\|_2^2\right)^{1/2} \leq \left(\sum_n 2^{-n\nu}\|f\|_2^2\right)^{1/2} \leq \|f\|_2.$$

We therefore restrict our attention to the second term in the above summand; we will prove the following

Proposition 3.5.1. *In the above notation,*

$$\sum_n \|\mathcal{V}_n^2(K_t * f_{\mathfrak{M}_n})\|_2^2 \lesssim \|f\|_2^2.$$

We make a few remarks before we turn to the proof proper:

For $t \approx 2^n$,

$$\mathcal{V}_n^2(K_t * f_{\mathfrak{M}_n}) \equiv \mathcal{V}_n^2(C_t * f_{\mathfrak{M}_n}),$$

since we are summing over *differences* of operators.

Since, roughly speaking, for $t \approx 2^n$, on $\mathfrak{M}_n^i(\frac{a}{q})$ we have

$$\begin{aligned} |\widehat{C}_t(\alpha)| \text{``}=\text{''} |S_P^i(\frac{a}{q})| \left| v_t(\{b_d\alpha\} - \frac{a}{q}) - v_{2^n}(\{b_d\alpha\} - \frac{a}{q})\right| \\ \lesssim q^{-\nu} \min\left\{ 2^{dn}|\{b_d\alpha\} - \frac{a}{q}|, \frac{1}{2^n|\{b_d\alpha\} - \frac{a}{q}|^{1/d}} \right\}, \end{aligned}$$

it makes sense to partition the major arcs according to both the size of the denominators of, and the distance to, our rationals a/q. We therefore further decompose our major arcs:

For $k \gg 2s$, we introduce

$$\mathfrak{M}_n = \bigcup_{s \leq n\delta} \bigcup_{l \geq -n\delta} \mathfrak{R}_{s,nd+l},$$

where

$$\mathfrak{R}_{s,k} := \bigcup_{a/q \in \mathcal{R}_s} \left\{ \alpha : |\{b_d\alpha\} - a/q| \approx 2^{-k} \right\}.$$

Proof. On each $\mathfrak{R}_{s,nd+l}$ we bound

$$|\widehat{C}_t(\alpha)| \lesssim 2^{-\nu s}\left(2^{-|l|/d} + 2^{-n/2}\right).$$

If we define the critical $l_n := \frac{nd}{2}$ to be the unique distance where $2^{-|l|/d} = 2^{-n/2}$, we have

$$|\widehat{C}_t(\alpha)| \lesssim \begin{cases} 2^{-\nu s} 2^{-|l|/d} & \text{if } l \leq l_n \\ 2^{-\nu s} 2^{-l_n/d} \equiv 2^{-\nu s} 2^{-n/2} & \text{if } l > l_n \end{cases}.$$

We now collect $\bigcup_{l>l_n} \mathfrak{R}_{s,nd+l} =: \mathfrak{R}_{s,n^*}$, and use Lemma 3.4.2 to majorize

$$\begin{aligned}
&\|\mathcal{V}_n^2(C_t * f_{\mathfrak{M}_n})\|_2^2 \\
&\quad \leq \left(\sum_{s \leq n\delta} \sum_{l=-n\nu}^{l_n} \|\mathcal{V}_n^2(C_t * f_{\mathfrak{R}_{s,nd+l}})\|_2 + \|\mathcal{V}_n^2(C_t * f_{\mathfrak{R}_{s,n^*}})\|_2 \right)^2 \\
&\quad \lesssim \left(\sum_{s \leq n\delta} \sum_{l=-n\nu}^{l_n} 2^{-s\nu/2} 2^{-|l|/2d} \|f_{\mathfrak{R}_{s,nd+l}}\|_2 + 2^{-s\nu/2} 2^{-l_n/2d} \|f_{\mathfrak{R}_{s,n^*}}\|_2 \right)^2 \\
&\quad \lesssim \sum_{s \leq n\delta} 2^{-s\nu/2} \sum_{l=-n\nu}^{l_n} 2^{-|l|/2d} \|f_{\mathfrak{R}_{s,nd+l}}\|_2^2 + 2^{-l_n/2d} \|f_{\mathfrak{R}_{s,n^*}}\|_2^2,
\end{aligned}$$

where $f_{\mathfrak{R}_{s,k}}$ denotes the fourier projection onto $\mathfrak{R}_{s,k}$, etc. and we used Cauchy-Schwarz in $\{s \leq n\nu, l \leq l_n\}$ in the final inequality.

Summing the foregoing over n, and interchanging the (n,s)-order of summation yields the upper estimate

$$\sum_n \|\mathcal{V}_n^2(K_t * f_{\mathfrak{M}_n})\|_2^2 \leq \sum_s 2^{-s\nu/2} \left(\sum_{n=\frac{s}{\delta}} \sum_{l=-n\nu}^{l_n} 2^{-|l|/2d} \|f_{\mathfrak{R}_{s,nd+l}}\|_2^2 + 2^{-n/4} \|f_{\mathfrak{R}_{s,n^*}}\|_2^2 \right);$$

we will show that each bracketed term is $\lesssim \|f\|_2^2$.

To do so, with s fixed, we expand the bracketed expression, make the change of variables

$k = nd + l$, and interchange (n, k) order of summation to obtain

$$\sum_{n=\frac{s}{\delta}} \sum_{l=-n\nu}^{l_n} 2^{-|l|/2d} \|f_{\mathfrak{R}_{s,nd+l}}\|_2^2 + 2^{-n/4} \|f_{\mathfrak{R}_{s,n^*}}\|_2^2$$

$$= \sum_{n=\frac{s}{\delta}} \sum_{l=-n\nu}^{l_n} 2^{-|l|/2d} \|f_{\mathfrak{R}_{s,nd+l}}\|_2^2 + 2^{-n/4} \sum_{l>l_n} \|f_{\mathfrak{R}_{s,nd+l}}\|_2^2$$

$$= \sum_{n=\frac{s}{\delta}} \sum_{k=nd-n\nu}^{\frac{3nd}{2}} 2^{-|k-nd|/2d} \|f_{\mathfrak{R}_{s,k}}\|_2^2 + 2^{-n/4} \sum_{k>\frac{3nd}{2}} \|f_{\mathfrak{R}_{s,k}}\|_2^2$$

$$= \sum_{k=\frac{s}{\delta}(d-\nu)} \sum_{\max\{\frac{s}{\delta}, \frac{k}{3d/2}\}}^{\frac{k}{d-\nu}} 2^{-|k-nd|/2d} \|f_{\mathfrak{R}_{s,k}}\|_2^2 + \sum_{k=\frac{s}{\delta}3d/2} \sum_{\frac{s}{\delta}}^{\frac{k}{3d/2}} 2^{-n/4} \|f_{\mathfrak{R}_{s,k}}\|_2^2$$

$$\lesssim \sum_{k=\frac{s}{\delta}(d-\nu)} \|f_{\mathfrak{R}_{s,k}}\|_2^2 + \sum_{k=\frac{s}{\delta}3d/2} \|f_{\mathfrak{R}_{s,k}}\|_2^2$$

$$\lesssim \|f\|_2^2,$$

as desired. $\square$

3.6 Proof of Proposition 3.4.1

In this section we prove Proposition 3.4.1, and thereby conclude the argument.

Proposition 3.6.1. *Suppose $\frac{a}{q} \in \mathcal{R}_s, s \leq n\delta$, and that $t \approx 2^n$. Then there exists $\nu > 0$ so that*

1. *For any $\alpha \in \mathbb{T}$, $|\widehat{C_t - C_{t+1}}(\alpha)| \equiv |\widehat{K_t - K_{t+1}}(\alpha)| \lesssim 2^{-n}$;*
2. *For $\alpha \in \mathfrak{m}_n$ a minor arc, $|\widehat{C}_t(\alpha)| \lesssim 2^{-n\nu}$; and*
3. *For $\alpha \in \mathfrak{M}_n^i(\frac{a}{q})$ each segment of a major arc*

$$|\widehat{C}_t(\alpha)| \lesssim 2^{-\nu s} \left(\min\left\{ 2^n |\{b_d\alpha\} - a/q|^{1/d}, \frac{1}{2^n |\{b_d\alpha\} - a/q|^{1/d}} \right\} + 2^{-n/2} \right).$$

Proof. The first point follows trivially from the triangle inequality, so we begin the proof proper with the minor arcs.

Here, whenever $|\{b_d\alpha\} - a/q| < 2^{-n(d-\delta)}$ we necessarily have $q \gtrsim 2^{n\delta}$ (the approximate inequality comes from the fact that we admit all $a/q \in \mathcal{R}_s, \ s \leq n\delta$ into the definition of our major arcs, rather than simply a/q such that $q < 2^{n\delta}$). By Dirichlet's theorem, we may choose a reduced fraction

$$x/y, \ (x,y) = 1, \ y \leq 2^{n(d-\delta)}$$

so that

$$|\{b_d\alpha\} - x/y| \leq \frac{1}{y 2^{n(d-\delta)}} \leq \frac{1}{y^2}.$$

By Weyl's inequality [41, Lemma 2.1], we have

$$\left| \frac{1}{t} \sum_{n=1}^{t} e(-b_d \alpha n^d - \cdots - b_1 \alpha n) \right| = |\widehat{K_t(\alpha)}| \lesssim_\epsilon t^\epsilon \left(1/y + 1/t + y/t^d\right)^{1/2^{d-1}},$$

which leads to the effective estimate $|\widehat{K_t}(\alpha)| \lesssim 2^{-n\nu}$ for some $\nu = \nu(\epsilon, \delta) > 0$. The triangle inequality yields the second point.

We next turn to the major arcs.

Suppose $\alpha \in \mathfrak{M}_n^i(a/q)$, so that we may express

$$b_d \alpha = i + a/q + \beta, \quad \text{where } |\beta| < 2^{-n(d-\delta)}, \ q \lesssim 2^{n\delta}.$$

With q_i as in §4 above, and $m \leq t \approx 2^n$, we express $m = pq_i + r$ and write

$$\begin{aligned} P(m)\alpha &= (b_d\alpha)(pq_i + r)^d + \frac{b_{d-1}}{b_d}(b_d\alpha)(pq_i+r)^{d-1} + \cdots + \frac{b_1}{b_d}(b_d\alpha)(pq_i + r) \\ &= (i + a/q + \beta)(pq_i+r)^d + \frac{b_{d-1}}{b_d}(i + a/q+\beta)(pq_i+r)^{d-1} + \cdots + \frac{b_1}{b_d}(i+a/q+\beta)(pq_i+r) \\ &\equiv \frac{a_d^i}{q_i}(pq_i+r)^d + \beta(pq_i)^d + \frac{a_{d-1}^i}{q_i}(pq_i+r)^{d-1} + \cdots + \frac{a_1^i}{q_i}(pq_i+r) + O_P(2^{n(2\delta-1)}) \mod 1 \\ &\equiv \beta(pq_i)^d + \left(\frac{a_d^i r^d + \cdots + a_1^i r}{q_i}\right) + O_P(2^{n(2\delta-1)}) \mod 1, \end{aligned}$$

so that for $m = pq_i + r$,

$$e(-P(m)\alpha) = e\left(-\frac{a_d^i r^d + \cdots + a_1^i r}{q_i}\right) e(-\beta(pq_i)^d) + O_P(2^{n(2\delta-1)}).$$

Now, with $t = p_t q_i + r_t$, we have

$$\begin{aligned}
\frac{1}{t}\sum_{m\le t} e(-P(m)\alpha) &= \frac{1}{t}\sum_{m\le p_t q_i} e(-P(m)\alpha) + O_P(2^{n(\delta-1)}) \\
&= \frac{q_i}{t}\sum_{p=0}^{p_t-1} e(-\beta(pq_i)^d)\cdot\frac{1}{q_i}\sum_{r=1}^{q_i} e\left(-\frac{a_d^i r^d + \cdots + a_1^i r}{q_i}\right) + O_P(2^{n(2\delta-1)}) \\
&= \frac{q_i}{t}\int_0^{p_t} e(-\beta(q_i)^d s^d)\ ds \cdot S_P^i(a/q) + O_P(2^{n(2\delta-1)}) \\
&= \frac{q_i}{t}\int_0^{q_i/t} e(-\beta(q_i)^d s^d)\ ds \cdot S_P^i(a/q) + O_P(2^{n(2\delta-1)}) \\
&= \int_0^{1} e(-\beta t^d s^d)\ ds \cdot S_P^i(a/q) + O_P(2^{n(2\delta-1)}) \\
&= v_t(\beta) S_P^i(a/q) + O_P(2^{n(2\delta-1)}),
\end{aligned}$$

where we used that for $|p-s| \le 1$

$$e(-\beta(q_i)^d\cdot p^d) = e(-\beta(q_i)^d\cdot s^d) + O_P(2^{n(2\delta-1)})$$

in passing to the third line. The upshot is that on $\mathfrak{M}_N^i(a/q)$ we have

$$\widehat{C}_t(\alpha) = S_P^i(a/q)\cdot(v_t(\{b_d\alpha\}-a/q) - v_{2^n}(\{b_d\alpha\}-a/q)) + O_P(2^{n(2\delta-1)}).$$

Taking into account our estimates on $S_P^i(a/q),\ v_t$ we have the upper bound

$$\begin{aligned}
|\widehat{C}_t(\alpha)| &\lesssim 2^{-s\nu}\left(\min\left\{2^{nd}|\{b_d\alpha\}-a/q|, \frac{1}{2^n|\{b_d\alpha\}-a/q|^{1/d}}\right\}\right) + O_P(2^{n(2\delta-1)}) \\
&\lesssim 2^{-s\nu}\left(\min\left\{2^{n}|\{b_d\alpha\}-a/q|^{1/d}, \frac{1}{2^n|\{b_d\alpha\}-a/q|^{1/d}}\right\} + 2^{-n/2}\right),
\end{aligned}$$

since $2^{s\nu} \le 2^{n\delta\nu}$ for $s \le n\delta$. $\square$

3.7 Interpolation

In this section, we will interpolate our L^2-based $\{\mathcal{V}^r\}_{r>2}$ estimates against Bourgain's $L^p,\ p>1$-based $\mathcal{V}^\infty$ estimate to some partial variational estimates in other L^p spaces, and thereby make further progress towards understanding Problem 3.1.2.

We begin with the following general mixed-norm (complex) interpolation lemma, whose proof follows the same lines as the classical Riesz-Thorin interpolation theorem.

Lemma 3.7.1. *Suppose that T is a linear operator, bounded*

$$T : l^{p_0}_x \to l^{p_0}_x(l^{r_0}_k)$$
$$T : l^{p_1}_x \to l^{p_1}_x(l^{r_1}_k).$$

If $\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$ and $\frac{1}{r} = \frac{1-\theta}{r_0} + \frac{\theta}{r_1}$ are determined by convexity, T is bounded

$$T : l^p_x \to l^p_x(l^r_k)$$

as well.

To use this lemma, we remark that we may linearize the variation, by setting

$$Tf(x,k) := (K_{N_k(x)} - K_{N_{k+1}(x)}) * f(x),$$

where $\{N_k\}$ are any (finite) collection of measurable functions. We are ready to prove our

Proposition 3.7.2. *Suppose $r > \max\{p,p'\}$. Then*

$$\|\mathcal{V}^r(K_N * f)\|_{l^p} \lesssim_{p,r} \|f\|_{l^p}.$$

By Bourgain's Theorem and the L^2 variational result, we know that

$$\|\mathcal{V}^r(K_N * f)\|_2 \lesssim_r \|f\|_2$$

for $r > 2$, and that

$$\|\mathcal{V}^\infty(K_N * f)\|_p \lesssim_p \|f\|_p$$

for $p > 1$.

For each p fixed, the task is to choose r as small as possible so that θ satisfies

$$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{2}$$

and

$$\frac{1}{r} = \frac{1-\theta}{\infty} + \frac{\theta}{r_1} = \frac{\theta}{r_1},$$

where $p_0 > 1$ and $r_1 > 2$. To minimize r, we want to take r_1 close to 2, and θ as large as possible. Consequently, it is in our interest to choose p_0 as far from p as possible (i.e. $p_0 = \infty$ for $p > 2$, and p_0 near 1 for $p < 2$).

With this strategy in mind, we turn to the:

Proof. We begin with the slightly more involved case, $p < 2$.

To this end, let $r > p'$ be arbitrary but fixed. In this case we may find $\delta < 1$ (but possibly very close) so that

$$\frac{1}{r} = \frac{1}{p'} + (\delta - 1).$$

We set $\gamma = \gamma(\delta) = (\delta - \frac{1}{2})^{-1}$, so that

$$\gamma(\delta - \frac{1}{2}) = 1.$$

If we choose θ to satisfy

$$\frac{1}{p} = \frac{1-\theta}{\delta^{-1}} + \frac{\theta}{2}$$

then we also have

$$\frac{1}{r} = \frac{\theta}{\gamma},$$

by direct computation.

The previous Lemma 3.7.1 allows us to interpolate the result.

We next turn to the case $p > 2$. With $r > p$, and θ satisfying

$$\frac{1}{p} = \frac{\theta}{2},$$

we may write

$$\frac{1}{r} = \frac{\theta}{\frac{2r}{p}},$$

where $\frac{2r}{p} > 2$.

Once again, Lemma 3.7.1 completes the proof. □

3.8 The 2-Variation is Unbounded on L^2

In this section, we shift our frame of reference from the integer setting to the torus.

For functions on the torus, $\mathbb{T}$, we (abuse notation and) define the convolution operators

$$K_N^\alpha * f = K_N * f(x) := \frac{1}{N} \sum_{n \leq N} f(x + n^2\alpha);$$

we will take $\alpha = 2^{-R}$, where R is a large parameter, and define the 2-variation associated to the family $\{K_N\}$ in the expected way, as

$$\mathcal{V}^2(K_N * f)(x) := \sup_{(N_k) \text{ increasing}} \left(\sum_k |K_{N_k} * f - K_{N_{k+1}} * f|^2 \right)^{1/2} (x).$$

We prove the following

Theorem 3.8.1 (The 2-Variation Operator is Unbounded on L^2)**.** *For any $C > 0$, there exists an $f = f_C$ of L^2-norm one, but so that*

$$\|\mathcal{V}^2(K_N * f)\|_{L^2(\mathbb{T})} \geq C.$$

Here's the set-up:

Let L be a large natural number, and let

$$k_1 > k_2 > \cdots > k_L$$

and

$$j_1 < j_2 < \cdots < j_L$$

be two sequences to be determined presently. We also set $R := 2^{2^L}$.

For functions in the (finite-dimensional) span of $\{e(2^{k_i}x) : 1 \leq i \leq L\}$

$$g(x) = \sum_{i=1}^{L} b_i e(2^{k_i}x),$$

we define the partial summation operators

$$S_m g(x) := \sum_{i=m}^{L} b_i e(2^{k_i}x).$$

For such g, we define the 2-variation of the $\{S_m\}$ in the natural way

$$\mathcal{V}^2(S_m(g))(x) := \sup_{(m_k) \text{ increasing}} \left(\sum_k |S_{m_k} g - S_{m_{k+1}} g|^2 \right)^{1/2} (x).$$

We use the following result of Lewko and Lewko [14]:

Proposition 3.8.2 ([14] Theorem 6)**.** *There exists some*

$$f(x) := \sum_{i=1}^{L} a_i e(2^{k_i}x)$$

with $\|f\|_{L^2(\mathbb{T})} = 1$*, but*

$$\|\mathcal{V}^2(S_m(f))\|_{L^2(\mathbb{T})} \gtrsim \log(\log(L)).$$

Fix this f; suppose we knew the following

Lemma 3.8.3. *The error function*

$$\eta(f) := \sum_{l=1}^{L} |S_l(f) - K_{2^{j_l}} * f|$$

is bounded uniformly on $\mathbb{T}$ (and in particular has L^2-norm $\lesssim 1$).

