New York Journal of Mathematics

New York J. Math. **3A** (1997) 31–67.

Ergodic Theory and Connections with Analysis and Probability

Roger L. Jones

Abstract. In this paper we establish a variety or results in ergodic theory by using techniques from probability and analysis. We discuss divergence of operators, including strong sweeping out and Bourgain's entropy method. We consider square functions, oscillation operators, and variational operators for ergodic averages. We also consider almost everywhere convergence of convolution powers.

Contents

1. Introduction 31
2. Divergence and Strong Sweeping Out 33
3. Proof of Bourgain's Entropy Theorem 38
4. Convolution Powers 47
5. Good-λ Inequalities for Convolution Powers 54
6. Oscillation Inequalities 58
7. Concluding Remarks 64
References 65

1. Introduction

There are many connections between ergodic theory, probability and harmonic analysis. In the discussion to follow, we will be mainly interested in ergodic theory results, but we will obtain these results by applying techniques or results from analysis or probability. In most cases, the results below appear elsewhere, in some cases in greater generality. However the idea here is to highlight some of the tools

Received November 24, 1997.

Mathematics Subject Classification. Primary: 28D05, 42B20; Secondary: 40A05, 42A50, 42B25, 60G42.

Key words and phrases. almost everywhere convergence, ergodic averages, strong sweeping out, convolution powers, oscillation inequalities, jump inequalities, variational inequalities, maximal functions, square functions, Calderón-Zygmund decomposition, Bourgain's entropy theorem.

R. Jones is partially supported by NSF Grant DMS—9531526.

This paper is based on a series of three talks given at the New York Journal of Mathematics Conference, which was held at Albany, N.Y. from June 9 – June 14, 1997.

ISSN 1076-9803/97

from analysis and probability that have proven to be most useful in the study of certain of questions in ergodic theory.

We begin Section 2 with a discussion of ways to show that a sequence of operators diverges a.e.. This discussion will include Bourgain's Entropy Theorem, which has already become a very useful tool in ergodic theory. To demonstrate this, we highlight some of the results that have already been obtained using his theorem. Since the proof involves a nice blend of analysis, probability and ergodic theory, Section 3 contains a discussion of the proof.

In Section 4 we discuss convolution powers of a single measure μ on $\mathbb{Z}$. This discussion leads to a discussion of the transfer principle, the Calderón-Zygmund decomposition, and some connections with singular integrals.

In Section 5 we discuss a "good–lambda" inequality that relates two maximal functions. These kinds of inequalities have proven useful in both probability and harmonic analysis. Currently only minimal work has been done regarding such inequalities in ergodic theory.

In Section 6 we discuss oscillation and variational inequalities. The proofs of these results involve an interesting interaction between martingales and certain ergodic averages. The variational inequalities give rise to inequalities for the number of "λ-jumps" that a sequence of ergodic averages can take. The results for "λ-jumps" were motivated by similar previously known results for martingales.

Throughout the paper, (X, Σ, m) denotes a complete non-atomic probability space, and τ denotes an ergodic, measurable, measure preserving point transformation from X to itself. To simplify the discussion, we will assume τ is invertible, although in many cases this assumption is not needed. If μ is a probability measure on $\mathbb{Z}$, we can define an associated operator

$$\mu(f)(x) = \sum_{k=-\infty}^{\infty} \mu(k) f(\tau^k x).$$

In this situation we will also refer to μ as an operator. It should be clear from the context if μ is being considered as a measure or as an operator.

The proofs of several of the results that will be discussed below are based on the same general principle. While the results are in ergodic theory, we can remove any discussion of ergodic theory from the proofs. The idea is to transfer the problem in ergodic theory to a problem in harmonic analysis. To understand what is going on, consider the orbit of a single point, $x, \tau x, \tau^2 x, \ldots$, and relabel the points as $0, 1, 2, \ldots$. We can then pretend we are working on $\mathbb{Z}$, with results obtained on $\mathbb{Z}$ implying results about the orbit of the point x. The details to make this precise are contained in a paper by A. P. Calderón [27] that appeared in the Proc. of the National Academy of Science in 1968. See Bellow's paper [8] for an excellent discussion of this general principle.

As a consequence of this transfer principle, when we see $A_k f(x) = \frac{1}{k}\sum_{j=0}^{k-1} f(\tau^j x)$ we can, without loss of generality, think of $x \in \mathbb{Z}$ and $\tau(x) = x+1$, so that $A_k f(x) = \frac{1}{k}\sum_{j=0}^{k-1} f(x+j)$, for $f \in \ell^1(\mathbb{Z})$. In fact, more is true. In many cases the results on $\mathbb{Z}$ imply the same results for averages associated with positive contractions. See [45] or [46] for details.

In general, we will not be interested in obtaining the best constants in the inequalities we consider. Hence c will often be used to denote a constant, but c may not be the same constant from one occurrence to the next.

2. Divergence and Strong Sweeping Out

In ergodic theory we are often interested in knowing if a given sequence of operators converges or diverges. Consequently, for a sequence of operators, $\{T_k\}$, each mapping $L^p(X, \Sigma, m)$ to itself for some p, $1 \le p \le \infty$, it is useful to have a collection of tools that can be used to establish a.e. divergence. Further, when divergence occurs, it is useful to know how badly the sequence diverges. In particular, sometimes it is possible that divergence occurs for some $f \in L^{p_0}$, but convergence occurs for all $f \in L^p$ for all $p > p_0$. Examples of this behavior, for averages along subsequences, were first constructed by Bellow [10] and later refinements were given by Reinhold [65].

For a sequence of $L^1 - L^\infty$ contractions, about the worst possible divergence that can occur is divergence of $\{T_k \chi_E\}$ where χ_E is the characteristic function of the measurable set E. For such operators, applied to χ_E, the possible values range from 0 to 1. The worst that can happen is for the operators to get arbitrarily close to both of these two extremes a.e.. This leads to the following definition.

Definition 2.1. A sequence $\{T_n\}$, of positive linear contractions on L^1 and L^∞, is said to be *strong sweeping out* if given $\epsilon > 0$ there is a set E such that $m(E) < \epsilon$ but $\limsup_n T_n \chi_E(x) = 1$ a.e. and $\liminf T_n \chi_E(x) = 0$ a.e..

There are several methods of showing that a sequence of operators is strong sweeping out. Moreover, many examples of operators that are strong sweeping out arise naturally in ergodic theory.

One family of operators that is of interest to ergodic theorists is the following. Let $\{w_k\}$ denote an increasing sequence of positive integers and let $\{\nu_n\}$ denote a dissipative sequence of probability measures with support on $\{w_k\}$. (A sequence of probability measures $\{\mu_n\}$ is said to be dissipative if $\lim_{n\to\infty} \mu_n(j) = 0$ for each integer j. If a sequence of measures is not dissipative, then certain early terms will continue to play a significant role in the value of the associated averages for arbitrarily late averages.) Define

$$\nu_n f(x) = \sum_{n=1}^{\infty} \nu_n(w_k) f(\tau^{w_k} x).$$

For example if we let $w_k = 2^k$, and $\nu_n(w_k) = \frac{1}{n}$ if $k \le n$, then

$$\nu_n f(x) = \frac{1}{n} \sum_{k=1}^{n} f(\tau^{2^k} x),$$

and we have the usual Cesaro averages along the subsequence $\{2^n\}$. We could also consider the Riesz harmonic means, given by

$$R_n f(x) = \frac{\alpha_n}{\ln n} \sum_{k=1}^{n-1} \frac{1}{n-k} f(\tau^{w_k} x),$$

where α_n is chosen so that $\frac{\alpha_n}{\ln n} \sum_{k=1}^{n-1} \frac{1}{n-k} = 1$, or many other similar weighted averages.

In [1] a condition was introduced that is sometimes useful to establish strong sweeping out. This condition is a condition on the Fourier transforms of the associated sequence of measures.

To fix notation, we define the Fourier transform as follows.

Definition 2.2. Let ν be a probability measure that is supported on the sequence $\{w_k\}$. The Fourier Transform of ν is given by

$$\hat{\nu}(t) = \sum_{k=1}^{\infty} \nu(w_k) e^{2\pi i w_k t}.$$

(In some cases it is convenient to let $\gamma = e^{2\pi i t}$ and write $\hat{\nu}(\gamma) = \sum_{k=1}^{\infty} \nu(w_k)\gamma^{w_k}$.)

Theorem 2.3. [1] *Let $\{\nu_n\}$ denote a dissipative sequence of measures on $\mathbb{Z}$. If there is a dense set $D \subset \mathbb{R}$ such that*

$$\lim_{n\to\infty} |\hat{\nu}_n(t)| = 1$$

for all $t \in D$, then the operators (ν_n), given by

$$\nu_n f(x) = \sum_{j\in\mathbb{Z}} \nu_n(w_j) f(\tau^{w_j} x),$$

are strong sweeping out.

To see how to apply this condition, we give a few examples.

Example 2.4. Let $\{\mu_n\}$ denote the sequence of Cesaro averages along the subsequence $\{2^k\}$. That is, let $\mu_n = \frac{1}{n}\sum_{j=1}^{n} \delta_{2^j}$. Then the associated operators are strong sweeping out. To see this just note that $\hat{\mu}_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{2\pi i 2^j t}$. If we take $t = \frac{r}{2^s}$ for some positive integers r and s, then for large enough j ($j \geq s$) we have $e^{2\pi i 2^j t} = 1$. Consequently, it is clear that we have $\lim_{n\to\infty} \hat{\mu}_n(t) = 1$ for each dyadic rational t. Since such t are dense, we can apply the theorem. (The fact that these averages diverge for $f \in L^p$, $p < \infty$, was first shown by Bellow [9], following Krengel's result [54] showing the existence of a subsequence for which a.e. divergence occurs. See also [42]. The study of strong sweeping out for the Cesaro averages along lacunary sequences, and for the Riemann sums, (that is, for averages of the form $A_n f(x) = \frac{1}{n}\sum_{k=1}^{n} f(x + \frac{k}{n})$, where f is periodic with period 1), started in the 1980's. The first proof of the strong sweeping out for Cesaro averages along lacunary sequences (a^k), as well as the first proof for strong sweeping out of the Riemann sums, was given by Bellow and Losert using a different technique from the one in Theorem 2.3. Their argument appears in the appendix to [1]. The fact that divergence occurs for bounded functions when we average along any lacunary sequence was first shown by Rosenblatt [68] using Bourgain's entropy condition, which we will discuss below.)

Remark 2.5. Actually the argument associated with Example 2.4 shows more. For any dissipative sequence of probability measures that are supported on the subsequence $\{2^k\}$, the associated operators are strong sweeping out. In particular, this means that the operators obtained from averaging the Cesaro averages, using other weighted averages, or selecting a subsequence of the Cesaro averages, will all result in a sequence of operators that is strong sweeping out. No matter what we do, if the dissipative sequence of measures is supported on the set of dyadic integers, strong sweeping out will always occur.

In some cases, even if the Fourier transform condition fails to show that we have strong sweeping out, we can still use the Fourier transform condition to obtain useful information. Consider the following.

Example 2.6. Let μ_n denote the Cesaro averages along the sequence of integers of the form $n[\log_2 n]$. (Here $[\cdot]$ is being used to denote the greatest integer function.) Then the associated sequence of operators diverges a.e. In fact, give $\epsilon > 0$, we can find a set E, $m(E) < \epsilon$, such that $\limsup_{n\to\infty} \mu_n \chi_E(x) \geq \frac{1}{2}$ a.e. To see this just note that $\mu_{2^{n+1}} \geq \frac{1}{2}\frac{1}{2^n}\sum_{k=2^n}^{2^{n+1}} \delta_{k[\log_2 k]} = \frac{1}{2}\nu_n$. Now just apply the Fourier transform condition to show that $\{\nu_n\}$ is strong sweeping out. This follows, since it is enough to look at the subsequence $\{\nu_{2^n}\}$. We have $\hat{\nu}_{2^n}(t) = \frac{1}{2^{2^n}}\sum_{j=2^{2^n}}^{2\times 2^{2^n}} e^{2\pi i j 2^n t}$. Now just take $t = \frac{r}{2^s}$ as before, and the result follows. (At this time it is not known if the averages associated with μ_n are strong sweeping out.)

Remark 2.7. It is interesting that as we saw above, the Cesaro averages along the subsequence $\{n[\log_2 n]\}$ can diverge even for $f \in L^\infty$ while Wierdl [75] has shown that the Cesaro averages along the subsequence $\{[n \log_2 n]\}$ converge a.e. for all $f \in L^p$, $p > 1$. This shows quite clearly that sequences can have the same rate of growth, and still have the associated averages behave quite differently. See also the paper by Boshernitzan and Wierdl, [22], where it is shown that many more subsequences yield convergent sequences of Cesaro averages.

In [1] an example is constructed to show that the Fourier transform condition is not necessary for a sequence of averages to have the strong sweeping out property. Consequently it becomes interesting to consider other conditions that imply strong sweeping out.

A second condition for establishing strong sweeping out was also introduced in [1]. This condition is sometimes satisfied when the Fourier transform condition fails.

Definition 2.8. A sequence of real numbers $\{w_k\}$ satisfies the $C(\alpha)$ condition, $0 < \alpha < \frac{1}{2}$, if given any finite sequence of real numbers, $x_1, \ldots, x_N$, there is a real number r such that

$$rw_k \in x_k + (\alpha, 1-\alpha) + \mathbb{Z},$$

for $k = 1, 2, \ldots, N$.

This condition can be thought of as a very weak form of the following Theorem due to Kronecker.

Theorem 2.9 (Kronecker). *Let $w_1, \ldots, w_N$ be real numbers such that $w_1, \ldots, w_N$ and 1 are linearly independent over the rationals. Let $\epsilon > 0$ and $x_1, \ldots, x_N$ be given. Then there is an integer r such that $rw_k \in x_k + (-\epsilon, \epsilon) + \mathbb{Z}$, for $k = 1, \ldots, N$.*

The reason for introducing the $C(\alpha)$ condition is clear from the following theorem.

Theorem 2.10 ([1]). *If $\{w_k\}$ satisfies the $C(\alpha)$ condition, and $\{\nu_n\}$ is a dissipative sequence of measures with support on $\{w_k\}$, then the associated operators $\{\nu_n\}$*

defined by $\nu_n f(x) = \sum_{k=1}^{\infty} \nu_n(w_k) f(\tau^{w_k} x)$ *are strong sweeping out. The same result holds if*

$$\nu_n f(x) = \sum_{k=1}^{\infty} \nu_n(w_k) f(U_{w_k} x),$$

where $U_t x$ *is a measure preserving flow.* (*In this case* $\{w_k\}$ *need not be an integer sequence.*)

We now easily have the following.

