Analysis and Ergodic Theory

Summer School, Lake Arrowhead *

September 17th - September 22nd 2006

Organizers:

Ciprian Demeter, University of California, Los Angeles

Christoph Thiele, University of California, Los Angeles

*supported by NSF grant DMS 0400879

Contents

1 The primes contain arbitrarily long arithmetic progressions II **5**

Tim Austin, UCLA 5

1.1 Orienting remarks 5

1.2 Warmup: two tricks to help handle the primes 6

1.2.1 The W-trick 6

1.2.2 Dealing with wraparound 6

1.2.3 Our goal, modified 7

1.3 The actual construction 7

1.4 Proving pseudorandomness 9

2 A generalization of Birkhoff's pointwise ergodic theorem **13**

Svetlana Butler, UIUC 13

2.1 Introduction 13

2.2 Main results 14

2.3 Remarks 15

2.4 Outline of the proofs 16

3 A pointwise ergodic theorem for amenable groups **19**

S. Zubin Gautam, UCLA 19

3.1 Preliminaries 19

3.2 Statement of the main result 20

3.3 Random selection of Følner sets 20

3.3.1 Proof sketch of Lemma 5 for G discrete 21

3.3.2 Proof sketch of Lemma 5 for general G 23

3.4 Maximal inequality and pointwise ergodic theorem 24

3.5 Example: The lamplighter group 25

4 Arithmetic progressions in primes I **27**

Alexander Gorodnik, CalTech 27

4.1 Introduction 27

4.1.1 Notation 30

4.1.2 Pseudorandom measures 31

4.2 Tools 32

4.2.1 Gowers uniformity norms 32

4.2.2 Obstructions to uniformity and dual functions 33

4.2.3 σ-Algebras generated by generalized Bohr sets 34
4.3 Structure theorem 35

5 Polynomial extensions of van der Waerden's and Szemerédi's theorems 38
Michael Johnson, Northwestern University 38
5.1 Introduction 38
5.2 Polynomial Expressions 39
5.3 Theorem 4 40
5.4 Theorem 2 41

6 Convergence of Conze-Lesigne Averages 46
Tamara Kucherenko, UCLA 46
6.1 Introduction 46
6.2 Statement of results 48
6.3 Reduction to a simpler system 48
6.3.1 Characteristic factors 48
6.3.2 Reduction to the isometric extension of the Kronecker . 49
6.3.3 Reduction to an abelian group extension 50
6.4 Sketch of the proof of the main theorem 50

7 Pointwise Ergodic Theorems for Arithmetic Sets, part 2 53
Victor Lie 53
7.1 Introduction 53
7.2 The L^2 case for Theorem 1 54
7.3 The almost sure convergence of $A_N f$ for $f \in l^2(Z)$ 57
7.4 The L^p case for Theorem 1 59
7.5 The outlines of the Theorem 2 61

8 The Ergodic Theoretical Proof of Szemerédi's Theorem II 63
Anne E. McCarthy, Temple University 63
8.1 Introduction 63
8.2 Factors 64
8.2.1 skew products 64
8.2.2 projecting functions 65
8.2.3 fiber squares (or relative squares) 65
8.3 Maximal SZ Factors 66

8.4 Relative Weak Mixing 67
8.5 Compact Extensions and Existence 68

9 Multiple recurrence and Szemerédi's theorem 69
Richard Oberlin,UW Madison 69
9.1 Introduction 69
9.2 Theorem 2 implies Theorem 3 70
9.3 Beginning the proof of Theorem 2 71
9.3.1 Weak-mixing systems 72
9.3.2 Compact systems 72
9.3.3 Weak-mixing and compact factors 74

10 Entropy of Convolutions on the Circle, I 75
Robert C. Rhoades, University of Wisconsin 75
10.1 Furstenberg's Conjecture 75
10.2 Uniform Distribution and $\{c_n\}$-genericity 77
10.3 The Role of Entropy 78

11 Entropy of convolutions on the circle 81
Shuanglin Shao, UCLA 81
11.1 Introduction 81
11.2 Dimension of sum sets 84
11.3 Uniform distribution on subgroups 84
11.4 Entropy and subgroups 85
11.5 The convolution theorem 85

12 Bourgain's Entropy Estimates 87
John Workman, Cornell 87
12.1 Preliminaries 87
12.2 The First Entropy Result 88
12.3 The Second Entropy Result 90
12.4 Application: Bellow's Averages 91

13 Pointwise Ergodic Theorems for Arithmetic Sets, part 1 93
Andrew Yingst, USC Columbia 93
13.1 Introduction 93
13.2 Reduction to Inequalities. 94
13.3 A Lemma 96

1 The primes contain arbitrarily long arithmetic progressions II

after B. Green and T. Tao [2]
A summary written by Tim Austin

Abstract

We cover the second half of Green and Tao's proof of the existence of arbitrarily long arithmetic progressions of prime numbers, by constructing a function and pseudorandom measure suitably associated to the primes.

1.1 Orienting remarks

The proof of the Green-Tao Theorem breaks conveniently into two distinct stages:

1. First, it is shown how the conclusion of Szemerédi's Theorem can be extended to cover also the cases of certain sets with asymptotically zero density, by assuming instead positive density with respect to a 'measure' satisfying certain conditions (wrapped up into a property called 'pseudorandomness');

2. Second, such a measure and function must actually be constructed with the function non-zero only on some set related to the primes in such a way that the presence of long arithmetic progressions in the primes themselves may be deduced. (Note that the slightly convoluted wording here is necessary: it will turn out that our function is not itself supported by the primes.)

Here we will be concerned chiefly with the second of the above stages, detailed in Sections 9 – 11 of [2]. Results from the first stage (Sections 1 – 8 of [2]) will mostly be assumed. One theme for this presentation is a *pragmatic* attitude towards the construction and use of a suitable pseudorandom measure for studying the primes. It seems that the exact construction of such a measure is somewhat arbitrary; the Green-Tao construction should be used because it is one for which existing techniques yield sufficiently good estimates to give a proof of the theorem, and not for any more high-minded reason. Indeed, Bernard Host [3] has recently given a slightly different construction based on a different methodology for some of these estimates.

1.2 Warmup: two tricks to help handle the primes

Even before constructing a particular function and measure, we can foresee certain issues when working with the primes.

1.2.1 The W-trick

Suppose that f is some function $\mathbb{N} \to \mathbb{R}$ that is non-zero only on primes. As it stands there can be no k-pseudorandom measure majorizing f. This is because the primes "notice and avoid" certain congruence classes: if a, q are not coprime, then there can be at most one prime $\equiv q \mod a$. Thus for any q the primes are concentrated into the $\phi(q)$ congruence classes modulo q that are coprime to q, and so this choice of f, and hence also any $\nu \geq f$, would be concentrated on these congruence classes.

However, the (m_0, t_0, L_0)-linear forms condition forces ν to be equidistributed among different congruence classes, *if q is sufficiently small*: for $q \leq L_0$, simply apply the condition to the one linear form $\psi(x) = qx + b$ for any b. We circumvent this problem by looking for arithmetic progressions among the primes that occur within one fixed congruence class that *is* coprime to all $q \leq L_0$: say among $\{n : Wn + 1 \text{ is a prime}\}$ for W a product of all small primes.

In order to find k-term arithmetic progressions, we need the $(k2^{k-1}, 3k - 4, k)$-linear forms condition (as specified in the definition of k-pseudorandomness) for our measure. Thus our choice of W only need depend on k, but it is easier to choose a function $w(N)$ growing slowly to ∞ with N (how slowly can safely be decided later), and set

$$W = W(N) := \prod_{p \leq w(N)} p;$$

clearly this is eventually big enough for any fixed k.

1.2.2 Dealing with wraparound

Suppose that we have found an arithmetic progression in $\mathbb{Z}_N$. Even upon identifying $\mathbb{Z}_N$ with $\{1, 2, \ldots, N\}$, our find need not correspond to a genuine arithmetic progression in $\mathbb{N}$, owing to the possibility of wraparound in the cyclic group $\mathbb{Z}_N$.

However, we can rule out this problem if we know that our length-k progression in $\mathbb{Z}_N$ is contained in some very short arc (how short depending

on k). It suffices to restrict attention to arcs of length less than $(1/k)N$, but for other reasons we will later take $\varepsilon_k N$ for $\varepsilon_k := 1/2^k(k+4)!$ (much smaller).

1.2.3 Our goal, modified

Combining the two tricks described above, we will search for a specific function and measure of the form

$$f(n) := \begin{cases} g(Wn+1) & \text{if } \varepsilon_k N \leq n \leq 2\varepsilon_k N \\ 0 & \text{else.} \end{cases}$$

(for some function g supported on the primes), and

$$\nu(n) := \begin{cases} h(Wn+1) & \text{if } \varepsilon_k N \leq n \leq 2\varepsilon_k N \\ 1 & \text{else.} \end{cases}$$

(for some other suitable function h).

1.3 The actual construction

It has long been known in number theory that sometimes it is easier to study the function $\mathbf{1}_{\text{primes}}(n)\log n$ than the indicator function $\mathbf{1}_{\text{primes}}(n)$ itself, and so this former seems a good ingredient for f. Thus, we set

$$\widetilde{\Lambda}(n) := \begin{cases} \frac{\phi(W)}{W}\log(Wn+1) & \text{if } Wn+1 \text{ is prime} \\ 0 & \text{else} \end{cases}$$

and

$$f(n) := \begin{cases} k^{-1}2^{-k-5}\widetilde{\Lambda}(n) & \text{if } \varepsilon_k N \leq n \leq 2\varepsilon_k N \\ 0 & \text{else} \end{cases}$$

for $0 \leq n < N$. The various factors in the above definition are simply convenient normalizers.

For our choice of ν we will need to make estimates to verify pseudorandomness. Before the appearance of [2], Goldston and Yıldırım [1] had studied sums of the form

$$\sum_{n \leq N} \Lambda_R(n+h_1)\Lambda_R(n+h_2)\cdots\Lambda_R(n+h_m)$$

for Λ_R the so-called **truncated divisor sum**

$$\Lambda_R(n) := \sum_{d\,|\,n,\ d\leq R} \mu(d)\log\frac{R}{d},$$

where μ is the Möbius function:

$$\mu(n) := \begin{cases} (-1)^r & \text{if } n \text{ is squarefree with } r \text{ prime factors} \\ 0 & \text{else,} \end{cases}$$

which in turn has long been central to the study of the distribution of the primes.

It is Goldston and Yıldırım's work that provides a measure for which pseudorandomness can be checked: we set $R := N^{k^{-1}2^{-k-4}}$, and

$$\nu(n) := \begin{cases} \frac{\phi(W)}{W\log R}\Lambda_R(Wn+1)^2 & \text{if } \varepsilon_k N \leq n \leq 2\varepsilon_k N \\ 1 & \text{else} \end{cases}$$

for $0 \leq n < N$.

Remark The choice of this law for R as a function of N is dictated, firstly, by the requirement that ν majorize a positive multiple of f, and then by the various considerations of how slow it needs to be in the proof of pseudorandomness. We take it to be as high a power of N as will allow all these estimates to go through easily; the value of that power is an artefact of these different parts of the proof. It turns out that that the growth of R in turn forces a bound on the growth of $w(N)$ (and so also $W(N)$), but this can then simply be absorbed into the requirement that $w(N)$ grow "slowly enough".

Assuming we have proved pseudorandomness for our ν, the Green-Tao Theorem now follows from:

Lemma 1 (Lemma 9.4 in [2])**.** *We have $\nu(n) \geq 0$ for all $n \in \mathbb{Z}_N$, and $\nu(n) \geq f(n)$ for all $n \in \mathbb{Z}_N$ whenever N is sufficiently large.*

Thus:

Proposition 2 (Proposition 9.1 in [2])**.** *Write $\varepsilon_k := 1/2^k(k+4)!$, and let N be a sufficiently large prime number. Then there is a k-pseudorandom measure $\nu : \mathbb{Z}_N \to [0,\infty)$ such that $\nu(n) \geq k^{-1}2^{-k-5}\widetilde{\Lambda}(n)$ for all $\varepsilon_k N \leq n \leq 2\varepsilon_k N$.*

Proof of the Green-Tao theorem assuming Proposition 2 Define f as above. From Dirichlet's theorem we observe that

$$\mathbb{E}(f) = \frac{k^{-1}2^{-k-5}}{N} \sum_{\varepsilon_k N \le n \le 2\varepsilon_k N} \widetilde{\Lambda}(n) = k^{-1}2^{-k-5}\varepsilon_k(1+\mathrm{o}(1)).$$

We can now apply Proposition 2 and Theorem 3.5 from [2] (as covered in the first half of that paper) to conclude that

$$\mathbb{E}\big(f(x)f(x+r)\cdots f(x+(k-1)r) \mid x, r \in \mathbb{Z}_N\big) \ge c(k, k^{-1}2^{-k-5}\varepsilon_k) - \mathrm{o}(1),$$

where $c(k, k^{-1}2^{-k-5}\varepsilon_k)$ is a positive constant guaranteed by Theorem 3.5 of [2]. Observe that the degenerate case $r = 0$ can contribute at most $\mathrm{O}(\frac{1}{N}\log^k N) = \mathrm{o}(1)$ to the left-hand side and can thus be discarded. Furthermore, by our trick of restricting attention to $[\varepsilon_k N, 2\varepsilon_k N]$, every progression counted by the expression on the left is a genuine arithmetic progression of integers. Since the right-hand side is positive for sufficiently large N, the claim follows from the definition of $\widetilde{\Lambda}$. □

Remark In fact, by a simple modification of the construction of f and ν above, one can show that there are arbitrarily long arithmetic progressions in any subset A of the primes with positive relative upper density; that is, such that

$$\limsup_{N\to\infty} \frac{|A \cap [1,N]|}{|\{n \in [1,N] : \ n \text{ is prime}\}|})0.$$

This is the full strength of Green and Tao's result.

1.4 Proving pseudorandomness

Pseudorandomness requires the linear forms condition and the correlation condition. These two conditions are proved via two other results, one for each; these intermediate results connect more closely with existing number-theoretic estimates, and then the onward journey to pseudorandomness is more combinatorial.

Proposition 3 (Proposition 9.5 in [2], leading to linear forms condition)**.** *Let m, t be positive integers. For each $1 \le j \le m$, let $\psi_j(\mathbf{x}) := \sum_{s=1}^{t} L_{jk}x_k + b_j$ be linear forms with integer coefficients L_{jk} such that $|L_{jk}| \le \sqrt{w(N)}/2$ for all $j = 1, \ldots, m$ and $k = 1, \ldots, t$. We assume that the t-tuples $(L_{jk})_{k=1}^{t}$ are*

never identically zero, and that no two t-tuples are rational multiples of each other. Write $\theta_j := W\psi_j + 1$. Suppose that B is a product $\prod_{k=1}^{t} I_k \subseteq \mathbb{R}^t$ of intervals I_k, each of which having length at least R^{10m}. Then (if the function $w(N)$ is sufficiently slowly growing in N)

$$\mathbb{E}\big(\Lambda_R(\theta_1(\mathbf{x}))^2 \cdots \Lambda_R(\theta_m(\mathbf{x}))^2 \;\big|\; \mathbf{x} \in B\big) = (1 + \mathrm{o}_{m,t}(1))\left(\frac{W \log R}{\phi(W)}\right)^m.$$

Proposition 4 (Proposition 9.6 in [2], leading to correlation condition)**.** *Let $m \geq 1$ be an integer, and let B be an interval of length at least R^{10m}. Suppose that $h_1, \ldots, h_m$ are distinct integers satisfying $|h_i| \leq N^2$ for all $1 \leq i \leq m$, and let Δ denote the integer*

$$\Delta := \prod_{1 \leq j < k \leq m} |h_j - h_k|.$$

Then (for N sufficiently large depending on m, and assuming the function $w(N)$ sufficiently slowly growing in N)

$$\begin{aligned}&\mathbb{E}\big(\Lambda_R(W(x+h_1)+1)^2 \cdots \Lambda_R(W(x+h_m)+1)^2 \;\big|\; x \in B\big)\\ &\leq (1 + \mathrm{o}_m(1))\left(\frac{W \log R}{\phi(W)}\right)^m \prod_{p \,|\, \Delta} (1 + \mathrm{O}_m(p^{-1/2})).\end{aligned}$$

Remarks

1. The rôles of the arbitrary sets B in the above propositions become clear during the deduction of pseudorandomness from them. Briefly, we need the sets B because we will be defining ν in two different ways inside and outside the interval $[\varepsilon_k N, 2\varepsilon_k N]$, following our second trick from Section 1.2; when verifying the linear forms and correlation conditions, the overall expectation is broken into several expectations over (products of) smaller intervals, such that most of these products of smaller intervals are "homogeneous" for the definition of ν.