Then, by the triangle inequality, we would be able to bound from below

$$\|\mathcal{V}^2(K_{2^{j_l}} * f)\|_2 \geq \|\mathcal{V}^2(S_l(f))\|_2 - \|\eta(f)\|_2 \gtrsim \log(\log(L)) - O(1)$$

which tends to ∞ with L, which would prove our theorem.

Before beginning the proof of our (technical) lemma, we sketch out our strategy.

We consider the interactions

$$K_{2^{j_l}} * e(2^{k_i}x)$$

in two separate regimes: when k_i is "R-small" relative to j_l ($i \geq l$), and

$$K_{2^{j_l}} * e(2^{k_i}x) \text{``} = \text{''} e(2^{k_i}x)$$

and when k_i is "R-large" relative to j_l ($i < l$), and

$$K_{2^{j_l}} * e(2^{k_i}x) \text{``} = \text{''} 0.$$

More precisely, in the first regime we use Weyl's Lemma for quadratic polynomials [15], to bound

$$\left| \frac{1}{2^{j_l}} \sum_{n \leq 2^{j_l}} e(\frac{2^{k_i}}{2^R} \cdot n^2) \right| \lesssim j_l \left(\frac{2^{k_i}}{2^R} + \frac{1}{2^{j_l}} + \frac{2^R}{2^{2j_l + k_i}} \right)^{1/2};$$

when j_l is large (as it will be), and

$$\frac{2^{k_i}}{2^R} \geq \frac{2^R}{2^{2j_l+k_i}}$$

(as it will be) we have the approximate uniform bound

$$\left| \frac{1}{2^{j_l}} \sum_{n \leq 2^{j_l}} e(\frac{2^{k_i}}{2^R} \cdot n^2) \right| \lesssim j_l \sqrt{\frac{2^{k_i}}{2^R}} + \text{ small errors.}$$

In the second regime, we simply use the mean-value theorem to estimate

$$\left| K_{2^{j_l}} * e(2^{k_i}x) - e(2^{k_i}x) \right| \lesssim \frac{2^{k_i+2j_l}}{2^R}.$$

Up to some careful optimization, these estimates allow us to uniformly bound

$$\begin{aligned}
\eta(f) &= \sum_{l=1}^{L} |S_l(f) - K_{2^{j_l}} * f| \\
&= \sum_{l=1}^{L} \left| \left(\sum_{i=1}^{l-1} a_i K_{2^{j_l}} * e(2^{k_i}x) \right) - \left(\sum_{i=l}^{L} a_i K_{2^{j_l}} * e(2^{k_i}x) - e(2^{k_i}x) \right) \right| \\
&\leq \sum_{l=1}^{L} \left(\sum_{i=1}^{l-1} |K_{2^{j_l}} * e(2^{k_i}x)| + \sum_{i=l}^{L} |K_{2^{j_l}} * e(2^{k_i}x) - e(2^{k_i}x)| \right),
\end{aligned}$$

taking into account the trivial $|a_i| \leq \sum_i |a_i|^2 = 1$.

Proof. We begin by recursively constructing our sequences $\{k_l : 1 \leq l \leq L\}$ and $\{j_l : 1 \leq l \leq L\}$:

Starting with our top terms

$$k_L := 0, \; j_L := \frac{R-L}{2},$$

we define

$$k_{l-1} := R - j_l, \quad \text{and } j_l := \frac{R - L - k_l}{2};$$

in particular,

$$k_{l-1} + j_l = R, \quad \text{and } k_l + 2j_l + L = R.$$

By induction, we find

$$k_{L-t} = \frac{2^t - 1}{2^t}(R + L), \quad \text{and } j_{L-t} = \frac{R - (2^{t+1} - 1)L}{2^{t+1}},$$

and so

$$k_1 = \frac{2^{L-1} - 1}{2^{L-1}}(R + L), \quad \text{and } j_1 = \frac{R - (2^L - 1)L}{2^L}.$$

Now, with $1 \leq l \leq L$ temporarily fixed, we estimate terms in the first regime

$$\begin{aligned} |K_{2^{j_i}} * e(2^{k_i}x)| &\lesssim j_l \left(\frac{2^{k_i}}{2^R} + \frac{1}{2^{j_l}} + \frac{2^R}{2^{2j_l + k_i}} \right)^{1/2} \\ &\lesssim \frac{j_l}{2^{j_l/2}} + j_l \left(\frac{2^{k_i}}{2^R} + \frac{2^R}{2^{2j_l + k_i}} \right)^{1/2} \\ &\leq \frac{j_1}{2^{j_1/2}} + j_l \sqrt{\frac{2^{k_i}}{2^R}}, \end{aligned}$$

since for $i \leq l - 1$, we have

$$\frac{2^{k_i}}{2^R} \geq \frac{2^R}{2^{2j_l + k_i}},$$

by our choice

$$k_{l-1} + j_l = R.$$

Summing $\frac{j_1}{2^{j_1/2}} + j_l \sqrt{\frac{2^{k_i}}{2^R}}$ over $1 \leq i < l$ leads to an upper estimate of no more than

$$L \cdot \frac{j_1}{2^{j_1/2}} + j_l \sqrt{\frac{2^{k_1}}{2^R}}.$$

We next estimate terms on our second regime, $l \leq i \leq L$,

$$\left|K_{2^{j_l}} * e(2^{k_i}x) - e(2^{k_i}x)\right| \lesssim \frac{2^{k_i+2j_l}}{2^R} \leq \frac{2^{k_l+2j_l}}{2^R} = \frac{1}{2^L},$$

using that $k_l + 2j_l + L = R$. Summing over $l \leq i \leq L$ contributes a further

$$\frac{L}{2^L},$$

so that for our given l, we have

$$|S_l(f) - K_{2^{j_l}} * f| \leq L \cdot \frac{j_1}{2^{j_1/2}} + j_l\sqrt{\frac{2^{k_1}}{2^R}} + \frac{L}{2^L}.$$

Finally, summing over $1 \leq l \leq L$ leads to the estimate

$$L^2\frac{j_1}{2^{j_1/2}} + j_L\sqrt{\frac{2^{k_1}}{2^R}} + \frac{L^2}{2^L},$$

which is $O(1)$ for all L sufficiently large.

The proof is complete. $\square$

Remark 3.8.4. *Although we have chosen for simplicity to work with a rational*

$$\alpha = 2^{-R},$$

this is purely a matter of taste: setting e.g.

$$\alpha' = 2^{-R} + \beta$$

where (say)

$$\beta = .\ \underbrace{0\ldots0}_{2^{2^R}\ zeroes}\ 1\ \underbrace{0\ldots0}_{2^{3^R}\ zeroes}\ 1\ \underbrace{0\ldots0}_{2^{4^R}\ zeroes}\ 1\ldots,$$

leaves our proof unchanged, since both the mean-value and Weyl estimates remain valid. The *key point is that for each* $\{k_i\}$

$$|2^{k_i}\alpha - 2^{k_i}\alpha'| = |2^{k_i}\beta| \ll \frac{1}{2^{2R}}.$$

Remark 3.8.5. *An interesting question concerns the behavior of the* $\mathcal{V}^2$ *operator on other* L^p *spaces. Since our function* f_C *is a linear combination of lacunary frequencies, it has* L^p *norm* $\approx_p 1$ *for each* $1 \leq p < \infty$*. The interesting question – beyond the scope of the present paper – is whether [14, Theorem 6] is generalizable to other* L^p *spaces. Given the poor behavior of the* $\mathcal{V}^2$ *operator associated to the standard Birkhoff averages (cf. e.g. [10]) we feel comfortable risking the following:*

Conjecture 1. *For each* $1 \leq p \leq \infty$*, and any integer-valued polynomial* $P(n)$*, the operator*

$$\mathcal{V}^2(M_N f)$$

is unbounded on $L^p(X)$*.*

References

[1] Bourgain, J. On the maximal ergodic theorem for certain subsets of the integers. Israel J. Math. 61 (1988), no. 1, 39-72.

[2] Bourgain, J. On the pointwise ergodic theorem on L^p for arithmetic sets. Israel J. Math. 61 (1988), no. 1, 73-84.

[3] Bourgain, J. Pointwise ergodic theorems for arithmetic sets. Inst. Hautes Études Sci. Publ. Math. (69):5-45, 1989. With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein.

[4] Birkhoff, G. D. Proof of the ergodic theorem. Proc Natl Acad Sci USA 17 (12): 656-660 (1931)

[5] Krause, B. On Higher-Dimensional Oscillation in Ergodic Theory. Preprint.

[6] Krause, B. Some Optimizations for Maximal Multipliers in L^p. http://arxiv.org/pdf/1402.1804.pdf.

[7] Buczolich, Z.; Mauldin, R. Divergent square averages. Ann. of Math. (2) 171 (2010), no. 3, 1479-1530.

[8] Calderón, A. Ergodic theory and translation invariant operators, Proc. Nat. Acad. Sci., USA 59 (1968), 349-353

[9] Hua, L-K. Introduction to Number Theory. Second edition. Springer-Verlag, Berlin, 1982.

[10] Jones, R.; Kaufman, R.; Rosenblatt, J.; Wierdl, M. Oscillation in ergodic theory. Ergodic Theory Dynam. Systems 18 (1998), no. 4, 889-935.

[11] Jones, R.; Rosenblatt, J.; Wierdl, M. Oscillation in ergodic theory: higher dimensional results. Israel J. Math. 135 (2003), 1-27.

[12] Jones, R.; Seeger, A.; Wright, J. Strong variational and jump inequalities in harmonic analysis. Trans. Amer. Math. Soc. 360 (2008), no. 12, 6711-6742.

[13] Jones, R.; Wang, G. Variation inequalities for the Fejr and Poisson kernels. Trans. Amer. Math. Soc. 356 (2004), no. 11, 4493-4518.

[14] Lewko, A; Lewko, M. Estimates for the Square Variation of Partial Sums of Fourier Series and their Rearrangements, J. Funct. Anal. 262 (2012) 2561-2607

[15] Lyall, N. The Weyl Inequality for Quadratic Polynomials. http://www.math.uga.edu/ lyall/Learning/WeylQuadratic.pdf

[16] Nazarov, F.; Oberlin, R.; Thiele, C. A Calderón-Zygmund decomposition for multiple frequencies and an application to an extension of a lemma of Bourgain. Math. Res. Lett. 17 (2010), no. 3, 529-545.

[17] Petersen, K. Ergodic theory. Corrected reprint of the 1983 original. Cambridge Studies in Advanced Mathematics, 2. Cambridge University Press, Cambridge, 1989. xii+329 pp.

[18] Quas, A; Wierdl, M. Combinatorial and Diophantine applications of ergodic theory, Appendix A. Handbook of dynamical systems. Vol. 1B, 745869, Elsevier B. V., Amsterdam, 2006.

[19] Stein, E. M. Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals. Princeton Mathematical Series, 43. Monographs in Harmonic Analysis, III. Princeton University Press, Princeton, NJ, 1993.

[20] Thouvenot, J-P. La convergence presque sre des moyennes ergodiques suivant certaines sous-suites d'entiers (d'après Jean Bourgain). Séminaire Bourbaki, Vol. 1989/90. Astérisque No. 189-190 (1990), Exp. No. 719, 133153.

[21] Vaughan, R. C. The Hardy-Littlewood method. Second edition. Cambridge Tracts in Mathematics, 125. Cambridge University Press, Cambridge, 1997.

CHAPTER 4

(Uniform) Convergence of Twisted Ergodic Averages – Joint with Tanja Eisner

4.1 Introduction

Let T be an ergodic invertible measure-preserving transformation on a non-atomic probability space (X, Σ, μ), and denote by T its Koopman operator given by

$$(Tf)(x) := f(Tx).$$

The study of pointwise convergence of averages formed from the iterates $\{T^n\}$ began in 1931 with the classical pointwise ergodic theorem of Birkhoff [6]:

Theorem 4.1.1. *For any $f \in L^1(X)$ the averages*

$$\left\{ \frac{1}{N} \sum_{n \leq N} T^n f(x) \right\}$$

converge μ-a.e. to the space mean $\int_X f \, d\mu$.

This result was extended to more general polynomial averages in the late eighties by Bourgain, who proved the following celebrated theorem [11]:

Theorem 4.1.2. *Let P be a polynomial with integer coefficients. For any $f \in L^p(X)$, $p > 1$, the averages*

$$\left\{ \frac{1}{N} \sum_{n \leq N} T^{P(n)} f(x) \right\}$$

converge μ-a.e..

Prior to Bourgain's polynomial ergodic theorem, another generalization of Birkhoff's theorem was announced by Wiener and Wintner [42], see Assani [1] for more information and various proofs (here and throughout the paper, $e(t) := e^{2\pi i t}$ denotes the exponential):

Theorem 4.1.3 (Wiener-Wintner). *For every $f \in L^1(X)$ there exists a subset $X' \subset X$ of full measure so that the weighted averages*

$$\left\{ \frac{1}{N} \sum_{n \leq N} e(n\theta) T^n f(x) \right\}$$

converge for all $x \in X'$ and every $\theta \in [0, 1]$.

Remark 4.1.4. *Bourgain [12] observed that for f orthogonal to eigenfunctions (i.e., to the Kronecker factor) of T the above averages converge* uniformly *in θ to zero. In addition, if (X, T) is uniquely ergodic and f is continuous, then convergence is uniform in $x \in X$ too, see Assani [1].*

This result was in turn generalized to polynomial weights by Lesigne in [34]:

Theorem 4.1.5 (Wiener-Wintner theorem for polynomial weights). *For every $f \in L^1(X)$ there exists a subset $X' \subset X$ of full measure so that the weighted averages*

$$\left\{ \frac{1}{N} \sum_{n \leq N} e(P(n)) T^n f(x) \right\}$$

converge for all $x \in X'$ and all real polynomials $P \in \mathbb{R}[\cdot]$.

Although this list is by no means comprehensive, we remark that Lesigne's result has been further generalized and "uniformized" by Frantzikinakis [18], Host, Kra [23], and Eisner, Zorin-Kranich [16]. We also mention here that the corresponding characteristic factor for Lesigne's averages, the *Abramov factor*, is induced by generalized eigenfunctions of T and is larger than the Kronecker factor; in particular, the limit of Lesigne's averages can be non-zero for functions which are orthogonal to the Kronecker factor.

There are two main aims of this paper.

We first extend Lesigne's result to weights coming from *Hardy field* functions (see §2 below for the precise definition and references). As the weights come from Hardy field functions which are "far" from polynomials (defined below) the corresponding weighted ergodic averages always converge to zero – unlike in the case of polynomial weights. More precisely, our first main result in this direction can be stated as follows. (We refer the reader to §2 below for the precise definition of the class $\mathcal{M}_{\delta,M,m}$; informally, these are smooth functions which are uniformly "(δ, M, m)-far" from the class of polynomials.)

Theorem 4.1.6. *For $0 < \delta < 1/2$, $m \geq 0$, and $M \geq 1$, consider the class of functions $\mathcal{M}_{\delta,M,m}$. Then for every $f \in L^1(X)$ there exists a subset $X' \subset X$ of full measure so that the averages*

$$\left\{ \frac{1}{N} \sum_{n \leq N} e(p(n)) T^n f(x) \right\} \tag{4.1}$$

converge to zero uniformly *in $p \in \mathcal{M}_{\delta,M,m}$ for all $x \in X'$, i.e.,*

$$\sup_{p \in \mathcal{M}_{\delta,M,m}} \left| \frac{1}{N} \sum_{n \leq N} e(p(n)) T^n f(x) \right| \to 0 \quad \mu - a.e..$$

Moreover, if (X, μ, T) is uniquely ergodic and $f \in C(X)$, one has moreover

$$\sup_{p \in \mathcal{M}_{\delta,M,m}} \left\| \frac{1}{N} \sum_{n \leq N} e(p(n)) T^n f \right\|_\infty \to 0.$$

Remark 4.1.7. *The requirement of invertibility is not used (or needed) in the proof of this theorem.*

We show moreover that for $p \in \mathcal{U}$ with type less than $m \in \mathbb{N}$, the averages (4.1) are bounded by the $(m+1)$st Gowers-Host-Kra uniformity seminorm of f with certain quantitative uniformity in p, see Theorem 4.2.11 below for the precise formulation. This implies that for functions f which are orthogonal to some Host-Kra factor, the averages (4.1) converge to 0 uniformly in $p \in \mathcal{U}$ with corresponding growth bound.

Our second aim is to study weighted averages with polynomial powers, i.e., a combination of the Wiener-Wintner and Bourgain averages. Specifically, we consider pointwise convergence of weighted averages of the form

$$\left\{ \frac{1}{N} \sum_{n \leq N} e(n\theta) T^{P(n)} f(x) \right\}$$

where $\theta \in [0, 1]$ and P is a polynomial with integer coefficients.

We state our main result in this direction in the case of the squares.

Theorem 4.1.8. *Let $0 < c < 1$ and $E = E_c \subset [0, 1]$ be a set of c-badly approximable numbers with upper Minkowski dimension strictly less than $1/16$. Then for every $f \in L^2(X)$ there exists a subset $X' \subset X$ of full measure so that*

$$\lim_{N \to \infty} \sup_{\theta \in E_c} \left| \frac{1}{N} \sum_{n \leq N} e(n\theta) T^{n^2} f(x) \right| = 0$$

holds for every $x \in X'$.

We recall that a number, θ, is said to be *c-badly approximable* if for all $\frac{p}{q}$ reduced fractions, we have the lower bound

$$\left| \theta - \frac{p}{q} \right| \geq \frac{c}{q^2}.$$

(Note that by Dirichlet's principle we automatically have $0 < c < 1$.)

A natural follow-up question concerns the behavior of the twisted square means in the case where θ is *not* badly approximable. We provide a new (see the remark below), elementary proof of the following

Theorem 4.1.9. *Let P be as above, and $\theta \in [0, 1]$ be arbitrary. For any $f \in L^p(X)$, $p > 1$,* *there exists a subset $X' \subset X$ of full measure so that the averages*

$$\left\{ \frac{1}{N} \sum_{n \leq N} e(n\theta) T^{P(n)} f(x) \right\}$$

converge pointwise on X'.

Remark 4.1.10. *By Bourgain's L^p-maximal inequality for polynomial ergodic averages [11, §7], the set of functions for which convergence holds is closed in L^p for each $p > 1$. Since Bourgain's [10, Theorem 6] establishes the theorem for L^2-functions, this result is already known (this argument was also recently recovered as a corollary of Mirek and Trojan [35]). However, Bourgain's (and Mirek and Trojan's) argument is based off a study of $\mathbb{Z}^2$-actions, and transference to $\mathbb{R}^2$; the novelty of our approach is that we conduct our analysis entirely in the one-dimensional setting.*

Note that related results concerning Wiener-Wintner type convergence for linear and polynomial weights for the double ergodic theorem were announced in Assani, Duncan, Moore [2] and Assani, Moore [3]. Moreover, Theorem 4.1.9 gives a partial answer to the general question of the polynomial return time convergence, i.e., whether for every (ergodic) invertible system (X, μ, T), $f \in L^\infty(X, \mu)$ and a polynomial P with integer coefficients, the sequence $\{f(T^{p(n)}x)\}$ is for a.e. x a good weight for the pointwise ergodic theorem, i.e., whether for every other system (Y, ν, S) and $g \in L^\infty(Y, \nu)$, the averages

$$\frac{1}{N} \sum_{n \leq N} f(T^{p(n)}x) g(S^n y)$$

converge for a.e. y, see Assani, Presser [4, Question 7.1].

A random ergodic theorem with Hardy fields weights is presented in the recent preprint by Krause, Zorin-Kranich [32].

The paper is organized as follows:

In §2 we recall relevant definitions of Hardy field functions and develop the machinery to prove Theorem 4.1.6, as well as provide a uniform estimate for Hardy field weights using the Gowers-Host-Kra uniformity norms of the function f, see Theorem 4.2.11;

In §3 we prove a quantitative estimate on Weyl sums, which we then combine with a metric-entropy argument to prove Theorem 4.1.8; and

In §4 we prove Theorem 4.1.9.