Example 2.11. Let $\{w_k\}$ be a sequence of real numbers that are linearly independent over the rationals. Let $\{\nu_n\}$ be a dissipative sequence of measures with support on $\{w_k\}$. Then the associated operators $\{\nu_n\}$ are strong sweeping out. This is immediate from Theorem 2.10, combined with Kronecker's Theorem. In particular, if we consider $w_k = \sqrt{k}$, it is not difficult to see that we can extract a subsequence which is linearly independent over the rationals, and has positive density in the sequence. Consequently we know that Cesaro averages $\frac{1}{n}\sum_{k=1}^{n} f(U_{\sqrt{k}} x)$ can diverge a.e. (See [52] for further discussion of this example as well as further examples of situations where Kronecker's Theorem can be used to imply strong sweeping out.)

There are many sequences which satisfy the $C(\alpha)$ condition. In [1] it is shown that any lacunary sequence of integers (after possibly neglecting the first few terms) satisfies the $C(\alpha)$ condition. In fact, any finite union of lacunary sequences (again after possibly neglecting the first few terms) satisfies the $C(\alpha)$ condition. (See [56].) Consequently there is the following theorem.

Theorem 2.12. *If* $\{w_k\}$ *is a lacunary sequence of positive integers,* (*or a finite union of lacunary sequences of positive integers*) *and* $\{\nu_n\}$ *is any dissipative sequence of probability measures with support on* $\{w_k\}$*, then the associated operators,* $\{\nu_n\}$ *are strong sweeping out.*

In [44] Theorem 2.12 is applied to obtain the following example.

Example 2.13. Let $\{Y_n\}$ be a sequence of Bernoulli random variables with $P(Y_n = 1) = \frac{1}{n}$ and $P(Y_n = 0) = 1 - \frac{1}{n}$. Then for a.e. ω, the Cesaro averages associated with the subsequence $A(\omega) = \{n : Y_n(\omega) = 1\}$ will be strong sweeping out. The idea of the proof is to show that by eliminating only a small proportion of the set $A(\omega)$, one can obtain a lacunary sequence, then apply Theorem 2.12 above.

The above theorem and example suggest some interesting questions. For an arbitrary sequence of integers, Λ, let C_Λ denote the space of all continuous function on $\mathbb{T}$ such that $\hat{f}(n) = 0$ for all $n \notin \Lambda$. The set Λ is a Sidon set if every $f \in C_\Lambda$ has an absolutely convergent Fourier series. It is not difficult to see that every Lacunary sequence is a Sidon set, and further, so are finite unions of lacunary sequences.

Question 2.14. *Is it true that the Cesaro averages associated with a Sidon set are strong sweeping out?* (*In Rosenblatt* [67] *it is shown that the Cesaro averages along Sidon sets fail to converge in mean.*) *Second, if we form the sets* $A(\omega)$ *as above, we see that they are close to lacunary for a.e.* ω*, but are these sets Sidon sets for a.e.* ω*?*

In [15] moving averages of the form $\nu_k = \frac{1}{\ell_k}\sum_{j=n_k}^{n_k+\ell_k-1}\delta_k$ are studied. Conditions, which are both necessary and sufficient for the a.e. convergence of these averages are given. In particular, if $n_k = k^2$ and $\ell_k = k$ then the averages are strong sweeping out. (Akcoglu and del Junco [2] had already shown that this particular sequence of averages diverge a.e.) To see that the Fourier transform condition will not help in this case, just note that $\hat{\nu}_k(t) = \frac{1}{\ell_k}\sum_{j=n_k}^{n_k+\ell_k-1} e^{2\pi ijt} = e^{2\pi in_k t}\frac{1}{\ell_k}\sum_{j=0}^{\ell_k-1} e^{2\pi ijt} = e^{2\pi in_k t}\frac{e^{\pi i\ell_k t}}{e^{\pi it}}\frac{\sin \pi\ell_k t}{\ell_k \sin \pi t}$. It is clear that this converges to zero for all $t \neq 0 \pmod 1$. It is also easy to see the $C(\alpha)$ condition fails to be satisfied for this example as well. (In a forthcoming paper, [4], the $C(\alpha)$ condition will be generalized, and that generalization can be used to show strong sweeping out for this class of examples, as well as other related examples.)

These moving averages are associated with the following example. Let $\mu = \frac{1}{2}(\delta_0+\delta_1)$, and consider $\mu^n = \mu\star\mu\cdots\star\mu$ where there are n terms in the convolution. Then $\mu^n = \frac{1}{2^n}\sum_{j=0}^{n}\binom{n}{j}\delta_j$. (See Rosenblatt's paper, [66], where these averages were first shown to be strong sweeping out.) By looking at the normal approximation to the binomial, we see that most of the "mass" of these measures will be in an interval proportional to $\sqrt{n}$ and centered at $\frac{n}{2}$. Consequently, it is like a moving average, for which we already know strong sweeping out. (See [15] for further discussion of this example.) However, neither the Fourier transform condition, nor the $C(\alpha)$ condition are satisfied. (It turns out that modifications of either condition can in fact be used to establish strong sweeping out for these examples. See [1], [3] and [4].)

This brings us to a third condition that can be used to show divergence of a sequence of operators.

Theorem 2.15 (Bourgain[18]). *Let (T_k) denote a sequence of uniformly bounded linear operators in L^2 of a probability space. Assume*

(1) *The T_k's commute with a sequence (R_j) of positive isometries on L^2 of the same probability space, satisfying $R_j(1) = 1$ and the mean ergodic theorem.*
(2) *There is a $\delta > 0$ such that given any $N > 0$ there is a function f, $\|f\|_2 \le 1$, and $n_1, n_2, \ldots, n_N$ such that*

$$\|T_{n_j} f - T_{n_k} f\|_2 \ge \delta$$

for $j \neq k$, $1 \le j,k \le N$.

Then there is a bounded function g such that $(T_k g)$ is not a.e. convergent.

Condition (2) in the above theorem is often referred to as Bourgain's entropy condition. Bourgain used this theorem to prove a number of very interesting results.

(1) The Bellow problem: Let $\{a_k\}$ denote a sequence of real numbers that converges to zero. Does there exist a bounded function f on the real line such that $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f(x+a_k)$ diverges a.e.?
(2) A problem of Marcinkiewicz and Zygmund on Riemann sums: Does there exist a bounded function f on $[0,1)$ such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} f(x+\frac{j}{n})$$

diverges? (Here we are using addition mod 1.) This was first solved by Rudin [70] by different techniques, and as mentioned earlier, strong sweeping out was first shown by Bellow and Losert (see the appendix to [1]).

(3) Khinchine's Problem: Does there exist a bounded function on $[0, 1)$ such that the averages $\frac{1}{n}\sum_{j=1}^{n} f(jx)$ diverge? (This was first solved by Marstrand [57], again by different techniques.)

In each of the above problems, Bourgain was able to use his theorem to show divergence of the averages being considered. While the second and third problems had been solved earlier, his method gives a unified way of solving such problems. Later Rosenblatt [68] used Bourgain's Entropy Theorem to show that for any dissipative sequence of measures supported on a lacunary sequence, divergence occurs when these averages are applied to the indicator function of some very small set. This was part of the motivation for later trying to prove strong sweeping out for such averages. In [13] Bourgain's Entropy Theorem is used to show that the convolution powers (see Section 4 below), applied to bounded functions, can diverge if a condition known as the bounded angular ratio condition fails to be satisfied. Applications to problems outside ergodic theory also occur. For example in [39] the operators defined by

$$T_\gamma f(x) = \sum_{n=-\infty}^{\infty} \hat{f}(n)|n|^{i\gamma} sgn(n) e^{inx}$$

are considered. These operators are related to the Hilbert transform, and convergence in mean is known. However, using Bourgain's Entropy Theorem it is possible to prove the following.

Theorem 2.16 ([39]). *Let (γ_n) denote a sequence of real numbers, converging to zero. Then there is a function $f \in L^\infty(\mathbb{T})$ such that $\lim_{j\to\infty} T_{\gamma_j} f(x)$ fails to exist on a set of positive measure.*

3. Proof of Bourgain's Entropy Theorem

In this section we want to sketch a proof of Bourgain's Entropy Theorem, and show its relation with strong sweeping out. Bourgain's Entropy Theorem does not imply strong sweeping out. However, the following more quantative version does. This version is a slight variation on a result that is contained in joint work with M. Akcoglu and M. D. Ha [3]. Most of the ideas involved in the proof are either in Bourgain's original paper or are the result of discussions with Akcoglu and Ha.

Theorem 3.1. *Let (T_k) denote a sequence of uniformly bounded linear operators on L^2 of a probability space. Assume*

(1) *The T_k's commute with a sequence (R_j) of positive isometries on L^2 of the same probability space, which satisfy $R_j(1) = 1$ and the mean ergodic theorem. (That is, $\|\frac{1}{J}\sum_{j=1}^{J} R_i f - \int_X f(x)dx\|_2 \to 0$.)*

(2) *There is a $0 < \rho < 1$ and $d > 0$ such that for any integer L we can find $(n_1, n_2, \ldots, n_L)$ and f, $\|f\|_2 \le 1$, such that*

$$\int_X T_{n_j} f(x) T_{n_k} f(x) dx \le \rho \|T_{n_j} f\|_2 \|T_{n_k} f\|_2$$

for all $k \neq j$, and

$$\int_X |T_{n_j} f(x)|^2 dx \geq d^2$$

for $1 \leq j \leq L$.

Then for any $\eta > 0$, $\epsilon > 0$, and $\epsilon' > 0$, we can find a function g such that

$$\|g\|_2 < \eta, \ \|g\|_\infty \leq 1 \text{ and}$$

$$m\Big\{x : \max_{1 \leq k \leq L} T_{n_k} g(x) > d(1-\epsilon)\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}\Big\} > 1 - \epsilon'.$$

REMARK 3.2. If in a particular application we can take d as close to 1 as we desire, and if ρ can be taken as close to zero as desired, then with g as above, and any $\epsilon > 0$, $\epsilon' > 0$ we have

$$m\{x : \max_{1 \leq k \leq L} T_{n_k} g > 1 - \epsilon\} > 1 - \epsilon'. \tag{1}$$

If in addition, the (T_k) are positive contractions on L^1 and L^∞, then, using inequality (1), strong sweeping out can be obtained. See [12] for a discussion of replacing a function of small norm by a characteristic function, and see [32], where del Junco and Rosenblatt show how to go from inequality (1), (with g the characteristic function of a set of small measure) to strong sweeping out. Also see [3] where further discussion of this, as well as related details, can be found.

We want to give a proof of this version of Bourgain's entropy theorem. The idea is to exploit the fact that condition (2) above implies a degree of orthogonality between the vectors in the family $\{T_{n_j} f\}_{j=1}^{L}$. However, before we give the proof, we want to show how Theorem 3.1 implies Theorem 2.15.

Proof of Theorem 2.15. If we could show that the hypothesis of Theorem 2.15 implies the hypothesis of Theorem 3.1 then we would have the conclusion of Theorem 3.1. We could then finish the proof by applying the Banach Principle for L^∞. (See [12] for a discussion of this principle and the necessary details to complete the argument.)

To see that the hypothesis of Theorem 2.15 implies the hypothesis of Theorem 3.1 we argue as follows. There are two ways that vectors can be far apart. One way is for them to have different lengths, and the other is for them to point in very different directions. If we can eliminate the possibility that they have different lengths, they must point in different directions. That is, we have the required orthogonality.

Let (T_k), f and δ be as in assumptions 1) and 2) of Theorem 2.15. In the following argument we can assume that the uniform bound on the operators is 1. If not, just divide by the uniform bound, and consider the new sequence of operators.

In the hypothesis of Theorem 3.1, we need to find L vectors with the orthogonality property (2). If there is a vector with norm less than $\frac{\delta}{2}$, then the rest must all have norm greater than $\frac{\delta}{2}$. Hence by taking one extra vector, we can assume that all L vectors have norm greater than $\frac{\delta}{2}$. Fix $\gamma < \frac{\delta}{4}$, and cut the interval $[\frac{\delta}{2}, 1]$ into a finite number, (say r), pieces of length less than γ. Call the kth interval I_k. In the hypothesis of Theorem 2.15 take $N > rL$.

Let $A_k = \{j | \, \|T_{n_j} f\|_2 \in I_k\}$. Since $N > rL$, and there are only r sets, A_k, at least one A_k, say A_{k_0} must have at least L elements. If we rename our terms, we have $(T_1 f, T_2 f, \dots, T_L f)$ such that $\|T_j f\|_2 \in I_{k_0}$, and $\|T_i f - T_j f\|_2 \geq \delta$. That is, the functions $(T_1 f, \dots, T_L f)$ are all about the same length, and satisfy condition (2) of Theorem 2.15.

Define $\tilde{T}_j = \frac{1}{\|T_j f\|_2} T_j$. We now have that $\|\tilde{T}_j f\|_2 = 1$ for each j, $1 \leq j \leq L$. We also have

$$\begin{aligned}
\delta &\leq \|T_i f - T_j f\|_2 \\
&= \Big\| \|T_i f\|_2 \tilde{T}_i f - \|T_j f\|_2 \tilde{T}_j f \Big\|_2 \\
&= \Big\| \|T_i f\|_2 \tilde{T}_i f - \|T_i f\|_2 \tilde{T}_j f + \|T_i f\|_2 \tilde{T}_j f - \|T_j f\|_2 \tilde{T}_j f \Big\|_2 \\
&\leq \Big\| \|T_i f\|_2 \tilde{T}_i f - \|T_i f\|_2 \tilde{T}_j f \Big\|_2 + \Big\| \|T_i f\|_2 \tilde{T}_j f - \|T_j f\|_2 \tilde{T}_j f \Big\|_2 \\
&\leq \|T_i f\|_2 \|\tilde{T}_i f - \tilde{T}_j f\|_2 + \|\tilde{T}_j f\|_2 (\|T_i f\|_2 - \|T_j f\|_2) \\
&\leq \|\tilde{T}_i f - \tilde{T}_j f\|_2 + \gamma.
\end{aligned}$$

Thus $\delta - \gamma \leq \|\tilde{T}_i f - \tilde{T}_j f\|_2$.

From this we see

$$\begin{aligned}
(\delta - \gamma)^2 &\leq \int_X |\tilde{T}_i f(x) - \tilde{T}_j f(x)|^2 dx \\
&\leq \int_X |\tilde{T}_i f(x)|^2 dx - 2 \int_X \tilde{T}_i f(x) \tilde{T}_j f(x) dx + \int_X |\tilde{T}_j f(x)|^2 dx \\
&\leq 2 - 2 \int_X \tilde{T}_i f(x) \tilde{T}_j f(x) dx.
\end{aligned}$$

Solving for the inner product, we see

$$\int_X \tilde{T}_i f(x) \tilde{T}_j f(x) dx \leq 1 - \frac{(\delta - \gamma)^2}{2}.$$

Thus if we take $\rho = 1 - \frac{(\delta-\gamma)^2}{2}$, we have $\langle T_i f, T_j f \rangle < \rho \|T_i f\|_2 \|T_j f\|_2$. Consequently the hypothesis of Theorem 2.15 is satisfied (with $\|T_j f\|_2 \geq \frac{\delta}{2}$). □

Before we prove Theorem 3.1 we will first prove a proposition. While this proposition appears too special to be useful for our purposes, it is in fact the key ingredient in the proof of Theorem 3.1. In the following, when we say a random vector has $N(0, \Sigma)$ distribution we mean it is a multivariate normal random vector, with mean given by the zero vector, and covariance given by the matrix Σ.