2. The extra factor
$$\prod_{p \,|\, \Delta} (1 + \mathrm{O}_m(p^{-1/2}))$$
arises naturally in the course of the calculations, and sheds a little light on the correlation condition itself. The correlation condition takes the form it does because the quantity
$$\mathbb{E}\big(\Lambda_R(W(x+h_1)+1)^2 \cdots \Lambda_R(W(x+h_m)+1)^2 \;\big|\; x \in B\big)$$

does not admit a uniform bound as a function of $(h_1, \ldots, h_m)$, but we *can* show that those tuples $(h_1, \ldots, h_m)$ for which its value is large are very few (notice that, typically, most p dividing Δ will be large, and so the product will usually *not* be too large). Recalling our view of the Green-Tao proof as split into two stages, we might say that the definition of the correlation condition is a staging post that has been placed precisely so that both halves of the journey are just possible.

We finish with a representative sketch of some of the proofs relating the above results.

Sketch proof of the linear forms condition from Proposition 3 Let ψ_j, $1 \leq j \leq m$, be our linear forms. Following the first remark above, we chop up the range of summation in the linear forms condition into $Q(N)^t$ almost equal-sized boxes, where Q is slowly growing in N. We now treat the sums over the different sets

$$B_{u_1,\ldots,u_t} := \{\mathbf{x} : \ x_k \in [\lfloor u_k N/Q \rfloor, \lfloor (u_k+1)N/Q \rfloor) \text{ for } k = 1, \ldots, t\}$$

for $u_1, \ldots, u_t \in \mathbb{Z}_Q$, and observe that asymptotically we may replace the expectation $\mathbb{E}(\,\cdot\,|\,\mathbf{x} \in \mathbb{Z}_N^t)$ with the iterated expectation

$$\mathbb{E}\big(\mathbb{E}(\,\cdot\,|\,\mathbf{x} \in B_{u_1,\ldots,u_t})|\, u_1, \ldots, u_t \in \mathbb{Z}_Q\big)$$

in the linear forms condition.

Call a t-tuple $(u_1, \ldots, u_t) \in \mathbb{Z}_Q^t$ **nice** if for every $1 \leq j \leq m$ the sets $\psi_j(B_{u_1,\ldots,u_t})$ are either completely contained in the interval $[\varepsilon_k N, 2\varepsilon_k N]$ or are completely disjoint from it. Then in these cases we can either replace every $\nu(\psi_j(\mathbf{x}))$ factor by $\frac{\phi(W)}{W \log R}\Lambda_R(\theta_j(\mathbf{x}))^2$, and apply Proposition 3 (note that if Q grows slowly enough then for N sufficiently large each $B_{u_1,\ldots,u_t}$ is large enough), or replace every such factor by 1. Hence we always obtain $1 + \mathrm{o}_{m,t}(1)$ for the inner of our iterated expectations when $(u_1, \ldots, u_t)$ is nice.

Finally, we prove by elementary estimates that the proportion of non-nice t-tuples is asymptotically zero as $N \to \infty$. $\square$

Sketch of proof method for Propositions 3 and 4 Adapting the methods of [1], these are proved by representing the summation in question as an iterated contour integral by applying the standard identity

$$\log^+ y = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma_1} \frac{y^z}{z^2}\, \mathrm{d}z,$$

where Γ_1 is the vertical contour $\Gamma_1(t) = 1 + \mathrm{i}t$, to the $\log(R/d)$ factors in the definition of $\widetilde{\Lambda}$. The integrand is then massaged into the form of a multiplicative function, which can then in turn be decomposed as

$$\big(\text{something relatively benign}\big) \times \big(\text{some product of zeta functions}\big).$$

The values of the contour integrals then depend most strongly on the second factor; since the zeta function is relatively well-understood, methods of the residue calculus can then be used to estimate the contour integrals, with small errors that can be tracked as the contours are moved around. The trick is to find estimates on the behaviour of the first factor that are good enough to allow the application of such contour integral methods. $\square$

Remark Bernard Host [3] has offered an alternative construction for ν. His follows the pattern of the Green-Tao definition except that he takes

$$\widetilde{\Lambda}(n) := \sum_{d \mid n} \mu(d) \chi\Big(\frac{\log d}{\log R}\Big)$$

for a C^∞ function $\chi : \mathbb{R} \to [0, \infty)$ satisfying certain extra conditions.

This allows him to write

$$\chi(x) = \int_{\mathbb{R}} \tau(t) \mathrm{e}^{-x(1+\mathrm{i}t)} \, \mathrm{d}t$$

for a rapidly decreasing function τ. He then mimics the proofs of Propositions 9.5 and 9.6 sketched above, except the use of the contour integrals is replaced by a (shorter) estimate using Fourier analysis.

References

[1] Goldston D.A. & Yıldırım C.Y., "Small gaps between primes I", preprint;

[2] Green B. & Tao T., "The primes contain arbitrarily long arithmetic progressions", to appear in Ann. Math.;

[3] Host B., "Progressions arithmétiques dans les nombres premiers [d'après B.Green et T.Tao]", Séminaire BOURBAKI, Mars 2005, 57ème année, 2004-2005, no. 944.

Tim Austin, UCLA
email: **timaustin@math.ucla.edu**

2 A generalization of Birkhoff's pointwise ergodic theorem

after Amos Nevo and Elias M. Stein [4]
A summary written by Svetlana Butler

Abstract

The paper generalizes Birkhoff's pointwise ergodic theorem by presenting ergodic theorems for free non-Abelian groups on finitely many generators.

2.1 Introduction

Given an arbitrary invertible measure preserving transformation T on a probability space X, Birkhoff's pointwise ergodic theorem says that for any $f \in L^1(X)$, the averages of f along the orbit of T

$$\frac{1}{2n+1}\sum_{k=-n}^{n} f(T^k x)$$

converge to the limit $\tilde{f}(x)$ for almost all $x \in X$, where $\tilde{f}$ is the conditional expectation of f with respect to the σ-algebra of T-invariant sets.

Consider two invertible, measure preserving commutative transformations T and S. It is known that the expressions

$$\frac{1}{(2n+1)^2}\sum_{-n\le n_1,n_2\le n} f(T^{n_1}S^{n_2}x)$$

converge for almost all $x \in X$, for any $f \in L^1(X)$, and the limit is the conditional expectation of f with respect to the σ-algebra of sets invariant under T and S. Thus the pointwise ergodic theorem holds for finite measure preserving actions of the of the free Abelian group on two generators, $\mathbb{Z}^2$.

It is natural to try to consider a pointwise ergodic theorem for measure preserving actions of the free non-Abelian group on 2 (or r) generators. This leads to the following setting:

For a countable group Γ let $l^1(\Gamma) = \{\mu = \sum_{\gamma\in\Gamma}\mu(\gamma)\gamma : \sum_{\gamma\in\Gamma}|\mu(\gamma| < \infty\}$ denote the group algebra. Let $(X, \mathcal{B}, m)$be a standard Lebesque probability

space, and assume Γ acts on X by measurable automorphisms preserving m. The action $(\gamma, x) \mapsto \gamma x$ induces a representation of Γ by isometries on the $L^p(X)$ spaces, $1 \leq p \leq \infty$, and this represenatation can be extended to the group algebra by $(\mu f)(x) = \sum_{\gamma \in \Gamma} \mu(\gamma) f(\gamma^{-1} x)$.

Let $\mathcal{B}_1 = \{A \in \mathcal{B} : m(\gamma A \triangle A) = 0 \ \ \forall \gamma \in \Gamma\}$ denote the sub-σ-algebra of invariant sets, and denote by E_1 the conditional expectation operator on $L^1(X)$ which is associated with $\mathcal{B}_1$.

A sequence $\nu_n \in l^1(\Gamma)$ is called *a pointwise ergodic sequence in L^p* if, for any action of Γ on a Lebesque space X which preserves a probability measure, and for every $f \in L^p(X)$, $\nu_n f(x) \to E_1 f(x)$ for almost all $x \in X$, and in the norm of $L^p(X)$.

In search for pointwise ergodic sequences we look at sequences in $l^1(\Gamma)$ with an explicit geometric form. Assume Γ is finitely generated, and let S be a finite generating set which is symmetric: $S = S^{-1}$ (we will assume $e \notin S$). When Γ is the free group F_r, the set of generators S will always be taken to be the set of free generators and their inverses. S induces a length function on Γ, given by $|\gamma| = min\{n : \gamma = s_1 \dots s_n, \ s_i \in S\}$, $|e| = 0$. The sphere and the ball of radius n are $S_n = \{w : |w| = n\}$ and $B_n = \{w : |w| \leq n\}$ respectively.

2.2 Main results

Consider the following sequences:

$$\sigma_n = \frac{1}{\sharp S_n} \sum_{w \in S_n} w \qquad \mu_n = \frac{1}{n+1} \sum_{k=0}^{n} \sigma_k$$

$$\sigma_n^{'} = \frac{1}{2}(\sigma_n + \sigma_{n+1}) \qquad \beta_n = \frac{1}{\sharp B_n} \sum_{w \in B_n} w$$

Theorem 1. *Consider the free group F_r, $r \geq 2$.*

1. *The sequence μ_n is a pointwise ergodic sequence in L^p, for $1 \leq p < \infty$.*
2. *The sequence $\sigma_n^{'}$ is a pointwise ergodic sequence in L^p, for $1 < p < \infty$.*
3. *σ_{2n} converges to an operator of conditional expectation with respect to an F_r-invariant sub-σ-algebra. β_{2n} converges to the operator $E_1 + \frac{r-1}{r} E$, where E is a projection disjoint from E_1. Given $f \in L^p(X)$, $1 < p < \infty$, the convergence is pointwise almost everywhere and in the L^p-norm.*

Note that the first part of this theorem is a direct analog of Birkhoff's pointwise ergodic theorem.

Given a sequence $\nu_n \in l^1(\Gamma)$, define the associated maximal function $f^*_\nu(x) = \sup_{n\geq 0} |\nu_n f(x)|$. Let $(X, \mathcal{B}, m)$ be an F_r-space with an invariant σ-finite measure. Then:

Theorem 2. *For each F_r, $r \geq 2$, there exist positive constants $C_p(r)$ such that for any $f \in L^p(X)$ the following inequalities hold:*

1. *$\| f^*_\mu \|_p$, $\| f^*_\sigma \|_p$, and $\| f^*_\beta \|_p$ are all bounded by $C_p(r) \| f \|_p$, for $1 < p < \infty$.*

2. *f^*_μ satisfies the maximal inequality of weak type $(1,1)$, namely:*

$$m\{x : |f^*_\mu(x)| \geq \delta\} \leq C_1(r)\delta^{-1} \| f \|_1$$

for every $\delta > 0$, and $f \in L^1(X)$.

2.3 Remarks

In pointwise ergodic theorems one seeks to establish, for a general sequence of Markov operators T_k acting in $L^p(X, \mathcal{B}, m)$, the existence of the limit $\lim_{k\to\infty} T_k f(x) = \tilde{f}(x)$ for almost all x and in the L^p norm. A natural and interesting choice is to consider the sequence of operators β_k of averaging a function on balls B_k of radius k. In the case of a finitely generated Abelian group, the main ingredients which figure in the proof of ergodic theorems are: the fact that the nested sequence of balls B_n is asymptotically invariant under translation; the transfer principle (which uses the fact that $\lim\limits_{N\to\infty} \dfrac{\sharp B_N}{\sharp B_{N-n}} = 1$); and the covering argument. None of these ingredients works if the group is a finitely generated free non-Abelian group F_r. For example, $\lim\limits_{N\to\infty} \dfrac{\sharp B_N}{\sharp B_{N-n}} = (2r-1)^n$, so the transfer principle does not apply.

In proving ergodic theorems for free non-abelian groups on r generators different methods are necessary. The following two observations lead to special methods used to prove the ergodic theorems in the paper.

(I) The Markov operators β_k and σ_k are comparable, in the sense that a maximal inequality for one sequence implies the same inequality for the other, up to a constant. (This fact follows since $\sigma_n \leq C_r\beta_n$ and β_n is a convex combination of σ_k, $k = 0, \ldots, n$.) This means that one might use

spheres rather than balls. The approach in the paper is to use methods similar to ones devised in [5] to handle singular means.

(II) The convolution identity

$$\sigma_1 * \sigma_n = \frac{1}{2r}\sigma_{n-1} + \left(1 - \frac{1}{2r}\right)\sigma_{n+1} \tag{1}$$

holds in the group algebra $l^1(F_r)$. It implies, by induction, that the elements σ_n are linear combinations of the convolution powers σ^k, $0 \le k \le n$. Therefore, the spheres σ_n generate a commutative convolution $*$-algebra, denoted by $A(F_r)$. The algebra $A(F_r)$, introduced in 1963 (see [1]), has an explicit spectral theory (see, for example, [3]). This spectral theory is important in proofs of both theorems.

2.4 Outline of the proofs

We outline the proof of Theorem 2 first.

Part (2.) of Theorem 2 as well as the strong L^p maximal inequality for the maximal function f^*_μ (which is Lemma 1 in [4]) follows from the inequality

$$\sum_{k=0}^{N} \sigma_k \le C_r \sum_{k=0}^{3N} \sigma_1^k$$

and application of the Hopf-Dunford-Schwartz theorem ([2], Ch.8). The above inequality is proved by using approximation of the binomial coefficients to estimate coefficients $a_n(k)$ in $\sigma_1^n = \sum_{k=0}^{n} a_n(k)\sigma_k$. Here writing σ_1^n as a convex combination of σ_k's, $0 \le k \le n$, follows from the convolution identity (1).

The proof of part (1.) is more elaborate. First observe that maximal inequalities for f^*_μ and f^*_β follow from the maximal inequality for f^*_σ, since μ_n and β_n are convex averages of σ_k's, $0 \le k \le n$. The proof of the maximal inequality for f^*_σ follows the method devised in [5].

For a sequence P_k of bounded linear operators on $L^p(X)$ and a function $f \in L^p(X)$ define a sequence of Cesaro sums:

$$S_n^\lambda f(x) = \sum_{k=0}^{n} A_{n-k}^\lambda P_k f(x)$$

where $A_n^\lambda = \prod_{k=1}^{n}(1+\frac{\lambda}{k})$ are complex binomial coefficients. The asssociated maximal functions are

$$S_*^\lambda f(x) = \sup_{n\geq 0}\left|\frac{S_n^\lambda f(x)}{(n+1)^{\lambda+1}}\right|$$

Notice that when $P_k = \sigma_k$, we get $S_*^0 f(x) = f_\mu^*(x)$, while the singular mean $S_*^{-1}f(x) = f_\sigma^*(x)$. When $P_k = \sigma_{2k}$, the singular mean $S_*^{-1}f(x) = \sup_{n\geq 0}\sigma_{2n}f(x)$.

Step 1. Establish the maximal L^p inequality

$$\| S_*^{\alpha+i\beta} f \|_p \leq C_\alpha exp(2\beta^2) \| S_*^0|f| \|_p, \quad 1\leq p<\infty, \quad \alpha>0$$

This is Lemma 4 (with $P_k = \sigma_k$) in [4] which is proved by using some estimates and the convolution formula for complex binomial coefficients.

Step 2. Prove the L^2 maximal inequality

$$\| S_*^{-m+i\beta} f \|_2 \leq C_m exp(3\beta^2) \| f \|_2$$

for every nonpositive integer $-m$ and $\beta\in\mathbb{R}$. Here we use $P_k = \sigma_{2k}$.

It is enough to enough to show the L^2 maximal inequality for S_*^{-m}. This inequality is proved using the Littlewood-Paley square-function method. The spectral theory for the algebra $A(F_r)$ plays a key role in estimating the Littlewood-Paley square-function.