4.1.1 Acknowledgements

This project began during the first author's research visit to the University of California, Los Angeles. Both authors are deeply grateful to Terence Tao for his great encouragement, input and support. They also thank Michael Boshernitzan, Kevin Hughes, and Pavel Zorin-Kranich for helpful conversations and comments as well as the referee for improvement suggestions. The support of the Hausdorff Research Institute and the Max Planck Institute for Mathematics in Bonn is also gratefully acknowledged.

4.1.2 Notation

We let $e(t) := e^{2\pi i t}$ denote the exponential.

We will make use of the modified Vinogradov notation. We use $X \lesssim Y$, or $Y \gtrsim X$ to denote the estimate $X \leq CY$ for an absolute constant C. If we need C to depend on a parameter, we shall indicate this by subscripts, thus for instance $X \lesssim_p Y$ denotes the estimate $X \leq C_p Y$ for some C_p depending on p. We use $X \approx Y$ as shorthand for $Y \lesssim X \lesssim Y$.

We also make use of big-O notation: we let $O(Y)$ denote a quantity that is $\lesssim Y$, and similarly $O_p(Y)$ a quantity that is $\lesssim_p Y$. Finally, we let $\widehat{O}(Y)$ denote a quantity which is $\leq Y$.

4.2 Wiener-Winter Convergence and Hardy Fields

In this section we prove our (uniform) Wiener-Wintner type Theorem 4.1.6 for weights which come from a Hardy field sequence as well as a uniform estimate of the averages (4.1) given in Theorem 4.2.11 below. If not explicitly stated otherwise, we will not need the assumption of invertibility. We begin with a few preliminary definitions and tools.

Definition 4.2.1. *We call a set A of uniformly bounded complex sequences a* set of (uniform) Wiener-Wintner weights *if for every ergodic measure-preserving system (X, μ, T) and every $f \in L^1(X, \mu)$ there is $X' \subset X$ with $\mu(X') = 1$ such that the averages*

$$\frac{1}{N} \sum_{n \leq N} a_n f(T^n x) \tag{4.2}$$

converge (uniformly) for every $x \in X'$ and every $\{a_n\} \in A$.

Recall further that for an ergodic measure-preserving system (X, μ, T), a point $x \in X$ is called *generic* for a function $f \in L^1(X, \mu)$ if it satisfies the assertion of Birkhoff's ergodic theorem, i.e.,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n \leq N} T^n f(x) \to \int_X f \, d\mu.$$

In order to be able to restrict ourselves to a dense subclass of functions we will need the following approximation lemma.

Lemma 4.2.2 (A Wiener-Wintner type Banach Principle)**.** *Let A be a set of uniformly bounded sequences $\{a_n\} \subset \mathbb{C}$. Then for every ergodic measure-preserving system (X, μ, T), the set of functions $f \in L^1(X, \mu)$ so that for almost every $x \in X$, the averages (4.2) converge for every $\{a_n\} \in A$ is closed in L^1-norm. The same holds for uniform convergence in $\{a_n\}$. Moreover, for $f = \lim_{j\to\infty} f_j$ in $L^1(X, \mu)$, the limits of the averages (4.2) for f_j converge to the limit of the averages (4.2) for f.*

Proof. Denote

$$C := \sup\{\|\{a_n\}\|_\infty, \ \{a_n\} \in A\}$$

and take (X,μ,T) an ergodic measure-preserving system and $f \in L^1(X,\mu)$. Observe first that

$$\left|\frac{1}{N}\sum_{n\le N} a_n T^n f\right| \le C\frac{1}{N}\sum_{n\le N} T^n|f|$$

holds for every $f \in L^1(x,\mu)$. Assume now that there is a sequence of functions

$$\{f_j\} \subset L^1(X,\mu)$$

with

$$\lim_{j\to\infty} \|f_j - f\|_{L^1(X)} = 0$$

so that for every f_j there is a set $X_j' \subset X$ with full measure such that for every $x \in X_j'$, the averages

$$\frac{1}{N}\sum_{n\le N} a_n T^n f_j$$

converge for every $x \in X_j'$ (uniformly) for every $\{a_n\} \in A$. Define

$$X' := \bigcap_j X_j' \cap \{x : x \text{ is generic for each } |f - f_j|\};$$

we then have $\mu(X') = 1$. Now, for every $x \in X'$ and $j \in \mathbb{N}$ we use the triangle inequality and the genericity assumption to majorize

$$\begin{aligned}
&\limsup_{N,M\to\infty} \left\{\sup_{\{a_n\}\in A} \left|\tfrac{1}{N}\textstyle\sum_{n\le N} a_n T^n f(x) - \tfrac{1}{M}\sum_{n\le M} a_n T^n f(x)\right|\right\} \\
&\le \limsup_{N,M\to\infty} \left\{\sup_{\{a_n\}\in A} \left|\tfrac{1}{N}\textstyle\sum_{n\le N} a_n T^n f_j(x) - \tfrac{1}{M}\sum_{n\le M} a_n T^n f_j(x)\right|\right\} \\
&\qquad + 2C\cdot \limsup_{N\to\infty} \tfrac{1}{N}\textstyle\sum_{n\le N} T^n|f - f_j|(x) \\
&\qquad = 2C\|f - f_j\|_{L^1(X)}.
\end{aligned}$$

Letting $j \to \infty$ finishes the argument. The assertion about the limit follows analogously, see e.g. [14, Lemma 21.7]. □

Remark 4.2.3. *By an inspection of the above argument and the fact that Birkhoff's averages* *converge uniformly in x for uniquely ergodic systems and continuous functions, one has the following variation of Lemma 4.2.2. Let A be a set of uniformly bouded sequences $\{a_n\} \subset \mathbb{C}$ and let (X, μ, T) be uniquely ergodic. Then the set of all continuous functions for which the averages (4.2) converge uniformly in $x \in X$ and $\{a_n\} \in A$ is closed in the L^1-norm.*

We will also need the following classical inequality, see e.g. Montgomery [36].

Lemma 4.2.4 (Van der Corput's inequality)**.** *Let $N \in \mathbb{N}$ and $u_1, \ldots, u_N \subset \mathbb{C}$ be with $|u_n| \leq 1$ for every $n = 1, \ldots, N$. Then for every $H \in \{1, \ldots, N\}$ the following inequality holds.*

$$\left|\frac{1}{N}\sum_{n\leq N} u_n\right|^2 \leq \frac{2(N+H)}{N^2(H+1)}\sum_{h=1}^{H}\left(1-\frac{h}{H+1}\right)\left|\sum_{n=1}^{N-h} u_{n+h}\overline{u_n}\right| + \frac{N+H}{N(H+1)}.$$

We now introduce Hardy fields and some of their properties. We refer the reader to Boshernitzan [7], Boshernitzan, Wierdl [9], Boshernitzan, Kolesnik, Quas, Wierdl [8], Frantzikinakis, Wierdl [22] and Frantzikinakis [19, 20] for further discussion of Hardy field functions and their applications to ergodic theory.

We call two real valued functions of one real variable that are continuous for large values of $s \in \mathbb{R}$ *equivalent* if they coincide for large $s \in \mathbb{R}$. Here and later, we say that a property holds for large s (or eventually) if it holds for every s in an interval of the form $[s_0, \infty)$. The equivalence classes under this relation are called *germs*. The set of all germs we denote by B which is a ring.

Definition 4.2.5. *A* Hardy field *is a subfield of B which is closed under differentiation. A Hardy field is called* maximal *if it is maximal among Hardy fields with respect to inclusion of sets. The union of all Hardy fields is denoted by $\mathcal{U}$.*

One can show that every maximal Hardy field contains the class $\mathcal{L}$ of logarithmico-exponential functions of Hardy, i.e., the class of functions which can be obtained by finitely many combinations of real constants, the variable s, log, exp, summation and multiplication. Thus, for example, it contains functions of the form s^α, $\alpha \in \mathbb{R}$.

Another property of Hardy fields is that each Hardy field is totally ordered with respect to the order $<_\infty$ defined by

$$f <_\infty g \quad \iff \quad f(s) < g(s) \quad \text{for all large } s.$$

Since the class $\mathcal{L}$ belongs to every maximal Hardy field, we conclude that every element of $\mathcal{U}$ is comparable to every logarithmico-exponential function. In particular, we can define the *type* of a function $p \in U$ to be

$$t(p) := \inf\{\alpha \in \mathbb{R} : \, |p(s)| < s^\alpha \text{ for large } s\}.$$

We say that p is *subpolynomial* if $t(p) < +\infty$, i.e., if $|p|$ is dominated by some polynomial. In particular, for eventually positive subpolynomial p with finite type there is $\alpha \in \mathbb{R}$ such that for every ε there is an s_0 so that

$$s^{\alpha-\varepsilon} < p(s) < s^{\alpha+\varepsilon}$$

holds for every $s > s_0$. Note that considering eventually positive p is not a restriction since every nonzero $p \in \mathcal{U}$ is either eventually positive or eventually negative.

We now consider subpolynomial elements of $\mathcal{U}$ with positive non-integer type, such as for example $p(s) = 5s^\pi + s\log s$. More precisely, we introduce following classes.

Definition 4.2.6. *For $\delta \in (0, 1/2)$, $M \geq 1$ and $m \in \mathbb{N}_0$ denote by $\mathcal{M}_{\delta,M,m}$ the set of all $p \in \mathcal{U}$ so that there exist $\alpha \in [\delta, 1-\delta]$, $k \leq m$ and $\varepsilon < \min\{(\alpha-\delta)/3, 1-\alpha-\delta\}$ with*

$$\frac{1}{M}s^{k+\alpha-\varepsilon-j} \leq p^{(j)}(s) \leq Ms^{k+\alpha+\varepsilon-j} \text{ for all } s \geq 1 \text{ and } j = 0, \ldots, k+1. \tag{4.3}$$

Remark 4.2.7. *By [8, Lemma 4.2], if $a(s), b(s) \in \mathcal{U}$ are non-polynomial with*

$$\lim_{x\to\infty} a(s)/b(s) = 0,$$

then $a'(s)$, and $b'(s)$ are non-polynomial and $\lim_{s\to\infty}\frac{a'(s)}{b'(s)} = 0$ too. Thus, by repeating the argument, every subpolynomial $p \in \mathcal{U}$ of positive non-integer type belongs to some class $\mathcal{M}_{\delta,M,m}$ after a possible left translation.

We begin our study of the classes $\mathcal{M}_{\delta,M,m}$ in the case where $m = 0$; this special case will anchor the inductive proof of our Theorem 4.1.6.

Lemma 4.2.8. *Let $\delta \in (0, 1/2)$ and $M \geq 1$. Then*

$$\lim_{N\to\infty} \sup_{p\in\mathcal{M}_{\delta,M,0}} \left| \frac{1}{N} \sum_{n\leq N} e(p(n)) \right| = 0.$$

Proof. Analogously to Kuipers, Niederreiter [33, Example 2.4] we use Euler's summation formula

$$\sum_{n=1}^{N} F(n) = \int_1^N F(t)\,dt + \frac{F(1)+F(N)}{2} + \int_1^N \left(\{t\} - \frac{1}{2}\right) F'(t)\,dt \tag{4.4}$$

for the function $F(t) := e(p(t))$ for $p \in \mathcal{M}_{\delta,M,0}$. We have

$$\frac{1}{N}\int_1^N |F'(t)|\,dt \leq \frac{2\pi M}{N}\int_1^N t^{\delta+\varepsilon-1}\,dt$$

converging to 0 uniformly in p. We also clearly have

$$\left|\frac{F(1)+F(N)}{2N}\right| \leq \frac{1}{N}.$$

To estimate the first summand on the right hand side of (4.4) observe

$$\left(\frac{1}{p'(t)}e(p(t))\right)' = 2\pi i e(p(t)) - \frac{p''(t)}{(p'(t))^2}e(p(t)).$$

This implies

$$\left|\frac{1}{N}\int_1^N F(t)\,dt\right| \;\leq\; \left(\frac{1}{p'(N)} + \frac{1}{p'(1)}\right)\frac{1}{2\pi N} + \frac{1}{2\pi N}\int_1^N \left|\frac{p''(t)}{(p'(t))^2}\right| dt =: I + II.$$

We have

$$\frac{1}{Np'(N)} \leq \frac{M}{N^{\alpha-\varepsilon}} < \frac{M}{N^{\delta}}$$

and therefore I converges to 0 uniformly in p. Moreover,

$$II \leq \frac{M^3}{2\pi N} \int_1^N \frac{t^{\alpha+\varepsilon-2}}{t^{2\alpha-2\varepsilon-2}}\, dt \leq \frac{M^3}{2\pi N} \int_1^N t^{-\delta}\, dt.$$

The assertion follows. □

Remark 4.2.9. *As pointed out to us by Michael Boshernitzan and follows from [7], for $p \in \mathcal{U}$ with subpolynomial growth, there are three different types of behavior of $\frac{1}{N}\sum_{n\leq N} e(p(n))$.*

1) Assume that there is a rational polynomial q such that $p-q$ is bounded (or, equivalently, has finite limit c) at ∞. Then one has

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n\leq N} e(p(n)) = e(c) \lim_{N\to\infty} \frac{1}{N} \sum_{n\leq N} e(q(n))$$

which exists but does not necessarily equal 0. For example, for p with negative type the above limit is 1.

2) Assume that there is a rational polynomial q so that $(p-q)(s)/\log s$ is bounded (or, equivalently, has finite limit) at ∞, but $p-q$ is unbounded. In this case the limit does not exist, see the proofs of [7, Theorem 1.3] and [33, Theorem I.2.6] based on the Hardy-Littlewood Tauberian theorem. Hence $(e(p(n)))$ is not a good weight even for the mean ergodic theorem.

3) For all other subpolynomial $p \in \mathcal{U}$, the sequence $(p(n))$ is uniformly distributed modulo 1, see [7, Theorem 1.3], and in particular one has

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n\leq N} e(p(n)) = 0.$$

This is for example the case for $p \in \mathcal{U}$ with finite non-integer type.

We now allow functions to have any positive non-integer type (i.e. we allow any $m \geq 0$), and prove our Theorem 4.1.6, restated below:

Theorem 4.2.10. *Let (X, μ, T) be an ergodic measure-preserving system, $f \in L^1(X, \mu)$, $\delta \in (0, 1/2)$, $m \in \mathbb{N}_0$ and $M \geq 1$. Then there is a subset $X' \subset X$ with $\mu(X') = 1$ such that the averages*

$$\sup_{p \in \mathcal{M}_{\delta,M,m}} \left| \frac{1}{N} \sum_{n \leq N} e(p(n)) T^n f(x) \right| \tag{4.5}$$

converge to 0 for every $x \in X'$. Moreover, if (X, μ, T) is uniquely ergodic and $f \in C(X)$, then one has

$$\lim_{N \to \infty} \sup_{p \in \mathcal{M}_{\delta,M,m}} \left\| \frac{1}{N} \sum_{n \leq N} e(p(n)) T^n f \right\|_\infty = 0.$$

In particular, the weights $(e(p(n)))$ are uniform Wiener-Wintner weights for $p \in \mathcal{M}_{\delta,M,m}$ and, if the system is invertible, Wiener-Wintner weights for subpolynomial $p \in \mathcal{U}$ with positive non-integer type. (For the last assertion recall that each such eventually positive p belongs to one of the classes $\mathcal{M}_{\delta,M,m}$ after a possible left translation.)

Proof. We will argue by induction on m and first discuss the case $m = 0$. Take $p \in \mathcal{M}_{\delta,M,0}$, i.e., assume that (4.3) holds for some

$$\alpha \in [\delta, 1 - \delta], \ \varepsilon < \min\{(\alpha - \delta)/3, 1 - \alpha - \delta\}$$

and $k = 0$. Let (X, μ, T) be an ergodic measure-preserving system. Recall the von Neumann decomposition

$$L^1(X, \mu) = \mathbb{C} \cdot 1 \oplus \overline{\{f - Tf, \ f \in L^\infty(X, \mu)\}}^{\|\cdot\|_1}. \tag{4.6}$$

For constant f, the averages (4.5) converge uniformly to 0 by Lemma 4.2.8. By (4.6) and Lemma 4.2.2, it remains to show the assertion for functions of the form $f - Tf$ for $f \in L^\infty(X, \mu)$.

Let $f \in L^\infty(X,\mu)$ with $\|f\|_\infty \le 1$ and observe by the telescoping sum argument

$$\begin{aligned}\frac{1}{N}\sum_{n\le N} e(p(n))(T^n f - T^{n+1}f)(x) &= \frac{e(p(1))Tf(x) - e(p(N))T^{N+1}f(x)}{N} \\ &+ \frac{1}{N}\sum_{n=1}^{N-1}(e(p(n+1)) - e(p(n)))T^n f(x).\end{aligned}$$

Take x with $|T^n f(x)| \le 1$ for every $n \in \mathbb{N}$. Then in the above, the first term is bounded by $2/N$ and the second by

$$\frac{1}{N}\sum_{n=1}^{N-1} |e(p(n+1)) - e(p(n))|. \tag{4.7}$$

By the mean value theorem and the assumption on p one has

$$|e(p(n+1)) - e(p(n))| \le 2\pi \sup_{s\in[n,n+1]} |p'(s)| \le 2\pi M n^{\alpha+\varepsilon-1}$$

and therefore (4.7) is bounded by $2\pi M$ times

$$\frac{1}{N}\sum_{n=1}^{N-1} n^{\alpha+\varepsilon-1} \le \frac{1}{N}\sum_{n=1}^{N-1} n^{-\delta}.$$

The uniform convergence to 0 follows.

The last assertion of the theorem follows analogously again using the fact that for a uniquely ergodic system (X,μ,T) and $f \in C(X)$, Birkhoff's ergodic averages converge uniformly in x and Remark 4.2.3.

After having established the case $m = 0$, assume that the theorem holds for $m \in \{1,\dots,k-1\}$ and we will show the assertion for $m = k$. Assume that p satisfies (4.3) for $m = k$. By Lemma 4.2.2 we can assume without loss of generality that $\|f\|_\infty \le 1$ and take x with $|T^n f(x)| \le 1$ for every $n \in \mathbb{N}$. We are going to use the van der Corput trick from Lemma 4.2.4 for $u_n := e(p(n))f(T^n x)$. Observe first that for $h \in \mathbb{N}$

$$u_{n+h}\overline{u_n} = e(p(n+h) - p(n))T^n(T^h f \cdot \overline{f})(x).$$

By Taylor's formula and the assumption on p we have

$$\begin{aligned}\left|p(n+h)-p(n)-p'(n)h-\ldots-\frac{p^{(k)}(n)h^k}{k!}\right| \\ \leq \frac{h^{k+1}\sup_{s\in[n,n+h]}|p^{(k+1)}(s)|}{(k+1)!} \\ \leq \frac{Mh^{k+1}n^{\alpha+\varepsilon-1}}{(k+1)!} \\ \leq \frac{Mh^{k+1}n^{-\delta}}{(k+1)!}.\end{aligned}$$

Define $q_h \in U$ by

$$q_h(s) := p'(s)h + \ldots + \frac{p^{(k)}(s)h^k}{k!}$$

and observe that $q_h \in M_{\delta,\tilde{M},k-1}$ for some $\tilde{M}$ depending on δ, M and h. Thus we have

$$u_{n+h}\overline{u_n} = e(q_h(n))T^n(T^h f \cdot \overline{f})(x) + O_{M,k}(h^{k+1}n^{-\delta}).$$

By Lemma 4.2.4, we thus have, for some constant $C_{M,k}$, depending only on M and k, and arbitrary $H, N \in \mathbb{N}$ with $H \leq N$

$$\begin{aligned}\left|\frac{1}{N}\sum_{n\leq N} u_n\right|^2 \quad &\leq \quad \frac{2(N+H)}{N(H+1)}\sum_{h=1}^{H}\left|\frac{1}{N-h}\sum_{n=1}^{N-h} e(q_h(n))T^n(T^h f\cdot\overline{f})(x)\right| \\ &+ \quad C_{M,k}\frac{(N+H)H^k}{N}\frac{1}{N}\sum_{n\leq N} n^{-\delta} + \frac{N+H}{N(H+1)}.\end{aligned}$$

Now take x for which additionally the assertion of the theorem is satisfied for functions $T^n f \cdot \overline{f}$ for every $n \in \mathbb{N}$ as well as δ, $m = k-1$ and $\tilde{M}$. Such x form a full measure set by the induction hypothesis and we conclude that for every $H \in \mathbb{N}$

$$\limsup_{N\to\infty} \sup_{p\in\mathcal{M}_{\delta,M,k}} \left|\frac{1}{N}\sum_{n\leq N} u_n\right|^2 \leq \frac{1}{(H+1)}.$$

Letting $H \to \infty$ finishes the argument. $\square$

It is natural to ask how restrictive the class of Hardy weights of positive non-integer type, or the classes $\mathcal{M}_{\delta,M,k}$, are for the (uniform) Wiener-Wintner convergence of the averages (4.1). For $p \in \mathcal{U}$ with transpolynomial growth, the behavior of $e(p(n))$ can be arbitrarily bad, see Boshernitzan [7, Theorem 1.6], so we restrict ourselves to subpolynomial $p \in \mathcal{U}$. Remark 4.2.9,2) shows that the family $(e(p(n))_{n\in\mathbb{N}})$, $p \in \mathcal{U}$ subpolynomial, is not a Wiener-Wintner family for the ergodic theorem. So one needs some restrictions on the class p to exclude functions from Remark 4.2.9,2) and, for convergence to *zero*, also to exclude functions from Remark 4.2.9,1). We remark that by Frantzikinakis [18, Section 3], the class of subpolynomial $p \in \mathcal{U}$ with positive non-integer type for which one has convergence of the averages (4.1) to zero cannot be enlarged to include all (even quadratic) irrational polynomials.