Proposition 3.3. *Fix ρ, $0 \leq \rho < 1$. Assume that for all large L, we can find $(T_{n_1}, T_{n_2}, \dots, T_{n_L})$ and f such that $(T_{n_1} f, T_{n_2} f, \dots, T_{n_L} f)$ has $N(0, \Sigma)$ distribution with $\sigma_{ij} \leq \rho$ for $1 \leq i < j \leq L$ and $\sigma_{ii} = 1$ for $1 \leq i \leq L$. Further, assume f is distributed $N(0,1)$. Let $0 < \epsilon < \frac{1}{4}$ and $\epsilon' > 0$ be given. Then for any $\eta > 0$ and for all L large enough, depending on ρ, ϵ and ϵ', we can find a function g such that $\|g\|_2 \leq \eta$, $\|g\|_\infty \leq 1$, and*

$$m\Big\{ x | \max_{1 \leq k \leq L} T_{n_k} g(x) > (1 - \epsilon) \frac{\sqrt{1-\rho}}{1 + \sqrt{\rho}} \Big\} > 1 - \epsilon'.$$

REMARK 3.4. We do not necessarily need to have the $\{T_{n_j}\}$ defined on all of L^2. It is enough for the operators to be defined and uniformly bounded on the linear subspace generated by functions of the form $f\chi_E$ where χ_E can be any L^2 function that takes on only the values 0 and 1.

REMARK 3.5. While the constant, $(1-\epsilon)\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}$ looks complicated, it will arise in the computation below. While we could replace it with a simpler expression, the proof would be slightly less natural.

To prove Proposition 3.3 we need a lemma about the multivariate normal and a lemma about the size of the tail of a normal random variable.

Lemma 3.6. *Let* $(Y_1, Y_2, \ldots, Y_L)$ *denote a random vector with multivariate normal distribution* $N(0,\Sigma)$ *with* $\sigma_{ij} \le \rho$ *for* $i \ne j$ *and* $\sigma_{ii} = 1$ *for* $1 \le i \le L$. *Let* $0 < \epsilon < 1$ *and* $0 < \epsilon'$ *be given. Then for all* L *large enough (depending on* ρ *and* ϵ *and* ϵ'*) we have the following estimate:*

$$P\Big\{\omega : \max_{1\le k\le L} Y_k(\omega) > (1-\epsilon)\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}\sqrt{2\ln L}\Big\} > 1-\epsilon'.$$

Proof. Slepian's lemma (see [74] for a proof) says that if we have 2 multivariate normal random vectors, X and Y, both with mean zero, and covariance Σ and Σ' respectively, and if each entry of Σ is less than or equal to the corresponding entry in Σ', then

$$P(\max_{1\le k\le L} X_k > \lambda) \ge P(\max_{1\le k\le L} Y_k > \lambda).$$

This is not too surprising since the more positive correlation, the less the vector should behave like an independent sequence, and so the maximum should be smaller.

Using Slepian's Lemma, we can increase some entries in Σ, so that without loss of generality we can assume $\sigma_{ij} = \rho$ for $i \ne j$. The special covariance structure makes it possible to obtain the necessary estimate of

$$P\Big\{\omega : \max_{1\le k\le L} Y_k(\omega) > (1-\epsilon)\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}\sqrt{2\ln L}\Big\}.$$

To make the required computation, we first consider a vector of independent standard normal random variables; $(Z_0, Z_1, \ldots, Z_L)$. Form the new vector

$$\big(\sqrt{1-\rho}Z_1 + \sqrt{\rho}Z_0, \sqrt{1-\rho}Z_2 + \sqrt{\rho}Z_0, \ldots, \sqrt{1-\rho}Z_L + \sqrt{\rho}Z_0\big).$$

This new random vector has the multivariate normal distribution. In fact the covariance matrix for this random vector is the same as the covariance matrix of the random vector $(Y_1, Y_2, \ldots, Y_L)$. Consequently, by Slepian's Lemma we have

$$\begin{aligned} P\{\max_{1\le k\le L} Y_k > \lambda\} &= 1 - P\{\cap_{k=1}^{L}(Y_k \le \lambda)\} \\ &= 1 - P\{\cap_{k=1}^{L}(\sqrt{1-\rho}Z_k + \sqrt{\rho}Z_0 \le \lambda)\}. \end{aligned}$$

We will now take advantage of the structure of this new sequence. We note that

$$P\big(\cap_{k=1}^{L}(\sqrt{1-\rho}Z_k + \sqrt{\rho}Z_0) \le \lambda\big) = P\big(\max_{1\le k\le L}\sqrt{1-\rho}Z_k + \sqrt{\rho}Z_0 \le \lambda\big).$$

Since $\max_{1\le k\le L}\sqrt{1-\rho}Z_k$ and $\sqrt{\rho}Z_0$ are independent, the distribution of the sum is given by a convolution. Thus for all λ we have

$$\begin{aligned}
P\Big\{\cap_{j=1}^{L}\big(\sqrt{1-\rho}Z_j+\sqrt{\rho}Z_0\le\lambda\big)\Big\}&\\
&=\int_{-\infty}^{\infty}P\left(\cap_{j=1}^{L}(\sqrt{1-\rho}Z_j+\sqrt{\rho}t\le\lambda)\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt\\
&=\int_{-\infty}^{\infty}P\left(\cap_{j=1}^{L}Z_j\le\frac{\lambda-\sqrt{\rho}t}{\sqrt{1-\rho}}\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt\\
&=\int_{-\infty}^{\infty}\prod_{j=1}^{L}\left(\int_{-\infty}^{\frac{\lambda-\sqrt{\rho}t}{\sqrt{1-\rho}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt.
\end{aligned}$$

We need to estimate this last integral, and show it is less than ϵ' if L is large enough.

Let $\lambda=(1-\epsilon)\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}\sqrt{2\ln L}$, and select L so large that $\int_{\lambda}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt<\frac{\epsilon'}{2}$. We now break the integral into 2 parts,

$$\begin{aligned}
&\int_{-\infty}^{-\lambda}\prod_{j=1}^{L}\left(\int_{-\infty}^{\frac{\lambda-\sqrt{\rho}t}{\sqrt{1-\rho}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt\\
&\qquad+\int_{-\lambda}^{\infty}\prod_{j=1}^{L}\left(\int_{-\infty}^{\frac{\lambda-\sqrt{\rho}t}{\sqrt{1-\rho}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt\\
&=I+II.
\end{aligned}$$

Because of our choice of λ, we clearly have that $I<\frac{\epsilon'}{2}$ if L is large enough. We now need an estimate for the integral II.

We have

$$\begin{aligned}
II&=\int_{-\lambda}^{\infty}\prod_{j=1}^{L}\left(\int_{-\infty}^{\frac{\lambda-\sqrt{\rho}t}{\sqrt{1-\rho}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt\\
&\le\int_{-\lambda}^{\infty}\prod_{j=1}^{L}\left(\int_{-\infty}^{\frac{\lambda(1+\sqrt{\rho})}{\sqrt{1-\rho}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt\\
&=\int_{-\lambda}^{\infty}\left(\int_{-\infty}^{\frac{\lambda(1+\sqrt{\rho})}{\sqrt{1-\rho}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\right)^{L}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt.
\end{aligned}$$

Thus if

$$\left(\int_{-\infty}^{\frac{\lambda(1+\sqrt{\rho})}{\sqrt{1-\rho}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\right)^{L}<\frac{\epsilon'}{2},$$

then we have the desired estimate.

Let $M=\frac{\lambda(1+\sqrt{\rho})}{\sqrt{1-\rho}}=(1-\epsilon)\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}\left(\frac{1+\sqrt{\rho}}{\sqrt{1-\rho}}\right)\sqrt{2\ln L}=(1-\epsilon)\sqrt{2\ln L}$. We need

$$\left(\int_{-\infty}^{M}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt\right)^{L}<\frac{\epsilon'}{2},$$

or

$$\left(1 - \int_M^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt\right)^L < \frac{\epsilon'}{2}.$$

Taking logs, we see that we need to establish the inequality

$$L \ln\left(1 - \int_M^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt\right) < \ln \frac{\epsilon'}{2}.$$

Since $\ln(1-x) \leq -x$, it will be enough to show

$$L\left(-\int_M^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt\right) < -\ln(2/\epsilon').$$

Estimating the integral, (see [34] page 175) we need to have

$$L\Big(\frac{1}{M} - \frac{1}{M^3}\Big)\frac{1}{\sqrt{2\pi}} e^{-\frac{M^2}{2}} > \ln(2/\epsilon').$$

Using the value of M given above, and the fact that $\epsilon > 0$, we can now easily check that we have the desired estimate if L is large enough. □

Lemma 3.7. *Let X have $N(0,1)$ distribution, and let $0 < \delta < \frac{1}{2}$ be given. Define $b = \delta X \chi_{\{|\delta X|>1\}}$. Then*

$$\|b\|_2^2 \leq \frac{3\delta}{\sqrt{2\pi}} e^{-\frac{1}{2\delta^2}}.$$

Proof. The proof is just a standard computation with the density function of a normal random variable, and is left as an exercise for the reader. □

Proof of Proposition 3.3. Let

$$\delta = \frac{\mu}{\sqrt{2\ln L}},$$

where $\mu < 1$ will be chosen later. Define

$$g(x) = \delta f(x)\chi_{\{x||\delta f(x)|\leq 1\}}(x).$$

Assume that $\sup_j \|T_j\|_2 \leq \beta$. For L large enough we have $\delta < \eta$, and hence $\|g\|_2 < \eta$. We need to estimate

$$P\Big\{x| \max_{1\leq k\leq L} |T_{n_k} g(x)| > (1-\epsilon)\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}\Big\}.$$

Write $g = \delta f - b$. We then have

$$\begin{aligned}
&P\Big\{x : \max_{1\leq k\leq L} |T_{n_k} g(x)| > (1-\epsilon)\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}\Big\} \\
&\quad = P\Big\{x : \max_{1\leq k\leq L} |T_{n_k}(\delta f(x) - b(x))| > (1-\epsilon)\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}\Big\} \\
&\quad \geq P\Big\{x : \max_{1\leq k\leq L} |T_{n_k}\delta f(x)| > (1-\frac{1}{2}\epsilon)\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}\Big\} \\
&\qquad\quad - P\Big\{x : \max_{1\leq k\leq L} |T_{n_k} b(x)| > \frac{1}{2}\epsilon\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}\Big\} \\
&\quad = I - II.
\end{aligned}$$

For I, using our value of δ above, and Lemma 3.6 (with $\frac{1-\frac{\epsilon}{2}}{\mu}$ in place of $1-\epsilon$ in Lemma 3.6, hence we need $\frac{1-\frac{\epsilon}{2}}{\mu} < 1$), we have

$$I = P\Big\{x : \max_{1\le k\le L} T_{n_k} f(x) > (\frac{1-\frac{1}{2}\epsilon}{\mu})\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}\sqrt{2\ln L}\Big\} \ge 1 - \frac{\epsilon'}{2}$$

for all L large enough.

We now need to estimate II. We will show that for L large enough, we have $II < \frac{\epsilon'}{2}$. By taking L large enough, both the estimate of I and the estimate of II will hold, and we will have the estimate.

$$P\Big\{x : \max_{1\le k\le L} |T_{n_k} g(x)| > (1-\epsilon)\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}\Big\} > 1-\epsilon'.$$

Let $\alpha = \frac{\epsilon}{2}\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}$. We have

$$\begin{aligned} P\Big\{x : \max_{1\le k\le L} |T_k b(x)| > \alpha\Big\} &\le \frac{1}{\alpha^2}\int_X \max_{1\le k\le L} |T_k b(x)|^2 dx \\ &\le \frac{1}{\alpha^2}\int_X \sum_{k=1}^{L} \chi_{E_k}(x)|T_k b(x)|^2 dx \end{aligned}$$

where E_k is the set of x where the maximum is achieved for that value of k. However,

$$\begin{aligned} \int_X \sum_{k=1}^{L} \chi_{E_k}(x)|T_k b(x)|^2 dx &\le \sum_{k=1}^{L} \int_X \chi_{E_k}(x)|T_k b(x)|^2 dx \\ &\le \sum_{k=1}^{N} \int_X |T_k b(x)|^2 dx \\ &\le \sum_{k=1}^{L} \beta^2 \int_X |b(x)|^2 dx \\ &\le L\beta^2 \|b\|_2^2. \end{aligned}$$

Thus we have $II \le \frac{1}{\alpha^2} L\beta^2\|b\|_2^2$. Applying Lemma 3.7, we see

$$II < \frac{1}{\alpha^2} L\beta^2 \frac{3\delta}{\sqrt{2\pi}} e^{-\frac{1}{2\delta^2}}.$$

The required estimate will follow if $\frac{1}{2\delta^2} \ge \gamma \ln L$ for some $\gamma > 1$, since then we would have $II < c\frac{\beta^2}{\alpha^2} L e^{-\gamma \ln L} < c\frac{\beta^2}{\alpha^2} L^{1-\gamma}$. Thus we could take L large enough so that $c\frac{\beta^2}{\alpha^2} L^{1-\gamma} < \frac{\epsilon'}{2}$. Since

$$\delta = \frac{\mu}{\sqrt{2\ln L}},$$

we see that we need $\frac{1}{\mu^2} > \gamma$. Thus we select μ such that $1 - \frac{\epsilon}{2} < \mu < 1$, and then select γ with $\frac{1}{\mu^2} > \gamma > 1$. With these selections, all the required estimates are satisfied, and the proof is complete. □

Proof of Theorem 3.1. Assume f and $T_{n_1}, T_{n_2}, \dots, T_{n_L}$ are given as in the statement of the theorem. We can assume that $\int T_{n_j} f(x) T_{n_k} f(x)\, dx \le \rho$ for $j \ne k$ and $\int |T_{n_k} f(x)|^2\, dx = 1$ for $1 \le k \le L$, since if not, replace each operator T_{n_k} by the new operator S_{n_k} defined by $S_{n_k}g(x) = T_{n_k}g(x)/\|T_{n_k}f\|_2$. These new operators will be uniformly bounded since $\|T_{n_k}f\|_2 \geq d$ for all k, and the T_{n_k} were uniformly bounded. Further, they clearly satisfy $\int S_{n_j}f(x)S_{n_k}f(x)\,dx \leq \rho$ for $j \neq k$ and $\int |S_{n_k}f(x)|^2\,dx = 1$, as required.