Step 3. Use maximal inequalities from Step 1 and Step 2 and apply an analytic interpolation theorem ([6]). The result is a maximal inequality for the singular mean $S_*^{-1}f(x) = \sup_{n\geq 0}\sigma_{2n}f(x)$ in every L^p, $p>1$. This implies the maximal inequality for f_σ^* in every L^p, $p>1$, since for $f\geq 0$

$$\frac{1}{2r}\sigma_{2n-1}f(x) \leq \sigma_{2n} * \sigma_1 f(x)$$

which follows from the convolution identity (1).

The proof of Theorem 1 is based on strong maximal L^p inequalities from Theorem 2 and the spectral theory of $A(F_r)$. It is shown first (using the spectral theory) that the pointwise limits of Theorem 1 are valid for functions in $L^2(X)$, in particular, for a dense set of bounded functions. The proof of the rest of Theorem 1 is a standard argument with an application of strong maximal inequalities from Theorem 2.

References

[1] Arnold, V. I. and Krylov, A. L., *Uniform distribution of points on a sphere and some ergodic properties of solutions of linear ordinary differential equations in the complex plane.* Soviet Math. Dokl., 4 (1962), 1-5.

[2] Dunford, N. and Schwartz, J. T., *Linear operators, vol. 1.* Interscience, New York, 1963.

[3] Figa-Talamanca, A. and Picardello, M. A., *Harmonic analysis on free groups.* Lecture Notes in Pure and Appl. Math., 87. Dekker, New York, 1983.

[4] Nevo, A. and Stein, E. M., *A generalization of Birkhoff's pointwise ergodic theorem.* Acta Math., 173 (1994), 135-154.

[5] Stein, E. M., *On the maximal ergodic theorem.* Proc. Nat. Acad. Sci. U.S.A., 47 (1961), 1894-1897.

[6] – *Interpolation of linear operators.* Trans. Amer. Math. Soc., 83 (1956), 482-492.

Svetlana Butler, UIUC
email: **svbutler@uiuc.edu**

3 A pointwise ergodic theorem for amenable groups

after E. Lindenstrauss [1]
A summary written by S. Zubin Gautam

Abstract

We summarize the proof of the L^1-pointwise ergodic theorem for actions of amenable groups in [1], focusing on the discrete case. We then present some results on the lamplighter group $\mathbb{Z}_2 \wr \mathbb{Z}$ as a demonstration of the scope of the main theorem.

3.1 Preliminaries

In the sequel, all groups are assumed to be locally compact and second countable; $|\cdot|$ and m_L will both denote the left Haar measure on a group G. $(X, \mathcal{M}, \mu)$ will denote a Lebesgue probability space. We begin by recalling the Følner characterization of amenability:

Definition 1. *A locally compact group G is* amenable *if for any compact set $K \subset G$ and any $\delta > 0$ there is another compact set $F \subset G$ such that*

$$|F \triangle KF| < \delta |F|.$$

Such an F is called (K, δ)-invariant.

Equivalently, G is amenable if it has a "Følner sequence":

Definition 2. *A* Følner sequence *in a group G is a sequence of compact subsets $F_n \subset G$ such that for every compact $K \subset G$ and $\delta > 0$, F_n is (K, δ)-invariant for all sufficiently large n.*

Now consider a left action $G \curvearrowright X$ by measure-preserving transformations with G amenable. For $f : X \to \mathbb{R}$ measurable, we consider the averages of f over sets $F \subset G$,

$$A(F, f)(x) = \frac{1}{|F|} \int_F f(gx) \, \mathrm{d}m_L(g).$$

Results on the pointwise convergence of these averages along suitable Følner sequences provide natural generalizations of the classical Birkhoff ergodic theorem for $\mathbb{Z}$-actions; however, care must be exercised in the choice of Følner sequence (even in the classical case).

Definition 3. *A sequence of subsets $F_n \subset G$* is tempered *(or* satisfies the Shulman condition*) if there is some $C > 0$ such that for all n*

$$\left| \bigcup_{k<n} F_k^{-1} F_n \right| \leq C|F_n|. \tag{1}$$

It is not too hard to show that *any* Følner sequence has a tempered subsequence; hence, any convergence result along tempered Følner sequences will apply to general amenable groups.

3.2 Statement of the main result

Theorem 4 (Pointwise Ergodic Theorem)**.** *Let G be an amenable group acting on $(X, \mathcal{M}, \mu)$ by measure-preserving transformations, and let $\{F_n\}$ be a tempered Følner sequence for G. Then for any $f \in L^1(\mu)$ there is a G-invariant $\bar{f} \in L^1(\mu)$ such that*

$$\lim_{n\to\infty} A(F_n, f)(x) = \bar{f}(x) \quad a.e.$$

In particular, if G acts ergodically,

$$\lim_{n\to\infty} A(F_n, f)(x) = \int_X f \, \mathrm{d}\mu \quad a.e.$$

(NB: This is both Theorem 1.2 and Theorem 3.3 of [1].)

3.3 Random selection of Følner sets

We will use the following "randomized" covering lemma, Lemma 2.1 of [1]:

Lemma 5. *Let $\delta > 0$ be given, and let G be a locally compact second countable group. Let $F_1, \ldots, F_N$ be a finite tempered sequence of compact subsets of G with constant $C > 0$, and let F be another compact subset of G. Take arbitrary sets A_j such that $F_j A_j \subset F$ for $1 \leq j \leq N$, and set*

$$\bar{\mathcal{F}} = \{F_j\, a \mid a \in A_j,\ 1 \leq j \leq N\}.$$

Then there is a probability space $(\Omega, \mathbb{P})$ and a map $\vec{\omega} \mapsto \mathcal{F}(\vec{\omega})$ from Ω to the set of subcollections of $\bar{\mathcal{F}}$ satisfying the following conditions:

1. *$\mathcal{F}(\vec{\omega}\,)$ is a.s. finite and the counting function*

$$\Lambda^{\vec{\omega}}(g) = \sum_{B \in \mathcal{F}(\vec{\omega}\,)} \mathbf{1}_B(g)$$

of $\mathcal{F}(\vec{\omega}\,)$ is a measurable function on $\Omega \times F$.

2. *For all $g \in F$,*

$$\mathbb{E}(\Lambda^{\vec{\omega}}(g) \,|\, \Lambda^{\vec{\omega}}(g) \geq 1) \;\leq\; 1 + \delta.$$

3. *With the notation $\|\,\mathcal{S}\| = \sum_{B \in \mathcal{S}} |B|$ for a collection of sets $\mathcal{S}$, we have*

$$\mathbb{E}(\|\mathcal{F}(\vec{\omega}\,)\|) = \mathbb{E}\bigg(\int_F \Lambda^{\vec{\omega}}(g)\, \mathrm{d}m_L(g) \bigg) \geq \gamma(\delta, C) \bigg| \bigcup_{j=1}^{N} A_j \bigg|,$$

where $\gamma(\delta, C) = \delta/(1 + C\delta)$.

(We use probabilistic notation for $(\Omega, \mathbb{P})$; $\mathbb{E}$ denotes expectation or conditional expectation. $\mathbb{E}(\cdot \,|\, A)$ is the value taken on A by the conditional expectation with respect to the sub-σ-algebra $\{\emptyset, A, A^c, \Omega\}$.)

Intuitively, condition 2 of the lemma says that "on average" the subcollections $\mathcal{F}(\vec{\omega}\,)$ are almost disjoint, while condition 3 guarantees that a typical $\mathcal{F}(\vec{\omega}\,)$ will cover a large enough portion of F.

3.3.1 Proof sketch of Lemma 5 for G discrete

For clarity, it is instructive first to consider the case where G is a discrete group. Take the sample space Ω to be

$$\Omega = \big\{ \vec{\omega} = \{\omega(j,a)\}_{1 \leq j \leq N,\, a \in G} \mid \omega(j,a) \in \{0,1\} \;\forall\, j, a \big\}.$$

(One may think of Ω as a coin-flipping experiment indexed by N copies of G.) Let $\mathbb{P}$ be the probability measure on Ω such that

$$\mathbb{P}\big(\{\vec{\omega} \mid \omega(j_0, a_0) = 1\}\big) = \frac{\delta}{|F_{j_0}|} =: p_{j_0}$$

for all j_0, a_0 (taking the σ-algebra generated by all $\omega(j,a)$ as the domain); this makes each $\omega(j,a)$ an independent $\{0,1\}$-valued random variable on $(\Omega, \mathbb{P})$ that is 1 with probability p_j.

Given $\vec{\omega} = \{\omega(j,a)\} \in \Omega$, $\mathcal{F}(\vec{\omega}\,)$ is given by the following N-step recursive algorithm:

Algorithm A:

1. Begin by setting $j = N$, and define sets $A_{i|N+1} = A_i$ for $1 \le i \le N$.
2. Set $\Sigma_j = \{a \in A_{j|j+1} \mid \omega(j,a) = 1\}$ and $\mathcal{F}_j(\vec{\omega}) = \{F_j a \mid a \in \Sigma_j\}$.
3. For all $i < j$, define $A_{i|j} = \{a \in A_{i|j+1} \mid F_i a \cap F_j \Sigma_j = \emptyset\}$.
4. If $j > 1$, return to step 2 and replace j with $j-1$; if j=1, proceed to step 5.
5. Set $\mathcal{F}(\vec{\omega}\,) = \bigcup_{j=1}^{N} \mathcal{F}_j(\vec{\omega})$.

The sets $A_{i|j} = A_{i|j}(\vec{\omega})$, $1 \le i < j \le N$ should be regarded as encoding the "admissible" translates of F_i at stage j of the recursion; their definition guarantees that the collections $\mathcal{F}_j(\vec{\omega})$ are mutually pairwise disjoint. Note that each $A_{i|j}$ only depends on the σ-algebra Φ_j generated by the random variables $\omega(j',a')$ for $j' \ge j$ and $a' \in G$. Similarly, the algorithm enjoys a useful recursive property: Once $\mathcal{F}_N(\vec{\omega})$ (or equivalently Φ_N) is given, the random collection $\bigcup_1^{N-1} \mathcal{F}_j(\vec{\omega})$ has the same distribution (i.e. pushforward measure) as that obtained by running the original algorithm on the sequence $\{F_j\}_1^{N-1}$ and the sets $A_{j|N}$, $1 \le j \le N-1$.

Condition 1 of the lemma is easily verified. To prove condition 2, we write $\Lambda^{\vec{\omega}}(g) = \sum_1^N \Lambda_j^{\vec{\omega}}(g)$, where

$$\Lambda_j^{\vec{\omega}}(g) = \sum_{B \in \mathcal{F}_j(\vec{\omega})} \mathbf{1}_B(g)$$

is the counting function of $\mathcal{F}_j(\vec{\omega})$. Now since the collections $\mathcal{F}_j(\vec{\omega})$ are mutually disjoint, it suffices to show that $\mathbb{E}(\Lambda_j^{\vec{\omega}}(g) \,|\, \Lambda_j^{\vec{\omega}}(g) \ge 1) \le 1 + \delta$ for all j. But, viewing $\Lambda_j^{\vec{\omega}}(g)$ as a random variable on Ω, once Φ_{j+1} is given, we can write

$$\Lambda_j^{\vec{\omega}}(g) = \sum_{a \in A_{j|j+1} \cap F_j^{-1} g} \omega(j,a),$$

which is a sum of fewer than $|F_j|$ i.i.d. $\{0,1\}$-valued random variables that are 1 with probability p_j. From here calculation yields

$$\begin{aligned} \mathbb{E}(\Lambda_j^{\vec{\omega}}(g) \,|\, \Lambda_j^{\vec{\omega}}(g) \ge 1) &= \mathbb{E}\big(\, \mathbb{E}(\Lambda_j^{\vec{\omega}}(g) \,|\, \Lambda_j^{\vec{\omega}}(g) \ge 1; \Phi_{j+1}) \,\big|\, \Lambda_j^{\vec{\omega}}(g) \ge 1 \,\big) \\ &\le 1 + p_j |F_j| \;=\; 1 + \delta. \end{aligned}$$

Condition 3 is proved by induction on N; it is easy for $N = 1$. By the recursive property of Algorithm A mentioned above, if the statement holds for $N-1$ we obtain

$$\begin{aligned}
\mathbb{E}\big(\|\mathcal{F}(\vec{\omega}\,)\| \,\big|\, \Phi_N\big) &\geq \|\mathcal{F}_N(\vec{\omega})\| + \gamma(\delta, C)\Big|\bigcup_{j=1}^{N-1} A_{j|N}(\vec{\omega})\Big| \\
&\geq \|\mathcal{F}_N(\vec{\omega})\| + \gamma(\delta, C)\left(\Big|\bigcup_{j=1}^{N-1} A_j\Big| - C|F_N| \cdot |\Sigma_N(\vec{\omega})|\right) \\
&= \|\mathcal{F}_N(\vec{\omega})\| + \gamma(\delta, C)\left(\Big|\bigcup_{j=1}^{N-1} A_j\Big| - C\|\mathcal{F}_n(\vec{\omega})\|\right),
\end{aligned}$$

where the second inequality follows from the Shulman condition and the fact that

$$\bigcup_{1}^{N-1} A_{j|N} \supset \bigcup_{1}^{N-1} A_j \setminus \left(\bigcup_{1}^{N-1} F_j^{-1} F_N\right)\Sigma_N.$$

Taking the expectation of both sides and computing $\mathbb{E}\big(\|\mathcal{F}_N(\vec{\omega})\|\big) = \delta|A_N|$ yields the claim.

3.3.2 Proof sketch of Lemma 5 for general G

For G a general locally compact, second countable group, we take

$$\Omega = \big\{\vec{\omega} = \{\Gamma_j\}_{j=1}^N \,|\, \Gamma_j \subset G \text{ locally finite } \forall\, j\big\},$$

where the random subsets Γ_j are chosen independently according to Poisson processes on G with respect to the rescaled *right* Haar measures $\frac{\delta}{|F_j|}\,\mathrm{d}m_R$ (this reduces to our original $(\Omega, \mathbb{P})$ when G is discrete). We then make the natural modification of step 2 in Algorithm A:

2'. Set $\Sigma_j = \Gamma_j \cap A_{j|j+1}$ and $\mathcal{F}_j(\vec{\omega}) = \{F_j\, a \,|\, a \in \Sigma_j\}$.

Once we have constructed Ω and the map $\mathcal{F}$, the proof proceeds much as in the discrete case, up to some minor modifications involving the basic properties of a Poisson process and the relationship between the left and right Haar measures on G. NB: Amenability is *not* necessary for Lemma 5.

3.4 Maximal inequality and pointwise ergodic theorem

Definition 6. *Given a sequence F_n of finite-measure subsets of G, the* maximal function *of $f \in L^1(G)$ is*

$$M[f](x) = \sup_n A(F_n, f)(x).$$

Theorem 7 (Weak (1,1) maximal inequality)**.** *Let $G \curvearrowright (X, \mathcal{M}, \mu)$ as above with G amenable, and let $F_n \subset G$ be a tempered sequence of compact sets. Then the associated maximal operator M is weak-type $(1,1)$; i.e., there exists $c > 0$ such that*

$$\mu\{M[f](x) > \lambda\} \leq \frac{c}{\lambda}\|f\|_1.$$

The constant c depends on $\{F_n\}$ but not on X.

Proof sketch: Fix n and $\varepsilon < 0$. Set $D_n = \{x \mid \max_{1\leq j\leq n} A(F_j, f)(x) > \lambda\}$. By the amenability of G, we can choose some large compact set $F' \subset G$ such that $F := \bigcup_1^n F_j F'$ has measure $|F| \leq (1+\varepsilon)|F'|$. The first step is to prove the "group-side maximal inequality"

$$\int_{F'} \mathbf{1}_{D_n}(gx)\,\mathrm{d}m_L(g) \leq \frac{c}{\lambda}\int_F |f(gx)|\,\mathrm{d}m_L(g) \tag{2}$$

for all $x \in X$ and $c = 2/\gamma(1, C)$, with C as in the Shulman condition (1).