We now look at this question from a different perspective. As mentioned above, one of the problems for general p is possible divergence of $\frac{1}{N}\sum_{n\leq N} e(p(n))$ and hence divergence of weighted ergodic averages (4.1) for $f = \mathbf{1}$. If we restrict ourselves to a smaller class of functions f than the whole $L^1(X,\mu)$, then the class of p can be enlarged, as the following shows.

Let us consider the following (large) classes: For $m \in \mathbb{N}_0$, $\delta, M > 0$ we denote by $\mathcal{L}_{\delta,M,m}$ the class of all $p \in \mathcal{U}$ such that there exist $k \leq m$ with

$$|p^{(j)}(s)| \leq M s^{k-\delta-j} \text{ for all } s \geq 1 \text{ and } j = 0, \ldots, k. \tag{4.8}$$

By Remark 4.2.7, every subpolynomial $p \in \mathcal{U}$ belongs to some $\mathcal{L}_{\delta,M,m}$.

We recall that for $f \in L^\infty(X,\mu)$ the *Gowers-Host-Kra* (or *uniformity*) *seminorms* $\|\cdot\|_{U^m}$ are defined inductively as follows:

$$\begin{aligned} \|f\|_{U^1} &:= \left|\int_X f\, d\mu\right|, \\ \|f\|_{U^m}^{2^m} &:= \limsup_{N\to\infty} \frac{1}{N}\sum_{n\leq N} \|T^n f \cdot \overline{f}\|_{U^{m-1}}^{2^{m-1}}, \quad m \geq 2. \end{aligned}$$

For an equivalent definition and properties of the Gowers-Host-Kra seminorms we refer to

Host, Kra [24] and Eisner, Tao [15].

Theorem 4.2.11 (Uniform estimate of averages (4.1))**.** *Let (X, μ, T) be as above, let $m \in \mathbb{N}_0$ and $\delta, M > 0$. Then for every $f \in L^\infty(X, \mu)$ the inequality*

$$\limsup_{N\to\infty} \sup_{p\in\mathcal{L}_{\delta,M,m}} \left| \frac{1}{N} \sum_{n\le N} e(p(n)) T^n f(x) \right| \le \|f\|_{U^{m+1}}$$

holds for a.e. $x \in X$. Moreover, if (X, μ, T) is uniquely ergodic and $f \in C(X)$, then one has

$$\lim_{N\to\infty} \sup_{p\in\mathcal{L}_{\delta,M,m}} \left\| \frac{1}{N} \sum_{n\le N} e(p(n)) T^n f \right\|_\infty \le \|f\|_{U^{m+1}}.$$

In particular, for every f orthogonal to the Host-Kra factor $\mathcal{Z}_m$, the averages (4.1) converge a.e. to zero in $p \in \mathcal{U}$ with type strictly less than m (uniformly in $p \in \mathcal{L}_{\delta,M,m}$ for every δ and M), and the convergence is uniform in x whenever (X, μ, T) is uniquely ergodic and $f \in C(X)$.

Proof. We proceed by induction in m and assume without loss of generality that $\|f\|_\infty \le 1$.

Assume that $p \in \mathcal{L}_{\delta,M,0}$, i.e., $|p(s)| \le M s^{-\delta}$ for every $s \ge 1$. Then we have for every generic x with $|T^n f(x)| \le 1$ $\forall n \in \mathbb{N}$

$$\begin{aligned} \left| \frac{1}{N} \sum_{n\le N} e(p(n)) T^n f(x) \right| &\le \frac{1}{N} \sum_{n\le N} |e(p(n)) - 1| |T^n f(x)| + \left| \frac{1}{N} \sum_{n\le N} T^n f(x) \right| \\ &\le 2\pi M \cdot \frac{1}{N} \sum_{n\le N} n^{-\delta} + \left| \frac{1}{N} \sum_{n\le N} T^n f(x) \right| \end{aligned}$$

which converges to $|\int_X f\, d\mu| = \|f\|_{U^1}$ uniformly in $p \in \mathcal{L}_{\delta,M,0}$.

Assume now that the assertion holds for m and we show that it holds for $m+1$. Take $p \in \mathcal{L}_{\delta,M,m+1}$ and denote $u_n := e(p(n)) T^n f(x)$. Take $x \in X$ such that $|T^n f(x)| \le 1$ holds for every $n \in \mathbb{N}$. As in the proof of Theorem 4.2.10 observe

$$\begin{aligned} u_{n+h}\overline{u_n} &= e(p(n+h)-p(n))T^n(T^h f \cdot \overline{f})(x) \\ &= e(q_h(n))T^n(T^h f \cdot \overline{f})(x) + O_{M,m}(h^{m+1}n^{-\delta}) \end{aligned}$$

for $q_h(s) = hp'(s) + \ldots + \frac{h^m p^{(m)}(s)}{m!}$. The function $q_h \in \mathcal{U}$ satisfies $q_h \in \mathcal{L}_{\delta,\tilde{M},m}$ for some constant $\tilde{M}$ depending on M and h.

By Lemma 4.2.4, we thus have for some constant $C_{M,m}$, depending only on M and m, and arbitrary $H, N \in \mathbb{N}$ with $H \leq N$

$$\begin{aligned} \left| \frac{1}{N} \sum_{n \leq N} u_n \right|^2 &\leq \frac{2(N+H)}{N^2(H+1)} \sum_{h=1}^{H} \left(1 - \frac{h}{H}\right) \left| \sum_{n=1}^{N-h} e(q_h(n))T^n(T^h f \cdot \overline{f})(x) \right| \\ &+ C_{M,m} \frac{(N+H)H^k}{N} \frac{1}{N} \sum_{n \leq N} n^{-\delta} + \frac{N+H}{N(H+1)}. \end{aligned}$$

Now take x for which in addition the assertion of the theorem is satisfied for functions $T^n f \cdot \overline{f}$ for every $n \in \mathbb{N}$ as well as δ, m and $\tilde{M}$. Such x form a full measure set by the induction hypothesis and we conclude that for every $H \in \mathbb{N}$ using the Cauchy-Schwarz inequality (or the convexity of the function $s \mapsto s^{2^m}$)

$$\begin{aligned} \limsup_{N\to\infty} \sup_{p \in \mathcal{M}_{\delta,M,k}} \left| \frac{1}{N} \sum_{n \leq N} u_n \right|^2 &\leq \frac{2}{(H+1)} \sum_{h=1}^{H} \left(1 - \frac{h}{H+1}\right) \|T^h f \cdot \overline{f}\|_{U^m} + \frac{1}{H+1} \\ &\leq \frac{H+1}{H} \left(\frac{2}{H(H+1)} \sum_{h=1}^{H} (H+1-h) \|T^h f \cdot \overline{f}\|_{U^m}^{2^m} \right)^{1/2^m} + \frac{1}{H+1}. \end{aligned} \tag{4.9}$$

Since for every bounded sequence $\{a_h\}_{h=1}^{\infty} \subset \mathbb{C}$ and its partial sums $s_h := \sum_{j=1}^{h} a_j$, $h \in \mathbb{N}$, we have

$$\limsup_{H\to\infty} \frac{2}{H(H+1)} \sum_{h=1}^{H} (H+1-j)a_h = \limsup_{H\to\infty} \frac{2}{H(H+1)} \sum_{h=1}^{H} s_h \leq \limsup_{H\to\infty} \frac{s_H}{H},$$

letting $H \to \infty$ in (4.9) finishes the proof. $\square$

Remark 4.2.12. *1) By Lemma 4.2.2 and the inequality $\|f\|_{U^m} \le \|f\|_{L^{p_m}}$ for $p_m := \frac{2^m}{m+1}$, see Eisner, Tao [15], one can extend the assertion of Theorem 4.2.11 to every $f \in L^{p_{m+1}}(X,\mu)$.*

2) Theorem 4.2.11, the Host-Kra structure theorem, see [24], and Lemma 4.2.2 imply that the question of finding the largest (or a maximal) family of Wiener-Wiener weights (with limit zero or not) for general ergodic measure-preserving systems restricts to the question of finding such a family for nilsystems. Analogously, a Hardy field weight is a good weight for the pointwise ergodic theorem if and only if it is a good weight for nilsystems.

4.3 Wiener-Wintner Convergence of Twisted Square Means

In this section we present a Wiener-Wintner type result for weighted polynomial averages for a subclass of badly approximable numbers θ.

It is well-known that the set of badly approximable numbers (i.e., numbers which are badly approximable for some $c > 0$) build a perfect compact subset of $[0,1]$ with zero Lebesgue measure and Hausdorff dimension 1, see Hutchinson [26]. Moreover, θ is badly approximable if and only if its coefficients in the continued fraction expansion are bounded. We recall that the sequence $\{a_j\} \in \mathbb{N}$ is called the continued fraction expansion of $\theta \in (0,1)$ if

$$\theta = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}}$$

holds. There is the following relation between the constant c and the bound of the continued fraction expansion coefficients $M := \sup\{a_j,\, j \in \mathbb{N}\}$:

$$\frac{1}{\inf\{a_j,\, j \in \mathbb{N}\}} \le c \le \frac{1}{(M+2)(M+1)^2},$$

see Khintchine [30, Proof of Theorem 23].

The starting point in our analysis of Weyl sums (involving badly approximable θ) is as always the classical estimate of exponential sums due to Weyl, see e.g. Vaughan [41, Lemma 2.4].

Lemma 4.3.1 (Weyl)**.** *Let P be a polynomial of degree d with leading coefficient α, and let p, q be relatively coprime with $|\alpha - p/q| < 1/q^2$. Then for every $\varepsilon > 0$ and $N \in \mathbb{N}$,*

$$\left|\sum_{n=1}^{N} e(P(n))\right| \lesssim_{\varepsilon} N^{1+\varepsilon}\left(\frac{1}{q} + \frac{1}{N} + \frac{q}{N^d}\right)^{1/2^{d-1}}.$$

We have the following quantitative estimation of Weyl's sums for square polynomials with non-leading coefficient being c-badly approximable, see [21, Lemma A.5] for the case of the golden ratio. The proof is a quantification of the argument in [21].

Proposition 4.3.2. *Let θ be c-badly approximable. Then for every $\varepsilon > 0$*

$$\sup_{\alpha\in\mathbb{R}}\left|\frac{1}{N}\sum_{n\leq N} e(n\theta + n^2\alpha)\right| \lesssim_{\varepsilon} \frac{1}{cN^{1/32-\varepsilon}}.$$

Proof. Fix $\alpha \in \mathbb{R}$, $N \in \mathbb{N}$ and $\varepsilon > 0$. Since the assertion is effective only for $\varepsilon < 1/32$, we can assume that $\varepsilon < 1/32$. The Dirichlet principle implies the existence of p, q relatively prime with $q < N^{2-1/16}$ such that

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{qN^{2-1/16}}.$$

If $N^{1/16} \leq q < N^{2-1/16}$, then Weyl's Lemma 4.3.1 implies that for every $\varepsilon > 0$

$$\left|\frac{1}{N}\sum_{n\leq N} e(n\theta + n^2\alpha)\right| \lesssim_{\varepsilon} N^{\varepsilon}\left(\frac{1}{q} + \frac{1}{N} + \frac{q}{N^2}\right)^{1/2} \leq N^{\varepsilon}\left(\frac{2}{N^{1/16}} + \frac{1}{N}\right)^{1/2} \lesssim \frac{1}{N^{1/32-\varepsilon}}.$$

Since $c < 1$, the assertion is proved for such q.

Thus we may assume from now on that $q < N^{1/16}$. Again the Dirichlet principle applied to $q\theta$ implies the existence of t, u relatively prime with $u \leq N^{1/2}$ such that

$$\left|\theta - \frac{t}{uq}\right| \leq \frac{1}{uqN^{1/2}}.$$

By the assumption on θ we have on the other hand

$$\left|\theta - \frac{t}{uq}\right| \geq \frac{c}{q^2u^2}$$

and we have $cN^{1/2} \leq uq < uN^{1/16}$ implying

$$u > cN^{7/16}.$$

Take $M \in \mathbb{N}$ with $N^{1-1/16} \leq M \leq N$. We now show that the sums

$$S(M) := \sum_{n=1}^{M} e(n\theta + n^2p/q),$$

where α is replaced by its rational approximation p/q, satisfy

$$|S(M)| \lesssim_\varepsilon \frac{N^{1+\varepsilon-3/8}}{c} \tag{4.10}$$

independently of M.

Observe

$$|S(M)| = \left|\sum_{j=1}^{q} \sum_{k\geq 0, kq+j\leq M} e((qk+j)\theta + j^2p/q)\right| \leq \sum_{j=1}^{q} \left|\sum_{k\geq 0, kq+j\leq M} e((qk+j)\theta)\right|. \tag{4.11}$$

By Weyl's Lemma 4.3.1 using $|q\theta - t/u| \leq 1/(N^{1/2}u) \leq 1/u^2$, the sum inside satisfies for every $\varepsilon > 0$

$$\left|\sum_{k\geq 0, k\leq (M-j)/q} e((qk+j)\theta)\right| \leq_\varepsilon \left(\frac{M}{q}\right)^{1+\varepsilon} \left(\frac{1}{u} + \frac{q}{M} + \frac{uq}{M}\right).$$

Remembering that $M \in [N^{1-1/16}, N]$, $q < N^{1/16}$ and $u \in [cN^{7/16}, N^{1/2}]$, this is estimated by above by

$$\frac{N^{1+\varepsilon}}{q}\left(\frac{1}{cN^{7/16}} + \frac{1}{N^{7/8}} + \frac{1}{N^{3/8}}\right) \lesssim \frac{N^{1+\varepsilon}}{q}\frac{1}{cN^{3/8}}.$$

This together with (4.11) proves (4.10).

We finally estimate the desired exponential sums using the "rational" sums $S(M)$ and summation by parts. Observe

$$\begin{aligned}\left|\sum_{n=1}^{N} e(n\theta + n^2\alpha)\right| &\leq N^{1-1/16} + \left|\sum_{N^{1-1/16}\leq n\leq N} e(n\theta + n^2\alpha)\right| \\ &= N^{1-1/16} + \left|\sum_{N^{1-1/16}\leq n\leq N} e(n^2(\alpha - p/q))[S(n) - S(n-1)]\right|\end{aligned}$$

Since the discrete derivative of $e(n^2\gamma)$ satisfies

$$|e((n+1)^2\gamma) - e(n^2\gamma)| = |e((2n+1)\gamma) - 1| \lesssim n|\gamma|,$$

summation by parts and (4.10) implies

$$\begin{aligned}\left|\sum_{n=1}^{N} e(n\theta + n^2\alpha)\right| &\lesssim N^{1-1/16} + |S(N)| + |S([N^{15/16}] + 1)| + \sum_{N^{1-1/16}\leq n\leq N} n|\alpha - p/q||S(n)| \\ &\leq N^{1-1/16} + |S(N)| + |S([N^{15/16}] + 1)| + \sum_{N^{1-1/16}\leq n\leq N} |S(n)|N^{1/16-1} \\ &\lesssim_\varepsilon N^{1-1/16} + N\frac{N^{1+\varepsilon-3/8}}{cN^{1-1/16}} \\ &= N^{1-1/16} + c^{-1}N^{1+\varepsilon-3/16} \\ &\lesssim c^{-1}N^{1-1/16},\end{aligned}$$

since by assumption $\varepsilon \leq 1/32$. We used that

$$n|\alpha - p/q| \leq \frac{N}{qN^{2-1/16}} < N^{1/16-1}$$

in passing to the second line.

Thus we have

$$\sup_{\alpha\in\mathbb{R}}\left|\frac{1}{N}\sum_{n\leq N} e(n\theta+n^2\alpha)\right| \lesssim_\varepsilon \max\left\{\frac{1}{N^{1/32-\varepsilon}}, \frac{1}{cN^{1/16}}\right\},$$

completing the proof. ☐

Recall that the *upper Minkowski (or box) dimension* of a set $E\subset\mathbb{R}$ is given by

$$\overline{\dim}_{\text{box}}(E) := \limsup_{r\to\infty}\frac{N(\varepsilon)}{\log(1/\varepsilon)},$$

where $N(\varepsilon)$ is the minimal number of intervals with length ε needed to cover E. The upper Minkowski dimension of a set is always bigger than or equal to its Hausdorff dimension. It is well-known that one can replace $N(\varepsilon)$ in the above definition by the so-called *ε-metric entropy number*, i.e., the cardinality of the largest ε-net in E, where an *ε-net* is a set with distances between any two different elements being larger than or equal to ε. See Tao [39, Section 1.15] for basic properties of the (upper) Minkowski dimension.

We are ready for our uniform Wiener-Wintner type result for subsets of c-badly approximable numbers with small Minkowski dimension. We restate our result below for the reader's convenience:

Theorem 4.3.3. *Let $0 < c < 1$ and $E = E_c \subset [0,1]$ be a set of c-badly approximable numbers with upper Minkowski dimension strictly less than $1/16$.*

For every $f\in L^2(X)$ there exists a subset $X_f\subset X$ of full measure so that the averages

$$\left\{\frac{1}{N}\sum_{n\leq N} e(n\theta)T^{n^2}f(x)\right\}$$

converge uniformly *(in $\theta\in E_c$) to zero for all $x\in X_f$, where X_f is independent of $\theta\in E_c$.*

Proof. We first prove the result for f being a simple function, i.e., a finite linear combination of characteristic functions of measurable sets. By the boundedness of f and Rosenblatt, Wierdl [37, Lemma 1.5], it is enough to prove that for any lacunary constant $\rho > 1$

$$\lim_{N\to\infty,\ N=\lfloor\rho^k\rfloor} \sup_{\theta\in E} \left| \frac{1}{N} \sum_{n\le N} e(n\theta) T^{n^2} f(x) \right| = 0$$

holds for almost every $x \in X$. Note that for a fixed N, the function defined by

$$g_N(x) := \sup_{\theta\in E} \left| \frac{1}{N} \sum_{n\le N} e(n\theta) T^{n^2} f(x) \right|$$

is again a simple function and hence measurable. By the Borel-Cantelli lemma, it is enough to show that

$$\sum_{N=\lfloor\rho^k\rfloor} \left\| \sup_{\theta\in E} \left| \frac{1}{N} \sum_{n\le N} e(n\theta) T^{n^2} f(x) \right| \right\|_{L^2(X)}^2 < \infty;$$

this will be accomplished by showing that, provided

$$\overline{\dim}_{\text{box}}(E) < 1/16,$$

we have

$$\left\| \sup_{\theta\in E} \left| \frac{1}{N} \sum_{n\le N} e(n\theta) T^{n^2} f(x) \right| \right\|_{L^2(X)} = O(N^{-\nu}) \tag{4.12}$$

for some $\nu = \nu(\overline{\dim}_{\text{box}}(E)) > 0$.