Fix α and γ so that $\rho < \alpha < \gamma < 1$ and $\frac{\alpha}{\gamma} < \gamma < 1$. Define $\tilde{\rho} = \frac{\alpha}{\gamma}$. (This is to give us some room, since the best we know is that certain averages are converging to a number less than or equal to ρ, but at any finite stage we may be above ρ, and other averages are converging to 1, but at any finite time may be below 1.) We will now introduce a product space $X \times \Omega$ and extend the operators to that space. In the product space we will see that for most fixed x, treating our function on the product space as a function of $\omega \in \Omega$, we will be able to get an example with the required properties. We will then apply Fubini's Theorem to see that for some fixed $\omega \in \Omega$, the associated function of x will have the required properties, concluding the proof. To do this, we need to first put ourselves into position to use Proposition 3.3. With this in mind, we introduce the following matrix.

For each $x \in X$ and positive integer J, let

$$A(x) = \frac{1}{\sqrt{J}} \begin{pmatrix} R_1T_{n_1}f(x) & R_1T_{n_2}f(x) & \dots & R_1T_{n_L}f(x) \\ R_2T_{n_1}f(x) & R_2T_{n_2}f(x) & \dots & R_2T_{n_L}f(x) \\ \dots & \dots & \dots & \dots \\ R_JT_{n_1}f(x) & R_JT_{n_2}f(x) & \dots & R_JT_{n_L}f(x) \end{pmatrix}.$$

Let $(Z_1, Z_2, \dots)$ be a sequence of independent standard normal random variables. Let $Z = (Z_1, Z_2, \dots, Z_J)$ denote the initial segment of length J, and let $Y_x(\omega) = (Y_1(\omega), Y_2(\omega), \dots, Y_L(\omega))$ be given by $Y_x(\omega) = Z(\omega)A(x)$.

Then Y_x has a multinormal distribution with covariance matrix given by $\Gamma = A^tA$. Thus

$$\begin{aligned} \gamma_{jk}(x) &= \frac{1}{J}\sum_{i=1}^{J} R_iT_{n_j}f(x)R_iT_{n_k}f(x) \\ &= \frac{1}{J}\sum_{i=1}^{J} R_i(T_{n_j}fT_{n_k}f)(x). \end{aligned}$$

Since the sequence (R_i) satisfies the mean ergodic theorem, we have that

$$\frac{1}{J}\sum_{i=1}^{J} R_i(T_{n_j}fT_{n_k}f)(x)$$

converges in mean to

$$\int_X T_{n_j}f(x)T_{n_k}f(x)dx.$$

Hence for each fixed function f and sequence $T_{n_1}, T_{n_2}, \dots, T_{n_L}$, we can let

$$E_J = \{x : \gamma_{jk}(x) < \alpha \text{ for all } 1 \leq j < k \leq L;\ \gamma_{jj}(x) > \gamma \text{ for } 1 \leq j \leq L\}.$$

Since

$$\int_X T_{n_j}f(x)T_{n_k}f(x)dx \leq \rho \text{ for all } 1 \leq j < k \leq L,$$

and

$$\int_X T_{n_j} f(x) T_{n_j} f(x) dx = 1 \text{ for } 1 \le j \le L,$$

we see that $m(E_J) \to 1$ as $J \to \infty$.

On the new space, $X \times \Omega$, we extend the definition of $\{T_k\}$ to functions of the form $\sum Z_j(\omega) f_j(x)$ by $T_k(\sum Z_j(\omega) f_j(x) = \sum Z_j(\omega) T_k f_j(x)$.

Define $F(x,\omega) = \frac{1}{\sqrt{J}} \sum_{j=1}^{J} Z_j(\omega) R_j f(x)$.

For $x \in E_J$ we now modify each T_{n_j} slightly. For $x \in E_J$ define $\tilde{T}_{n_j} F(x,\omega) = \frac{1}{\sqrt{\gamma_{jj}}} T_{n_j} F(x,\omega)$. For $x \notin E_J$ we can let $\tilde{T}_{n_j} F(x,\omega) = T_{n_j} F(x,\omega)$.

For each fixed x we can now think of our operators as operators on Ω.

For each fixed $x \in E_J$, since $\tilde{\rho} = \frac{\alpha}{\gamma} < 1$, we are now in the situation of Proposition 3.3, and hence given $\eta > 0$, $\epsilon > 0$ and $\epsilon_1 > 0$, we know if L is large enough, depending only on η, ϵ, ϵ_1 and $\tilde{\rho}$ then we can find a function $G(x,\omega)$ such that $\int_\Omega |G(x,\omega)|^2 dP(\omega) < (\frac{\eta}{2})^2$, $\|G(x,\cdot)\|_\infty \le 1$, and

$$P\Big\{\omega : \max_{1 \le k \le L} \tilde{T}_{n_k} G(x,\omega) > (1-\epsilon)\frac{\sqrt{1-\tilde{\rho}}}{1+\sqrt{\tilde{\rho}}}\Big\} > 1 - \epsilon_1.$$

For $x \notin E_J$ let $G(x,\omega) = 0$. An examination of the proof of Proposition 3.3 shows that $G(x,\omega)$ is obtained from $F(x,\omega)$ by multiplication by a constant, truncation by a function which is measurable on the product σ-field, and multiplication by χ_{E_J}, a function measurable on the product σ-field. Thus we see that $G(x,\omega)$ is measurable on the product σ-field.

We have

$$m \otimes P\Big\{(x,\omega) : \max_{1 \le k \le L} \tilde{T}_{n_k} G(x,\omega) > (1-\epsilon)\frac{\sqrt{1-\tilde{\rho}}}{1+\sqrt{\tilde{\rho}}}\Big\} > (1-\epsilon_1) m(E_J).$$

We now consider $G(x,\omega)$ as a function of x for ω fixed. Let

$$\Omega_1 = \{\omega : \int_X |G(x,\omega)|^2 dx \le \eta^2\}.$$

Clearly $P(\Omega_1) \ge \frac{3}{4}$ since we have

$$\int_\Omega \int_X |G(x,\omega)|^2 dx dP(\omega) = \int_X \int_\Omega |G(x,\omega)|^2 dP(\omega) dx < (\frac{\eta}{2})^2$$

and if $P(\Omega_1) < \frac{3}{4}$ then we would have

$$\int_\Omega \int_X |G(x,\omega)|^2 dx dP(\omega) \ge \frac{1}{4}\eta^2,$$

a contradiction.

Let

$$\Omega_2 = \Big\{\omega : m\{x : \max_{1 \le k \le L} \tilde{T}_{n_k} G(x,\omega) > (1-\epsilon)\frac{\sqrt{1-\tilde{\rho}}}{1+\sqrt{\tilde{\rho}}}\} > 1 - \epsilon'\Big\}.$$

We see that $P(\Omega_2) \ge \frac{3}{4}$ since if not, we would have

$$m \otimes P\Big\{(x,\omega) : \max_{1 \le k \le L} \tilde{T}_{n_k} G(x,\omega) > (1-\epsilon)\frac{\sqrt{1-\tilde{\rho}}}{1+\sqrt{\tilde{\rho}}}\Big\} < \frac{3}{4} + \frac{1}{4}(1-\epsilon').$$

Since we have

$$m \otimes P\Big\{(x,\omega) : \max_{1\le k\le L} \tilde{T}_{n_k} G(x,\omega) > (1-\epsilon)\frac{\sqrt{1-\tilde{\rho}}}{1+\sqrt{\tilde{\rho}}}\Big\} > (1-\epsilon_1)m(E_J),$$

we see we will have a contradiction if

$$(1-\epsilon_1)m(E_J) > 1 - \frac{\epsilon'}{4}.$$

By taking J large enough, and ϵ_1 small enough, this can be achieved.

Since $P(\Omega_1) \ge \frac{3}{4}$ and $P(\Omega_2) \ge \frac{3}{4}$, we clearly can find $\omega_0 \in \Omega_1 \cap \Omega_2$. The required function is $G(x,\omega_0)$. For that ω_0, because of our definition of $\tilde{T}_{n_j}$, we also have

$$m\Big\{x : \max_{1\le k\le L} T_{n_k} G(x,\omega_0) > \gamma(1-\epsilon)\frac{\sqrt{1-\tilde{\rho}}}{1+\sqrt{\tilde{\rho}}}\Big\} > 1-\epsilon'.$$

By taking α close enough to ρ and γ close enough to 1, (which we can do by taking J large enough), we can make $\tilde{\rho}$ as close to ρ as desired. In particular we can have

$$\gamma(1-\epsilon)\frac{\sqrt{1-\tilde{\rho}}}{1+\sqrt{\tilde{\rho}}} > (1-2\epsilon)\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}.$$

Thus we can have

$$m\Big\{x : \max_{1\le k\le L} T_{n_k} G(x,\omega_0) > (1-2\epsilon)\frac{\sqrt{1-\rho}}{1+\sqrt{\rho}}\Big\} > 1-\epsilon'.$$

We now just replace ϵ by $\epsilon/2$ to achieve the desired conclusion. □

4. Convolution Powers

Let μ be a probability measure on $\mathbb{Z}$. Define

$$\mu(f)(x) = \sum_{k=-\infty}^{\infty} \mu(k) f(\tau^k x).$$

and for $n > 1$,

$$\mu^n(f)(x) = \mu(\mu^{n-1}(f))(x).$$

It is quite clear that the operator μ is a positive contraction on L^1 and L^∞, and hence is a Dunford Schwartz operator. Consequently we know the averages $\displaystyle\frac{1}{n}\sum_{k=0}^{n-1} \mu^k(f)(x)$ converges a.e. for all $f \in L^p$, $1 \le p \le \infty$. However, much more is true. If the measure μ is symmetric, that is, $\mu(k) = \mu(-k)$ for all k, then the operator μ is self-adjoint. For self-adjoint operators, Burkholder and Chow [24], with a very nice argument involving a square function, proved the following theorem.

Theorem 4.1 (Burkholder and Chow, [24]). *Let T be a linear self-adjoint operator in $L^2(X)$ such that for each $f \in L^2(X)$ we have $\int_X |Tf(x)|dx \le \int_X |f(x)|dx$ and $\|Tf\|_\infty \le \|f\|_\infty$. Then for each $f \in L^2$, $\lim_{n\to\infty} T^{2n} f(x)$ exists a.e.*

Applying the result to Tf, we also have convergence for $T^{2n+1}f$. For symmetric measures, μ, the associated operator μ satisfies the hypothesis of Theorem 4.1, and an easy square function argument shows that μ^{2n} and μ^{2n+1} are close, so either both converge or both diverge a.e.. Thus we have the following theorem.

Theorem 4.2. *If μ is a symmetric measure on $\mathbb{Z}$, then $\mu^n(f)(x)$ converges a.e. for all $f \in L^2$.*

Later Stein [71], using his complex interpolation method, was able to extend Theorem 4.1 to obtain convergence of $\mu^n f$ for all $f \in L^p$, $1 < p \leq \infty$.

Working with A. Bellow and J. Rosenblatt [14], we were able to extend Stein's Theorem to measures that were not necessarily symmetric. This generalization depends on the following definition.

Definition 4.3. A probability measure μ on $\mathbb{Z}$ has *bounded angular ratio* if $|\hat{\mu}(\gamma)| = 1$ only for $\gamma = 1$, and

$$\sup_{|\gamma|=1} \frac{|\hat{\mu}(\gamma) - 1|}{1 - |\hat{\mu}(\gamma)|} < \infty.$$

We obtained the following theorem.

Theorem 4.4 ([15])**.** *Let μ have bounded angular ratio.*

(1) *For $f \in L^p$, $1 < p \leq \infty$, $\mu^n f(x)$ converges a.e..*

(2) *For $1 < p \leq \infty$ we have*

$$\| \sup_n |\mu^n f| \|_p \leq c(p) \|f\|_p.$$

Proof. Below is a sketch of the proof of this result in the case $p = 2$. The general result follows in a similar way, using Stein's complex interpolation method to reach the values of $p < 2$. (For values of $p > 2$, the result follows from the case $p = 2$, the trivial estimate $\| \sup_n |\mu^n f| \|_\infty \leq \|f\|_\infty$, and an application of the Marcinikewicz interpolation theorem.)

We have (for $N > 2M$)

$$\begin{aligned}
&\frac{1}{N-M} \sum_{k=M}^{N} \mu^k f(x) \\
&\quad = \frac{1}{N-M} \sum_{k=M}^{N} \Big(\sum_{j=k}^{N} \big(\mu^j f(x) - \mu^{j+1} f(x)\big) + \mu^{N+1} f(x) \Big) \\
&\quad = \frac{1}{N-M} \sum_{j=M}^{N} \Big(\mu^j f(x) - \mu^{j+1} f(x) \Big) \sum_{k=M}^{j} 1 + \mu^{N+1} f(x) \\
&\quad = \frac{1}{N-M} \sum_{j=M}^{N} (j - M + 1) \Big(\mu^j f(x) - \mu^{j+1} f(x) \Big) + \mu^{N+1} f(x).
\end{aligned}$$

Consequently for $N > 2M$, we have

$$\begin{aligned}
\Big| \frac{1}{N-M} \sum_{k=M}^{N} \mu^k f(x) - \mu^{N+1} f(x) \Big| &\leq \frac{1}{N-M} \sum_{j=M}^{N} j \Big| \mu^j f(x) - \mu^{j+1} f(x) \Big| \\
&\leq \frac{1}{N-M} \Big(\sum_{j=M}^{N} (\sqrt{j})^2 \Big)^{\frac{1}{2}} \Big(\sum_{j=M}^{N} (\sqrt{j})^2 \Big| \mu^j f(x) - \mu^{j+1} f(x) \Big|^2 \Big)^{\frac{1}{2}} \\
&\leq c \Big(\sum_{j=M}^{\infty} j \Big| \mu^j f(x) - \mu^{j+1} f(x) \Big|^2 \Big)^{\frac{1}{2}}.
\end{aligned}$$

Since the operator μ is a positive contraction on L^1 and L^∞, by the Dunford–Schwartz ergodic theorem, we know $\frac{1}{n}\sum_{k=0}^{n-1}\mu^k f(x)$ converges a.e.

Clearly for each fixed M, $\frac{1}{N-M}\sum_{k=M}^{N}\mu^k(f)(x)$ converges a.e. If we could show that the right hand side goes to zero as M goes to ∞, so would the left hand side. Consequently, for M large enough, and $N>2M$, we have that $\frac{1}{N-M}\sum_{k=M}^{N}\mu^k(f)(x)$ and $\mu^{N+1}f(x)$ are close. This implies $\mu^{N+1}f(x)$ must converge a.e. too.