Fix x and set $A_j = \{g \in F' \mid A(F_j\, g, f)(x) > \lambda\}$ for $1 \leq j \leq n$. We will apply Lemma 5 to these sets with $\delta = 1$. Let $\mathcal{F}(\vec{\omega}\,)$ be a random collection of subsets of F as in the lemma with counting function $\Lambda^{\vec{\omega}}(g)$. Firstly, the definition of the A_j implies the inequality

$$\lambda|F_j\, a| \leq \int_{F_j\, a} f(gx)\,\mathrm{d}m_L(g)$$

for all $a \in A_j$. Combining this with condition 3 of Lemma 5 yields

$$\begin{aligned}
\lambda\gamma(1,C)\int_{F'} \mathbf{1}_{D_n}(gx)\,\mathrm{d}m_L(g) &= \lambda\gamma(1,C)\left|\bigcup_{j=1}^n A_j\right| \\
&\leq \mathbb{E}\big(\|\mathcal{F}(\vec{\omega}\,)\|\big) \\
&\leq \mathbb{E}\bigg(\int_F \Lambda^{\vec{\omega}}(g) f(gx)\mathrm{d}m_L(g)\bigg).
\end{aligned} \tag{3}$$

But condition 2 of Lemma 5 can be used to get

$$\mathbb{E}\left(\int_F \Lambda^{\vec{\omega}}(g) f(gx) \mathrm{d}m_L(g)\right) \leq 2 \int_F |f(gx)| \, \mathrm{d}m_L(g). \tag{4}$$

Combining (3) and (4) yields (2), and from there integration yields

$$\mu(D_n) \leq \frac{c}{\lambda}(1+\varepsilon) \int_X |f(x)| \, \mathrm{d}\mu(x).$$

Letting $n \to \infty$ and $\varepsilon \to 0$ yields the maximal inequality. $\square$

The main theorem 4 now follows from the maximal inequality along the standard lines (see e.g. Chapters 5 and 6 of [3] or pp. 92-93 of [2]). It is worth noting, however, that the amenability of G comes into play again when showing that the theorem holds for the coboundaries $h(gx) - h(x)$ for $g \in G$ and $h \in L^\infty(G)$; we need to use the fact that F_n is a Følner sequence.

3.5 Example: The lamplighter group

The *lamplighter group* $\tilde{G}$ is the wreath product

$$\tilde{G} = \mathbb{Z}_2 \wr \mathbb{Z} := \left(\bigoplus_{i \in \mathbb{Z}} \mathbb{Z}_2\right) \rtimes_\sigma \mathbb{Z},$$

where σ is the right shift action of $\mathbb{Z}$ on $\bigoplus \mathbb{Z}_2$. $\tilde{G}$ is well-known to be an amenable group of exponential growth.

Theorem 8. *For all $C > 0$ there are a finite $K \subset \tilde{G}$ and $\delta > 0$ such that if A is (K, δ)-invariant and $C^{-1}|A| \leq |B| \leq |A|$, then $|B^{-1}A| \geq C|A|$.*

We omit the proof of this theorem, from which we immediately obtain:

Corollary 9. *The lamplighter group has no Følner sequences satisfying the* Tempelman condition

$$|F_n^{-1} F_n| \lesssim |F_n|.$$

Corollary 10. *If F_n is a tempered Følner sequence for the lamplighter group,*

$$\frac{|F_{n+1}|}{|F_n|} \to \infty.$$

In particular, the growth of $|F_n|$ is super-exponential.

Corollary 9 reveals the limited scope of earlier ergodic theorems along similar lines that require the Tempelman condition (see e.g. section 6.3 of [3]). However, Corollary 10 shows that Theorem 4 only guarantees the convergence of ergodic averages taken along a potentially "sparsely spaced" sequence of Følner sets; moreover, it is conjectured that the corollary should hold for *any* group of exponential growth.

References

[1] Lindenstrauss, E. 2001. Pointwise theorems for amenable groups. Invent. Math. **146**, no. 2, 259-295.

[2] Petersen, K. 1983. *Ergodic theory.* Cambridge, UK: Cambridge UP.

[3] Tempelman, A. 1992. *Ergodic theorems for group actions: informational and thermodynamical aspects.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

S. Zubin Gautam, UCLA
email: **sgautam@math.ucla.edu**

4 Arithmetic progressions in primes I

after B. Green and T. Tao [3]
A summary written by Alexander Gorodnik[1]

4.1 Introduction

This chapter contains the first part of an exposition of

Theorem 1 (Green,Tao [3]). *Let A be a subset of positive upper density in the set $\mathcal{P}$ of prime numbers, i.e.,*

$$\limsup_{N\to\infty} \frac{|A \cap [1,N]|}{|\mathcal{P} \cap [1,N]|} > 0.$$

Then for any $k \geq 3$, A contains an arithmetic progression of length k.

We start our discussion with the celebrated theorem of Szemerédi, which can be stated in several equivalent forms (see Section 4.1.1 below for basic notation):

Theorem 2 (Szemerédi). *1. Let A be a set of positive integers such that*

$$\limsup_{N\to\infty} \frac{1}{N}|A \cap [1,N]| > 0.$$

Then for any $k \geq 3$, A contains an arithmetic progression of length k.

2. Given $k \geq 3$ and $\delta > 0$, there exists $N_0 = N_0(k,\delta) > 0$ such that for every $N > N_0$ and every $A \subset [1,N]$ with $|A| > \delta N$, the set A contains an arithmetic progression of length k.

3. Given $k \geq 3$ and $\delta > 0$, there exists $c = c(k,\delta) > 0$ such that for a function $f : \mathbb{Z}_N \to \mathbb{R}$ satisfying $0 \leq f \leq 1$ and $\mathbb{E}(f) \geq \delta$, we have

$$\mathbb{E}(f(x)f(x+r)\cdots f(x+(k-1)r)|x,r \in \mathbb{Z}_N) \geq c$$

for sufficiently large N.

[1]The author is partially supported by NSF 0400631

Although there are explicit estimates on the constant N_0 in Theorem 2(2), the set $\mathcal{P}$ of prime numbers is too sparse to deduce existance of arithmetic progressions in $\mathcal{P}$ directly using known estimates. Recall that according to the prime number theorem,

$$|\mathcal{P} \cap [1, N]| \sim \frac{N}{\log N} \quad \text{as } N \to \infty.$$

The starting point of the Green-Tao argument is to consider $\mathcal{P}$ as a subset of the set $\mathcal{P}_R$ of almost primes, which consists of numbers all of whose prime factors are at least R. When $R = N^\alpha$ for small $\alpha > 0$, the set $\mathcal{P}_R$ is relatively well understood (for example, one can prove Theorem 1 for $\mathcal{P}_R$ using sieve methods), and by Mertens' theorem,

$$\limsup_{N\to\infty} \frac{|\mathcal{P} \cap [1, N]|}{|\mathcal{P}_R \cap [1, N]|} > 0.$$

The proof of Theorem 1 uses information about the structure of the set of almost primes ingeniously combined with positive density argument as in Theorem 2. In fact, it is proved that a subset of positive upper density in a so-called pseudorandom set, which will be defined in Section 4.1.2, contains arbitrary long arithmetic progressions.

Theorem 3 (relative Szemerédi theorem)**.** *Fix $k \geq 3$ and $\delta > 0$, and let $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ be a k-pseudorandom function such that $\mathbb{E}(\nu) = 1 + o(1)$. Then for any function $f : \mathbb{Z}_N \to \mathbb{R}^+$ satisfying $0 \leq f \leq \nu$ and $\mathbb{E}(f) \geq \delta$, we have*

$$\mathbb{E}(f(x)f(x+r)\cdots f(x+(k-1)r)|x, r \in \mathbb{Z}_N) \geq c - o_{k,\delta}(1) \tag{1}$$

as $N \to \infty$, where $c > 0$ is the same as in Theorem 2(3).

At this stage, we allow the reader to think naively that ν is the normalized characteristic function of the set of almost primes in $\mathbb{Z}_N$ and f is the normalized characteristic function of the set of primes in $\mathbb{Z}_N$. However, to give a rigorous argument, one needs to consider "smoothened" version of these function and to eliminate the irregularities coming for congruence properties of primes. The construction of functions ν and f for which Theorem 3 implies Theorem 1 will be given in the following chapter. In this chapter, we discuss the proof of Theorem 3.

Sketch of the proof of Theorem 3. The main ingredient of the proof is Theorem 8 below which implies that any nonnegative function f which is majorated by a k-pseudorandom ν has a decomposition

$$f = f_U + f_{U^\perp} + E \tag{2}$$

with the error term E satisfying

$$E \geq 0 \quad \text{and} \quad \mathbb{E}(E) = o(1). \tag{3}$$

Decomposition (2) is analogous to the ordinary Koopman–von-Neumann decomposition (see Remark 4), and it has the following properties:

$$f_U + f_{U^\perp} \geq 0, \tag{4}$$
$$0 \leq f_{U^\perp} \leq 1 + o(1), \tag{5}$$
$$\mathbb{E}(f) = \mathbb{E}(f_{U^\perp}) + o(1), \tag{6}$$
$$\mathbb{E}(f_0(x)f_1(x+r)\cdots f_{k-1}(x+(k-1)r)|x, r \in \mathbb{Z}_N) \quad \text{is small}, \tag{7}$$

where each f_i is either f_U or $f_{U^\perp}$, and $f_i \neq f_{U^\perp}$ for some i.

Now it follows from (3) and (4) that

$$\begin{aligned}&\mathbb{E}(f(x)\cdots f(x+(k-1)r)|x, r \in \mathbb{Z}_N)\\ \geq&\mathbb{E}((f_U + f_{U^\perp})(x)\cdots(f_U + f_{U^\perp})(x+(k-1)r)|x, r \in \mathbb{Z}_N).\end{aligned}$$

Because of (7), it suffices to prove a lower estimate for

$$\mathbb{E}(f_{U^\perp}(x)f_{U^\perp}(x+r)\cdots f_{U^\perp}(x+(k-1)r)|x, r \in \mathbb{Z}_N).$$

Using that the function $f_{U^\perp}$ satisfies (5) and (6), this lower estimate follows from the Szemerédi theorem (Theorem 2(3)). □

Remark 4 (comparison with ergodic theory)**.** When $k = 3$, decomposition (2) is a finitary analog of the Koopman–von-Neumann decomposition of $L^2(X)$, defined for a probability measure-preserving system $(X, \mathcal{B}, \mu, T)$, into weakly mixing and compact (almost periodic) parts. Note that the later decomposition can be used to show that for any nonnegative $f \in L^\infty(X)$ with $\int_X f\, d\mu > 0$, one has

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \int_X f(x)f(T^i x)f(T^{2i}x)d\mu(x) > 0.$$

This implies the Szemerédi theorem (Theorem 2) with $k = 3$.

For general k, similar decompositions appeared in the Furstenberg's proof of the Szemerédi theorem [1] and in the works of Host, Kra [2] and Ziegler [4]. These decompositions are of the form

$$f = f_U + f_{U^\perp} \quad \text{with} \quad f_U = f - \mathbb{E}(f|\mathcal{B}') \ \text{ and } \ f_{U^\perp} = \mathbb{E}(f|\mathcal{B}'),$$

where $\mathcal{B}'$ is a T-invariant sub-σ-algebra of $\mathcal{B}$, such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \int_X f_0(x) f_1(T^i x) \cdots f_{k-1}(T^{(k-1)i} x)\, d\mu(x) = 0,$$

where each f_i is either f_U or $f_{U^\perp}$, and $f_i \neq f_{U^\perp}$ for some i, and the average

$$\frac{1}{N} \sum_{i=0}^{N-1} \int_X f_{U^\perp}(x) f_{U^\perp}(T^i x) \cdots f_{U^\perp}(T^{(k-1)i} x)\, d\mu(x)$$

can be analyzed using some additional structure.

4.1.1 Notation

- $k \geq 3$ is the length of the arithmetic progression (fixed).
- N is a large prime number ($N \to \infty$).
- $o(1)$ and $O(1)$ denote the quantities that go to zero and bounded respectively as $N \to \infty$.
- $\mathbb{Z}_N$ is the field of residues mod N.
- For a function $f : \mathbb{Z}_N^l \to \mathbb{R}$ and $A \in \mathbb{Z}_N^l$,

  $$\mathbb{E}(f(x) \,|\, x \in A) := \frac{1}{|A|} \sum_{x \in A} f(x) \quad \text{and} \quad \|f\|_{L^q} := \mathbb{E}(|f(x)|^q \,|\, x \in \mathbb{Z}_N^l)^{1/q}.$$

  In particular, $\mathbb{E}(f) := \mathbb{E}(f(x)|x \in \mathbb{Z}_N^l)$. Also, $\|\cdot\|_{L^\infty}$ denotes the sup-norm.

- Given a σ-algebra $\mathcal{B}$ on $\mathbb{Z}_N$, we define the conditional expectation $\mathbb{E}(f|\mathcal{B})$, $f \in L^2(\mathbb{Z}_N)$, to be the orthogonal projection on the space of $\mathcal{B}$-measurable functions, i.e.,

$$\mathbb{E}(f|\mathcal{B})(x) = \mathbb{E}(f(y) \,|\, y \in \mathcal{B}(x))$$

  where $\mathcal{B}(x)$ denotes the atom of the σ-algebra $\mathcal{B}$ containing x.

- For σ-algebras $\mathcal{B}_1$ and $\mathcal{B}_2$, $\mathcal{B}_1 \vee \mathcal{B}_2$ denotes the smallest σ-algebra containing both $\mathcal{B}_1$ and $\mathcal{B}_2$.

4.1.2 Pseudorandom measures

A function $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ is called a *measure* if

$$\mathbb{E}(\nu) = 1 + o(1).$$

A measure ν satisfies (m_0, t_0, L_0)-*linear form condition* if for any $m \leq m_0$, $t \leq t_0$, and linear forms

$$\psi_i(x) = \sum_{j=1}^{t} L_{ij} x_j + b_i, \quad i = 1, \ldots, m,$$

where L_{ij} are rational numbers[2] with the numerator and denominator bounded by L_0 in absolute value and $b \in \mathbb{Z}_N$, such that the t-tuples $(L_{ij}; j = 1, \ldots, t)$ are not zero and not rational multiples of each other, we have

$$\mathbb{E}(\nu(\psi_1(x)) \cdots \nu(\psi_m(x)) | x \in \mathbb{Z}_N^t) = 1 + o_{m_0, t_0, L_0}(1).$$

Roughly speaking, the measure ν will be supported on almost primes and the linear form condition says that events "$\psi_j(x)$ is almost prime" are essentially independent.

The linear form condition is motivated by the Hardy-Littlewood prime tuples conjecture. Let $\Lambda(n)$ denote the Mangoldt function, which is equal to $\log p$ if n is a power of prime p and zero otherwise, and $\Lambda_p(n)$ denote the local Mangoldt function, which is equal to $\frac{p}{p-1}$ if $(n, p) = 1$ and zero otherwise. Let ψ_i's be the linear forms as above with positive L_{ij} and b_i.

[2]Recalling that N is a large prime, we can view L_{ij}'s as elements of $\mathbb{Z}_N$.

Conjecture 1 (Hardy-Littlewood prime tuple conjecure)**.**

$$\mathbb{E}(\Lambda(\psi_1(x))\cdots\Lambda(\psi_m(x))|x\in\mathbb{Z}_N^t)=\prod_p \alpha_p + o_{m_0,t_0,L_0}(1)$$

where

$$\alpha_p=\mathbb{E}(\Lambda_p(\psi_1(x))\cdots\Lambda_p(\psi_m(x))|x\in\mathbb{Z}_p^t).$$

The linear form condition is analogous to the following property that holds for any weakly mixing probability measure-preserving system $(X,\mathcal{B},\mu,T)$: for any $f_1,\dots,f_m\in L^\infty(X)$,

$$\frac{1}{N}\sum_{i=0}^{N-1}\int_X f_1(T^ix)\cdots f_m(T^{mi}x)\,d\mu(x)\to\left(\int_X f_1 d\mu\right)\cdots\left(\int_X f_m d\mu\right). \quad (8)$$

A measure ν satisfies *m_0-correlation condition* if for every $m=2,\dots,m_0$, there exists a function $\tau=\tau_m:\mathbb{Z}_N\to\mathbb{R}^+$ such that

$$\mathbb{E}(\tau^q)=O_{m,q}(1)$$

for all $q\geq 1$ and

$$\mathbb{E}(\nu(x+h_1)\cdots\nu(x+h_m)|x\in\mathbb{Z}_N)\leq\sum_{1\leq i<j\leq m}\tau(h_i-h_j)$$

for all $h_1,\dots,h_m\in\mathbb{Z}_N$.