Assume without loss of generality that $\|f\|_\infty \le 1$. Note first that for a fixed $\theta \in E$, the spectral theorem implies

$$\begin{aligned} \left\| \frac{1}{N} \sum_{n\le N} e(n\theta) T^{n^2} f(x) \right\|_{L^2(X)}^2 &= \int_{[0,1]} \left| \frac{1}{N} \sum_{n\le N} e(n\theta + n^2\alpha) \right|^2 d\mu_f(\alpha) \\ &\le \sup_{\alpha\in[0,1]} \left| \frac{1}{N} \sum_{n\le N} e(n\theta + n^2\alpha) \right|^2 \end{aligned}$$

for the corresponding spectral measure μ_f on $[0,1]$. Thus by Proposition 4.3.2 we have

$$\left\| \frac{1}{N} \sum_{n \leq N} e(n\theta) T^{n^2} f(x) \right\|_{L^2(X)} \lesssim_{\varepsilon} \frac{1}{cN^{1/32-\varepsilon}} \quad \text{for every } \theta \in E. \tag{4.13}$$

To show (4.12), suppose $\gamma < 1$ is such that

$$\overline{\dim}_{\text{box}}(E) = \gamma/16 < 1/16,$$

and choose $\beta > 1$ but so near to it that $\gamma\beta =: \kappa < 1$. Now, let Δ_N be a maximal $\frac{1}{10N^\beta}$-net in E, i.e., a set of maximal cardinality of points in E with distance between any two distinct points being larger than $\frac{1}{10N^\beta}$.

Since $\overline{\dim}_{\text{box}}(E) = \gamma/16$, we remark that for all N sufficiently large

$$\log |\Delta_N| \leq \overline{\dim}_{\text{box}}(E) \cdot \log(N^\beta) = \log(N^{\gamma\beta/16}),$$

or

$$|\Delta_N| \leq N^{\kappa/16}. \tag{4.14}$$

Now, take $x \in X$ such that $|T^{n^2} f(x)| \leq \|f\|_\infty \leq 1$ holds for every $n \in \mathbb{N}$; note that the set of such x has full μ-measure. The polynomial p_N defined by

$$p_N(\theta) := \frac{1}{N} \sum_{n \leq N} e(n\theta) T^{n^2} f(x)$$

satisfies $|p_N(\theta)| \leq 1$ and by inspection (or Bernstein's polynomial inequality [5])

$$\|p_N'\|_\infty \leq 2\pi N.$$

Take now $\theta \in E$ and denote by τ the nearest element to θ from Δ_N. By the maximality of Δ_N we have

$$|\theta - \tau| \leq \frac{1}{10N^{\beta}}.$$

By the mean-value theorem we may therefore estimate

$$\begin{aligned}
|p_N(\theta)| &\leq |p_N(\tau)| + |\theta-\tau| 2\pi N \\
&\leq \max_{\theta' \in \Delta_N} |p_N(\theta')| + N^{1-\beta} \\
&\leq \left(\sum_{\theta' \in \Delta_N} |p_N(\theta')|^2 \right)^{1/2} + N^{1-\beta},
\end{aligned}$$

replacing the l^{∞}-norm with the larger l^2-norm.

We may consequently estimate using (4.13) and (4.14)

$$\begin{aligned}
\left\| \sup_{\theta \in E} \left| \frac{1}{N} \sum_{n \leq N} e(n\theta) T^{n^2} f(x) \right| \right\|_{L^2(X)} &\leq \left\| \left(\sum_{\theta' \in \Delta_N} |p_N(\theta')|^2 \right)^{1/2} \right\|_{L^2(X)} + N^{1-\beta} \\
&= \left(\sum_{\theta' \in \Delta_N} \|p_N(\theta')\|_{L^2(X)}^2 \right)^{1/2} + N^{1-\beta} \\
&\lesssim_{\varepsilon} c^{-1} |\Delta_N|^{1/2} \cdot N^{-1/32+\varepsilon} + N^{1-\beta} \\
&\leq c^{-1} N^{\kappa/32} \cdot N^{-1/32+\varepsilon} + N^{1-\beta} \\
&\leq c^{-1} N^{\frac{\kappa-1}{32}+\varepsilon} + N^{1-\beta}.
\end{aligned}$$

Choosing $\varepsilon < \frac{1-\kappa}{32}$ finishes the proof of (4.12).

This concludes the argument in the case where f is a simple function.

To extend the result to all of $L^2(X)$, we observe that the maximal function satisfies

$$\sup_{\theta \in E} \left| \frac{1}{N} \sum_{n \leq N} e(n\theta) T^{n^2} f(x) \right| \leq \mathcal{M}_{N,\mathrm{sq}} f := \frac{1}{N} \sum_{n \leq N} T^{n^2} |f|(x),$$

where the maximal function for squares on the right hand side is L^2-bounded by the celebrated result of Bourgain [10].

Take now $f \in L^2(X)$ is arbitrary, and g a simple function. For any $\epsilon > 0$, we have the containment

$$\begin{aligned}
&\left\{x : \limsup_N \sup_{\theta\in E} \left|\frac{1}{N}\sum_{n\le N} e(n\theta)T^{n^2}f(x)\right| > \epsilon\right\} \\
&\quad\subset \left\{x : \limsup_N \sup_{\theta\in E} \left|\frac{1}{N}\sum_{n\le N} e(n\theta)T^{n^2}g(x)\right| > \epsilon/2\right\} \cup \left\{x : \limsup_N \mathcal{M}_{N,\mathrm{sq}}(f-g)(x) > \epsilon/2\right\} \\
&\quad\subset \left\{x : \sup_N \mathcal{M}_{N,\mathrm{sq}}(f-g)(x) > \epsilon/2\right\}
\end{aligned}$$

by the previous considerations for simple functions. We thus have by Bourgain's maximal inequality the upper estimate

$$\begin{aligned}
&\mu\left(\left\{x : \limsup_N \sup_{\theta\in E} \left|\frac{1}{N}\sum_{n\le N} e(n\theta)T^{n^2}f(x)\right| > \epsilon\right\}\right) \\
&\quad\le \mu\left(\left\{x : \sup_N \mathcal{M}_{N,\mathrm{sq}}(f-g)(x) > \epsilon/2\right\}\right) \\
&\quad\le \frac{4}{\epsilon^2}\left\|\sup_N \mathcal{M}_{N,\mathrm{sq}}(f-g)\right\|_{L^2(X)}^2 \\
&\quad\lesssim \frac{1}{\epsilon^2}\|f-g\|_{L^2(X)}^2.
\end{aligned}$$

Since the last quantity can be made as small as we wish (independent of ϵ) the result follows. □

Example 1 (Sets of badly approximable numbers with small Minkowski dimension)**.** *Let $A = \{a_1, \dots, a_k\} \subset \mathbb{N}$ and denote by E_A the set of all $\theta \in [0,1]$ which coefficients in the continued fraction expansion all belong to A. By Jenkinson [28], the set $\{\dim_H(E_A)\}$ of Hausdorff's dimensions of the sets E_A is dense in $[0,1/2]$ (in fact, it is even dense in [0,1] as showed by Kesseböhmer, Zhu [29]). Moreover, if $1 \notin A$, then the Hausdorff and the Minkowski dimension of E_A concide, see Falconer [17, Theorem 23]. Therefore, infinitely many sets E_A satisfy the condition in Theorem 4.3.3.*

4.4 Pointwise Convergence of the Twisted Polynomial Means

In this section we consider the behavior of the twisted means corresponding to more general polynomial shifts. We also relax our *badly approximable* hypothesis and now allow θ to be arbitrary. Although establishing a Wiener-Wintner type theorem for arbitrary θ seems very difficult under these general assumptions, we are able to establish pointwise convergence for such means. Specifically, we prove the following

Theorem 4.4.1. *For any measure-preserving system, (X, μ, τ), any $\theta \in [0,1]$, and any polynomial $P(n)$ with integer coefficients, the twisted means*

$$M_t^\theta f(x) := \frac{1}{t} \sum_{n \leq N} e(n\theta) \tau^{P(n)} f(x),$$

converge μ-a.e. for any $f \in L^p(X),\ p > 1$,

4.4.1 Strategy

Our strategy is as follows:

As previously remarked, using the majorization

$$\sup_t |M_t^\theta f| \leq \mathcal{M}_P |f| := \sup_t \frac{1}{t} \sum_{n \leq t} \tau^{P(n)} |f|(x),$$

we see that the set of functions in $L^p,\ p > 1$ for which μ-a.e. convergence holds is closed, since Bourgain's square polynomial maximal function $\mathcal{M}_P$ is bounded on L^p [11].

Consequently, in proving convergence we may work exclusively with bounded $f \in L^\infty$.

For such functions, it suffices to prove that for any $\rho > 1$, the means

$$\{M_t^\theta f : t \in \lfloor \rho^{\mathbb{N}} \rfloor\} =: \{M_t^\theta f : t \in I_\rho\}$$

converge μ-a.e.; we establish this result through a (long) *variational estimate* on the $\{M_t^\theta f\}$:

Definition 4.4.2. *For $0 < r < \infty$, we define the r-variation of the means $\{M_t^\theta f\}$*

$$\mathcal{V}^r(M_t^\theta f)(x) := \sup_{(t_k) \text{ increasing}} \left(\sum_k |M_{t_k}^\theta f - M_{t_{k+1}}^\theta f|^r \right)^{1/r} (x).$$

(The endpoint $\mathcal{V}^\infty(M_t^\theta f)(x) := \sup_{t,s} |M_t^\theta f - M_s^\theta f|(x)$ is comparable to the maximal function $\sup_t |M_t^\theta f|(x)$, and so is typically not introduced.)

By the nesting of little l^p spaces, we see that the variation operators grow more sensitive to oscillation as r decreases. (We shall restrict our attention to the range $2 < r < \infty$.) This sensitivity is reflected in the fact that although having bounded r-variation, $r < \infty$, is enough to imply pointwise convergence, there are collections of functions which converge, but which have unbounded r variation for *any* $r < \infty$. (e.g. $\{(-1)^i \frac{1}{\log i+1}\}$)

To prove that the $\mathcal{V}^r(M_t^\theta f) < \infty$ converges almost everywhere, we will show that they are *bounded* operators on $L^2(X)$. In particular, we will prove the following

Proposition 4.4.3. *There exists an absolute $C_{r,\rho,\theta,P}$ so that*

$$\|\mathcal{V}^r(M_t^\theta f : t \in I_\rho)\|_{L^2(X)} \leq C_{r,\rho,\theta,P} \|f\|_{L^2(X)}$$

provided $r > 2$.

By the transference principle of Calderón [13], this result will follow from the analogous one on the integer lattice:

Proposition 4.4.4. *Suppose $r > 2$. Then there exists an absolute constant $C_{r,\rho,\theta,P}$ so that for any $f \in l^2(\mathbb{Z})$,*

$$\|\mathcal{V}^r(K_N^\theta * f)\|_{l^2(\mathbb{Z})} \leq C_{r,\rho,\theta,P} \|f\|_{l^2(\mathbb{Z})},$$

where

$$K_N^\theta * f(x) := \frac{1}{N} \sum_{n \leq N} e(-n\theta) f(x - P(n))$$

is the discrete convolution operator.

Remark 4.4.5. *For notational ease, we have chosen to work with the above convolution kernels; by replacing θ and P with $-\theta$, $-P$, we prove an analogous result for*

$$\bar{K}_N^\theta * f(x) := \frac{1}{N} \sum_{n \leq N} e(n\theta) f(x + P(n))$$

which may be transferred appropriately.

It is this proposition to which we now turn.

4.4.2 Preliminaries

Fix throughout $0 < \delta \ll 1$.

Let $\rho > 1$ be a temporarily fixed lacunarity constant, and set

$$I = I_\rho := \{\lfloor \rho^k \rfloor : k \in \mathbb{N}\};$$

we will enumerate such elements using capital N, M, K, etc.

This is an L^2 problem, so we will use the Fourier transform. Our strategy will be to replace the twisted multipliers

$$\begin{aligned}\widehat{K_N}(\alpha) &:= \frac{1}{N} \sum_{n \leq N} e(P(n)\alpha - n\theta) \\ &= \frac{1}{N} \sum_{n \leq N} e(m_d \alpha \cdot n^d + \cdots + m_d \alpha \cdot n^2 + (m_1 \alpha - \theta) \cdot n)\end{aligned}$$

with increasingly tractable families of multipliers which remain L^2-close. [1]

The arithmetic properties of θ figure centrally in this approximation, so to organize our approach we introduce N-θ *rational approximates*; their significance is that the so-called N-major boxes for $\widehat{K_N}(\alpha)$ (introduced below) only appear "arithmetically near" such approximates.

[1] Our projection-based approach is based off of the excellent exposition of [37].

4.4.3 N-θ Rational Approximates

Regarding $\theta \in [0,1]$ as given, for $N \in I_\rho = I$ we define the N-θ rational approximates $\{\frac{x_N}{y_N}\} \subset [0,1]$ to be rational numbers (in reduced form) with

- $y_N \leq m_d N^\delta$;
- $|\gamma_N(\theta)| = |\gamma_N| := |\frac{x_N}{y_N} - \theta| \leq 2N^{\delta-1}$.

Some remarks are in order:

For many N there need not exist N-θ rational approximates, but – for sufficiently large N – such N-θ approximates are unique if they exist: if x/y, p/q were distinct N-θ approximates, we would have

$$\left(\frac{1}{m_d N^\delta}\right)^2 \leq \frac{1}{yq} \leq |x/y - p/q| \leq |x/y - \theta| + |\theta - p/q| \leq 4N^{\delta-1},$$

for the desired contradiction.

Many N may share the same N-θ rational approximate; for example, if θ is very close to 0, then $0/1$ will serve as an N-θ approximate many times over.

To make matters clearer, we enumerate the *distinct* N-θ rational approximates $\{\frac{x_{N_j}}{y_{N_j}} : j\}$ according to the size of pertaining N_j. Note that if θ is rational, or more generally *badly approximable*, there are only finitely many distinct N-θ rational approximates.

Moreover, distinct N-θ approximates are necessarily sparsely spaced.

Suppose $N < N_j$; then arguing as above we see

$$\frac{1}{m_d N^\delta y_{N_j}} < |x_N/y_N - x_{N_j}/y_{N_j}| < 4N^{\delta-1},$$

and thus $N^{1-2\delta} \lesssim_P y_{N_j}$.

4.4.4 Major Boxes

For $a/b, p_{d-1}/q_{d-1}, \ldots, p_2/q_2, x/y \in [0,1]$ with

$$\mathrm{lcm}(b, q_{d-1}, \ldots, q_2, y) \leq N^{\delta},$$

we define the N-major box

$$\begin{aligned}&\mathfrak{M}_N(a/b, p_{d-1}/q_{d-1}, \ldots, p_2/q_2, x/y)\\ &\quad := \left\{(\alpha_d, \alpha_{d-1}, \ldots, \alpha_2, \alpha_1) : |\alpha_d - a/b| \leq N^{\delta-d},\ |\alpha_i - p_i/q_i| \leq N^{\delta-i},\ |\alpha_1 - x/y| \leq N^{\delta-1}\right\}.\end{aligned}$$

By Bourgain's [11, Lemma 5.6], we know that if $(\alpha_d, \ldots, \alpha_1)$ does not lie in some major box, there exists some (small) $\kappa > 0$ so that

$$\left|\frac{1}{N}\sum_{n\leq N} e(\alpha_d n^d + \cdots + \alpha_1 n)\right| \lesssim N^{-\kappa}.$$

In our present context, this means that

$$\widehat{K_N}(\alpha) := \frac{1}{N}\sum_{n\leq N} e(m_d\alpha\cdot n^d + \cdots + m_2\alpha\cdot n^2 + (m_1\alpha - \theta)\cdot n)$$

is $O(N^{-\kappa})$ unless there exist some tuple

$$(a/b, p_{d-1}/q_{d-1}, \ldots, p_2/q_2, x/y),\ \mathrm{lcm}(b, q_{d-1}, \ldots, q_2, y) \leq N^{\delta},$$

so

$$\begin{aligned} m_d\alpha &\equiv a/b + \widehat{O}(N^{\delta-d}) \mod 1,\\ m_i\alpha &\equiv p_i/q_i + \widehat{O}(N^{\delta-i}) \mod 1,\ 2\leq i\leq d-1, \quad \text{and}\\ m_1\alpha - \theta &\equiv x/y + \widehat{O}(N^{\delta-1}) \mod 1.\end{aligned}$$

The first point means that there exists some $0 \leq j \leq m_d - 1$

$$\alpha \equiv \frac{1}{m_d}(j + a/b) + \widehat{O}\left(\frac{1}{m_d} N^{\delta - d}\right) \mod 1$$

which in turn forces

$$m_i \alpha \equiv \frac{m_i}{m_d}(j + a/b) + \widehat{O}\left(\frac{m_i}{m_d} N^{\delta - d}\right) \mod 1$$

and thus, for $2 \leq i \leq d-1$,

$$\frac{m_i}{m_d}(j + a/b) \equiv p_i/q_i + \widehat{O}(2N^{\delta-1}) \mod 1$$

for N sufficiently large. Since the left side of the foregoing has a priori denominator $\leq m_d N^\delta$, while the right side has denominator $\leq N^\delta$, we in fact must have

$$\frac{m_i}{m_d}(j + a/b) \equiv p_i/q_i \mod 1.$$

Turning to the $i = 1$ term: we have on the one hand

$$m_1 \alpha \equiv \frac{m_1}{m_d}(j + a/b) + \widehat{O}\left(\frac{m_1}{m_d} N^{\delta - d}\right) \mod 1$$

while on the other hand

$$m_1 \alpha - \theta \equiv x/y + \widehat{O}(N^{\delta-1}) \mod 1.$$

This forces (for N sufficiently large)

$$\frac{m_1}{m_d}(j + a/b) \equiv x/y + \theta + \widehat{O}(2N^{\delta-1}) \mod 1$$

and thus

$$\frac{m_1}{m_d}(j + a/b) - x/y \equiv \theta + \widehat{O}(2N^{\delta-1}) \mod 1.$$

Since the left hand side of the above expression has denominator of size $\leq m_d N^\delta$, it must be $\equiv x_N/y_N \mod 1$, i.e. we must have

$$\frac{m_1}{m_d}(j + a/b) - x/y \equiv x_N/y_N \mod 1.$$

This forces

$$\frac{m_1}{m_d}(j + a/b) - x_N/y_N \equiv x/y \mod 1.$$

The upshot is that, beginning with the assumption that

$$(m_d\alpha, \ldots, m_2\alpha, m_1\alpha - \theta) \in \mathfrak{M}_N(a/b, p_{d-1}/q_{d-1}, \ldots, p_2/q_2, x/y)$$

for *some* major box, with

$$\alpha \equiv \frac{1}{m_d}(j + a/b) + \widehat{O}\left(\frac{1}{m_d}N^{\delta-d}\right) \mod 1,$$

we have deduced the relationship

$$\begin{aligned} \frac{p_i}{q_i} &\equiv \frac{m_i}{m_d}(j + a/b) \mod 1 \\ \frac{x}{y} &\equiv \frac{m_1}{m_d}(j + a/b) - x_N/y_N \mod 1. \end{aligned}$$

Remark 4.4.6. *If θ is badly approximable, so there are only finitely many N-major boxes, then for all N large we see that the twisted polynomial means are $O(N^{-\kappa})$. By arguing as in the previous section, uniform Wiener-Wintner theorems can be proven for the twisted polynomial means for badly approximable θ which live in sets of upper Minkowski dimension $< 2\kappa$. We will not focus on attaining the best possible numerical constant κ in our present context.*

In fact, a little more is true: for $\alpha \in \mathfrak{M}_N(a/b, p_{d-1}/q_{d-1}, \ldots, p_2/q_2, x/y)$ we knew a priori that for $2 \leq i \leq d$

$$m_i\alpha \equiv p_i/q_i + \widehat{O}(N^{\delta-i}) \mod 1;$$

using the above determined structure, i.e. $\alpha \equiv \frac{1}{m_d}(j + a/b) + \widehat{O}\left(\frac{1}{m_d}N^{\delta-d}\right)$, we in fact find that

$$m_i\alpha \equiv \frac{m_i}{m_d}(j + a/b) + \widehat{O}\left(\frac{m_i}{m_d}N^{\delta-d}\right) \mod 1.$$

We're now ready for a working characterization of our major boxes.