The argument to show that $\Big(\sum_{j=M}^{\infty} j\Big|\mu^j f(x)-\mu^{j+1}f(x)\Big|^2\Big)^{\frac{1}{2}}$ converges to zero involves first transferring the problem to $\mathbb{Z}$. (See Bellow's paper [8].) A Fourier transform argument is then used to show that the multiplier is bounded, with $M=1$, and consequently, goes to zero as M goes to ∞. The details of the Fourier transform argument are as follows.

$$
\begin{aligned}
\Big\|\Big(\sum_{j=1}^{\infty} j\Big|\mu^j f(x)-\mu^{j+1}f(x)\Big|^2\Big)^{\frac{1}{2}}\Big\|_2^2 &= \int_{|\gamma|=1}\sum_{j=1}^{\infty} j\Big|\widehat{\mu^j f}(\gamma)-\widehat{\mu^{j+1}f}(\gamma)\Big|^2\,d\gamma\\
&= \int_{|\gamma|=1}\sum_{j=1}^{\infty} j\Big|\big(\hat\mu(\gamma)\big)^j\hat f(\gamma)-\big(\hat\mu(\gamma)\big)^{j+1}\hat f(\gamma)\Big|^2\,d\gamma\\
&= \int_{|\gamma|=1}\sum_{j=1}^{\infty} j|\hat\mu(\gamma)|^{2j}|1-\hat\mu(\gamma)|^2|\hat f(\gamma)|^2\,d\gamma\\
&\le \int_{|\gamma|=1}\Big(\frac{1}{1-|\hat\mu(\gamma)|^2}\Big)^2|1-\hat\mu(\gamma)|^2|\hat f(\gamma)|^2\,d\gamma\\
&\le \sup_{|\gamma|=1}\Big(\frac{|1-\hat\mu(\gamma)|}{|1-|\hat\mu(\gamma)||1+|\hat\mu(\gamma)|}\Big)^2\int_{|\gamma|=1}|\hat f(\gamma)|^2\,d\gamma\\
&\le c\sup_{|\gamma|=1}\Big(\frac{|1-\hat\mu(\gamma)|}{|1-|\hat\mu(\gamma)|}\Big)^2\|f\|_2^2.
\end{aligned}
$$

Hence, with the assumption that μ has bounded angular ratio, the first part of the theorem follows. As we see, the bounded angular ratio condition arose naturally in the computation.

The maximal inequality follows from the fact that

$$|\mu^{N+1}f(x)|\le|\frac{1}{N}\sum_{k=1}^{N}\mu^k f(x)|+c\Big(\sum_{k=1}^{\infty} j|\mu^j f(x)-\mu^{j+1}f(x)|^2\Big)^{\frac{1}{2}}.$$

Now take the supremum over N on both sides. Note that

$$\|\sup_N|\frac{1}{N}\sum_{k=1}^{N}\mu^k f(x)|\|p\le c_p\|f\|_p,$$

since μ is a contraction on all L^p, and we just saw that at least for $p=2$, the square function is a bounded operator. The general case follows by Stein's complex interpolation.

$\Box$

Of course, the theorem is only useful if there are measures that satisfy the bounded angular ratio condition. It turns out that there are many examples of such measures. The following theorem from [13] shows how to find a large class of examples.

Theorem 4.5. *If*

$$\sum_{k=-\infty}^{\infty} k\mu(k) = 0$$

and

$$\sum_{k=-\infty}^{\infty} k^2\mu(k) < \infty$$

then μ has bounded angular ratio, and $\mu^n f(x)$ converges a.e. for $f \in L^p$, $1 < p < \infty$.

Remark 4.6. If μ has finite first moment (in particular, if μ has finite second moment) and $E(\mu) = \sum_{k=-\infty}^{\infty} k\mu(k) \neq 0$ then μ has unbounded angular ratio. Hence if $E(\mu) = 0$ then the statement μ has finite second moment, and the statement μ has bounded angular ratio are equivalent.

It turns out that the bounded angular ratio condition is not only sufficient for a.e. convergence of the convolution powers, but it also is (almost) necessary. We have the following result. (See [1], [3] and [15].)

Theorem 4.7. *If $\lim_{\gamma\to 1} \dfrac{|\hat{\mu}(\gamma) - 1|}{1 - |\hat{\mu}(\gamma)|} = \infty$ then (μ^n) has the strong sweeping out property*

It is possible that $\lim_{\gamma\to 1} \dfrac{|\hat{\mu}(\gamma) - 1|}{1 - |\hat{\mu}(\gamma)|}$ fails to exist. In that case, there are situations where we are still uncertain if divergence occurs.

Remark 4.8. In [13] Bourgain's entropy method was used to show that divergence occurs. Later in [1] a modification of the Fourier transform condition for strong sweeping out was used to obtain the stronger result. In [3] a modification of the entropy method was used to also obtain strong sweeping out.

In the above discussion we have only considered the case $p > 1$. The case $p = 1$ is much more difficult. The first non-trivial result in this direction was obtained by Karin Reinhold [64].

Theorem 4.9 (Reinhold [64])**.** *If μ has bounded angular ratio and*

$$\sum_{k=-\infty}^{\infty} |k|^{2+\delta}\mu(k) < \infty$$

for some $\delta > \frac{(\sqrt{17}-3)}{2}$, then $\mu^n f(x)$ converges a.e. for all $f \in L^1$.

(Recall that for μ as above, since μ has bounded angular ratio, we are restricted to μ such that $E(\mu) = 0$.)

The argument used by Reinhold involved a comparison with the appropriate normal density, and used the fact that the maximal function associated with convolution of dilates of the normal density function will satisfy a weak (1,1) inequality. However, it seemed unlikely that the term $\frac{(\sqrt{17}-3)}{2}$ could be sharp.

Encouraged by Reinhold's result, the problem was later studied by A. Bellow and A. P. Calderón. They knew about the Calderón transfer principle, and they realized that in some ways, these convolution powers behaved like singular integrals. Moreover they knew how to use the Calderón-Zygmund decomposition to obtain results about singular integrals. With these tools, they improved Reinhold's result to obtain the following.

Theorem 4.10 (Bellow–Calderón [11]). *If μ has bounded angular ratio and*

$$\sum_{k=-\infty}^{\infty} |k|^2 \mu(k) < \infty,$$

then:

- *For each $\lambda > 0$ we have*

$$m\{x : \sup_n |\mu^n f(x) > \lambda\} \leq \frac{c}{\lambda}\|f\|_1;$$

- *For all $f \in L^1(X)$, $\mu^n f(x)$ converges a.e..*

A sketch of their proof is given below. They first transfer the problem to $\mathbb{Z}$, and write $\mu^n(x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{\mu}(e^{2\pi i\theta})^n e^{-2\pi i x\theta}\, d\theta$. They then use a clever integration by parts argument to obtain the following Lemma.

Lemma 4.11 (Bellow–Calderón). *If μ has bounded angular ratio and*

$$\sum_{k=-\infty}^{\infty} |k|^2 \mu(k) < \infty,$$

then

$$\sup_n |\mu^n(x-y) - \mu^n(x)| \leq \frac{c|y|}{|x|^2}.$$

This lemma gave them exactly the same control of the "smoothness" of the convolution powers that one has for the Hilbert transform. Once Bellow and Calderón established this control, they knew the proof could be finished in exactly the same way that one completes the proof of the weak type (1,1) inequality for the Hilbert transform. Convergence on a dense class (all of L^2) was known. Consequently, by the Banach principle, to establish convergence for all $f \in L^1$, it was enough to prove this weak type (1,1) inequality. To prove such inequalities, with either the convolution powers or the Hilbert transform, one begins with the Calderón-Zygmund decomposition. The version of this important decomposition stated below is for $\mathbb{Z}$, however, with nothing more than a change in notation, the same decomposition holds on $\mathbb{R}$. It is the key to prove the weak type (1,1) result for the Hilbert transform, and as Bellow and Calderón showed, for convolution powers as well.

Theorem 4.12 (Calderón-Zygmund decomposition). *Given a nonnegative function $f \in \ell^1(\mathbb{Z})$ and $\lambda > 0$, we can write $f = g + b$, where $g \in \ell^2(\mathbb{Z})$ and*

(1) $\|g\|_{\ell^1} \leq \|f\|_{\ell^1}$,
(2) $\|g\|_\infty \leq 2\lambda$,
(3) $b = \sum_i b_i$ *where each b_i satisfies*:
  a) *each b_i is supported on a dyadic interval B_i*,

b) $\sum_j b_i(j) = 0$ *for each* i,
c) $\frac{1}{|B_i|}\sum_{j\in B_i} |b_i(j)| \le 4\lambda$ *and* $\lambda \le \frac{1}{|B_i|}\sum_{j\in B_i} |f(j)|$,
d) *For each* $i \neq j$ *we have* $B_i \cap B_j = \emptyset$.

REMARK 4.13.

- We are using $|S|$ to denote the number of integers in the set S.
- The conditions above imply

$$\sum_i |B_i| \le \frac{1}{\lambda}\sum_i \|b_i\|_{\ell^1} \le \frac{1}{\lambda}\|f\|_{\ell^1}.$$

- In the above decomposition, if $\lambda > \|f\|_\infty$ then we can take $f = g$ and $b = 0$.

Proof of 4.12. The proof is not difficult. First find a dyadic interval, I, so large that $\frac{1}{|I|}\sum_{j\in I} |f(j)| \le \lambda$, and $|f(j)| \le \lambda$ for $j \notin I$. Now divide I into two equal dyadic subintervals, I_1 and I_2. Look at the average over each piece. If in either case the averages is more than λ, keep that interval and it will become one of the B_i's. If the average is less than λ, divide that interval into two equal dyadic subintervals, and repeat the procedure. The process clearly stops after only a finite number of steps. We then have a collection of disjoint dyadic intervals. The average over any one of the selected intervals is dominated by 2λ. The value of f off the union of these selected intervals is at most λ. Denote the selected intervals by $B_1, B_2, \ldots$. On B_i define $g = \frac{1}{|B_i|}\sum_{j\in B_i} f(j)$, and off $\cup B_i$ define $g(j) = f(j)$. Define $b(j) = f(j) - g(j)$, and let $b_i = b\chi_{B_i}$. It is easy to check that all the desired properties are satisfied. $\square$

REMARK 4.14. There is also an analog of the Calderón–Zygmund decomposition for martingales. This was discovered by Gundy, [38], and has been useful for proving several results about martingales.

Proof of 4.10. The following proof is due to Bellow and Calderón. We first use the strong type (2,2) inequality to maintain control of g. We then use the smoothness of the operator to control b when we are far from the support of b, and use the fact that b has small support to control what happens near the support of b. The details are as follows.

Note that

$$|\{\sup_n |\mu^n f(x)| > 2\lambda\}| \le |\{\sup_n |\mu^n g(x)| > \lambda\}| + |\{\sup_n |\mu^n b(x)| > \lambda\}|.$$

To handle the first term, recall that by Theorem 4.4 we have

$$\|\sup_n |\mu^n g\|_{\ell^2} \le c\|g\|_{\ell^2}.$$

Consequently,

$$\begin{aligned}
|\{\sup_n |\mu^n g(x)| > \lambda\}| &\le \frac{1}{\lambda^2} \| \sup_n |\mu^n g| \|_{\ell^2}^2 \\
&\le \frac{c}{\lambda^2} \|g\|_{\ell^2}^2 \\
&\le \frac{c}{\lambda^2} \sum_j |g(j)|^2 \\
&\le \frac{c}{\lambda^2} \sum_j |g(j)|\lambda \\
&\le \frac{c}{\lambda} \|g\|_{\ell^1} \\
&\le \frac{c}{\lambda} \|f\|_{\ell^1}.
\end{aligned}$$

To handle the second term, we first write $\tilde{B}_i$ for the interval with the same center as B_i, but with five times the length. We also let $\tilde{B} = \cup_i \tilde{B}_i$. Let $\tilde{B}^c$ denote the complement of $\tilde{B}$. We then have

$$|\{\sup_n |\mu^n b(x)| > \lambda\}| \le |\{\sup_n |\mu^n b(x)| > \lambda\} \cap \tilde{B}^c| + |\tilde{B}|.$$

This time the second term is easy. We have

$$|\tilde{B}| \le 5 \sum_i |B_i| \le \frac{5}{\lambda} \|f\|_{\ell^1}.$$

For the remaining term, we need some additional notation. Let r_i denote a point near the center of the interval B_i. Recall that $\mu^n b(x) = \sum_r \mu^n(x-r)b(r)$, and that $\sum_r b(r) = 0$. Consequently,

$$\sum_{r \in B_i} \mu^n(x - r_i) b(r) = 0.$$

For $x \in \tilde{B}_i^c$, we know $|x - r_i| > 2|B|$ and for $r \in B_i$, we know $|r - r_i| \le |B|$. Hence

$$\sum_{x \in \tilde{B}_i^c} \frac{|r - r_i|}{|x - r_i|^2} \le 2 \sum_{k=|B_i|}^{\infty} \frac{|B|}{k^2} \le c.$$

We have

$$
\begin{aligned}
|\{\sup_n |\mu^n b(x)| > \lambda\} \cap \tilde{B}^c| &\le \frac{1}{\lambda} \sum_{x \in \tilde{B}^c} \sup_n |\mu^n b(x)| \\
&\le \frac{1}{\lambda} \sum_i \sum_{x \in \tilde{B}_i^c} \sup_n |\mu^n b(x)| \\
&\le \frac{1}{\lambda} \sum_i \sum_{x \in \tilde{B}_i^c} \sup_n \Big| \sum_r \mu^n(x-r) b(r) - \sum_r \mu^n(x-r_i) b(r) \Big| \\
&\le \frac{1}{\lambda} \sum_i \sum_{x \in \tilde{B}_i^c} \sum_r \sup_n \Big| \mu^n(x-r) - \mu^n(x-r_i) \Big| |b(r)| \\
&\le \frac{1}{\lambda} \sum_i \sum_r |b(r)| \sum_{x \in \tilde{B}_i^c} c \frac{|r-r_i|}{|x-r_i|^2} \\
&\le \sum_i \frac{c}{\lambda} \sum_r |b(r)| \\
&\le c \sum_i |B_i| \\
&\le \frac{c}{\lambda} \|f\|_{\ell^1}.
\end{aligned}
$$

□

5. Good-λ Inequalities for Convolution Powers

The above discussion suggested that convolution powers exhibit a behavior similar to that of a singular integral. In this section we give an additional result in the same spirit. That is, the result is proved in the same way as a related result for singular integrals. Let μ be a probability measure on $\mathbb{Z}$. As before, let

$$\mu f(x) = \sum_{k=-\infty}^{\infty} \mu(k) f(\tau^k x),$$

and for $n > 1$, $\mu^n f(x) = \mu(\mu^{n-1} f)(x)$. Let $\mu^\star f(x) = \sup_{n \ge 0} |\mu^n f(x)|$ and

$$f^\star(x) = \sup_{a \le 0 \le b} \frac{1}{b-a+1} \sum_{j=a}^{b} |f(\tau^j x)|.$$

Theorem 5.1. *Assume μ is a probability measure on $\mathbb{Z}$ which has finite second moment and $\sum_k k\mu(k) = 0$. There are constants $C > 0$ and $C' > 0$ (which depend only on μ) such that if $\beta > 1$ and $0 < \gamma \le 1$ satisfy $\beta - 1 - \gamma C' > 0$ then*

$$m\{x : \mu^\star f(x) > \beta\lambda, f^\star(x) < \gamma\lambda\} \le \frac{C\gamma}{\beta - 1 - \gamma C'} m\{x : \mu^\star f(x) > \lambda\}.$$

Remark 5.2. Joint distribution function inequalities such as given in the above theorem are often called "good–lambda" inequalities. They were introduced by Burkholder and Gundy, and have played an important role both in probability and in harmonic analysis. For example, see Coifman's article [28], where he proves a similar inequality between certain singular integrals and the Hardy-Littlewood maximal function.