A measure ν is called *k-pseudorandom* if it satisfies $(k2^{k-1},3k-4,k)$-linear form condition and 2^{k-1}-correlation condition.

4.2 Tools

4.2.1 Gowers uniformity norms

We define *d-th Gowers uniformity norm* $\|\cdot\|_{U^d}$ inductively. Denote by T_h the shift operator: $(T_hf)(x)=f(x+h)$. For $f:\mathbb{Z}_N\to\mathbb{R}$, we set

$$\|f\|_{U^1}=|\mathbb{E}(f(x)|x\in\mathbb{Z}_N)|,$$
$$\|f\|_{U^d}=\mathbb{E}\left(\|f\cdot(T_hf)\|_{U^{d-1}}^{2^{d-1}}|h\in\mathbb{Z}_N\right)^{1/2^d}.$$

Explicitly,

$$\|f\|_{U^d} = \mathbb{E}\left(\prod_{\omega\in\{0,1\}^d} f(x+\omega h) \,|\, x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^d\right)^{1/2^d}.$$

One can show that for $d \geq 2$, $\|\cdot\|_{U^d}$ is a genuine norm.

Given a family of functions f_ω, $\omega \in \{0,1\}^d$, we define *d-dimensional Gowers inner product*:

$$\langle (f_\omega)_{\omega\in\{0,1\}^d}\rangle_{U^d} = \mathbb{E}\left(\prod_{\omega\in\{0,1\}^d} f_\omega(x+\omega h) \,|\, x\in\mathbb{Z}_N, h\in\mathbb{Z}_N^d\right).$$

Then we have *Gowers Cauchy-Schwarz inequality:*

$$|\langle (f_\omega)_{\omega\in\{0,1\}^d}\rangle_{U^d}| \leq \prod_{\omega\in\{0,1\}^d} \|f_\omega\|_{U^d}.$$

Gowers uniformity norms can be used to control the averages as in (1):

Theorem 5 (generalized von Neumann theorem)**.** *For a k-pseudorandom measure $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ and functions $f_0, \ldots, f_{k-1} : \mathbb{Z}_N \to \mathbb{R}$ such that $|f_i| \leq \nu + 1$, we have*

$$\mathbb{E}\left(\prod_{i=0}^{k-1} f_i(x+ih) \,|\, x,h\in\mathbb{Z}_N\right) = O\left(\min_{0\leq i\leq k-1} \|f_i\|_{U^{k-1}}\right) + o(1).$$

Theorem 5 is a finitary analog of (8). The proof uses van der Corput type argument and the linear form condition.

We call a function f *Gowers uniform* if $\|f\|_{U^{k-1}}$ is small. Theorem 5 shows that if at least one of f_i's is Gowers uniform functions, the average $\mathbb{E}\left(\prod_{i=0}^{k-1} f_i(x+ih)|x,h\in\mathbb{Z}_N\right)$ is negligible. This is used to arrange (7).

4.2.2 Obstructions to uniformity and dual functions

For a function $F : \mathbb{Z}_N \to \mathbb{R}$, define the *dual function* of F:

$$\mathcal{D}F(x) = \mathbb{E}\left(\prod_{\omega\in\{0,1\}^{k-1}:\omega\neq 0^{k-1}} F(x+\omega h) \mid h\in\mathbb{Z}_N^{k-1}\right).$$

Note that

$$\langle F, \mathcal{D}F \rangle = \|F\|_{U^{k-1}}^{2^{k-1}}. \tag{9}$$

Hence, if the Gowers norm of F is large, F correlates with its dual function, and the dual functions provide obstructions to uniformity.

The following two properties of dual functions are used in the proof.

Proposition 6. *For a function $F : \mathbb{Z}_N \to \mathbb{R}$ such that $|F| \leq \nu + 1$, we have*

$$\|\mathcal{D}F\|_{L^\infty} \leq 2^{2^{k-1}-1} + o(1).$$

Proposition 6 is deduced from the linear form condition.

Proposition 7. *Let $I = [-2^{2^k}, 2^{2^k}]$. Given function $F_1, \ldots, F_n : \mathbb{Z}_N \to \mathbb{R}$ such that $|\mathcal{D}F_i| \leq 2^{2^k}$ and a continuous function $\Phi : I^n \to \mathbb{R}$, we define*

$$\psi(x) = \Phi(\mathcal{D}F_1(x), \ldots, \mathcal{D}F_n(x)).$$

Then

$$\langle \nu - 1, \psi \rangle = o_{n,\Psi}(1).$$

Proposition 7 is deduced from the correlation condition using the Gowers Cauchy-Schwarz inequality.

4.2.3 σ-Algebras generated by generalized Bohr sets

Fix $\varepsilon, \eta > 0$. Given a function $G : \mathbb{Z}_N \to I := [-2^{2^k}, 2^{2^k}]$, one defines a σ-algebra $\mathcal{B}_{\varepsilon,\eta}(G)$ on $\mathbb{Z}_N$ that satisfies the following properties:

1. For any σ-algebra $\mathcal{B}$ on $\mathbb{Z}_N$,

$$\|G - \mathbb{E}(G|\mathcal{B} \vee \mathcal{B}_{\varepsilon,\eta}(G))\|_{L^\infty} \leq \varepsilon. \tag{10}$$

2. The σ-algebra $\mathcal{B}_{\varepsilon,\eta}(G)$ is generated by at most $O(1/\varepsilon)$ atoms.

3. If A is any atom of $\mathcal{B}_{\varepsilon,\eta}(G)$, then there exists a continuous function $\Psi_A : I \to [0,1]$ such that

$$\|(1_A - \Psi_A(G))(\nu + 1)\|_{L^1} = O(\eta). \tag{11}$$

   In fact, this implies that if $G_1, \ldots, G_n : \mathbb{Z}_N \to I$ and A is an atom of $\mathcal{B}_{\varepsilon,\eta}(G_1) \vee \cdots \vee \mathcal{B}_{\varepsilon,\eta}(G_n)$, then there exists a continuous function $\Psi_A : I^n \to [0,1]$ such that

$$\|(1_A - \Psi_A(G_1, \ldots, G_n))(\nu + 1)\|_{L^1} = O_n(\eta). \tag{12}$$

Roughly speaking, the σ-algebra $\mathcal{B}_{\varepsilon,\eta}(G)$ is generated by the level sets of the function G: the atoms of $\mathcal{B}_{\varepsilon,\eta}(G)$ are $G^{-1}([\varepsilon(n+\alpha), \varepsilon(n+1+\alpha)))$, $n \in \mathbb{Z}$, for suitably chosen α.

4.3 Structure theorem

Theorem 8 (generalized Koopman–von Neumann structure theorem)**.** *Let ν be a k-pseudorandom measure and $f : \mathbb{Z}_N \to \mathbb{R}$ such that $0 \le f \le \nu$. Let $\varepsilon > 0$ be a small parameter and $N > N_0(\varepsilon)$ sufficiently large. Then there exists a σ-algebra $\mathcal{B}$ and an exceptional set $\Omega \in \mathcal{B}$ such that*

$$\mathbb{E}(1_\Omega \nu) = o_\varepsilon(1), \tag{13}$$

$$\|(1 - 1_\Omega)\mathbb{E}(\nu - 1|\mathcal{B})\|_{L^\infty} = o_\varepsilon(1), \tag{14}$$

$$\|(1 - 1_\Omega)(f - \mathbb{E}(f|\mathcal{B}))\|_{U^{k-1}} \le \varepsilon^{1/2^k}. \tag{15}$$

Now setting

$$f_U = (1 - 1_\Omega)(f - \mathbb{E}(f|\mathcal{B})),\ f_{U^\perp} = (1 - 1_\Omega)\mathbb{E}(f|\mathcal{B}),\ E = 1_\Omega f,$$

we have decomposition (2) satisfying (3)–(7). Note that (3)–(6) follow directly from Theorem 8, and (7) is derived using Theorem 5.

Sketch of the proof of Theorem 8. In the proof, we use a parameter $\eta \to 0^+$.

First, we set $\mathcal{B}_0 = \{\emptyset, \mathbb{Z}_N\}$ and $\Omega_0 = \emptyset$. Then (13) and (14) obviously hold. If (15) fails, we set

$$F_1 := (1 - 1_{\Omega_0})(f - \mathbb{E}(f|\mathcal{B}_0)),$$
$$\mathcal{B}_1 := \mathcal{B}_0 \vee \mathcal{B}_{\varepsilon,\eta}(\mathcal{D}F_1),$$

and define the exceptional set Ω_1 to be the union of the atoms $A \in \mathcal{B}_1$ such that $\mathbb{E}(1_A(\nu + 1)) \le \eta^{1/2}$. Then

$$\mathbb{E}(1_{\Omega_1}(\nu + 1)) = O_\varepsilon(\eta^{1/2}),$$

and using (11) and Proposition 7, one shows that

$$\|(1 - 1_{\Omega_1})\mathbb{E}(\nu - 1|\mathcal{B}_1)\|_{L^\infty} = O_\varepsilon(\eta^{1/2}).$$

Continuing this procedure, we construct sequences of functions

$$F_n := (1 - 1_{\Omega_{n-1}})(f - \mathbb{E}(f|\mathcal{B}_{n-1})),$$

σ-algebras

$$\mathcal{B}_n := \mathcal{B}_{n-1} \vee \mathcal{B}_{\varepsilon,\eta}(\mathcal{D}F_n),$$

and exceptional sets $\Omega_n \in \mathcal{B}_n$ satisfying

$$\mathbb{E}(1_{\Omega_n}(\nu+1)) = O_{n,\varepsilon}(\eta^{1/2}), \tag{16}$$

$$\|(1-1_{\Omega_n})\mathbb{E}(\nu-1|\mathcal{B}_n)\|_{L^\infty} = O_{n,\varepsilon}(\eta^{1/2}). \tag{17}$$

Note that one can check inductively that

$$\|(1-1_{\Omega_n})\mathbb{E}(f|\mathcal{B}_n)\|_{L^\infty} \leq 1 + O_{n,\varepsilon}(\eta^{1/2}), \tag{18}$$

$$|F_n| \leq (1 + O_{n,\varepsilon}(\eta^{1/2}))(\nu+1). \tag{19}$$

By (19), (12) and Proposition 7 can be applied at every step to deduce (17).

It remains to show that after finitely many steps, we get

$$\|F_n\|_{U^{k-1}} \leq \varepsilon^{1/2^k}.$$

This follows from the following claim compared with estimate (18) when η is chosen sufficiently small.

Claim (energy increment property). *If* $\|F_n\|_{U^{k-1}} > \varepsilon^{1/2^k}$, *then*

$$\|(1-1_{\Omega_n})\mathbb{E}(f|\mathcal{B}_n)\|_{L^2}^2 > \|(1-1_{\Omega_{n-1}})\mathbb{E}(f|\mathcal{B}_{n-1})\|_{L^2}^2 + 2^{-2^k+1}\varepsilon.$$

Heuristically, the claim follows from the observation that if F_n is not Gowers uniform, then by (9) it has a nontrivial correlation with the dual function $\mathcal{D}F_n$.

Since the contribution of the exceptional sets can be controlled using (16), we sketch the proof assuming that the exceptional sets are empty. Using that by (9),

$$\langle F_n, \mathcal{D}F_n\rangle = \|F_n\|_{U^{k-1}}^{2^{k-1}} > \varepsilon^{1/2},$$

and by (10),

$$\langle F_n, \mathcal{D}F_n - \mathbb{E}(\mathcal{D}F_n|\mathcal{B}_n)\rangle = O(\varepsilon),$$

we deduce that

$$\langle F_n, \mathbb{E}(\mathcal{D}F_n|\mathcal{B}_n)\rangle = \langle f - \mathbb{E}(f|\mathcal{B}_{n-1}), \mathbb{E}(\mathcal{D}F_n|\mathcal{B}_n)\rangle > \varepsilon^{1/2} + O(\varepsilon).$$

In the last inequality, we can replace f by $\mathbb{E}(f|\mathcal{B}_n)$. Then by the Cauchy-Schwarz inequality and Proposition 6,

$$\|\mathbb{E}(f|\mathcal{B}_n) - \mathbb{E}(f|\mathcal{B}_{n-1})\|_{L^2} \cdot (2^{2^{k-1}-1} + O_{n,\varepsilon}(\eta^{1/2})) > \varepsilon^{1/2} + O(\varepsilon).$$

Since $\mathbb{E}(f|\mathcal{B}_n) - \mathbb{E}(f|\mathcal{B}_{n-1}) \perp \mathbb{E}(f|\mathcal{B}_{n-1})$, the claim now follows from the Pythagoras theorem (with an appropriate choice of parameter η). $\square$

References

[1] H. Furstenberg, Recurrence in ergodic theory and combinatorial number theory. M. B. Porter Lectures. Princeton University Press, Princeton, N.J., 1981.

[2] B. Host and B. Kra, Nonconventional ergodic averages and nilmanifolds. Ann. of Math. (2) 161 (2005), no. 1, 397–488.

[3] B. Green and T. Tao, The primes contain arbitrary long arithmetic progressions. To appear in Ann. Math.

[4] T. Ziegler, Universal characteristic factors and Furstenberg averages. To appear in JAMS.

Alexander Gorodnik, CalTech
email: **gorodnik@caltech.edu**

5 Polynomial extensions of van der Waerden's and Szemerédi's theorems

after V. Bergelson and A Leibman [BL]
A summary written by Michael Johnson

Abstract

We summarize the main ideas of V. Bergelson and A. Leibman's multidimensional polynomial versions of Szemerédi's and van der Waerden's theorems, emphasizing the use of PET-induction in both proofs.

5.1 Introduction

In 1976, H. Furstenberg proved the following ergodic version of Szemerédi's theorem.

Theorem 1. *[Fu1] Let $(X, \mathcal{B}, \mu)$ be a probability space, and let T be a measure preserving transformation of X. Then for all $A \in \mathcal{B}$ with $\mu(A) > 0$,*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(T^{-n}A \cap \ldots \cap T^{-kn}A) > 0.$$

In 1996, V. Bergelson and A. Leibman generalized this result in following manner: first they allowed for polynomial powers on the transformations, and second, their result deals with a commuting group of invertible transformations. We call polynomials that take integer values on the integers, **integer polynomials**.

Theorem 2. *[BL] Let $(X, \mathcal{B}, \mu)$ be a probability space, let $T_1, \ldots, T_l$ be commuting invertible measure preserving transformations of X, let $p_{i,j}$ be integer polynomials satisfying $p_{i,j}(0) = 0$ for all $1 \leq i \leq k, 1 \leq j \leq l$, and let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(\bigcap_{i=1}^{k} T_1^{-p_{i,1}(n)} \ldots T_l^{-p_{i,l}(n)} A) > 0.$$

By translating Theorem 2 back to combinatorics, we get the following result as a corollary.

Theorem 3. *[BL] Let $S \subseteq \mathbb{Z}^l$ have positive upper Banach density, and let $p_1, \ldots, p_k$ be integer polynomials with $p_i(0) = 0$ for all $i = 1, 2, \ldots, k$. Then for each finite subset $\mathbf{v}_1, \ldots, \mathbf{v}_k$ of $\mathbb{Z}^l$, there exists an $n \in \mathbb{N}$, and $\mathbf{u} \in \mathbb{Z}^l$ such that $\mathbf{u} + p_i(n)\mathbf{v}_i \in S$ for all $i = 1, 2, \ldots, k$.*

The following topological version of Theorem 2 corresponds to the multidimensional polynomial version of Van der Waerden's Theorem.

Theorem 4. *[BL] Let (X, d) be a compact metric space, $T_1, \ldots, T_l$ be commuting homeomorphisms of X, and p_{ij} be integer polynomials satisfying $p_{i,j}(0) = 0$ for all $1 \leq i \leq k, 1 \leq j \leq l$. Then for every $\varepsilon > 0$, there exists $x \in X$ such that $d(T_1^{p_{i,1}(n)} \ldots T_l^{p_{i,l}(n)} x, x) < \varepsilon$ for all $i = 1, 2, \ldots, k$.*

In this summary paper, we present the main ideas in the proofs of Theorem 2 and 4, emphasizing the use of PET-induction in both proofs.