For $a/b \in \mathbb{Q} \cap [0, 1]$ in reduced form, define by equivalence $\mod 1$

$$\frac{\widehat{a_i^j}}{\widehat{b_i^j}} :\equiv \frac{m_i}{m_d}(j + a/b) \text{ for } 2 \leq i \leq d-1$$
$$\frac{\widehat{a_{1,N}^j}}{\widehat{b_{1,N}^j}} :\equiv \frac{m_1}{m_d}(j + a/b) - x_N/y_N,$$

and set

$$b_N^j := \operatorname{lcm}(b, \widehat{b_{d-1}^j}, \ldots, \widehat{b_2^j}, \widehat{b_{1,N}^j}).$$

Collect all $b_N^j \leq N^\delta$ in

$$A_N^0 := \{a/b \neq 0/1 : b_N^0 \leq N^\delta\}$$
$$A_N^j := \{a/b : b_N^j \leq N^\delta\}, \ 1 \leq j \leq m_{d-1},$$

(these sets are possibly empty) and set

$$A_N := \bigcup_{j=0}^{m_d-1} A_N^j.$$

If we define

$$\mathfrak{M}_N^j(a/b) := \left\{\alpha : \alpha \equiv \frac{1}{m_d}(j + a/b) + \widehat{O}\left(\frac{1}{m_d}N^{\delta-d}\right) \mod 1\right\}$$

then we have shown that the union of the major boxes

$$\bigcup_{(a/b,p_{d-1}/q_{d-1},\ldots,p_2/q_2,x/y):\mathrm{lcm}(b,q_{d-1},\ldots,q_2,y)\leq N^\delta} \mathfrak{M}_N(a/b,p_{d-1}/q_{d-1},\ldots,p_2/q_2,x/y)$$

is (simply) contained in

$$\bigcup_{j=0}^{m_d-1} \bigcup_{a/b\in A_N^j} \mathfrak{M}_N^j(a/b).$$

Before proceeding, we further decompose our sets A_N^j, A_N according to the size of denominators of its elements. Specifically, for $2^t \leq N^\delta$, we define

$$A_{N,t}^j := \{a/b : b_N^j \approx 2^t\},$$

and $A_{N,t} := \bigcup_{j=0}^{m_d-1} A_{N,t}^j$. (Note that $|A_{N,t}| \lesssim_P 4^t$.)

We isolate the following simple subdivision lemma:

Lemma 4.4.7. *For each t, there is at most one scale $l(t)$ so that if $A_{N,t}^j \neq \emptyset$ for any $0 \leq j \leq m_d - 1$, we necessarily have $N_{l(t)} \leq N < N_{l(t)+1}$.*

Proof. Seeking a contradiction, suppose there existed some $N_k < N_l$ so that

$$N_k \leq N < N_{k+1} \leq N_l \leq M < N_{l+1},$$

and $2^t \leq N^\delta < M^\delta$, but

$$A_N^j,\ A_M^h \neq \emptyset$$

for some $0 \leq j, h \leq m_d - 1$. This means we may find some $a/b \in A_N^j$ so that

$$\widehat{b_{1,N}^j} \leq b_N^j \lesssim 2^t \leq N^\delta,$$

which forces $y_{N_k} = y_N \lesssim m_d 2^t$.

Similarly, we may find some $p/q \in A^h_M$ so that

$$\widehat{q^h_{1,M}} \leq q^h_M \lesssim 2^t \leq N^\delta,$$

which forces $y_{N_l} = y_M \lesssim m_d 2^t$, which means we have

$$y_{N_l} \lesssim m_d 2^t \leq m_d N^\delta.$$

But this contradicts the sparsity condition, $N^{1-2\delta} \lesssim_P y_{N_l}$, and completes the proof. $\square$

4.4.5 The Multiplier on Major Arcs

We now proceed to describe the shape of the twisted multiplier

$$\widehat{K_N}(\alpha) = \frac{1}{N} \sum_{n \leq N} e(m_d \alpha \cdot n^d + \cdots + m_d \alpha \cdot n^2 + (m_1 \alpha - \theta) \cdot n)$$

on $\mathfrak{M}^j_N(a/b)$.

For $a/b \in A^j_N$ define the weight

$$S^j_N(a/b) := \frac{1}{b^j_N} \sum_{r=1}^{b^j_N} e\left(r^d \frac{a}{b} + r^{d-1} \frac{\widehat{a^j_{d-1}}}{\widehat{b^j_{d-1}}} + \cdots + r \frac{\widehat{a^j_{1,N}}}{\widehat{b^j_{1,N}}} \right);$$

we remark that

$$S^0_N(0/1) = \frac{1}{y_N} \sum_{r=1}^{y_N} e(-r x_N / y_N).$$

We note that for any fixed $0 \ll \nu < 1/d$ we have

$$|S^j_N(a/b)| \lesssim (b^j_N)^{-\nu}$$

by Hua's [25, §7, Theorem 10.1].

Define further the oscillatory (twisted) "pseudo-projection"

$$V_N(\beta) := \int_0^1 e(N^d m_d t^d \beta + N t \gamma_N)\, dt,$$

and observe that

$$\left| V_N(\beta) - \int_0^1 e(N t \gamma_N)\, dt \right| =: |V_N(\beta) - \omega_N| \lesssim N^d m_d |\beta|$$

by the mean-value theorem, while

$$|V_N(\beta)| \lesssim \frac{1}{N m_d^{1/d} |\beta|^{1/d}}$$

by van der Corput's estimate on oscillatory integrals [38, §8].

We have the following approximation lemma:

Lemma 4.4.8. *On $\mathfrak{M}_N^j(a/b)$, if $a/b \in A_N^j$*

$$\widehat{K_N}(\alpha) = S_N^j(a/b) V_N\left(\alpha - \frac{1}{m_d}(j + a/b)\right) + O(N^{2\delta - 1}).$$

If $|\alpha| \leq \frac{1}{m_d} N^{\delta - d}$,

$$\widehat{K_N}(\alpha) = S_N^0(0/1) V_N(\alpha) + O(N^{2\delta-1}).$$

Proof. This follows by arguing as in the proof of [31, Proposition 4.1]. □

Let $\chi := 1_{[-1,1]}$ denote the indicator function of $[-1, 1]$. If we define via the Fourier transform

$$\begin{aligned} \widehat{{}^0R_N}(\alpha) &:= S_N^0(0/1) V_N(\alpha) \chi(N^{d-\delta} m_d \alpha) \\ \widehat{R_N}(\alpha) &:= \sum_{j=0}^{m_d - 1} \sum_{a/b \in A_N^j} S_N^j(a/b) V_N\left(\alpha - \frac{1}{m_d}(j + a/b)\right) \chi\left(N^{d-\delta} m_d \left(\alpha - \frac{1}{m_d}(j + a/b)\right)\right) \end{aligned}$$

then in light of the previous lemma we have

$$\widehat{K_N}(\alpha) = \widehat{{}^0R_N}(\alpha) + \widehat{R_N}(\alpha) + O(N^{2\delta-1}),$$

and thus

$$\mathcal{V}^r(K_N * f) \leq \mathcal{V}^r\left({}^0R_N * f\right) + \mathcal{V}^r(R_N * f) + \left(\sum_N \left|\left({}^0R_N * f + R_N * f\right) - K_N * f\right|^2\right)^{1/2}.$$

By Parseval's inequality, the square sum has l^2 norm $\lesssim_\rho 1$, so we need only l^2-bound

$$\mathcal{V}^r\left({}^0R_N * f\right)$$

and

$$\mathcal{V}^r(R_N * f).$$

The following transference lemma, essentially due to Bourgain [11, Lemma 4.4] will allow us to view our operators as acting on the real line $\mathbb{R}$:

Lemma 4.4.9. *Suppose $r > 2$, and $\{m_N\}$ are uniformly bounded multipliers on $A \subset [-1,1]$, a fundamental domain for $\mathbb{T} \cong \mathbb{R}/\mathbb{Z}$. If*

$$\left\| \mathcal{V}^r\left(\left(m_N \widehat{f}\right)^\vee\right)\right\|_{L^2(\mathbb{R})} \lesssim \|f\|_{L^2(\mathbb{R})}$$

then

$$\left\| \mathcal{V}^r\left(\left(m_N \widehat{f}\right)^\vee\right)\right\|_{l^2(\mathbb{Z})} \lesssim \|f\|_{l^2(\mathbb{Z})}$$

as well.

Since $\widehat{{}^0R_N}$ are supported in (say) $[-1/2, 1/2]$, while $\widehat{R_N}$ are supported in $[0,1]$, the above transference lemma applies. In what follows, we therefore (abuse notation and) view our operators as acting on $f \in L^2(\mathbb{R})$.

We begin our analysis with the simpler case $\mathcal{V}^r\left({}^0R_N * f\right)$ which already contains the core of our method.

4.4.6 Proof of Proposition 4.4.4

$\mathcal{V}^r({}^0R_N * f)$ is L^2-Bounded. We define, again via the Fourier transform, the more tractable approximate operators

$$\widehat{{}^0T_N}(\alpha) := S_N^0(0/1)\omega_N\chi(N^d m_d\alpha).$$

The claim is that

$$\sup_\alpha \sum_N \left| \widehat{{}^0R_N} - \widehat{{}^0T_N} \right|^2 (\alpha) \lesssim 1;$$

assuming this result, a square function argument as above will reduce the problem to bounding $\mathcal{V}^r({}^0T_N * f)$ in L^2.

Viewing α as fixed, we expand the foregoing as

$$\begin{aligned}
&\sum_N |S_N^0(0/1)|^2 |V_N(\alpha) - \omega_N\chi(N^d m_d\alpha)|^2 |\chi(N^{d-\delta}m_d\alpha)|^2 \\
&\quad = \sum_j |S_{N_j}^0(0/1)|^2 \sum_{N:\frac{x_N}{y_N}=\frac{x_{N_j}}{y_{N_j}}} |V_N(\alpha) - \omega_N\chi(N^d m_d\alpha)|^2 |\chi(N^{d-\delta}m_d\alpha)|^2 \\
&\quad \lesssim \sum_j |S_{N_j}^0(0/1)|^2 \sum_{N:\frac{x_N}{y_N}=\frac{x_{N_j}}{y_{N_j}}} \min\left\{ N^d m_d|\alpha|, \frac{1}{N m_d^{1/d}|\alpha|^{1/d}} \right\},
\end{aligned}$$

by our twisted "pseudo-projective" estimates. We majorize the previous sum:

$$\lesssim_\rho \sum_j |S_{N_j}^0(0/1)|^2 \lesssim \sum_j y_{N_j}^{-2\nu}$$

for any $\nu < 1/d$, which converges by our sparsity condition on the $\{y_{N_j}\}$. The claim is proved.

We now majorize

$$\mathcal{V}^r\left({}^0T_N * f\right) \lesssim \left(\sum_j \left|\mathcal{V}^r\left({}^0T_N * f : N_j \leq N < N_{j+1}\right)\right|^2\right)^{1/2} + \mathcal{V}^r\left({}^0T_{N_j} * f : j\right).$$

The second sum is majorized $\lesssim \left(\sum_j \left|{}^0T_{N_j} * f\right|^2\right)^{1/2}$; we use Parseval to estimate its L^2-size by

$$\left(\sum_j y_{N_j}^{-2\nu}\right)^{1/2} \|f\|_{L^2} \lesssim \|f\|_{L^2},$$

where we used the trivial bound $|\omega_{N_j}| \leq 1$.

We now estimate the L^2 norm of

$$\mathcal{V}^r\left({}^0T_N * f : N_j \leq N < N_{j+1}\right)$$

for each j.

With

$$\widehat{g_j}(\alpha) := S^0_{N_j}(0/1)\chi(N_j^d m_d\alpha)\widehat{f}(\alpha),$$

we have the pointwise majorization

$$\mathcal{V}^r({}^0T_N * f : N_j \leq N < N_{j+1}) \lesssim \mathcal{V}^1(\omega_N : N_j \leq N < N_{j+1}) \cdot \mathcal{V}^r\left(\left(\chi(N^d m_d\alpha)\widehat{g_j}(\alpha)\right)^\vee\right).$$

But $\mathcal{V}^1(\omega_N : N_j \leq N < N_{j+1}) \lesssim_\rho 1$, since if $N < M$ are successive elements of I, we have the bounds

$$\begin{aligned}
|\omega_N - \omega_M| &= \left|\int_0^1 e(Nt\gamma_{N_j}) - e(Mt\gamma_{N_j})\, dt\right| \\
&\lesssim \min\left\{\frac{1}{N|\gamma_{N_j}|}, (M-N)|\gamma_{N_j}|\right\} \\
&\lesssim \min\left\{\frac{1}{N|\gamma_{N_j}|}, N|\gamma_{N_j}|\right\}.
\end{aligned}$$

Using a square function argument and Bourgain's [11, Lemma 3.28], we have that

$$\|\mathcal{V}^r(\big(\chi(N^d m_d \alpha)\widehat{g_j}(\alpha)\big)^\vee)\|_{L^2} \lesssim \frac{r}{r-2}\|g_j\|_{L^2} \lesssim_r y_{N_j}^{-\nu}\|f\|_{L^2}.$$

Since this is square-summable in j, the result is proved. □

$\mathcal{V}^r(R_N * f)$ *is* L^2*-Bounded.* We begin by subdividing

$$\begin{aligned}
\widehat{R_N}(\alpha) &:= \sum_{t:2^t \leq N^\delta} \widehat{R_{N,t}}(\alpha) \\
&:= \sum_{t:2^t \leq N^\delta} \left(\sum_{j=0}^{m_d-1} \sum_{a/b \in A^j_{N,t}} S^j_N(a/b) V_N\left(\alpha - \frac{1}{m_d}(j + a/b)\right) \chi\left(N^{d-\delta} m_d \left(\alpha - \frac{1}{m_d}(j+a/b)\right)\right)\right) \\
&= \sum_{t:2^t \leq N^\delta} \left(\sum_{j=0}^{m_d-1} \sum_{a/b \in A^j_{N_{l(t)},t}} S^j_N(a/b) V_N\left(\alpha - \frac{1}{m_d}(j + a/b)\right) \chi\left(N^{d-\delta} m_d \left(\alpha - \frac{1}{m_d}(j+a/b)\right)\right)\right).
\end{aligned}$$

Our task will be to show that

$$\sum_t \mathcal{V}^r(R_{N,t} * f : N^\delta \geq 2^t) = \sum_t \mathcal{V}^r(R_{N,t} * f : N^\delta \geq 2^t, N_{l(t)} \leq N < N_{l(t)+1})$$

is bounded in L^2. For notational ease, let $N_{m(t)} := \min\{N \in I : N \geq N_{l(t)}, 2^{t/\delta}\}$; we're interested in showing that

$$\|\mathcal{V}^r(R_{N,t} * f : N_{m(t)} \leq N \leq N_{l(t)+1})\|_{L^2(\mathbb{R})}$$

is summable in t. (If no scale $N_{l(t)}$ exists, then the pertaining $R_{N,t}$s are just zero operators.)

To this end, we similarly define for $N_{m(t)} \leq N < N_{l(t)+1}$

$$\begin{aligned}
\widehat{T_{N,t}}(\alpha) &:= \sum_{j=0}^{m_d-1} \sum_{A^j_{N,t}} S^j_N(a/b)\omega_N \chi\left(N^d m_d\left(\alpha - \frac{1}{m_d}(j+a/b)\right)\right) \\
&= \sum_{j=0}^{m_d-1} \sum_{A^j_{N_{l(t)},t}} S^j_N(a/b)\omega_N \chi\left(N^d m_d\left(\alpha - \frac{1}{m_d}(j+a/b)\right)\right).
\end{aligned}$$

We again estimate

$$\sup_{\alpha} \sum_{N_{m(t)} \leq N < N_{l(t)+1}} |R_{N,t} - T_{N,t}|^2(\alpha).$$

To do so, we fix α, set

$$E_N\left(\alpha - \frac{1}{m_d}(j+a/b)\right) := V_N\left(\alpha - \frac{1}{m_d}(j+a/b)\right) - \omega_N \chi\left(N^d m_d\left(\alpha - \frac{1}{m_d}(j+a/b)\right)\right),$$

where

$$\left|E_N\left(\alpha - \frac{1}{m_d}(j+a/b)\right)\right| \lesssim \min\left\{N^d m_d|\alpha - \frac{1}{m_d}(j+a/b)|, \frac{1}{N m_d^{1/d}|\alpha - \frac{1}{m_d}(j+a/b)|^{1/d}}\right\},$$

and consider

$$\begin{aligned}
&|R_{N,t} - T_{N,t}|^2(\alpha) \\
&= \left|\sum_{j=0}^{m_d-1} \sum_{A^j_{N_{l(t)},t}} S^j_{N_{l(t)}}(a/b)\left(E_N\left(\alpha - \frac{1}{m_d}(j+a/b)\right)\right) \cdot \chi\left(N^{d-\delta} m_d\left(\alpha - \frac{1}{m_d}(j+a/b)\right)\right)\right|^2 \\
&= \sum_{j=0}^{m_d-1} \sum_{A^j_{N_{l(t)},t}} |S^j_{N_{l(t)}}(a/b)|^2 \left|E_N\left(\alpha - \frac{1}{m_d}(j+a/b)\right)\right|^2 \cdot \chi\left(N^{d-\delta} m_d\left(\alpha - \frac{1}{m_d}(j+a/b)\right)\right)^2,
\end{aligned}$$

using the fact that for N sufficiently large the supports

$$\chi\left(N^{d-\delta} m_d\left(\alpha - \frac{1}{m_d}(j+a/b)\right)\right)$$

are disjoint in $a/b \in A_N^j$. Again using the disjoint support assumptions, we may estimate the foregoing by

$$\max_{a/b \in A_{N_{l(t)},t}} |S^j_{N_{l(t)}}(a/b)|^2 \left| E_N\left(\alpha - \frac{1}{m_d}(j + a/b)\right)\right|^2 \cdot \chi\left(N^{d-\delta} m_d \left(\alpha - \frac{1}{m_d}(j+a/b)\right)\right)^2.$$

Now, if there exists no $a/b \in A_{N_{l(t)},t}$ for which

$$\left|\alpha - \frac{1}{m_d}(j + a/b)\right| \leq \frac{1}{m_d} N_{m(t)}^{\delta - d},$$

then the above is just zero. Otherwise, let

$$p/q = p(\alpha)/q(\alpha) \in A_{N_{l(t)},t}$$

be the unique such element satisfying the above inequality with pertaining index k, i.e.

$$\left|\alpha - \frac{1}{m_d}(k + p/q)\right| \leq \frac{1}{m_d} N_{m(t)}^{\delta - d}.$$

In this case we may estimate the above maximum by

$$\begin{aligned} &|S^k_{N_{l(t)}}(p/q)|^2 \left| E_N\left(\alpha - \frac{1}{m_d}(k + p/q)\right)\right|^2 \\ &\qquad \lesssim |S^k_{N_{l(t)}}(p/q)|^2 \cdot \min\left\{ N^d m_d |\alpha - \frac{1}{m_d}(k+p/q)|, \frac{1}{N m_d^{1/d} |\alpha - \frac{1}{m_d}(k+p/q)|^{1/d}} \right\}. \end{aligned}$$

Summing now the above over $N_{m(t)} \leq N < N_{l(t)+1}$ accrues an upper estimate of

$$\lesssim_\rho |S^k_{N_{l(t)}}(p/q)|^2 \lesssim 2^{-2t\nu}$$

for $\nu < 1/d$.