Proof. First we will prove the result on $\mathbb{Z}$ and then transfer the result to the ergodic theory setting in the standard way. We will use the same notation for the operators on $\mathbb{Z}$ with the transformation being translation.

Let $\phi \in \ell^1$. We can assume without loss of generality that ϕ is non-negative. Let $S = \{j : \mu^\star\phi(j) > \lambda\}$. Then we can write $S = \cup I_i$, where the I_i are disjoint intervals of integers, and the intervals are of maximal length. That is, if $I_i = \{p_i+1,\dots,p_i+d_i\}$, then $\mu^\star\phi(p_i+j) > \lambda$ for $j = 1,2,\dots,d_i$, $\mu^\star\phi(p_i) \le \lambda$ and $\mu^\star\phi(p_i+d_i+1) \le \lambda$.

Let $I = \{p+1,\dots,p+d\}$ be one of these intervals. It will be enough to prove

$$|\{j \in I : \mu^\star\phi(j) > \beta\lambda, \phi^\star(j) < \gamma\lambda\}| \le \frac{C\gamma}{\beta - 1 - \gamma C'}|I|. \tag{2}$$

We can assume that there is a point $p_0 \in I$ so that $\phi^\star(p_0) < \gamma\lambda$ since otherwise the left side of (2) is zero and the result is obvious.

Let $\tilde{I} = \{p-d,\dots,p+2d\}$. Then $I \subset \tilde{I}$, and I is at least a distance d from the boundary of $\tilde{I}$. Let $\phi = \phi_1 + \phi_2$ where $\phi_1(j) = \phi(j)\chi_{\tilde{I}}(j)$.

Although we only need to estimate $\mu^\star\phi(j)$ for $j \in I \cap \{k : \phi^\star(k) < \gamma\lambda\}$, we will actually show first that there is a constant C' such that we have the uniform estimate $\mu^\star\phi_2(j) < (1+C'\gamma)\lambda$ for all $j \in I$.

For fixed n, we have for $j \in I$,

$$\begin{aligned}
\mu^n\phi_2(j) &= \sum_{k=-\infty}^{\infty} \mu^n(j-k)\phi_2(k) \\
&= \sum_{k=-\infty}^{\infty} \mu^n(p-k)\phi_2(k) + \sum_{k=-\infty}^{\infty} \Big(\mu^n(j-k) - \mu^n(p-k)\Big)\phi_2(k) \\
&= A + B.
\end{aligned}$$

We have

$$A \le \sum_{k=-\infty}^{\infty} \mu^n(p-k)\phi(k) = \mu^n\phi(p) \le \lambda$$

since $\phi_2 \le \phi$ and $\mu^\star\phi(p) \le \lambda$ by assumption.

For B we need to work harder. We can write

$$\begin{aligned}
B &= \sum_{k=-\infty}^{\infty} \Big(\mu^n(j-k) - \mu^n(p_0-k)\Big)\phi_2(k) \\
&\qquad + \sum_{k=-\infty}^{\infty} \Big(\mu^n(p_0-k) - \mu^n(p-k)\Big)\phi_2(k) \\
&= B_1 + B_2.
\end{aligned}$$

Recalling that ϕ_2 is supported outside $\tilde{I}$, that $\phi^\star(p_0) \le \gamma\lambda$, and the smoothness condition on μ^n, from Lemma 4.11, that is

$$|\mu^n(j-k) - \mu^n(p_0-k)| \le c\frac{|p_0-j|}{|p_0-k|^2},$$

which was established in [11], we now estimate

$$\begin{aligned}
B_1 &= \sum_{k=-\infty}^{\infty} \Big(\mu^n(j-k) - \mu^n(p_0-k)\Big)\phi_2(k) \\
&\le \sum_{|p_0-k|>d}^{\infty} \Big(\mu^n(j-k) - \mu^n(p_0-k)\Big)\phi_2(k) \\
&\le \sum_{|p_0-k|>d}^{\infty} c\frac{|j-p_0|}{|k-p_0|^2}\phi(k) \\
&\le cd\sum_{k=d}^{\infty} \frac{1}{k^2}\phi(p_0-k) + cd\sum_{k=d}^{\infty}\frac{1}{k^2}\phi(p_0+k) \\
&\le cd\sum_{k=d}^{\infty}\phi(p_0-k)\sum_{j=k}^{\infty}\Big(\frac{1}{j^2} - \frac{1}{(j+1)^2}\Big) \\
&\qquad + cd\sum_{k=d}^{\infty}\phi(p_0+k)\sum_{j=k}^{\infty}\Big(\frac{1}{j^2} - \frac{1}{(j+1)^2}\Big) \\
&\le \sum_{j=d}^{\infty}\frac{1}{j^3}\sum_{k=0}^{j}\phi(p_0-k) + cd\sum_{j=d}^{\infty}\frac{1}{j_3}\sum_{k=0}^{j}\phi(p_0+k) \\
&\le 2cd\phi^\star(p_0)\sum_{j=d}^{\infty}\frac{1}{j^2} \\
&\le 2c\gamma\lambda.
\end{aligned}$$

The term B_2 is estimated in exactly the same way, and we also obtain $B_2 \le 2c\gamma\lambda$. Hence if we let $C' = 4c$, then we have $\mu^\star\phi_2(j) \le (1 + C'\gamma)\lambda$ for all $j \in I$.

We now return to estimating the left side of (2). Here we will use Theorem 4.10, the fact established by Bellow and Calderón [11], that $\mu^\star$ is a weak type (1,1) operator. We have

$$\begin{aligned}
&|\{j \in I : \mu^\star\phi(j) > \beta\lambda, \phi^\star(j) < \gamma\lambda\}| \\
&\qquad \le |\{j \in I : \mu^\star\phi_1(j) + \mu^\star\phi_2(j) > (\beta - 1 - \gamma C')\lambda + (1+\gamma C')\lambda\}| \\
&\qquad \le |\{j \in I : \mu^\star\phi_1(j) > (\beta - 1 - \gamma C')\lambda\}| \\
&\qquad \le \frac{C}{(\beta - 1 - \gamma C')\lambda}\|\phi_1\|_1 \\
&\qquad \le \frac{C}{(\beta - 1 - \gamma C')\lambda}|I|\frac{1}{|I|}\sum_{k\in I}|\phi_(k)| \\
&\qquad \le \frac{C}{(\beta - 1 - \gamma C')\lambda}|I|\phi^\star(p_0) \\
&\qquad \le \frac{C}{(\beta - 1 - \gamma C')\lambda}|I|\gamma\lambda \\
&\qquad \le \frac{C\gamma}{\beta - 1 - \gamma C'}|I|
\end{aligned}$$

as required. □

This implies the following corollary.

Corollary 5.3. *Assume μ is a probability measure on $\mathbb{Z}$ which has finite second moment and $\sum_k k\mu(k) = 0$. Let $\Phi(t) \geq 0$ be increasing and such that $\Phi(2t) \leq c\Phi(t)$ for $t \geq 0$. Then there is a constant $\tilde{c}$ such that*

$$\int_X \Phi\big(\mu^\star f(x)\big)dx \leq \tilde{c}\int_X \Phi\big(f^\star(x)\big)dx.$$

In particular, taking $\Phi(t) = t^p$, we have

$$\int_X |\mu^\star f(x)|^p \leq c_p \int_X |f^\star(x)|^p dx$$

for $0 < p < \infty$.

Remark 5.4. Note that the range of p is $0 < p < \infty$, we do not have the usual restriction that $p > 1$.

Proof. Take $\beta = 2$ and $\gamma = \frac{1}{2^N}$ for some large enough integer N. We have

$$\begin{aligned}
\int_X \Phi(\frac{1}{2}\mu^\star f(x))dx &\leq \int_0^\infty m\{x : \mu^\star f(x) > 2\lambda\}d\Phi(\lambda)\\
&\leq \int_0^\infty m\{x : \mu^\star f(x) > 2\lambda, f^\star(x) < \frac{1}{2^N}\lambda\}d\Phi(\lambda)\\
&\quad + \int_0^\infty m\{x : \mu^\star f(x) > 2\lambda, f^\star(x) > \frac{1}{2^N}\lambda\}d\Phi(\lambda)\\
&\leq \int_0^\infty \frac{C\frac{1}{2^N}}{2 - 1 - \frac{C'}{2^N}} m\{x : \mu^\star f(x) > \frac{1}{2^N}\lambda\}d\Phi(\lambda)\\
&\quad + \int_0^\infty m\{x : f^\star(x) > \frac{1}{2^N}\lambda\}d\Phi(\lambda)\\
&\leq \frac{C\frac{1}{2^N}}{1 - \frac{C'}{2^N}} \int_X \Phi\big(\mu^\star f(x)\big)dx\\
&\quad + \int_X \Phi\big(2^N f^\star(x)\big)dx.
\end{aligned}$$

Hence we have

$$\int_X \Phi(\frac{1}{2}\mu^\star f(x))dx \leq \frac{C\frac{1}{2^n}}{1 - \frac{C'}{2^N}} \int_X \Phi\big(\mu^\star f(x)\big)dx + \int_X \Phi\big(2^N f^\star(x)\big)dx.$$

Using the properties of Φ we see that this implies

$$\int_X \Phi\big(\frac{1}{2}\mu^\star f(x)\big)dx \leq \frac{C\frac{1}{2^N}}{1 - \frac{C'}{2^N}}c \int_X \Phi\big(\frac{1}{2}\mu^\star f(x)\big)dx + c^{N+1}\int_X \Phi\big(\frac{1}{2}f^\star(x)\big)dx.$$

Now we solve, and obtain

$$\Big(1 - c\frac{C\frac{1}{2^N}}{1 - \frac{C'}{2^N}}\Big)\int_X \Phi(\frac{1}{2}\mu^\star f(x))dx \leq c^{N+1}\int_X \Phi\big(\frac{1}{2}f^\star(x)\big)dx.$$

Replacing f by $2f$, and taking N large enough so that $C' < 2^N$, we have the conclusion of the theorem.

(Actually there is an additional step. We first establish the theorem with $\mu^\star$ replaced by $\sup_{0\le k\le K}|\mu^k|$. For this operator we know that the term $\int_X \Phi\big(\frac{1}{2}\mu^\star f(x)\big)\,dx$ on the right hand side will be finite, so we can solve. Then we note that the inequality does not depend on K, so we have the result for the maximal operator, $\mu^\star$.)

$\square$

6. Oscillation Inequalities

To understand some of the discussion in this section, it helps to have a little background about martingales. We will not need the full theory, and consequently, will restrict our attention to two special cases: the dyadic martingale on $[0,1)$, and the dyadic reverse martingale on $\mathbb{Z}$.

For $f \in L^1([0,1))$ and $n = 0, 1, \dots$, define

$$f_n(x) = 2^n \int_{\frac{k-1}{2^n}}^{\frac{k}{2^n}} f(t)\,dt \text{ for } \frac{k}{2^n} \le x < \frac{k+1}{2^n},$$

with $k = 1, 2, \dots, 2^n$.

Some important operators include the martingale maximal function

$$f^\star(x) = \sup_n |f_n(x)|,$$

and the martingale square function

$$S_M f(x) = \Big(\sum_{k=0}^{\infty} |f_k(x) - f_{k+1}(x)|^2\Big)^{\frac{1}{2}}.$$

Doob, in his classical 1953 book, [33], proved that $\|f^\star\|_p \le c_p\|f\|_p$ and also proved the associated weak type (1,1) inequality, that is, $m\{x : |f^\star(x)| > \lambda\} \le \frac{c}{\lambda}\|f\|_1$. (See Doob's book, [33] for further discussion, as well as many other important facts about martingales.) Austin [7] first proved the weak type (1,1) inequality for $S_M f$ in 1966. As in the case of the maximal function, we have $\|S_M f\|_p \le c_p\|f\|_p$ for $1 < p < \infty$. It was Burkholder who first realized the importance of the martingale square function, and later working with Gundy and Silverstein, they were able to use the martingale square function to give the first real variable characterization of the Hardy space H^1 [26]. (See Petersen's book [61], "Brownian Motion, Hardy Spaces and Bounded Mean Oscillation", for an exposition.)

We can also consider two additional operators. First, if we fix an increasing sequence (n_k) we can define the oscillation operator,

$$O_M f(x) = \Big(\sum_{k=1}^{\infty} \sup_{n_k < n \le n_{k+1}} |f_{n_k}(x) - f_n(x)|^2\Big)^{\frac{1}{2}}.$$

For any $\varrho > 2$, we define the ϱ-variational operator for martingales by

$$V_{M,\varrho} f(x) = \sup_{n_1 < n_2 < \dots} \Big(\sum_{k=1}^{\infty} |f_{n_k}(x) - f_{n_{k+1}}(x)|^\varrho\Big)^{\frac{1}{\varrho}},$$

where the supremum is taken over all increasing sequences of positive integers.

It is easy to see that $\|O_M f\|_2 \le c\|f\|_2$. Just note that

$$\int \sup_{n_k<n\le n_{k+1}} |f_{n_k}(x) - f_n(x)|^2 dx \le c \int |f_{n_k}(x) - f_{n_{k+1}}(x)|^2 dx,$$

which follows from the maximal inequality applied to the martingale that starts at n_k. The strong (p,p) result for $V_{M,\varrho} f$ is a result of Lepingle [55]. See also [60] and [43] for further discussion of the operator $V_{M,\varrho}$. (If $\varrho \le 2$ the operator can diverge. That is why we restrict our attention to $\varrho > 2$. See Monroe's paper [58]. If we use an exponent $\varrho < 2$, then even the analog of the simple square function can diverge. See [5].)