5.2 Polynomial Expressions

Expressions of the form $T_1^{p_1(n)} \ldots T_l^{p_l(n)}$ with each p_i an integer polynomial of degree less than D are called **polynomial expressions**. We note that products and inverses of polynomial expressions are again polynomial expressions. Thus the set of all polynomial expressions, PE forms a group.

We define the **degree** of a polynomial expression $T_1^{p_1(n)} \ldots T_l^{p_l(n)}$ to be $\max_{1,\ldots,l} \deg(p_i)$. Its **weight** is the pair (r, d) where $\deg(p_{r+1}) = \ldots = \deg(p_l) = 0$ but $\deg(p_r) = d \geq 1$. We say $(r, d) > (s, e)$ if $r > s$ or if $r = s$ and $d > e$.

Example 5. *The polynomial expression $T_1^{n^2+n} T_2^{n^3-4n} T_3^{n^2+3n} T_4^{3n^2+n} T_5^0$ has degree 3 and weight $(4, 2)$.*

The expressions $T_1^{p_1(n)} \ldots T_l^{p_l(n)}$ and $T_1^{q_1(n)} \ldots T_l^{q_l(n)}$ are called **equivalent** if they have the same weight and $\deg(p_r - q_r) < d$.

A finite subset of PE is called a **system**. The degree of a system is the maximal degree of its elements. For a system A with l transformations and degree D, we form the weight matrix $\begin{pmatrix} N_{1,1} & \ldots & N_{1,D} \\ \vdots & & \vdots \\ N_{l,1} & \ldots & N_{l,D} \end{pmatrix}$ where $N_{r,d}$ is the the number of equivalence classes in A whose weight is (r, d). We say the weight matrix M precedes the weight matrix N if for some (r, d), $M_{rd} < N_{rd}$, and $M_{re} = N_{re}$ for all $e > d$ and $M_{se} > N_{se}$ for all $s > r$, $e = 1, \ldots, D$.

Example 6. *The system* $\{T_1^{n^2+n}T_2^{3n}, T_1^{3n^3+n}T_2^0, T_1^{3n^3}T_2^0, T_1^0T_2^{2n^2+2n}, T_1^{4n}T_2^{n^2}\}$ *has degree 3 and weight matrix* $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 2 & 0 \end{pmatrix}$.

5.3 Theorem 4

Given a polynomial system $P = \{g_1(n), \ldots, g_l(n)\}$ with weight matrix A where each $g_i(n) = T_1^{p_{i,1}(n)} \ldots T_l^{p_{i,l}(n)}$. Suppose that Theorem 4 holds for every system whose weight matrix precedes A. If we show that Theorem 4 holds for system P, we are done. As an example, we prove Theorem 4 in the case where $l = k = 1$ and $p_{1,1} = n^2$.

Let (X, d) be a compact metric space and T a homeomorphism of X. Without loss of generality, we assume that T is minimal. Let $\varepsilon > 0$. We need to find $x \in X$ and $n \in \mathbb{N}$ such that $d(T^{n^2}x, x) < \varepsilon$. We will find a sequence of points $x_0, x_1, \ldots$ and integers $n_1, n_2, \ldots$ such that

$$d(T^{(n_m+\ldots+n_{l+1})^2}x_m, x_l) < \varepsilon/2 \text{ for every l, m} \in \mathbb{N} \text{ with } l < m.$$

Here we state the appropriate PET-induction hypothesis, which is the linear topologial version of van der Waerden's theorem.

Proposition 7. *[FW] Let (X, d) be a compact metric space and let T be a homeomorphism of X. Then for any $\varepsilon > 0$, and $p \in \mathbb{N}$, and any $c_0, \ldots, c_{p-1} \in \mathbb{Z}$, there exists $x \in X$ and $n \in \mathbb{N}$ such that $d(T^{c_i n}x, x) < \varepsilon$ for all $i = 0, \ldots, p-1$. If (X, T) is minimal, the set of such x is dense in X.*

Since X is compact for some $l < m$ we will have $d(x_m, x_l) < \varepsilon/2$. Thus we will have $d(T^{(n_m+\ldots+n_{l+1})^2}x_m, x_m) < \varepsilon$.

Pick x_0 arbitrarily and set $n_1 = 1$ and $x_1 = T^{-n^2}x_0$. By continuity, choose $\varepsilon_1 < \varepsilon/2$ be such that $d(T^{n_1^2}y, x_0) < \varepsilon/2$ for every y for which $d(y, x_1) < \varepsilon_1$. Using Proposition 7 (with $\varepsilon = \varepsilon_1/2$, $p = 1$, and $c_0 = 2n_1$), find $y_1 \in X$, $n_2 \in \mathbb{N}$, such that $d(y_1, x_1) < \varepsilon_1/2$ and $d(T^{2n_1n_2}y_1, y_1) < \varepsilon_1/2$. Set $x_2 = T^{-n_2^2}y_1$. Thus we have

$$d(T^{n_2^2}x_2, x_1) < \varepsilon_1/2 < \varepsilon/2$$

and

$$d(T^{2n_1n_2+n_2^2}x_2, x_1) \leq d(T^{2n_1n_2}y_1, y_1) < \varepsilon_1.$$

Thus by choice of ε_1, we have

$$d(T^{(n_1+n_2)^2}x_2, x_0) = d(T^{n_1^2}T^{2n_1n_2+n_2^2}x_2, T^{n_1^2}x_1) < \varepsilon/2.$$

Given x_m, n_m, we inductively find x_{m+1}, n_{m+1}. By continuity, choose $0 < \varepsilon_m < \varepsilon/2$ such that for all $l = 0, 1, \ldots, m-1$, $d(T^{(n_m+\ldots+n_{l+1})^2}y, x_l) < \varepsilon/2$ for every y for which $d(y, x_m) < \varepsilon_1$. Using Proposition 7 (with $\varepsilon = \varepsilon_m/2$, $p = m$, and $c_l = 2(n_m + \ldots + n_{l+1})$) and $l = 0, 1, \ldots, m-1$), find $y_m \in X$, $n_{m+1} \in \mathbb{N}$, such that $d(y_m, x_m) < \varepsilon_m/2$, and $d(T^{2(n_m+\ldots+n_{l+1})n_{m+1}}y_m, y_m) < \varepsilon_m/2$ for $l = 0, \ldots, m-1$. Set $x_{m+1} = T^{-n_{m+1}^2}y_m$.

$$d(T^{2(n_m+\ldots+n_{l+1})n_{m+1}+n_{m+1}^2}x_{m+1}, x_m)$$

$$\leq d(T^{2(n_m+\ldots+n_{l+1})n_{m+1}}y_m, y_m) + d(y_m, x_m) < \varepsilon_m, for l = 0, \ldots, m-1.$$

Thus by choice of ε_m, we have

$$d(T^{n_{m+1}^2}x_{m+1}, x_m) < \varepsilon/2$$

and

$$d(T^{(n_{m+1}+\ldots+n_{l+1})^2}x_{m+1}, x_l) < \varepsilon/2, l = 0, \ldots, m-1.$$

5.4 Theorem 2

We now direct our attention to ideas for Theorem 2. The following proposition is instrumental in the proof of Theorem 2. We prove it using PET-induction.

Proposition 8. *[BL] Let $\alpha\colon (X, \mathcal{B}, \mu, \Gamma) \to (Y, \mathcal{D}, \nu, \Gamma)$ be a weakly mixing extension relative to Γ, let $\mu = \int \mu_y d\nu(y)$, let $T_1, \ldots, T_l \in \Gamma$, and let $g_i(n) = \prod_{j=1}^{l} T_j^{p_{ij}(n)}$, be such that g_i and $g_i(n)g_t^{-1}(n)$ depend nontrivially on n. Then for any $f_1, \ldots, f_k \in L^\infty(\mu)$*

$$\lim_{N\to\infty} \Big\| \sum_{n=0}^{N-1} \Big(\prod_{i=1}^{k} g_i(n) f_i - \prod_{i=1}^{k} \alpha^*(\int f_i d\mu_y) \Big) \Big\|_{L^2(\mu)} = 0.$$

We use the following version of the "van der Corput" Lemma.

Lemma 9. *[Be] Let $w_0, w_1, \ldots$ be a bounded sequence in a Hilbert space $(\mathcal{H}, \langle , \rangle)$. If*

$$\mathrm{D}-\lim_h \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \langle w_n, w_{n+h} \rangle = 0,$$

then $\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} w_n \right\| = 0.$

Using standard arguments we make the following reductions. First we assume that $p_{ij}(0) = 0$ for all i, j, and hence $\{g_1(n), \ldots, g_k(n)\}$ form a system A. Using the multilinearity of the integral, we also assume that $\int f_i d\mu_y = 0$ for all $i = 1, \ldots, k$. We must then show

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=1}^{k} g_i(n) f_i \right\|_{L^2(\mu)} = 0.$$

Lastly, we assume that our group Γ is a finitely generated free abelian group with basis $T_1, \ldots, T_l$.

Proof. Set $w_n = \prod_{i=1}^{k} g_i(n) f_i$.

We define

$$L(n,h) = \langle \prod_{i=1}^{k} g_i(n) f_i, \prod_{i=1}^{k} g_i(n+h) f_i \rangle = \int \prod_{i=1}^{k} g_i(n) f_i \cdot g_i(n+h) f_i d\mu. \quad (1)$$

Thus by Lemma 9 we need to show,

$$\mathrm{D}-\lim_{h} \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} L(n,h) = 0.$$

By reordering the polynomial expressions, we may assume that for some $q \leq k$, $\deg g_1(n) = \ldots = \deg g_q(n) = 1$ and $\deg g_i(n) \geq 2$ for $q+1 \leq i \leq k$. Thus $g_i(n+h) = g_i(n) g_i(h)$ when $i \leq q$. For a fixed $h \in \mathbb{N}$, we define a new system, $\tilde{A}_h = \{g_i(n), g_i(n+h) g_i^{-1}(h) \colon i = 1, \ldots, k\}$. We note that $g_i(n) = g_i(n+h) g_i^{-1}(h)$ if and only if $i \leq q$ (ie $\deg g_i(n) = 1$). Thus we can write equation (1) as follows:

$$L(n,h) = \int \prod_{i=1}^{q} g_i(n) f_i \cdot g_1(n+h) f_i \prod_{i=q+1}^{k} g_i(n) f_i \prod_{i=q+1}^{k} g_i(n+h) f_i d\mu$$

$$= \int \prod_{i=1}^{q} g_i(n)(f_i \cdot g_i(h) f_i) \prod_{i=q+1}^{k} g_i(n) f_i \prod_{i=q+1}^{k} g_i(n+h) g_i^{-1}(h)(g_i(h) f_i) d\mu$$

$$= \int \prod_{i=1}^{k'} \tilde{g}_i(n) \tilde{f}_i d\mu$$

where $k' = 2k - q$, $\tilde{f}_i$ is either f_m, $g_m(h)f_m$, or $f_m \cdot g_m(h)f_m$ for some $1 \leq m \leq k$, and $\tilde{g}_i$ is either g_m for some $1 \leq m \leq k$ or $g_m(n+h)g_m^{-1}(h)$ for some $q+1 \leq m \leq k$. In this way, we order the system $\tilde{A}_h = \{\tilde{g}_i \colon i = 1, \ldots, k'\}$.

By reordering, assume that $\tilde{g}_1(n)$ has minimal weight in $\tilde{A}_h$. Set $\hat{g}_i = \tilde{g}_i(n)\tilde{g}_1^{-1}(n)$, for $i = 1, \ldots, k'$. We consider the new system $A_h = \{\hat{g}_i \colon 1 = 2, \ldots, k'\}$. We claim that the weight matrix for A_h precedes that of A. We note that replacing $g_i(n)$ with $g_i(n+h)g_i^{-1}(h)$ does not change the family of equivalence classes. Thus, $\tilde{A}_h$ has the same weight matrix as A. Next, we multiplied each element in $\tilde{A}_h$ by the inverse of an element with the smallest weight $\tilde{g}_1(n)$. Thus, all polynomial expressions that are not equivalent to $\tilde{g}_1(n)$ do not change their weights when multiplied by $\tilde{g}_i^{-1}(n)$. However, those expressions that are equivalent to $\tilde{g}_i(n)$, decrease in weight when multiplied by $\tilde{g}_i^{-1}(n)$. Thus, the weight matrix for A_h precedes that of A.

We now apply PET-induction. Suppose that Proposition 8 is true for all systems whose weight matrix precedes that of A (namely A_h). Thus

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} L(n,h)$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \int \tilde{f}_1 \cdot \prod_{i=2}^{k'} \hat{g}_i(n)\tilde{f}_i d\mu$$

$$= \int \tilde{f}_1 \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=2}^{k'} \hat{g}_i(n)\tilde{f}_i d\mu$$

$$= \int \tilde{f}_1 \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=2}^{k'} \hat{g}_i(n)\alpha^*(\int \tilde{f}_i d\mu_y) d\mu$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=1}^{k'} \hat{g}_i(n)(\int \tilde{f}_i d\mu_y) d\nu$$

$$\leq \prod_{i=1}^{k'} \Big\| \int \tilde{f}_i d\mu_y \Big\|_{L^2(\nu)}$$

for h large enough.

If A has degree at least 2, then $\deg g_k(n) \geq 2$ and $\tilde{f}_i = f_k$ for some $i \leq k'$. Thus by assumption on f_k, the last product is zero. Thus D−

$\lim_h \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} L(n,h) = 0$. The linear case when A has degree 1, is done previously in [FW] to prove the linear multidimensional version of Szemerédi's Theorem. $\square$

An extension $\alpha\colon \mathbf{X} \to \mathbf{Y}$ is primitive if $\Gamma = \Gamma_c \times \Gamma_w$ where α is compact relative to Γ_c and weakly mixing relative to Γ_w.

Proposition 10. *[FK] If $\gamma\colon \mathbf{X} \to \mathbf{Z}$ is a nontrivial extension, one can find homomorphisms $\alpha\colon \mathbf{X} \to \mathbf{Y}$ and $\alpha'\colon \mathbf{Y} \to \mathbf{Z}$ with $\gamma = \alpha\alpha'$ such that $\mathbf{Y}$ is a nontrivial primitive extension of $\mathbf{Z}$.*

Proposition 11. *The family of Γ invariant factors which are SZP-systems have a maximal element.*

Thus to prove Theorem 2 we need only check that the property of being a SZP system is preserved under primitive extensions.

Proposition 12. *Given a primitive extension $\alpha\colon \mathbf{X} \to \mathbf{Y}$, if $\mathbf{Y}$ is SZP, then so is $\mathbf{X}$.*

Sketch of Proof. We first break each $g_i(n)$ into its compact and weakly mixing parts as follows: $g_i(n) = R_{(i)}(n)S_{(i)}(n)$ where $R_{(i)}(n) \in \Gamma_c$ and $S_{(i)}(n) \in \Gamma_w$. Let $\{R_1, \ldots, R_r\}$ be the set of all the compact components of $g_i(n)$, $i = 1, \ldots, k$ including the identity. Likewise let $\{S_1, \ldots, S_s\}$ be the set of all pairwise distinct weakly mixing components of $g_i(n)$, $i = 1, \ldots, k$.

It is sufficient to find a set $P \subseteq \mathbb{N}$ with $\underline{d}(P) > 0$ (where $\underline{d}$ is the lower density) and $c > 0$ such that

$$\mu\Big(\cap R_i^{-1}(n)S_j^{-1}(n)A\Big) > c$$

for all $n \in P$. To do this we first use a corollary of Proposition 8 to find a subset $P' \subseteq \mathbb{N}$ such that $\mu\Big(S_j^{-1}(n)A\Big) > c'$. This would prove the theorem in the case that $\mathbf{X}$ is weakly mixing relative to $\mathbf{Y}$.

To deal with the compact portion, we find a measurable subset $A' \subseteq A$ with $\mu(A') > 0$ such that $S_j^{-1}(n)A$ and $R_i^{-1}(n)S_j^{-1}(n)A$ are sufficiently close for all $i = 1, \ldots, r$, $j = 1, \ldots, s$. We use the following lemma.