Since this is summable in t, by another square function argument it suffices to show that

$$\|\mathcal{V}^r(T_{N,t} * f : N_{m(t)} \leq N < N_{l(t)+1})\|_{L^2(\mathbb{R})}$$

is summable in t.

With

$$\widehat{g_t}(\alpha) := \sum_{j=0}^{m_d-1} \sum_{a/b \in A^j_{N_{l(t)},t}} S^j_{N_{l(t)}}(a/b) \chi\left(N^d_{m(t)} m_d \left(\alpha - \frac{1}{m_d}(j + a/b)\right)\right) \widehat{f}(\alpha),$$

so

$$\|g_t\|_{L^2(\mathbb{R})} \lesssim 2^{-t\nu} \|f\|_{L^2(\mathbb{R})}$$

we majorize

$$\begin{aligned}
&\mathcal{V}^r(T_{N,t} * f : N_{m(t)} \leq N < N_{l(t)+1}) \\
&\qquad \leq \mathcal{V}^1(\omega_N : N_{m(t)} \leq N < N_{l(t)+1}) \cdot \mathcal{V}^r(B_{N,t} * g_t : N_{m(t)} \leq N < N_{l(t)+1}),
\end{aligned}$$

where

$$\widehat{B_{N,t}}(\alpha) := \sum_{j=0}^{m_d-1} \sum_{a/b \in A^j_{N_{l(t)},t}} \chi\left(N^d m_d \left(\alpha - \frac{1}{m_d}(j + a/b)\right)\right).$$

By arguing as above, we have that

$$\mathcal{V}^1(\omega_N : N_{m(t)} \leq N < N_{l(t)+1}) \lesssim_\rho 1.$$

By [31, Lemma 3.7], we may further estimate

$$\begin{aligned}
\|\mathcal{V}^r(B_{N,t} * g_t : N_{m(t)} \leq N < N_{l(t)+1})\| &\lesssim_\rho \left(\frac{r}{r-2} \log |A_{N_{l(t)},t}|\right)^2 \|g_t\|_{L^2(\mathbb{R})} \\
&\lesssim \left(\frac{r}{r-2} \cdot t\right)^2 2^{-t\nu} \|f\|_{L^2(\mathbb{R})}.
\end{aligned}$$

Since this is summable in t, the proof is complete. □

References

[1] Assani, I. Wiener Wintner Ergodic Theorems, World Scientific Publishing Co. Inc., River Edge, NJ, 2003.

[2] Assani, I.; Duncan, D. ; Moore R. *Pointwise characteristic factors for Wiener Wintner double recurrence theorem*, preprint, http://arxiv.org/abs/1402.7094.

[3] Assani, I.; Moore, R. *Extension of Wiener-Wintner double recurrence theorem to polynomials II*, preprint, http://arxiv.org/abs/1409.0463.

[4] Assani, I.; Presser, K. *A Survey of the return times theorem*, Ergodic Theory Dyn. Systems, Proceedings of the 2011-2012, UNC- Chapel Hill workshops, Walter De Gruyter, 2013.

[5] Bernstein, S. Lecons sur les Propriétés Extrêmales et la Meilleure Approximation des Fonctions Analytiques d'une Variable Réele, Gauthier-Villars, Paris, 1926; in: l'Approximation, Chelsea, New York, 1970.

[6] Birkhoff, G. *Proof of the ergodic theorem*, Proc. Natl. Acad. Sci. USA **17** (1931), 656–660.

[7] Boshernitzan, M. *Uniform distribution and Hardy fields*, J. Anal. Math. **62** (1994), 225–240.

[8] Boshernitzan, M.; Kolesnik,G.; Quas, A.; Wierdl, M. *Ergodic averaging sequences*, J. Anal. Math. **95** (2005), 63–103.

[9] Boshernitzan, M.; Wierdl, M. *Ergodic theorems along sequences and Hardy fields*, Proc. Nat. Acad. Sci. U.S.A. **93** (1996), 8205–8207.

[10] Bourgain, J. *On the maximal ergodic theorem for certain subsets of the integers*, Israel J. Math. **61** (1988), no. 1, 39–72.

[11] Bourgain, J. Pointwise ergodic theorems for arithmetic sets. Inst. Hautes Études Sci. Publ. Math. **69** (1989), 5–45. With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein.

[12] Bourgain, J. *Double recurrence and almost sure convergence*, J. Reine Angew. Math. **404** (1990), 140–161.

[13] Calderón, A. *Ergodic theory and translation invariant operators*, Proc. Nat. Acad. Sci., USA **59** (1968), 349–353.

[14] Eisner, T.; Farkas, B.; Haase, M.; Nagel, R. Operator Theoretic Aspects of Ergodic Theory, Graduate Texts in Mathematics, Springer, to appear.

[15] Eisner, T.; Tao, T. *Large values of the Gowers-Host-Kra seminorms*, J. Anal. Math. **117** (2012), 133–186.

[16] Eisner, T.; Zorin-Kranich, P. *Uniformity in the Wiener-Wintner theorem for nilsequences*, Discrete Contin. Dyn. Syst. **33** (2013), 3497–3516.

[17] Falconer, K. Fractal Geometry: Mathematical Foundations and Applications, 2nd ed., John Wiley & Sons Inc., Hoboken, NJ, 2003.

[18] Frantzikinakis, N. *Uniformity in the polynomial Wiener-Wintner theorem*, Ergodic Theory Dynam. Systems **26** (2006), 1061–1071.

[19] Frantzikinakis, N. *Equidistribution of sparse sequences on nilmanifolds* J. Anal. Math. **117** (2012), 133–186.

[20] Frantzikinakis, N. *Multiple recurrence and convergence for Hardy sequences of polynomial growth*, J. Anal. Math. **112** (2010), 79–135.

[21] Frantzikinakis, N.; Johnson, M.; Lesigne, E.; Wierdl M. *Powers of sequences and convergence of ergodic averages*, Ergodic Theory Dynam. Systems 30 (2010), 1431–1456.

[22] Frantzikinakis, N.; Wierdl, M. *A Hardy field extension of Szemerédi's theorem*, Adv. Math. **222** (2009), 1–43.

[23] Host, B.; Kra, B. *Uniformity seminorms on l^∞ and applications*, J. Anal. Math. **108** (2009), 219–276.

[24] Host, B.; Kra, B. *Nonconventional ergodic averages and nilmanifolds*, Ann. Math. (2) **161** (2005), no. 1, 397–488.

[25] Hua, L.-K. Introduction to Number Theory. Second edition. Springer-Verlag, Berlin, 1982.

[26] Hutchinson, J.-E. *Fractals and self-similarity*, Indiana Univ. Math. J. **30** (1981), 713–747.

[27] Jarnik, V. *Zur metrischen Theorie der diophantischen Approximationen*, Prace Mat.-Fiz. **36** (1928/9), 91–106.

[28] Jenkinson, O. *On the density of Hausdorff dimensions of bounded type continued fraction sets: the Texan conjecture*, Stochastics Dyn. **4** (2004), 63–76.

[29] Kesseböhmer, M.; Zhu, S. *Dimension sets for infinite IFSs: the Texan conjecture*, J. Number Theory **116** (2006), 230–246.

[30] Khintchine, A. Continued fractions, English transl. by P. Wynn, Noordhoff, Groningen, 1963.

[31] Krause, B. *Polynomial Ergodic Averages Converge Rapidly: Variations on a Theorem of Bourgain*. Preprint, http://arxiv.org/pdf/1402.1803v1.pdf.

[32] Krause, B.; Zorin-Kranich, P. *A random pointwise ergodic theorem with Hardy field weights*. Preprint, http://arxiv.org/pdf/1410.0806v1.pdf.

[33] Kuipers, L.; Niederreiter, H. Uniform Distribution of Sequences. Pure and Applied Mathematics, Wiley-Interscience, New York-London-Sydney, 1974.

[34] Lesigne, E. *Spectre quasi-discret et théorème ergodique de Wiener-Wintner pour les polynômes*, Ergodic Theory Dynam. Systems **13** (1993), 767–784.

[35] Mirek, M.; Trojan, B. *Discrete maximal functions in higher dimensions and applications to ergodic theory*. Preprint, http://arxiv.org/abs/1405.5566.

[36] Montgomery, H. *Harmonic Analysis as Found in Analytic Number Theory*, in Twentieth Century Harmonic Analysis–A Celebration. Proceedings of the NATO Advanced Study Institute Held in Il Ciocco, July 2-15, 2000 (Ed. J. S. Byrnes). Dordrecht, Netherlands: Kluwer, pp. 271–293, 2001.

[37] Rosenblatt, J.; Wierdl, M. *Pointwise ergodic theorems via harmonic analysis*. Ergodic theory and its connections with harmonic analysis (Alexandria, 1993). London Math. Soc. Lecture Note Ser., 205, Cambridge Univ. Press, Cambridge, (1995), 3–151.

[38] Stein, E. Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals. Princeton Mathematical Series, 43. Monographs in Harmonic Analysis, III. Princeton University Press, Princeton, NJ, 1993.

[39] Tao, T. An Epsilon of Room, Real Analysis: Pages from Year Three of a Mathematical Blog. Graduate Studies in Mathematics 117, American Mathematical Society, 2010.

[40] Thouvenot, J.-P. *La convergence presque sûre des moyennes ergodiques suivant certaines sous-suites d'entiers (d'après Jean Bourgain)*, Séminaire Bourbaki, Vol. 1989/90. Astérisque No. 189-190 (1990), Exp. No. 719, 133–153.

[41] Vaughan, R. The Hardy-Littlewood method. Second edition. Cambridge Tracts in Mathematics, 125. Cambridge University Press, Cambridge, 1997.

[42] Wiener, N.; Wintner, A. *Harmonic analysis and ergodic theory*, Amer. J. Math. **63** (1941), 415–426.

CHAPTER 5

A Weighted Random Pointwise Ergodic Theorem in L^1 – Joint with Pavel Zorin-Kranich

5.1 Introduction

A sequence of integers $\{n_k\} \subset \mathbb{Z}$ is said to be *universally L^p-good* if for every measure-preserving system (X, μ, T) and every $f \in L^p(X)$ the subsequence averages

$$A_N^{\{n_k\}} f := \frac{1}{N} \sum_{k=1}^{N} T^{n_k} f$$

converge pointwise almost everywhere. In this language, Birkhoff's classical pointwise ergodic theorem [2] states that the full sequence of integers is universally L^1-good.

Obtaining pointwise convergence results for rougher, sparser sequences is much more challenging. For instance, Bourgain's Polynomial Ergodic Theorem [7] states that the sequence $\{P(n)\}$, P integer polynomial, is universally L^p-good for each $p > 1$. Note that $\{P(n)\}$ are zero-Banach-density subsequences of the integers; in fact, Bourgain used a probabilistic method to find *extremely* sparse universally good sequences. From now on $\{X_n\}$ will denote a sequences of independent $\{0, 1\}$ valued random variables (on a probability space Ω) with expectations σ_n. The *counting function* $a_n(\omega)$ is the smallest integer subject to the constraint

$$X_1(\omega) + \cdots + X_{a_n(\omega)}(\omega) = n.$$

Theorem 5.1.1 ([6, Proposition 8.2]). *Suppose*

$$\sigma_n = \frac{(\log\log n)^{B_p}}{n},\ B_p > \frac{1}{p-1},\ 1 < p \le 2.$$

Then, almost surely, $\{a_n\}$ is universally L^p-good.

In the years to follow random sequences became a widely used model for pointwise ergodic theorems. One indication at their amenability to analysis is LaVictoire's L^1 random ergodic theorem.

Theorem 5.1.2 ([10]). *Suppose $\sigma_n = n^{-a}$ with $0 < a < 1/2$. Then, almost surely, $\{a_n\}$ is universally L^1-good.*

Here, by the strong law of large numbers, almost surely

$$a_n(\omega)/n^{\frac{1}{1-a}}$$

converges to a non-zero number. For comparison, it is known that the sequences of d-th powers, $d > 1$ integer, are universally L^1 bad [5, 11].

Random sequences have also been used as a model for multiple ergodic averages. Frantzikinakis, Lesigne, and Wierdl recently showed the following.

Theorem 5.1.3 ([9, Theorem 1.1]). *Suppose $\sigma_n = n^{-a}$, $0 < a < 1/14$. Then, almost surely, $\{a_n\}$ has the following property: for every pair of measure preserving transformations T, S on a probability space X and any functions $f, g \in L^\infty(X)$ the averages*

$$\sum_{n=1}^{N} g(S^n x) f(T^{a_n} x)$$

converge pointwise almost everywhere.

It is noted in their paper that the linear sequence of powers S^n can likely be replaced by other deterministic sequences, but their method of proof did not seem to allow this. In this article we prove a related result in which we are able to replace the linear sequence of powers by a sequence drawn from a more general class at the cost of weakening the result in several other respects. More precisely, with $0 < \epsilon < 1$ arbitrary but fixed, suppose $p : \mathbb{R} \to \mathbb{R}$ is a *logarithmico-exponential* function which satisfies

1. the second-order difference relationship

$$p(x+y+z) - p(x+y) - p(x+z) + p(x) = O(x^{\epsilon-1}yz)$$

   for $x, y, z > 0$ ("big-O" notation is recalled in the section on notation below); and

2. for all $a(x) \in C \cdot \mathbb{Q}[x]$, the set of real constant multiples of rational polynomials, $\frac{|a(x)-p(x)|}{\log x} \to \infty$.

Good examples of such functions are $p(x) = x^{1+\epsilon}$. We refer the reader to [4] for a more complete discussion of *logarithmico-exponential* functions; informally, these are all the functions one can get by combining real constants, the variable x, and the symbols exp, log, $\cdot$, and $+$. (e.g. $x^{1/2} = \exp(1/2 \cdot \log x)$ and $x^{\pi}/\log\log x$ are both logarithmico-exponential.)

Our main result is the following

Theorem 5.1.4. *Suppose $\sigma_n = n^{-a}$, $0 < a < 1/2$, and p is as above. Then, almost surely, the following holds:*

For each measure-preserving system (X, μ, T) and each $f \in L^1(X)$ the averages

$$\frac{1}{N}\sum_{n=1}^{N} e(p(n))T^{a_n(\omega)}f$$

converge to zero pointwise almost everywhere (here and later $e(t) := e^{2\pi i t}$).

Pointwise ergodic theorems with exponential polynomial weights are collectively known as Wiener–Wintner type theorems, see e.g. [1] for linear polynomials and [12] for general polynomials. If the random sequence $\{a_n\}$ is replaced by the linear sequence $\{n\}$ in Theorem 5.1.4, the result follows from the Wiener–Wintner theorem for Hardy field functions due to Eisner and the first author [8]. However, note that the full measure sets in our result depend on the choice of p. It would be interesting to remove this dependence. Also, the second order difference relation in the hypothesis of Theorem 5.1.4 can likely be replaced by a polynomial growth assumption; this would require an inductive application of van der Corput's inequality.

The structure of this paper is as follows:

In §5.2 we introduce a few preliminary tools, discuss our proof strategy, and reduce our theorem to proving Proposition 5.2.3; and

In §5.3 we prove Proposition 5.2.3, thereby completing the proof of Theorem 5.1.4.

5.1.1 Acknowledgments

Both authors are grateful to Nikos Frantzikinakis, Rowan Killip, and Alexander Volberg for helpful conversations, and to Tanja Eisner for suggesting the collaboration. We thank the Hausdorff Research Institute for Mathematics for hospitality during the Trimester Program "Harmonic Analysis and Partial Differential Equations".

5.1.2 Notation

With $X_n,\ \sigma_n$ as above, we let $Y_n := X_n - \sigma_n$.

We will be dealing with sums of random variables, so we introduce the following compact notation:

$$S_N = \sum_{n=1}^{N} X_N \quad \text{and} \quad S_{M,N} = \sum_{n=M}^{N} X_n.$$

We also let

$$W_N := \sum_{n=1}^{N} \sigma_n,$$

so that $W_N \approx N^{1-a}$ (see below).

We say that N is an element of a "lacunary sequence" if there exists some constant, $\rho > 1$, so that $N = \lfloor \rho^k \rfloor$ for some $k \in \mathbb{N}$. We say that a sequence of complex numbers $\{b_n\}$

"sums over lacunary sequences" if for each $\rho > 1$,

$$\sum_k |b_{\lfloor \rho^k \rfloor}| < \infty.$$

Two probabilistic sequences $\{b_N(\omega)\}$ and $\{b'_N(\omega)\}$ are *almost surely asymptotically equivalent*,

$$b_N(\omega) \approx_N b'_N(\omega)$$

if for each ω in a set of full probability, $\frac{b_N(\omega)}{b'_N(\omega)} \to 1$.

We will make use of the modified Vinogradov notation. We use $X \lesssim Y$, or $Y \gtrsim X$ to denote the estimate $X \leq CY$ for an absolute constant C. If we need C to depend on a parameter, we shall indicate this by subscripts, thus for instance $X \lesssim_\omega Y$ denotes the estimate $X \leq C_\omega Y$ for some C_ω depending on ω. We use $X \approx Y$ as shorthand for $Y \lesssim X \lesssim Y$.

We also make use of big-O notation: we let $O(Y)$ denote a quantity that is $\lesssim Y$, and similarly $O_\omega(Y)$ a quantity that is $\lesssim_\omega Y$.

5.2 Preliminaries

We begin with a few preliminary

5.2.1 Tools

Lemma 5.2.1 (van der Corput inequality, see e.g. [9]). *Let $\{v_n\}$ be a sequence in a Hilbert space H and $1 \leq M \leq N$. Then*

$$\left\| \sum_{n=1}^{N} v_n \right\|^2 \leq 2\frac{N}{M} \sum_{n=1}^{N} \|v_n\|^2 + 4\frac{N}{M} \sum_{m=1}^{M} \left| \sum_{n=1}^{N-m} \langle v_{n+m}, v_n \rangle \right| \tag{5.1}$$

Lemma 5.2.2 (Chernoff's Inequality – Special Case [14]). *Let $\{X_n\}$, $\{\sigma_n\}$ be as above.*

There exists an absolute constant $c > 0$ so that for each $A > 0$,

$$\mathbb{P}\left(|S_n - W_n| \geq A\right) \lesssim \max\left\{\exp\left(-c\frac{A^2}{W_n}\right), \exp(-cA)\right\}.$$

Consequently,

$$\mathbb{P}\left(|S_n - W_n| \geq \frac{1}{2}W_n\right) \lesssim \exp(-cW_n) \lesssim \exp(-cn^{1-a}).$$

With these in hand, we turn to our

5.2.2 Strategy

In proving his Random Ergodic Theorem, LaVictoire showed that on a set of full probability, $\Omega' \subset \Omega$, the maximal function

$$f \mapsto \sup_N \frac{1}{N}\sum_{n=1}^N T^{a_n(\omega)}|f|,$$

is weakly bounded on $L^1(X)$. We will henceforth condition on Ω'. For such ω, we see that the set of $f \in L^1(X)$ for which the averages

$$\frac{1}{N}\sum_{n=1}^N e(p(n))T^{a_n(\omega)}f$$

tend to zero pointwise is closed in L^1; for $\omega \in \Omega'$, it will be enough to prove pointwise convergence for $f \in L^\infty(X)$. In what follows, we shall without loss of generality assume that each function we introduce is almost everywhere bounded in magnitude by 1. Now, for bounded functions, by [13], it is enough to prove that the averages

$$\frac{1}{N}\sum_{n=1}^N e(p(n))T^{a_n(\omega)}f$$

tend to zero along lacunary subsequences. We will fix some $\rho > 1$ throughout, so that each averaging parameter is implicitly of the form $N = \lfloor \rho^k \rfloor$ for some $k \in \mathbb{N}$.