It is also useful to note another square function that has played an important role in analysis. That is, the Littlewood Paley g-function. Let $P_t(x) = \frac{2}{\pi}\frac{t}{|x|^2+t^2}$ denote the Poisson kernel on $\mathbb{R}$. Define $u(x,t) = f \star P_t(x)$, and let

$$gf(x) = \Big(\int_0^\infty t|\nabla u(x,t)|^2\, dt\Big)^{\frac{1}{2}}.$$

A closely related (and slightly easier) operator is

$$g_1 f(x) = \Big(\int_0^\infty t|\frac{\partial}{\partial t} f \star P_t(x)|^2 dt\Big)^{\frac{1}{2}}.$$

If we were look at a discrete version of the operator g_1, we might try to break up the region of integration, estimate $\frac{\partial}{\partial t} f \star P_t(x)$ on each piece, and add up the results in the appropriate way. Consider the following heuristic[1] argument.

$$\begin{aligned}
g_1 f(x) &= \Big(\int_0^\infty t|\frac{\partial}{\partial t} P_t(x) \star f|^2 dt\Big)^{\frac{1}{2}} \\
&= \Big(\sum_{k=-\infty}^{\infty} \int_{2^k}^{2^{k+1}} t|\frac{\partial}{\partial t} P_t(x) \star f|^2 dt\Big)^{\frac{1}{2}} \\
&\approx \Big(\sum_{k=-\infty}^{\infty} \int_{2^k}^{2^{k+1}} 2^k \Big|\frac{P_{2^{k+1}} \star f(x) - P_{2^k} \star f(x)}{2^k}\Big|^2 dt\Big)^{\frac{1}{2}} \\
&= \Big(\sum_{k=-\infty}^{\infty} |P_{2^{k+1}} \star f(x) - P_{2^k} \star f(x)|^2 \int_{2^k}^{2^{k+1}} \frac{1}{2^k} dt\Big)^{\frac{1}{2}} \\
&= \Big(\sum_{k=-\infty}^{\infty} |P_{2^{k+1}} \star f(x) - P_{2^k} \star f(x)|^2\Big)^{\frac{1}{2}}.
\end{aligned}$$

[1] Actually one direction of this heuristic argument can be made precise. Write

$$\begin{aligned}
\big(\sum_k |P_{2^{k+1}} \star f(x) - P_{2^k} \star f(x)|^2\big)^{\frac{1}{2}} &= \big(\sum_k \Big|\int_{2^k}^{2^{k+1}} \frac{\partial}{\partial t} P_t \star f(x)\, dt\Big|^2\big)^{\frac{1}{2}} \\
&\le \big(\sum_k 2^k \int_{2^k}^{2^{k+1}} |\frac{\partial}{\partial t} P_t \star f(x)|^2\, dt\big)^{\frac{1}{2}} \\
&\le \big(\int_0^\infty t|\frac{\partial}{\partial t} P_t \star f(x)|^2\, dt\big)^{\frac{1}{2}} = g_1 f(x).
\end{aligned}$$

With the above martingale square function, and the Littlewood-Paley g function as motivation, we define the ergodic analogs, and see to what extent they behave like the previously studied square functions.

Fix a sequence $n_1 < n_2 < \dots$. We define the ergodic square function associated with this sequence by

$$Sf(x) = \Big(\sum_{k=1}^{\infty} |A_{n_k} f(x) - A_{n_{k+1}} f(x)|^2 \Big)^{\frac{1}{2}}.$$

The theory for this operator developed as follows:

(1) The first special case was $n_k = k$, which was shown to be a bounded operator on L^p for $p > 1$, and weak type (1,1) in 1974. (See [40] and [41].)
(2) The second special case was $n_k = 2^k$ considered by de la Torre in 1975 [31].
(3) The general case was studied in a paper that appeared in 1996 in joint work with Ostrovskii and Rosenblatt [47]. We showed that in particular the operators are bounded on L^p, $1 < p \leq 2$ and are weak (1,1). Later, in joint work with Kaufmann, Rosenblatt and Wierdl, we showed that they map L^∞ into BMO and hence by interpolation, are bounded on all L^p, $1 < p < \infty$.

To prove these square function inequalities, we argue as follows. First, using the Calderón transfer principle, as usual, we transfer the problem to $\mathbb{Z}$. To prove the L^2 result, we use a Fourier transform argument. After writing things in terms of Fourier transforms, we make two different types of estimates, depending on the growth from n_k to n_{k+1}. In particular, for fixed θ, we look at those k so $|n_k\theta - n_{k+1}\theta| \leq 1$ and those where $|n_k\theta - n_{k+1}\theta| > 1$. It is then a matter of adding up the resulting estimates. This takes some work, but is not too unpleasant.

For the weak (1,1) result, we continue to work on $\mathbb{Z}$, and do a Calderón-Zygmund decomposition of f. (Recall Theorem 4.12.) Following the standard technique, as in the proof of Theorem 4.10, write $f = g + b$. Use the L^2 estimate on g, and then try to use an L^1 estimate on b, staying far away from the support of b. Unfortunately, even in some simple cases, (say $n_k = k$ and $b = \delta_0 - \delta_1$) the square function will not be in $\ell^1(\tilde{B}^c)$, that is, even if we only sum far away from the support of b, we can still diverge. Consequently, the standard technique fails, and we need to modify the method.

We avoid the above difficulty by using the following result:

Theorem 6.1. *Let S be a sublinear operator, and assume that S has the following properties*

(1) $|\{x : Sf > \lambda\}| \leq \frac{c}{\lambda^2}\|f\|_2^2$.
(2) *For some $\varrho \geq 1$, we have*
  a) $|S(\sum_j b_j)(x)| \leq c\Big(\sum_j |S(b_j)(x)|^\varrho\Big)^{\frac{1}{\varrho}}$ *for $x \notin \tilde{B}$.*
  b) $\frac{1}{\lambda^\varrho}\sum_{x\in\tilde{B}_j^c} |S(b_j)|^\varrho \leq c|B_j|$.

Then S is weak $(1,1)$.

Proof. First, using the decomposition, we can write

$$m\{x : Sf(x) > 2\lambda\} \leq m\{x : Sg(x) > \lambda\} + m\{x : Sb(x) > \lambda\}.$$

For the "good function", g, the argument is easy, as before. We have

$$\begin{aligned}|\{x : Sg(x) > \lambda\}| &< \frac{c}{\lambda^2}\|g\|_2^2 \\ &= \frac{c}{\lambda^2}\int |g(x)|^2\,dx \\ &\le \frac{c}{\lambda^2}\int |g(x)|\lambda\,dx \\ &\le \frac{c}{\lambda}\int |f(x)|\,dx = \frac{c}{\lambda}\|f\|_1,\end{aligned}$$

as required.

For the "bad function", b, we have the following estimate

$$\begin{aligned}|\{x : Sb(x) > \lambda\}| &\le |\{x \in \tilde{B}^c : Sb(x) > \lambda\}| + |\tilde{B}| \\ &\le \frac{1}{\lambda^\varrho}\sum_{x\in\tilde{B}^c} |S(\sum_j b_j)(x)|^\varrho + |\tilde{B}| \\ &\le \frac{c}{\lambda^\varrho}\sum_{x\in\tilde{B}}\sum_j |S(b_j)(x)|^\varrho + |\tilde{B}| \\ &\le \sum_j \frac{1}{\lambda^\varrho}\sum_{x\in\tilde{B}_j^c} |S(b_j)(x)|^\varrho + |\tilde{B}| \\ &\le c\sum_j |B_j| + 5\sum_j |B_j| \\ &\le \frac{c}{\lambda}\|f\|_1.\end{aligned}$$

□

To prove $S(f)$ is weak type (1,1) it will be enough to show that our square function satisfies the properties in Theorem 6.1.

The required weak (2,2) result is just a consequence of the strong (2,2) result. For the second condition we use $\varrho = 2$, and note that for fixed x and k, at most two of the terms in the squared part of the expression are non-zero, and that $(a+b)^2 \le 2(a^2+b^2)$.

The third condition requires some work to estimate, but is just a computation.

It turns out that these square functions map ℓ^∞ into BMO, and consequently, we can interpolate, and get strong type (p,p) for $1 < p < \infty$.

Remark 6.2. There are some sequences, such as $n_k = 2^k$ for which a different argument will work. When n_k increases fast enough, a condition sometimes referred to as the Hörmander condition is satisfied. In this case, the more standard argument will work. However, in general the Hörmander condition is not satisfied, and the above argument is required.

To give a flavor of the kind of computations necessary to prove the square function is bounded on L^2, we illustrate the technique with a very special case.

Let $n_k = k$, that is we are looking at the square function

$$Sf(x) = \Big(\sum_{k=1}^{\infty} |A_k f(x) - A_{k+1} f(x)|^2\Big)^{\frac{1}{2}}.$$

Rewriting the problem on $\mathbb{Z}$, we have

$$\hat{A}_n f(\theta) = \frac{1}{n}\sum_{k=0}^{n-1} e^{2\pi i k\theta} = \frac{1-e^{2\pi i n\theta}}{n(1-e^{2\pi i\theta})}.$$

Using Fourier transforms, we have

$$\begin{aligned}
\|Sf\|_{\ell^2}^2 &= \int_{\mathbb{Z}} |Sf(x)|^2\, dx \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} |\widehat{Sf}(\theta)|^2\, d\theta \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{k=1}^{\infty} |\widehat{A_k f}(\theta) - \widehat{A_{k+1} f}(\theta)|^2\, d\theta \\
&\le \| \sum_{k=1}^{\infty} |\widehat{A_k} - \widehat{A_{k+1}}|^2 \|_\infty \int_{-\frac{1}{2}}^{\frac{1}{2}} |\hat{f}(\theta)|^2\, d\theta \\
&\le | \sum_{k=1}^{\infty} |\widehat{A_k} - \widehat{A_{k+1}}|^2 \|_\infty \|f\|_{\ell^2}^2.
\end{aligned}$$

Hence it is enough to show that

$$\| \sum_{k=1}^{\infty} |\widehat{A_k} - \widehat{A_{k+1}}|^2 \|_\infty \le c.$$

We have

$$\begin{aligned}
\| \sum_{k=1}^{\infty} |\widehat{A_k} - \widehat{A_{k+1}}|^2 \|_\infty &= \Big\| \sum_{k=1}^{\infty} |(\frac{1}{k}) - (\frac{1}{k+1}) \sum_{j=0}^{k-1} e^{2\pi i j\theta} - \frac{1}{k+1} e^{2\pi i k\theta}|^2 \Big\|_\infty \\
&= \Big\| \sum_{k=1}^{\infty} \frac{1}{(k+1)^2} |\frac{1}{k}\sum_{j=0}^{k-1} e^{2\pi i j\theta} - e^{2\pi i k\theta}|^2 \Big\|_\infty \\
&\le \Big\| \sum_{k=1}^{\infty} \frac{1}{(k+1)^2} \Big(2|\widehat{A_k}(\theta)|^2 + 2\Big) \Big\|_\infty
\end{aligned}$$

Because of the term $\frac{1}{(k+1)^2}$, we will have a bounded sum if

$$|\widehat{A_k}(\theta)| \le c$$

for some constant c, independent of θ. The required estimate is trivial. Just estimate each term in the sum defining $\hat{A}_k$ by 1. Thus, the average is no more than 1.

Remark 6.3. In the case special case considered above, i.e., the case $n_k = k$, a different argument can be used. Following the argument above, but working directly with the ergodic average, we can dominate the square function by a constant times the maximal function, plus the operator

$$\tilde{S}f(x) = \Big(\sum_{k=1}^{\infty} (\frac{f(\tau^k x)}{k})^2\Big)^{\frac{1}{2}}.$$

It is easy to see that this last operator is bounded in L^2. A variant of this operator, when we look at $f(\tau^{n_k}x)$ instead of $f(\tau^k x)$, is studied in [50].

We now consider the oscillation operator

$$Of(x) = \Big(\sum_{k=1}^{\infty} \sup_{n_k < n \le n_{k+1}} |A_{n_k} f(x) - A_n f(x)|^2\Big)^{\frac{1}{2}}.$$

(1) In 1977, Gaposhkin [36] showed that $\|Of\|_2 \le c\|f\|_2$ in the case where we have $1 < \alpha \le \frac{n_{k+1}}{n_k} \le \beta < \infty$ for some constants α and β. Also see Gaposhkin's later paper [37].
(2) In 1989 Homer White showed that there was in L^2 inequality in case there are constants $\alpha > 0$ and $\beta > 1$ such that $n_{k+1} > \alpha n_k^{\beta}$. (See [6].)
(3) In joint work with R. Kaufman, J. Rosenblatt and M. Wierdl, [43] we showed that $\|Of\|_2 \le c\|f\|_2$, with no restriction on the increasing sequence (n_k), and in fact we showed, $\|Of\|_p \le c_p\|f\|_p$ for $1 < p < \infty$, and the operator is weak type (1,1). Further, the constant c_p depends only on p, and does not depend on the sequence (n_k).

We will now show how to prove a weaker version of this last result. We will show how to get the L^2 boundedness without the restriction imposed by Gaposhkin, that $\frac{n_{k+1}}{n_k} \le \beta < \infty$. (However, the argument will still require the restriction that $1 < \alpha < \frac{n_{k+1}}{n_k}$.) To do this we need one more result that is also contained in the joint paper with Kaufman, Rosenblatt and Wierdl. On $\mathbb{Z}$ we can consider the reverse martingale given by $E_n f(x) = \frac{1}{2^n}\sum_{j=r2^n}^{(r+1)2^n-1} f(j)$ where $r2^n \le x < (r+1)2^n$. (This reverse martingale satisfies the same results as the standard martingale, in particular we have $\|O_M f\|_2 \le c\|f\|_2$.) We can also define the square function

$$S_D f(x) = \Big(\sum_{k=1}^{\infty} |A_{2^k} f(x) - E_k f(x)|^2\Big)^{\frac{1}{2}}.$$

This operator, which gives us a way to transfer martingale results to the ergodic theory setting, is strong (p,p) for $1 < p < \infty$, and is weak (1,1).

We can now estimate as follows. For fixed $n_k < n$, let 2^{ℓ_k} denote the largest dyadic that is to the left of n_k, and let 2^{ℓ} denote the largest dyadic that is to the left of n. Let $f_n = A_n f(x)$. We have

$$\begin{aligned}
&|f_{n_k} - f_n| \\
&\quad = |f_{n_k} + (f_{2^{\ell_k}} - f_{2^{\ell_k}}) + (E_{\ell_k} f - E_{\ell_k} f) + (E_\ell f - E_\ell f)) + (f_{2^\ell} - f_{2^\ell}) - f_n| \\
&\quad = |(f_{n_k} - f_{2^{\ell_k}}) + (f_{2^{\ell_k}} - E_{\ell_k} f) + (E_{\ell_k} f - E_\ell f) + (E_\ell f - f_{2^\ell}) + (f_{2^\ell} - f_n)| \\
&\qquad \le |f_{n_k} - f_{2^{\ell_k}}| + |f_{2^{\ell_k}} - E_{\ell_k} f| + |E_{\ell_k} f - E_\ell f| + |E_\ell f - f_{2^\ell}| + |f_{2^\ell} - f_n|.
\end{aligned}$$

Using this observation, we can estimate $(\sum_{k=1}^{\infty} |f_{n_k}(x) - f_n(x)|^2)^{1/2}$ by first using the triangle inequality. From the resulting expressions, the first one we can estimate by using Sf, the second by $S_D f$, the third by $O_M f$, the fourth by $S_D f$ again, and for the last expression, use Gaposhkin's result, since $2^\ell < n \le 2^{\ell+1}$.

We can also define an analog of the operator $V_M^{\varrho} f$, by

$$V_\varrho f(x) = \sup_{n_1 < n_2 < \dots} \Big(\sum_{k=1}^{\infty} |A_{n_k} f(x) - A_{n_{k+1}} f(x)|^{\varrho}\Big)^{\frac{1}{\varrho}}.$$

The proof is much in the same spirit, but we need to use Lepingle's result in place of the result for $O_M f$.