Lemma 13. *[BL] Let $f, h_1, \ldots, h_k \in L^2(\mu)$ and $\varepsilon > 0$ be such that for almost all $y \in Y$ and all $R \in \Gamma_c$, there exists $m \in \{1, \ldots, K\}$ such that $\|Rf - h_m\|_{L^2(\mu_y)} < \varepsilon$. Then for all $B \in \mathcal{D}$ with $\nu(B) > 0$ there exists $P \subseteq \mathbb{N}$ with $\underline{d}(P) > 0$, a family of sets $B_n \in \mathcal{D}$, and $b > 0$ such that, for any $n \in P$, $1 \le i \le r$, $1 \le j \le s$*

1. $\nu(B_n) > b$,

2. $S_j(n)B_n \subseteq B$,

3. $\|R_i(n)S_i(n)f - S_j(n)f\|_{L^2(\mu_y)} < 2\varepsilon$ *for all* $y \in B_n$.

□

References

[Be] Bergelson, V. Weakly mixing PET. *Erg. Th. and Dyn. Sys.* **7** (1987), 337-349.

[BL] Bergelson, V., Leibman, A. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.*. **9** (1996), 725-753.

[Fu1] Furstenberg, H. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. d'Analyse Math.* **31** (1977), 204-256.

[FK] Furstenberg, H., Katznelson, Y., An ergodic Szemerédi theorem for commuting transformations. *J. d'Analyse Math.* **45** (1978) 275-291.

[FW] Furstenberg, H., Weiss, B., Topological dynamics and combinatorial number theory. *J. d'Analyse Math.* **34** (1978), 61-85.

Michael Johnson, Northwestern University
email: **johnson@math.northwestern.edu**

6 Convergence of Conze-Lesigne Averages

after B. Host and B. Kra [5]
A summary written by Tamara Kucherenko

Abstract

For a measure preserving system $(X, \mathcal{B}, \mu, T)$ we investigate convergence of

$$\frac{1}{N}\sum_{n=0}^{N-1} f_1(T^{a_1 n}) f_2(T^{a_1 n}) f_3(T^{a_3 n})$$

We are able to generalize a theorem of Conze and Lesigne and provide an alternative proof.

6.1 Introduction

Suppose that $(X, \mathcal{B}, \mu, T)$ is a measure preserving system and $l \geq 1$ is an integer. Given a collection of bounded measurable functions $f_1, \ldots, f_l \in L^\infty(\mu)$ and $a_1, \ldots, a_l \in \mathbb{N}$ one can ask the question whether the multiple ergodic average

$$\frac{1}{N}\sum_{n=0}^{N-1} f_1(T^{a_1 n}x) f_2(T^{a_2 n}x) \cdots f_l(T^{a_l n}x) \tag{1}$$

converges and in what sense. The case $l = 1$, the existence of this limit in $L^2(\mu)$ is the von Neuman Ergodic Theorem.

A measure preserving transformation $T : X \longrightarrow X$ induces an operator U_T, on functions in $L^2(\mu)$ defined by $U_T f(x) = f(Tx)$. We simply write T in place of U_T and hence $Tf(x) = f(Tx)$. The measure preserving system $(X, \mathcal{B}, \mu, T)$ is assumed to be *ergodic*, i.e. the only sets $A \in \mathcal{B}$ such that $T^{-1}A \subset A$ have either full or zero measure. This setting suffices for most of the theorems we consider since a general system can be decomposed into its ergodic components.

When a system is ergodic, for $l = 1$ the limit of (1) in $L^2(\mu)$ is the integral $\int_X f_1 \, d\mu$ and in particular is constant. On the other hand, without that assumption and $l \geq 2$, the limit of (1) is not necessarily constant. For

example, if $X = \mathbb{T}, T : \mathbb{T} \longrightarrow \mathbb{T}$ is the rotation $Tx = x + \alpha \mod 1$ for some $\alpha \in \mathbb{T}, f_1(x) = e^{4\pi i x}$ and $f_2(x) = e^{2\pi i x}$, then $f_1(T^n x) f_2(T^{2n} x) = f_2^{-1}(x)$ for all $n \in \mathbb{N}$. Consequently, the average

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) f_2(T^{2n} x)$$

converges to a nonconstant function.

A *factor* of $(X, \mathcal{B}, \mu, T)$ can be defined in several ways.

(a) a T-invariant sub-σ-algebra $\mathcal{A}$ of $\mathcal{B}$

(b) a measure preserving system $(Y, \mathcal{A}, \nu, S)$ and a measurable map $\pi : X \longrightarrow Y$ such that $\mu \circ \pi^{-1} = \nu$ and $S \circ \pi(x) = \pi \circ T(x)$ for μ-almost all $x \in X$

(c) a T-invariant subspace $\mathcal{F}$ of $L^\infty(\mu)$

The first two definitons are equivalent by identifying $\pi^{-1}(\mathcal{A})$ with a T-invariant sub-σ-algebra of $\mathcal{A}$ and noting that any T-invariant sub-σ-algebra arises in this way. The first definition implies the third by setting $\mathcal{F} = L^\infty(\mathcal{A})$, and the converse is obtained by taking $\mathcal{A}$ to be the σ-algebra generated by $\mathcal{F}$-measurable sets.

If $(Y, \mathcal{A}, \nu, S)$ is a factor of $(X, \mathcal{B}, \mu, T)$ and $f \in L^2(\mu)$, the *conditional expectation* $\mathbb{E}(f|\mathcal{A})$ of f with respect to $\mathcal{A}$ is the orthogonal projection of f onto $L^2(\nu)$.

Taking the rotational behavior into account, the double averages

$$\frac{1}{N} \sum_{n=0}^{N-1} T^{a_1 n} f_1 T^{a_2 n} f_2$$

can be understood. Furstenberg showed that to prove convergence of the double averages it suffices to replace each f_i for $i = 1, 2$ by its conditional expectation $\mathbb{E}(f_i|\mathcal{Z})$ on the Kronecker factor (see definition in a subsequent sections). The Kronecker factor is said to be *characteristic* for the double average. Projection to the Kronecker factor does not capture the limiting behavior for $l \geq 3$ (see [3]). Instead, we use a an abelian group extension of the Kronecker factor.

6.2 Statement of results

Existence of the limit (1) for $l = 3$ with the hypothesis of total ergodicity, meaning that T and all its powers are ergodic, was proven by Conze and Lesigne (see [1]). Here, convergence is shown directly through an alternative approach. Moreover, the assumption of total ergodicity can be dropped.

Theorem 1. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system and a_1, a_2, a_3 three distinct integers. Suppose that $f_1, f_2, f_3 \in L^\infty(\mu)$. Then the limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f_1(T^{a_1 n}x) f_2(T^{a_2 n}x) f_3(T^{a_3 n}x) \tag{2}$$

exists in $L^2(\mu)$.

The proof separates into two parts. First, we reduce the problem to study convergence on a simpler system following the approach of Furstenberg and Weiss [4]. The second part shows convergence of that simpler system an involves techniques from harmonic analysis.

6.3 Reduction to a simpler system

6.3.1 Characteristic factors

Definition 2. *Given distinct integers $a_1, a_2, \ldots, a_l$ and a factor $(Y, \mathcal{A}, \nu, S)$ of a system $(X, \mathcal{B}, \mu, T)$, we say that Y is a characteristic factor of X for the scheme $a_1, a_2, \ldots, a_l$ if for any $f_1, f_2, \ldots, f_l \in L^\infty(\mu)$ we have*

$$\lim_{N\to\infty} \left(\frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=1}^{l} f_i \circ T^{a_i n} - \frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=1}^{l} \mathbb{E}(f_i \circ T^{a_i n} | \mathcal{A}) \right) = 0$$

in $L^2(\mu)$.

Finding a characteristic factor Y for a system X allows us to restrict to functions defined only on Y, and this restriction simplifies computations when Y has a simple form.

For a double average the characteristic factor is the Kronecker factor which we denote by $(Z, \mathcal{Z}, m, S)$. Most specifically, $S : Z \longrightarrow Z$ is the rotation defined by $Z = z + \alpha$, and we use $\pi : X \longrightarrow Z$ for the natural projection.

For $f \in L^2(\mu)$, we write $\tilde{f}$ the functon on Z defined by $\tilde{f} \circ \pi = \mathbb{E}(f|\mathcal{Z})$.

Furstenberg and Weiss showed the following (see [4]).

Theorem 3. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system, with notations as above. Let b_1, b_2 be integers. Then for any $f_1, f_2 \in L^\infty(\mu)$ we have that*

$$\frac{1}{N}\sum_{n=0}^{N-1} f_1(T^{b_1 n}x) f_2(T^{b_2 n}x)$$

exists in $L^2(\mu)$ and equals

$$\int_Z \tilde{f}_1(z + b_1\theta)\tilde{f}_2(z + b_2\theta)\, dm(\theta)$$

where $z = \pi(x)$.

Let $\tilde{Z}$ be the closed subgroup

$$\tilde{Z} = \{(z + a_1 t, z + a_2 t, z + a_3 t) \,:\, z, t \in Z\}$$

of Z^3 and let $\tilde{m}$ be its Haar measure. The subgroup $\tilde{Z}$ is invariant under the transformation $\tilde{S} = S^{a_1} \times S^{a_2} \times S^{a_3}$. Then $\tilde{S}\tilde{z} = \tilde{z} + \tilde{\alpha}$ where $\tilde{\alpha} = (a_1\alpha, a_2\alpha, a_3\alpha)$.

We denote by $(\tilde{X}, \tilde{m}, \tilde{T})$ the product of the systems (X, μ, T^{a_i}) over $(\tilde{Z}, \tilde{m}, \tilde{S})$. Now by Theorem 3 for $f_1, f_2, f_3 \in L^\infty(\mu)$ we have

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \int_X \prod_{i=1}^{3} f_i(T^{a_i n}x)\, dm(x) &= \int_{\tilde{Z}} \prod_{i=1}^{3} \tilde{f}_i(z_i)\, d\tilde{m}(\tilde{z}) \\ &= \int_{\tilde{X}} \prod_{i=1}^{3} f_i(x_i)\, d\tilde{\mu}(x_1, x_2, x_3) \end{aligned} \tag{3}$$

6.3.2 Reduction to the isometric extension of the Kronecker

For an ergodic system $(X, \mathcal{B}, \mu, T)$ with Kronecker factor (Z, α), let $(\hat{Z}, \hat{\mathcal{D}}, \hat{\nu}, T)$ denote the maximal isometric extension of (Z, α) in (X, T).

Theorem 4. *(Furstenberg and Weiss [4]). $\hat{Z}$ is a characteristic factor of X for all schemes $\{a_1, a_2, a_3\}$.*

Thus in order to prove the existence of (1) in $L^2(m)$, it suffices to show the convergence for functions defined on the isometric extension $\hat{Z}$ of Z. We express the extension $\hat{Z}$ in such a way that the corresponding group extension $X_1 = Z \times L$ (for suitable metrizable compact group L) is ergodic. Since it suffices to prove the convergence for the system X_1 we can from now on assume that X is itself a group extension of its Kronecker factor Z.

6.3.3 Reduction to an abelian group extension

Following the paper of Furstenberg an Weiss [4] we use the theory of Mackey groups to obtain the next theorem.

Theorem 5. *(Furstenberg and Weiss [4]). X has a characteristic factor for all schemes a_1, a_2, a_3 that is an abelian group extension of its Kronecker factor Z.*

¿From now on we can assume that our system X is a compact abelian group extension $X = Z \times G$ of its Kronecker with the natural projection $X \longrightarrow Z$ given by $\pi(z, g) = z$. The transformation T on X is given by the cocycle $\sigma : Z \longrightarrow G$ and can be written as

$$T(z, g) = (Sz, g + \sigma(z))$$

For every integer n we have $T^n(z, g) = (S^n z, g + \sigma^{(n)}(z))$ where $\sigma^{(n)}(z) = \sigma(z) + \sigma(Sz) + \cdots \sigma(S^{n-1}z)$.

Now, $\tilde{X} = \tilde{Z} \times G^3$ and the measure $\tilde{\mu}$ is the product $\tilde{m} \times m_G \times m_G \times m_G$ of Haar measures. The transformation $\tilde{T}$ on $\tilde{X}$ is given by

$$\tilde{T}(\tilde{z}, g_1, g_2, g_3) = (\tilde{z} + \tilde{\alpha}, g_1 + \sigma^{(a_1)}(z_1), g_2 + \sigma^{(a_2)}(z_2), g_3 + \sigma^{(a_3)}(z_3)).$$

Thus, the system $(\tilde{X}, \tilde{\mu}, \tilde{T})$ is a compact abelian group extension of $(\tilde{Z}, \tilde{m}, \tilde{S})$ by the group G^3, given by the cocycle $\tilde{\sigma} : \tilde{Z} \longrightarrow G^3$, where $\tilde{\sigma} = \sigma^{(a_1)} \times \sigma^{(a_2)} \times \sigma^{(a_3)}$.

6.4 Sketch of the proof of the main theorem

It now suffices to prove the existence of the limit (2) for the modified system described above. By density, it is enough to consider the case when the functions f_i are of the form

$$f_i(z, g) = w_i(z)\chi_i(g)$$

for $i = 1, 2, 3$ and $w_i \in L^\infty(m), \chi_i \in \hat{G}$ and $x = (z, g)$.

We denote by M the Mackey group of the cocycle $\tilde{\sigma}$ on $\tilde{Z}$. The proof separates into two cases depending on whether or not the character $\tilde{\chi} = (\chi_1, \chi_2, \chi_3)$ belongs to $M^\perp$. The case $\tilde{\chi} \notin M^\perp$ is covered by the following lemma.

Lemma 6. *Let the functions $f_i = w_i(z)\chi_i(g)$ be given and assume that $\chi \notin M^\perp$. Then*

$$\lim_{N\to} \frac{1}{N} \sum_{n=0}^{N-1} f_1(T^{a_1 n}x) f_2(T^{a_2 n}x) f_3(T^{a_3 n}x) \tag{4}$$

exists in $L^2(\mu)$ and equals zero.

For the second case $\tilde{\chi} \in M^\perp$ we write

$$\frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=1}^{3} f_i(T^{a_i n x}) = \prod_{i=1}^{3} \chi_i(g) \frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=1}^{3} w_i(z + n a_i \alpha) \chi_i(\sigma^{(n a_i)}(z)).$$

First, we show that there exists a continuous mapping $t \mapsto \phi(t)$ from Z to $L^2(m)$ such that

$$\phi_{n\alpha}(z) = \prod_{i=1}^{3} \chi_i(\sigma^{(a_i n)}(z)).$$

Using the Hilbert space Van der Corput lemma we obtain the following result.

Lemma 7. *Let Z be a compact metric space, $S : Z \longrightarrow Z$ a homomorphism so that Z is uniquely ergodic with invariant measure m. Let $f : Z \longrightarrow \mathcal{H}$ be a continuous map into a Hilbert space $\mathcal{H}$. Then, for all $z \in Z$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n z) = \int_Z f(u)\, dm(u)$$

in $\mathcal{H}$.

By the lemma and the fact that (Z, S) is uniquely ergodic,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \prod_{i=1}^{3} w_i(z + n a_i \alpha) \chi_i(\sigma^{(n a_i)}(z)) = \int_Z \prod_{i=1}^{3} \phi_t(z) w_i(z + a_i t)\, dm(t)$$

in $L^2(m)$.
This completes the proof of the main theorem.

References

[1] J.-P. Conze and E. Lesigne, *Sur un théorème ergodique pour des measures diagonales.* C. R. Acad. Sci. Paris, Série I, **306** (1988), pp. 491–493.

[2] H. Furstenberg, *Nonconventional ergodic averages.* Proc. Sympos. Pure Math. , **31** (1990), pp. 43-56.

[3] H. Furstenberg, *Ergodic behaviour of diagonal measures and a theorem of Szeméredi on arithmetic progressions.* J. d'Analyse Math., **31** (1977), pp. 204-556.