We follow a similar plan to [9]. Roughly speaking, we will prove Theorem 5.1.4 by showing that for each fixed $x \in X$:

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N} e(p(n))T^{a_n} f &\approx_N \frac{1}{S_N}\sum_{n=1}^{N} X_n(\omega)e(p(S_n))T^n f \\
&\approx_N \frac{1}{W_N}\sum_{n=1}^{N} X_n(\omega)e(p(S_n))T^n f \\
&\approx_N \frac{1}{W_N}\sum_{n=1}^{N} \sigma_n e(p(S_n))T^n f \\
&\approx_N \frac{1}{N}\sum_{n=1}^{N} e(p(S_n))T^n f \\
&\approx_N \bar{f}\cdot\frac{1}{N}\sum_{n=1}^{N} e(p(S_n)) \\
&\approx_N \bar{f}\cdot\frac{1}{N}\sum_{n=1}^{N} e(p(n)) \to 0.
\end{aligned}$$

Here, $\bar{f} := \lim_N \frac{1}{N}\sum_{n=1}^{N} T^n f$ is the projection of f onto the invariant factor of T.

Now, the third approximation is the key to our argument. We isolate this crucial step in the following

Proposition 5.2.3. *In the above setting, almost surely the following holds: for each measure-preserving system, (X,μ,T), and each $f \in L^2(X)$ (and thus for each $f \in L^\infty(X)$),*

$$\left\| \frac{1}{W_N}\sum_{n=1}^{N} Y_n(\omega)e(p(S_n))T^n f \right\|_{L^2(X)}^2$$

is summable over lacunary N.

In particular, almost surely, for every (X,μ,T) and each $f \in L^2(X)$,

$$\frac{1}{W_N}\sum_{n=1}^{N} Y_n(\omega)e(p(S_n))T^n f \to 0 \quad \mu\text{-}a.e.$$

In the remainder of this section, we reduce the proof of Theorem 5.1.4 to Proposition

5.2.3.

5.2.3 The Reduction

1. The first equivalence is by definition;
2. The second is the strong law of large numbers;
3. The third is Proposition 5.2.3;
4. The fourth is summation by parts;
5. The fifth will be discussed below;
6. The sixth follows by applying the above steps in reverse, with $f = 1_X$; and
7. The convergence of the averages $\frac{1}{N}\sum_{n=1}^{N} e(p(n)) \to 0$ was proved in [3, Theorem 1.3].

So, the goal of this section will be to prove the following

Lemma 5.2.4. *In the setting of Theorem 5.1.4, we have the asymptotic equivalence*

$$\frac{1}{N}\sum_{n=1}^{N} e(p(S_n))T^n f \approx_N \bar{f} \cdot \frac{1}{N}\sum_{n=1}^{N} e(p(S_n)).$$

The proof here follows exactly that of [9, Lemma 2.2]; we reproduce their argument here for completeness.

Proof. This is clear when $Tf = f$ a.e., so we need to show that, almost surely, for each $f \in L^\infty(X)$ with $\bar{f} = 0$ a.e., we have

$$\frac{1}{N}\sum_{n=1}^{N} e(p(S_n))T^n f \to 0 \text{ a.e.}$$

We begin with the case when $f = h - Th$ is a coboundary:

Here, by summation by parts, we have

$$\frac{1}{N}\sum_{n=1}^{N} e(p(S_n))T^n(h - Th) = O(1/N) + \frac{1}{N}\sum_{n=1}^{N-1} \left(e(p(S_n)) - e(p(S_{n+1}))1_{X_{n+1}=1}\right) T^n h,$$

which in modulus is dominated by a constant multiple of

$$\frac{1}{N} + \frac{1}{N}\sum_{n=1}^{N-1} 1_{X_{n+1}=1}.$$

The first term clearly converges to zero, so we focus on the sum, which we seek to show almost surely converges to 0 as $N \to \infty$ along lacunary subsequences. We accomplish this by showing

$$\sum_{N \text{ lacunary}} \mathbb{E}\frac{1}{N}\sum_{n=1}^{N-1} 1_{X_{n+1}=1} = \sum_{N \text{ lacunary}} \frac{W_N}{N} \lesssim \sum_{N \text{ lacunary}} N^{-a} < \infty,$$

from which the result follows.

Since every $f \in L^\infty(X)$ with $\bar{f} = 0$ can be arbitrarily well approximated in $L^1(X)$ be coboundaries, the proof of the Lemma will be complete once we show that the set of (say) L^∞ functions for which we have pointwise convergence is closed in $L^1(X)$. In particular, it is enough to show that

$$\left\| \limsup_N \left| \frac{1}{N}\sum_{n=1}^{N} e(p(S_n))T^n f \right| \right\|_{L^1(X)} \leq \left\| \limsup_N \frac{1}{N}\sum_{n=1}^{N} T^n|f| \right\|_{L^1(X)} \leq \|f\|_{L^1(X)}.$$

But this is just the pointwise ergodic theorem! $\square$

With this reduction complete, we now turn to the proof of Proposition 5.2.3.

5.3 Proof of Proposition 5.2.3

Throughout this section, we will view $0 < \delta \ll 1$ as a (small) floating parameter, whose precise value will be fixed at the end of the proof; $0 < \nu = \nu(\delta) = O(\delta)$ will be used to denote (possibly different) parameters (all of which grow linearly in δ); $0 < \kappa = O(\delta)$ will be used similarly.

We begin with a Lemma that guarantees that a bounded sequence $\{c_n\}$ is a good sequence of weights for a pointwise ergodic theorem along (lacunary) times N:

Lemma 5.3.1. *Let $0 < a < 1/2$ and fix a lacunary sequence of N's. Let $\{c_n\}$ be a sequence such that the following holds:*

$$\sum_{n=1}^{N} |c_n| \lesssim N^{1-a} \tag{5.2}$$

$$N^{2a-1-b} \sum_{m=1}^{N^b} \left| \sum_{n=N^{1-\delta}}^{N-m} c_{n+m}\bar{c}_n \right| \text{ is summable in } N, \text{ where } b = a + \delta' < 1/2 \text{ as well.} \tag{5.3}$$

Then for every measure-preserving system (X, μ, T) and $f \in L^2(X)$ the sequence

$$\left\| \frac{1}{N^{1-a}} \sum_{n=1}^{N} c_n T^n f \right\|_{L^2(X)}^2$$

is summable over N.

Consequently, $\frac{1}{N^{1-a}} \sum_{n=1}^{N} c_n T^n f \to 0$ (along lacunary subsequences).

Proof. By (5.2) we have

$$\begin{aligned}
\left\| \frac{1}{N^{1-a}} \sum_{n=1}^{N^{1-\delta}} c_n T^n f \right\|_{L^2(X)} &\leq \frac{1}{N^{1-a}} \sum_{n=1}^{N^{1-\delta}} |c_n| \|f\|_{L^2(X)} \\
&\lesssim \frac{1}{N^{1-a}} N^{(1-\delta)(1-a)} \|f\|_{L^2(X)} \\
&= N^{-\kappa} \|f\|_{L^2(X)}
\end{aligned}$$

almost surely, so we may replace the sum in the conclusion of the theorem by $\sum_{n=N^{1-\delta}}^{N}$.

Using van der Corput inequality (5.1) in $L^2(X)$ with $M = N^b$, $b = a + \delta'$, estimate that new sequence by

$$N^{2a-2}\frac{N}{N^b}\sum_{n=N^{1-\delta}}^{N}\|c_nT^nf\|_{L^2(X)}^2 + N^{2a-2}\frac{N}{N^b}\sum_{m=1}^{M}\left|\sum_{n=N^{1-\delta}}^{N-m}\int_X c_{n+m}T^{n+m}f\bar{c}_nT^n\bar{f}\right|. \tag{5.4}$$

The first term in (5.4) is bounded by

$$N^{2a-2}\frac{N}{N^b}\sum_{n=N^{1-\delta}}^{N}|c_n|^2\|f\|_{L^2(X)}^2,$$

and by (5.2) this is $O(N^{-\kappa})$ provided $b > a$ (and (c_n) is bounded). By precomposing with T^{-n} and Cauchy-Schwartz, the second term in (5.4) is bounded by

$$N^{2a-1-b}\sum_{m=1}^{M}\left|\sum_{n=N^{1-\delta}}^{N-m}c_{n+m}\bar{c}_n\right|\cdot\|f\|_{L^2(X)}^2,$$

and this is summable by (5.3). □

Proposition 5.3.2. *Let $p : \mathbb{R} \to \mathbb{R}$ be a function such that*

$$p(x + y + z) - p(x + y) = p(x + z) - p(x) + O(x^{\epsilon-1}yz) \tag{5.5}$$

for $x, y, z > 0$. Suppose $0 < a < 1/2$. Then, almost surely, the sequence $c_n = Y_ne(p(S_n))$ satisfies (5.3).

Proof. By Fubini's theorem it suffices to show that the expectation

$$N^{2a-1-b}\sum_{m=1}^{N^b}\mathbb{E}\left|\sum_{n=N^{1-\delta}}^{N-m}Y_{n+m}e(p(S_{n+m}))Y_ne(-p(S_n))\right|$$

is summable along the lacunary sequence of N's. We use Cauchy–Schwarz to estimate the foregoing by

$$N^{2a-1-b}\sum_{m=1}^{N^b}\left(\mathbb{E}\left|\sum_{n=N^{1-\delta}}^{N-m}Y_nY_{n+m}e(p(S_{n+m})-p(S_n))\right|^2\right)^{1/2} := N^{2a-1}\frac{1}{N^b}\sum_{m=1}^{N^b}I(m), \tag{5.6}$$

and use van der Corput (5.1) with values in $L^2(\Omega)$ and $R = N^c$ inside the bracket with $0 < c < 1$ to be chosen later. This gives the estimate

$$\begin{aligned}
&I(m)^2 \le I_1(m)^2 + I_2(m)^2 + I_3(m)^2 := \\
&\quad \frac{N-m}{R}\sum_{n=N^{1-\delta}}^{N-m}\|Y_nY_{n+m}e(p(S_{n+m})-p(S_n))\|^2_{L^2(\Omega)} \\
&+\frac{N-m}{R}\left|\mathbb{E}\sum_{n=N^{1-\delta}}^{N-2m}Y_{n+2m}Y_{n+m}Y_{n+m}Y_ne(p(S_{n+2m})-p(S_{n+m})-p(S_{n+m})+p(S_n))\right| \\
&+\frac{N-m}{R}\sum_{r=1,r\neq m}^{R}\left|\mathbb{E}\sum_{n=N^{1-\delta}}^{N-m-r}Y_{n+r}Y_{n+r+m}Y_nY_{n+m}e(p(S_{n+r+m})-p(S_{n+r})-p(S_{n+m})+p(S_n))\right|.
\end{aligned}$$

The task is now to show that, uniformly in $m \le N^b$, we have

$$I_j(m)^2 \lesssim N^{2-4a-\kappa} \text{ for each } 1 \le j \le 3,$$

for some $\kappa = \kappa(\delta, a, b, c) > 0$.

To this end, we estimate the first term, $I_1(m)^2$, by

$$\frac{N}{R}\sum_{n=N^{1-\delta}}^{N-m}\|Y_n\|^2_{L^2(\Omega)}\|Y_{n+m}\|^2_{L^2(\Omega)}$$

by independence; this is bounded by

$$\frac{N}{R}\sum_{n=N^{1-\delta}}^{N-m}\sigma_n\sigma_{n+m} \lesssim N^{1-c}N^{1-2a} < N^{2-4a-\kappa}$$

provided we take $2a < c < 1$.

We next turn to $I_2(m)^2$, which contributes at most

$$\frac{N}{R}\mathbb{E}\sum_{n=N^{1-\delta}}^{N-2m}|Y_{n+m}Y_{n+2m}Y_nY_{n+m}|,$$

and by independence this is bounded by

$$\begin{aligned}\frac{N}{R}\sum_{n=N^{1-\delta}}^{N-2m}\mathbb{E}|Y_{n+m}|^2\cdot\mathbb{E}|Y_{n+2m}|\cdot\mathbb{E}|Y_n| &\lesssim N^{1-c}\sum_{n=N^{1-\delta}}^{N-2m}\sigma_n\sigma_{n+m}\sigma_{n+2m}\\ &\lesssim N^{1-c}N^{1-3a+\nu}\\ &\lesssim N^{2-4a-\kappa},\end{aligned}$$

provided $c>2a$ (from above) and $\nu=\nu(\delta)>0$ is taken sufficiently small. (ν arises from the possibility that $3a>1$, in which case we may take e.g. $\nu=(3a-1)\delta$.)

The contribution of this term is also acceptable.

It remains to estimate $I_3(m)^2$

To recover independence we apply (5.5) with

$$x=S_{n+t-1},\ y=X_{n+t},\ z=S_{n+t+1,n+t+s},$$

where $s=\min(r,m)$ and $t=\max(r,m)$. This gives the estimate

$$\begin{aligned}&\frac{N-m}{R}\sum_{r=1,r\neq m}^{R}\left|\mathbb{E}\sum_{n=N^{1-\delta}}^{N-m-r}Y_{n+r}Y_{n+r+m}Y_nY_{n+m}e(p(S_{n+t-1}+S_{n+t+1,n+t+s})-p(S_{n+t-1})-p(S_{n+s})+p(S_n))\right|\\ &+\frac{N-m}{R}\sum_{r=1,r\neq m}^{R}\mathbb{E}\sum_{n=N^{1-\delta}}^{N-m-r}\left|Y_{n+r}Y_{n+r+m}Y_nY_{n+m}\min(O(S_{n+t-1}^{\epsilon-1}X_{n+t}S_{n+t+1,n+t+s}),1)\right|.\end{aligned}$$

The main feature of this splitting is that the exponential in the first term does not depend on X_{n+t}, so Y_{n+t} is independent from all other terms. Therefore the first term vanishes identically. The second term is estimated by

$$\frac{N}{R}\sum_{r=1,r\neq m}^{R}\sum_{n=N^{1-\delta}}^{N-m-r}\mathbb{E}\left(|Y_{n+r}Y_{n+r+m}Y_nY_{n+m}|\cdot\min(S_{n+t-1}^{\epsilon-1}S_{n+t+1,n+t+s},1)\right)$$
$$\leq\frac{N}{R}\sum_{r=1,r\neq m}^{R}\sum_{n=N^{1-\delta}}^{N-m-r}\mathbb{E}\left(|Y_{n+r}Y_{n+r+m}Y_nY_{n+m}|\cdot\min(S_{n+t-1}^{\epsilon-1}(S_{n+t+1,n+t+s-1}+1),1)\right)$$

By Lemma 5.2.2 this is bounded by

$$\frac{N}{R}\sum_{r=1,r\neq m}^{R}\sum_{n=N^{1-\delta}}^{N-m-r}\mathbb{E}\left(|Y_{n+r}Y_{n+r+m}Y_nY_{n+m}|\cdot((W_{n+t-1})^{\epsilon-1}(S_{n+t+1,n+t+s-1}+1)\cdot 1_X+1_{E_{n+t-1}})\right),$$

where E_n is an exceptional set of measure $\lesssim\exp(-cn^{1-a})$. We consider first the non-exceptional part. All remaining random variables are independent, so we get the estimate

$$\frac{N}{R}\sum_{r=1,r\neq m}^{R}\sum_{n=N^{1-\delta}}^{N-m-r}\sigma_{n+r}\sigma_{n+r+m}\sigma_n\sigma_{n+m}(n+t-1)^{(1-a)(\epsilon-1)}\left(\sum_{j=n+t+1}^{n+t+s-1}\sigma_j+1\right)$$
$$\leq\frac{N}{R}\sum_{r=1,r\neq m}^{R}\sum_{n=N^{1-\delta}}^{N-m-r}n^{-4a}n^{(1-a)(\epsilon-1)}(n^{-a}N^b+1)$$
$$\lesssim\frac{N}{R}\sum_{r=1,r\neq m}^{R}N^b\sum_{n=N^{1-\delta}}^{N-m-r}n^{-4a}n^{(1-a)(\epsilon-1)}n^{-a}$$
$$\lesssim N^{2-4a+(b-a+\nu)+(1-a)(\epsilon-1)}$$
$$\lesssim N^{2-4a-\kappa},$$

provided that b is taken sufficiently close to a and δ is sufficiently small, since $(1-a)(\epsilon-1)<0$.

Finally, for the exceptional part we have superpolynomial decay in N.

The proof is complete.

References

[1] Assani, I. Wiener Wintner ergodic theorems. World Scientific Publishing Co., River Edge, NJ, 2003, pp. xii+216. ISBN: 981-02-4439-8.

[2] Birkhoff, G. Proof of the ergodic theorem. In: Proc. Natl. Acad. Sci. USA 17 (1931), pp. 656 660.

[3] Boshernitzan, M. Uniform distribution and Hardy fields. In: J. Anal. Math. 62 (1994), pp. 225240. ISSN: 0021-7670.

[4] Boshernitzan, M.; Kolesnik, G.; Quas, A.; Wierdl, M. Ergodic averaging sequences. In: J. Anal. Math. 95 (2005), pp. 63103. ISSN: 0021-7670.

[5] Buczolich, Z.; Mauldin, R. Concepts behind divergent ergodic averages along the squares. In: Ergodic theory and related fields. Vol. 430. Contemp. Math. Amer. Math. Soc., Providence, RI, 2007, pp. 4156.

[6] Bourgain, J. On the maximal ergodic theorem for certain subsets of the integers. In: Israel J. Math. 61.1 (1988), pp. 3972. ISSN: 0021-2172.

[7] Bourgain, J. Pointwise ergodic theorems for arithmetic sets. In: Inst. Hautes tudes Sci. Publ. Math. 69 (1989). With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein, pp. 545. ISSN: 0073-8301.

[8] Eisner, T.; Krause, B. (Uniform) Convergence of Twisted Ergodic Averages. Preprint. 2014. arXiv:1407.4736 [math.DS].

[9] Frantzikinakis, N.; Lesigne, E.; Wierdl, M. Random sequences and pointwise convergence of multiple ergodic averages. In: Indiana Univ. Math. J. 61.2 (2012), pp. 585617. ISSN: 0022-2518. arXiv:1012.1130.

[10] LaVictoire, P. An L1 ergodic theorem for sparse random subsequences. In: Math. Res. Lett. 16.5 (2009), pp. 849859. ISSN: 1073-2780. arXiv:0812.3175.

[11] LaVictoire, P. Universally L1 -bad arithmetic sequences. In: J. Anal. Math. 113 (2011), pp. 241 263. ISSN: 0021-7670. arXiv:0905.3865.

[12] Lesigne, E. Spectre quasi-discret et thorme ergodique de Wiener-Wintner pour les polynmes. In: Ergodic Theory Dynam. Systems 13.4 (1993), pp. 767784. ISSN: 0143-3857.

[13] Rosenblatt, J.; Wierdl, M. Pointwise ergodic theorems via harmonic analysis. In: Ergodic theory and its connections with harmonic analysis (Alexandria, 1993). Vol. 205. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, 1995, pp. 3151.

[14] Tao, T.; Vu, V. Additive combinatorics. Vol. 105. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2010, pp. xviii+512. ISBN: 978-0-521-13656-3.

$\square$