This result gives us information about "jumps" of the ergodic averages. Define the operator

$$N(f,\lambda,x) = \max\{n \mid \text{there exists } s_1 < t_1 \le s_2 < t_2 \cdots \le s_n < t_n, \\ \text{such that } |A_{t_k}f(x) - A_{s_k}f| > \lambda \text{ for each } 1 \le k \le n\}.$$

We can now show that

$$|\{x : N(f,\lambda,x) > n\}| \le \frac{c}{\lambda n^{\frac{1}{\varrho}}}\|f\|_1.$$

We just note that

$$\lambda\big(N(f,\lambda,x)\big)^{\frac{1}{\varrho}} \le V_\varrho f(x).$$

Hence

$$\begin{aligned} |\{x : N(f,\lambda,x) > n\}| &\le |\{x : \Big(\frac{V_\varrho f(x)}{\lambda}\Big)^{\varrho} > n\}| \\ &\le |\{x : V_\varrho f(x) > \lambda n^{\frac{1}{\varrho}}\}| \\ &\le \frac{c}{\lambda n^{\frac{1}{\varrho}}}\|f\|_1. \end{aligned}$$

With a little more work we can in fact replace the exponent $\frac{1}{\varrho}$ by $\frac{1}{2}$ (see [43]) and we can show that this is sharp (see [50]).

There are also higher dimensional results. For example we can use squares, rectangles, etc. See [51].

7. Concluding Remarks

There are many interesting and important areas that are related to the above discussion, but that we have mentioned only briefly, or in some cases, not at all. For example we did not discuss the extensive work on good subsequences for the pointwise ergodic theorem. This started with the block sequences of Bellow and Losert [16], followed by the important and difficult work of Bourgain ([19], [20]), and Wierdl ([76], [75]). For an excellent (and very readable) exposition of this work, the reader should see the article by Rosenblatt and Wierdl [69].

There is an important open problem associated with subsequences. At this time, there is no known example of a subsequence that is good for a.e. convergence for all $f \in L^1$, and has successive gaps increasing to infinity. In particular, the question of a.e. convergence for $f \in L^1$ along the sequence of squares is open, and probably very difficult. The techniques used in Section 4 and Section 6, including the Calderón-Zygmund decomposition, do not seem to apply.

We have not discussed the ergodic Hilbert transform, which has many of the same properties as the ordinary Hilbert transform. Petersen's book, "Ergodic Theory" [62] contains a good introduction to this operator. It turns out that the moving ergodic averages studied in [15], have an analog for the ergodic Hilbert transform. See [35].

We could consider oscillation inequalities and variational inequalities for several operators other than those discussed above. For example we could consider oscillation and variational inequalities for averages along subsequences, convolution powers, or other weighted averages. (Some results on oscillation and variational norms for convolution powers are contained in [48], but there is more work to do.)

While as mentioned above, Bellow and Calderón have shown a.e. convergence of $\mu^n f$ for $f \in L^1$, if μ has mean value zero and finite second moment However, there are many examples of μ for which we know a.e. convergence for $f \in L^p$, $p > 1$, but which do not satisfy the conditions necessary to apply Bellow and Calderón's result. Do the convolution powers of these measures converge a.e. for $f \in L^1$? Again, this problem seems difficult.

It is clear that there are many interesting questions that remain in this area, and it seems that the more we discover, the more questions arise.

References

[1] M. Akcoglu, A. Bellow, R. L. Jones, V. Losert, K. Reinhold-Larsson, and M. Wierdl, *The strong sweeping out property for lacunary sequences, Riemann sums, convolution powers, and related matters*, Ergodic Theory and Dynamical Systems, **16** (1996), 207–253.

[2] M. Akcoglu, and A. del Junco, *Convergence of averages of point transformations*, P.A.M.S. **49** (1975), 265–266.

[3] M. Akcoglu, M. D. Ha, and R. L. Jones, *Sweeping out properties for operator sequences*, Canadian J. Math., **49** (1997), 3–23.

[4] M. Akcoglu and R. L. Jones, *Strong sweeping out for block sequences and related ergodic averages*, preprint.

[5] M. Akcoglu, R. L. Jones, and P. Schwartz, *Variation in probability, ergodic theory and analysis*, to appear in Il. J. Math.

[6] I. Assani, K. Petersen, and H. White, *Some connections between ergodic theory and harmonic analysis*, Almost Everywhere Convergence, II, (Proceedings of the Conference on A.E. Convergence and Probability Theory, Northwestern University, Fall 1989) (A. Bellow and R. Jones, eds.), Academic Press, New York, 1991, pp. 17–40.

[7] D. G. Austin, *A sample function property of martingales*, Ann. Math. Stat., **37** (1966), 1396–1397.

[8] A. Bellow, *Transfer principles in ergodic theory*, to appear.

[9] A. Bellow, *On "Bad Universal" Sequences in Ergodic Theory, II*, Lecture Notes in Mathematics, no. 1033, Springer-Verlag, Berlin, 1983.

[10] A. Bellow, *Perturbation of a sequence*, Adv. Math. **78** (1989), 131–139.

[11] A. Bellow, and A. P. Calderón, *A weak type inequality for convolution products*, to appear.

[12] A. Bellow, and R. L. Jones, *A Banach principle for L^∞*, Advances in Math., **120** (1996), 155–172.

[13] A. Bellow, R. L. Jones, and J. Rosenblatt, *Almost everywhere convergence of convolution powers*, Ergodic Theory and Dyn. Sys., **14** (1994), 415–432.

[14] A. Bellow, R. L. Jones, and J. Rosenblatt, *Almost everywhere convergence of weighted averages*, Math Ann., **293** (1992), 399–426.

[15] A. Bellow, R. L. Jones, and J. Rosenblatt, *Convergence for moving averages*, Ergodic Theory and Dyn. Sys., **10** (1990), 45–62.

[16] A. Bellow, and V. Losert, *On sequences of density zero in ergodic theory*, Contemp. Math. **26** (1984), 49–60.

[17] E. Bishop, *Foundations of Constructive Analysis*, Mc Graw Hill, 1967.

[18] J. Bourgain, *Almost sure convergence and bounded entropy*, Israel Journal, **62** (1988), 79–97.

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

[20] J. Bourgain, *On the pointwise ergodic theorem on L^p for arithmetic sequences*, Israel J. Math., **61** (1988), 73–84.

[21] J. Bourgain, *Pointwise ergodic theorems for arithmetic sets*, Publications Mathematiques, Institut des Hautes Etudes Scientifiques, **69** (1989), 5–45.

[22] M. Boshernitzan, and M. Wierdl, *Ergodic theorems along sequences and Hardy fields*, Proc. Nat. Acad. Sci., **93** (1996), 8205–8207.

[23] D. Burkholder, *Distribution function inequalities for martingales*, Ann. of Prob., **1** (1973), 19–42.

[24] D. Burkholder and Y. S. Chow, *Iterates of conditional expectation operators*, Proc. A.M.S., **12** (1961), 490–495.

[25] D. Burkholder and R. Gundy, *Extrapolation and interpolation of quasi–linear operators on martingales*, Acta Math., **124** (1970), 249–304.

[26] D. Burkholder, R. Gundy, and M. Silverstein, *A maximal function characterization of the class H^p*, Trans. Amer. Math. Soc., **157** (1971), 137–153.

[27] A. P. Calderón, *Ergodic theory and translation invariant operators*, Proc. Nat. Acad. of Sci., USA **59** (1968), 349–353.

[28] R. R. Coifman, *Distribution function inequalities for singular integrals*, Proc. Nat. Acad. Sci., USA **69** (1972), 2838–2839.

[29] R. R. Coifman and G. Weiss, *Operators associated with representations of amenable groups, singular integrals induced by ergodic flows, the rotation method and multipliers*, Studia Math., **47** (1973), 285–303.

[30] M. Cotlar, *A unified theory of Hilbert transforms and ergodic theorems*, Revista Mat. Cuyana **1** (1955), 105–167.

[31] A. de la Torre, *Hardy Spaces Induced by an Ergodic Flow*, Ph.D Thesis, Washington University, 1975.

[32] A. Del Junco and J. Rosenblatt, *Counterexamples in ergodic theory and number theory*, Math. Ann., **245** (1979), 185–197.

[33] J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.

[34] W. Feller, *An Introduction to Probability Theory and its Applications* (Third ed.), Wiley, New York, 1950.

[35] S. Ferrando, R. L. Jones, and K. Reinhold, *On approach regions for the conjugate Poisson integral and singular integrals*, Studia Math. **120** (1996), 169–182.

[36] V. F. Gaposhkin, *A theorem on the convergence almost everywhere of a sequence of measurable functions and its applications*, Math. USSR Sbornik **33** (1977), 1–17.

[37] V. F. Gaposhkin, *Individual ergodic theorem for normal operators on L^2*, Functional Analysis and Applications **15** (1981), 14–18.

[38] R. F. Gundy, *A decomposition of L^1 bounded martingales*, Ann. Math. Statist. **39** (1968), 134–138.

[39] R. L. Jones, *A remark on singular integrals with complex homogeneity*, Proc. A.M.S. **114** (1992), 763–768.

[40] R. L. Jones, *Inequalities for the ergodic maximal function*, PhD Thesis, Rutgers University, 1974.

[41] R. L. Jones, *Inequalities for the ergodic maximal function*, Studia Math. **40** (1977), 111–129.

[42] R. L. Jones, *Maximal inequalities for pairs of ergodic transformations*, Radovi Matematicki, **4** (1988), 55–61.

[43] R. L. Jones, R. Kaufman, J. Rosenblatt, and M. Wierdl, *Oscillation in ergodic theory*, to appear in Ergodic Theory and Dyn. Sys.

[44] R. L. Jones, M. Lacey, and M. Wierdl, *Random sequences of integers and the pointwise ergodic theorem*, preprint.

[45] R. L. Jones and J. Olsen, *Subsequences pointwise ergodic theorems for operators in L^p*, Israel J. Math., **77** (1992), 33–54.

[46] R. L. Jones, J. Olsen, and M. Wierdl, *Subsequence ergodic theorems for L^p contractions*, Trans. A.M.S. **351** (1992), 837–850.

[47] R. L. Jones, I. Ostrovskii, and J. Rosenblatt, *Square functions in ergodic theory*, Ergodic Theory and Dyn. Sys., **16** (1996), 267–305.

[48] R. L. Jones and K. Reinhold, *Oscillation inequalities for convolution operators*, preprint.

[49] R. L. Jones, J. Rosenblatt, and A. Tempelman, *Ergodic theorems for convolutions of a measure on a group*, Il. J. Math., **38** (1994), 521–553.

[50] R. L. Jones, J. Rosenblatt, and M. Wierdl, *Counting in ergodic theory*, preprint.

[51] R. L. Jones, J. Rosenblatt, and M. Wierdl, *Oscillation in ergodic theory: higher dimensional results*, preprint.

[52] R. L. Jones and M. Wierdl, *Convergence and divergence of ergodic averages*, Ergodic Theory and Dyn. Sys., **14** (1994), 515–535.

[53] R. L. Jones and M. Wierdl, *Convergence of ergodic averages*, Convergence in Ergodic Theory and Probability, (Bergelson, March and Rosenblatt, ed.), de Gruyter, Berlin, 1996, pp. 229–247.
[54] U. Krengel, *On the individual ergodic theorem for subsequences*, Ann. Math. Stat. **42** (1971), 1091–1095.
[55] D. Lepingle, *La variation d'order p des semi-martingales*, Z. Wahrscheinlichkeitstheor. Verw. Geb., **36** (1976), 295–316.
[56] V. Losert, *A remark on the strong sweeping out property*, Convergence in Ergodic Theory and Probability, (Bergelson, March and Rosenblatt, ed.), de Gruyter, Berlin, 1996, pp. 291–294.
[57] J. M. Marstrand, *On Kinchine's conjecture about strong uniform distribution*, Proc. London Math. Soc., **21** (1970), 540–556.
[58] I. Monroe, *Almost sure convergence of the quadratic variation of martingales: a counterexample*, Ann. of Prob. **4** (1976), 133–138.
[59] A. Nagel and E. M. Stein, *On certain maximal functions and approach regions*, Adv. Math. **54** (1984), 83–106.
[60] G. Pisier and Q. Xu, *The strong p-variation of martingales and orthogonal series*, Prob. Theory **77** (1988), 497–514.
[61] K. Petersen, *Brownian Motion, Hardy Spaces and Bounded Mean Oscillation*, London Mathematical Society Lecture Series **28** Cambridge University Press, New York, 1977.
[62] K. Petersen, *Ergodic Theory*, Cambridge University Press, 1983.
[63] A. D. Pollington, *On the density of sequences* $\{n_k\psi\}$, Illinois J. Math. **23** (1979), 511–515.
[64] K. Reinhold, *Convolution powers in* L^1, Illinois J. Math., **37** (1993), 666–679.
[65] K. Reinhold, *Discrepancy of behavior of perturbed sequences in* L^p *spaces*, AMS Proc. **120** (1994), 865–874.
[66] J. Rosenblatt, *Ergodic group actions*, Arch. Math., **47** (1986), 263–269.
[67] J. Rosenblatt, *Norm convergence in ergodic theory and the behavior of Fourier transforms*, Canadian Journal of Math., **46** (1994), 184–199.
[68] J. Rosenblatt, *Universally bad sequences in ergodic theory*, Almost Everywhere Convergence, II (A. Bellow and R. Jones ed.), Academic Press, 1991, pp. 227–245.
[69] J. Rosenblatt and M. Wierdl, *Pointwise ergodic theorems via harmonic analysis*, Proc. Conference on Ergodic Theory (Alexandria, Egypt, 1993), London Mathematical Society Lecture Notes, no. 205, Cambridge University Press, 1995, pp. 3–151.
[70] W. Rudin, *An arithmetic property of Riemann sums*, Proc. A.M.S., **15** (1964), 321–324.
[71] E. M. Stein, *On the maximal ergodic theorem*, Proc. Natl Acad. Sci., **47** (1961), 1894–1897.
[72] E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton, N.J., 1970.
[73] J. Suerio, *A note on maximal operators of Hardy–Littlewood type*, Proc Camb. Phil. Soc., **102** (1987), 131–134.
[74] Y. L. Tong, *Probability Inequalities in Multivariate Distributions*, Academic Press, New York, 1980.
[75] M. Wierdl, *Almost Everywhere and Recurrence along Subsequences in Ergodic Theory*, PhD Thesis, The Ohio State University, 1989.
[76] M. Wierdl, *Pointwise ergodic theorem along the prime numbers*, Israel J. Math **64** (1988), 315–336.

Department of Mathematics, DePaul University, 2219 N. Kenmore, Chicago IL 60614
rjones@condor.depaul.edu http://www.depaul.edu/~rjones/