[4] H. Furstenberg and B. Weiss, *A mean ergodic theorem for* $\frac{1}{N}\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$. Convergence in Ergodic Theory and Probability, Walter de Gruyter & Co, Berlin, New York (1996), pp. 193–227.

[5] B. Host and B. Kra, *Convergence of Conze-Lesigne Averages.* Isr. J. Math., **149** (2005), pp. 1–19.

Tamara Kucherenko, UCLA
email: **tamara@math.ucla.edu**

7 Pointwise Ergodic Theorems for Arithmetic Sets, part 2

after J. Bourgain [1] A summary written by Victor Lie

Abstract

We discuss the Shift model associated to a pointwise Ergodic theorem along polynomial iterates

7.1 Introduction

In what follows we will prove the Shift model for the following two results:

Theorem 1. *Let $(\Omega, \mathcal{B}, \mu, T)$ be a dynamical system and let p be a polynomial with integer coefficients. Given $r > 1$, there is a constant C so that*

$$\| \sup_{N \geq 1} |A_N f| \|_r \leq C \|f\|_r$$

holds for all $f \in L^r(\Omega, \mu)$, where A_N is given by

$$A_N f = \frac{1}{N} \sum_{n=1}^{N} f \circ T^{p(n)}.$$

Furthermore, the sequence $A_N f$ converges pointwise almost everywhere.

Theorem 2. *Let $(\Omega, \mathcal{B}, \mu, T)$ be a dynamical system and let p be a polynomial with real coefficients. For $N \geq 1$, let*

$$A_N f = \frac{1}{N} \sum_{n=1}^{N} f \circ T^{\lfloor p(n) \rfloor}.$$

If f is any bounded measurable function on Ω, then $A_N f$ converges pointwise almost everywhere.

It is worth to mention that the second result can also be extended with some effort to L^r functions ; we won't treat this extension in this presentation.

7.2 The L^2 case for Theorem 1

According with the first part of the presentation (see Andy Yingst),the maximal inequality and convergence problem for the averages:

$$A_N f = \frac{1}{N}\sum_{n=1}^{N} f \circ T^{p(n)}.$$

with

$$p(x) = b_1 x \ + \ b_2 x^2 + \ \ldots \ + \ b_d x^d \quad b_j \in Z \ \ \& \ \ b_d > 0$$

can be reduced to proving certain inequalities for the Shift model (Z, S).In the last case , one has :

$$A_N f = f * K_N \quad \text{where} \quad K_N = \frac{1}{N}\sum_{n=1}^{N} \delta_{\{p(n)\}}$$

(here δ_x stands for the Dirac measure at $x \in Z$)
Now since this is an L^2 problem involving convolution operators it is natural to expect that the multiplier theory will play an important role;consequently,the ideea will be to split the Fourier transform of each K_N in well localized pieces and after that to use maximal and orthogonal methods to estimate each such piece and combining all these estimates to obtain a global control for our operator.
Following now these lines we write:

$$A_N f = \mathcal{F}^{-1}\left[\mathcal{F}\left[K_N\right]\mathcal{F}\left[f\right]\right]$$

where $\mathcal{F}$ stands for the Fourier transform:

$$\mathcal{F} \ : \ L^2(Z) \ \rightarrow \ L^2(\Pi) \quad \mathcal{F}(f)(\alpha) = \int_Z f(n) e^{-2\pi i n \alpha} d\mu(n)$$

with $\alpha \in \Pi$ ($\Pi = [0, 1]$) and μ the canonical measure on Z.

STEP 1 Obtaining informations about the multiplier $\mathcal{F}[K_N]$

Clear for $\alpha \in \Pi$ we have:

$$\mathcal{F}[K_N](\alpha) = \frac{1}{N}\sum_{n=1}^{N} e^{-2\pi i p(n)\alpha}$$

Now if $\theta\, b_j \equiv a'_j/q'$ (mod 1) and $(a'_1, ..., a'_d, q') = 1$ then defining :

$$S(\theta) = S(q', a'_1, ...a'_d) = \frac{1}{q'} \sum_{r=0}^{q'-1} e^{2\pi i \left\{ \sum_{j=1}^{d} r^j \frac{a'_j}{q'} \right\}}$$

$$w_N(\beta) = \frac{1}{N} \int_0^N e^{-2\pi i \beta \left\{ \sum_{j=1}^{d} b_j y^j \right\}} dy$$

it can be shown that for some $\delta = \delta(d) > 0$ and $\delta' = \delta'(d) > 0$ and for some fixed α we have:
- if $\exists\, \theta = a/q$ st $q < N^{\delta}$ and $|\alpha - q| < N^{-d+\delta}$ then:

$$\mathcal{F}[K_N](\alpha) = S(\theta) w_N(\alpha - \theta) \; + \; O(N^{-1/2})$$

- else:

$$|\mathcal{F}[K_N](\alpha)| \lesssim N^{-\delta'}$$

Now for $\theta = a/q$ ($(a, q) = 1$) we have:

$$|S(\theta)| \lesssim q^{-\delta'}$$

Also $\forall\, \beta \in \Pi$ the function $w_N(\beta)$ obeys the relations:

$$|1 - w_N(\beta)| \lesssim |\beta| N^d \quad \& \quad |w_N(\beta)| \lesssim (1 + |\beta| N^d)^{-1/d}$$

(we use the notation $A \lesssim B$ for $A \leq CB$ where C is an absolute constant independent of A and B)

STEP 2 Breaking $\mathcal{F}[K_N]$ into well localized pieces

For $s \geq 0$, define an exhaustion of the rationals in Π

$$\mathcal{R}_s = \left\{ \theta \in Q \cap [0, 1] \mid \theta = \frac{a}{q} \,,\; (a, q) = 1 \;\&\; 2^s \leq q < 2^{s+1} \right\}$$

Now using the notations from the previous step we introduce:

$$\psi_{s,N} = \sum_{\theta \in \mathcal{R}_s} S(\theta) w_N(\alpha - \theta) \zeta(10^s(\alpha - \theta))$$

where by ζ we understand a smooth function on R with $\zeta = 1$ on $[-\frac{1}{10}, \frac{1}{10}]$ and $\zeta = 0$ outside $[-\frac{1}{5}, \frac{1}{5}]$.
With these notations , using the estimates from the step 1 we have the following result:

Lemma1 1. *There exists $\delta_1 > 0$ such that the uniform estimates:*

$$\left| \mathcal{F}[K_N](\alpha) - \sum_{s \geq 0} \psi_{s,N}(\alpha) \right| \lesssim N^{-\delta^1}$$

holds.

STEP 3 Combining the estimations on the well localized pieces into a global estimate

- we claim that :

$$\left\| \sup_{N \in Z_1} |\mathcal{F}^{-1}[\psi_{s,N} \mathcal{F} f]| \right\|_2 \lesssim s^2 2^{-s\delta'} \|f\|_2$$

where $Z_1 = \left\{ 2^k \mid k = 1, 2, ... \right\}$.
If we believe for the moment this fact , setting $\psi_N = \sum_s \psi_{s,N}$ and using Lemma 1 we have:

$$\left\| \sup_{N \in Z_1} |f * K_N| \right\|_2 \leq \left\| \sup_{N \in Z_1} |\mathcal{F}^{-1}[\psi_N \mathcal{F} f]| \right\|_2 +$$

$$\left(\sum_{n \in Z_1} \|\mathcal{F}[K_N] - \psi_N\|_\infty^2 \right)^{1/2} \|f\|_2 \lesssim \sum_{s=0}^{\infty} \left\| \sup_{Z_1} |\mathcal{F}^{-1}[\psi_{s,N} \mathcal{F} f]| \right\|_2 + \|f\|_2 \lesssim \|f\|_2$$

and due to the positiveness of our maximal operator this will imply our l^2 variant of the theorem 1:

$$\left\| \sup_N |f * K_N| \right\|_{l^2(Z)} \lesssim \|f\|_{l^2(Z)}$$

-returning now at our claim , we want first to 'cut' the tiles of $\psi_{s,N}$; for this we define :

$$\tilde{\psi}_{s,N}(\alpha) = \sum_{\theta \in \mathcal{R}_s} S(\theta) \chi(N^d(\alpha - \theta)) \zeta(10^s(\alpha - \theta))$$

(with $\chi = \chi_{[-1,1]}$ considered as function on R) and remark that we have the uniform estimate:

$$\sum_{N \in Z_1} |\psi_{s,N} - \tilde{\psi}_{s,N}| \lesssim 2^{-s\delta'}$$

Therefore , by a square function argument:

$$\left\| \sup_{N\in Z_1} |\mathcal{F}^{-1}[\psi_{s,N}\mathcal{F}f]| \right\|_2 \lesssim \left\| \sup_{N\in Z_1} |\mathcal{F}^{-1}[\tilde{\psi}_{s,N}\mathcal{F}f]| \right\|_2 + 2^{-s\delta'} \|f\|_2$$

Now for $N \in Z_1$ with $N^d \approx 2^j$ let R_j be the 2^{-j} - neighborhood of $\mathcal{R}_s \subset \Pi$.Thus,setting:

$$\mathcal{F}[g_s] = \mathcal{F}[f] \sum_{\theta\in\mathcal{R}_s} S(\theta)\zeta(10^s(\alpha-\theta)) \quad \& \quad \tilde{\psi}_{s,N}\mathcal{F}[f] = \mathcal{F}[\gamma_s]\chi_{R_j}$$

it follows from the main lemma of this paper (see section 1.3 in the first part of this presentation) that :

$$\left\| \sup_{N\in Z_1} |\mathcal{F}^{-1}[\tilde{\psi}_{s,N}\mathcal{F}f]| \right\|_2 \leq \left\| \sup_{j\in Z_+} |\mathcal{F}^{-1}[\mathcal{F}[g_s]\chi_{R_j}]| \right\|_2 \lesssim (\log|\mathcal{R}_s|)^2 \|g\|_2$$

Now $|\mathcal{R}_s| < 4^s$, and from the Step 1 and Parseval we have $\|g\|_2 \lesssim 2^{-s\delta'} \|f\|_2$; putting these facts together we end the proof of our claim.

7.3 The almost sure convergence of $A_N f$ for $f \in l^2(Z)$

From the first part of the presentation it is enough to prove in the (Z, S) setting ,that:

$$\sum_{j=1}^{J} \|\mathcal{M}_j f\|_2 \leq o(J) \|f\|_2$$

with

$$\mathcal{M}_j f = \sup_{\substack{N_j<N<N_{j+1} \\ N\in Z_\epsilon}} |f * (K_N - K_{N_j})|$$

where $Z_\epsilon = \{(1+\epsilon)^n \mid n = 1, 2, ...\}$ for $\epsilon > 0$ fixed , and N_j is any rapidly increasing sequence ($N_{j+1} > 2N_j$).
Again the main tool will be the Fourier transform used for a better localization of the kernel K_N ; indeed from Lemma 1 we deduce that $f * (K_N - K_{N_j})$ may be replaced by $\mathcal{F}^{-1}[(\psi_N - \psi_{N_j})\mathcal{F}f]$ when defining $\mathcal{M}_j$;fixing s_0 it follows from the claim in the Step 3 that:

$$\|\mathcal{M}_j f\|_2 \leq \sum_{s\leq s_0} \left\| \sup_{\substack{N_j<N<N_{j+1} \\ N\in Z_\epsilon}} \left|\mathcal{F}^{-1}[(\psi_N - \psi_{N_j})\mathcal{F}f]\right| \right\|_2 + C\epsilon^{-1}2^{-\delta'' s_0} \|f\|_2$$

where the second term will be $o(\|f\|_2)$ for apprpriate s_0 (you may think at $s_0 = J^{\frac{1}{100}}$) .Thus is suffice to verify our statement for:

$$\mathcal{M}_j f = \sup_{\substack{N_j < N < N_{j+1} \\ N \in Z_\epsilon}} |f * (w_N - w_{N_j})|$$

with w_N defined at the section 2 , step 1.
Now cutting as before the tiles of w we define the auxiliary maximal function:

$$\tilde{\mathcal{M}}_j f = \sup_{N_j < N < N_{j+1}} |f * (\chi_{N^d} - \chi_{N_j^d})|$$

where $\chi = \chi_{[0,1]}$ and $\chi_t = \frac{1}{t}\chi_{[0,1]}$. Now since:

$$\left\{ \sum_{j=1}^{J} \tilde{\mathcal{M}}_j f \right\}^{1/2} \leq J^{1/4} \|\{f * \chi_N \mid N = 1, 2, ...\}\|_{v_4}$$

it follows from the Part 1 that:

$$\sum_{j=1}^{J} \left\| \tilde{\mathcal{M}}_j f \right\|_2^2 \lesssim J^{1/2} \|f\|_2^2$$

hence:

$$\sum_{j=1}^{J} \|\mathcal{M}_j f\|_2^2 \lesssim \sum_{N \in Z_\epsilon} \left\| \mathcal{F}^{-1}[(w_N - \mathcal{F}[\chi_{N^d}])\mathcal{F} f] \right\|_2^2 + J^{1/2} \|f\|_2 \leq$$

$$\sup_\alpha [\sum_{N \in Z_\epsilon} |w_N(\alpha) - \hat{\chi}(N^d \alpha)|^2] \|f\|_2 + J^{1/2} \|f\|_2 \lesssim_\epsilon J^{1/2} \|f\|_2$$

Now using Cauchy-Scwartz we conclude:

$$\sum_{j=1}^{J} \|\mathcal{M}_j f\|_2 \lesssim_\epsilon J^{3/4} \|f\|_2$$

fact that ends our proof.

7.4 The L^p case for Theorem 1

In this section we extend the L^2 theory to L^p , with $p > 1$.Since the a.s. convergence for functions f in L^p reduces to bounded functions and hence is taken care of by the L^2 result , we need to show only the maximal inequality:

$$\|\sup |A_N f|\|_p \lesssim \|f\|_p$$

and only for the shift model (Z, S) .
More than that , due to the good properties of the kernels $\{K_N\}$ our problem reduces to:

$$\left\| \sup_{k \in Z_+} |f * K_{2^k}| \right\|_p \lesssim \|f\|_p$$

In what follows we will treat only the case $1 < p < 2$; the remaining range was established earlier in a paper written by the same autor - see [2]. Now for dealing with our problem we will need a much more careful analysis of our operators and this is because the Fourier multipliers involved in the argument need to have good bounds in L^p .
Mentaining the previous notations we define the following expressions:
-for s a positive integer: $\mathcal{Q}_s = 2^s!$ & $\mathcal{K}_s = \{k \in Z \mid 4^s \leq k < 4^{s+1}\}$
-for $s' \leq s$ and $k \in \mathcal{K}_s$, the functions:

$$\Omega_{k,s'} = \sum_{0 \leq a < \mathcal{Q}_{s'}} S\left(\frac{a}{\mathcal{Q}_{s'}}\right) w_{2^k}\left(\alpha - \frac{a}{\mathcal{Q}_{s'}}\right) \zeta\left(\mathcal{Q}_{s'}^2\left(\alpha - \frac{a}{\mathcal{Q}_{s'}}\right)\right)$$

The multiplier $\Omega_{k,s'}$ will substitute in L^p the role played by the expression $\sum_{r \leq s'} \psi_{r,2^k}$ in the L^2 case ; indeed the former one has a better distribution of the support with respect to the rational numbers fact that will imply a good control in the L^p norm ; more over this multiplier will not lose the good property of the last expression in approximating the $\mathcal{F}(K_N)$.More exactly we have:
- if $s' \leq s$ and $k \in \mathcal{K}_s$ then $|\mathcal{F}[K_{2^k}](\alpha) - \Omega_{k,s'}(\alpha)| < 2^{-\delta' s'}$
- for $1 < p_0 < p < 2$ $\| \sup_k |\mathcal{F}^{-1}[\Omega_{k,s'} \mathcal{F} f]| \|_{p_0} \lesssim \|f\|_{p_0}$
The first estimation is a simple consequence of the properties mentioned at step 1 (section 2) while the second one is obtained by a carefully study of each component of $\Omega_{k,s'}$.
A very important intermediate step consists in obtaining the relation:

$$\left\| \sup_{k \in \mathcal{K}_s} |f * K_{2^k}| \right\|_p \lesssim s \|f\|_p$$

where here $p > 1$. This is solved by using duality and reducing it to an L^2 problem for which the Fourier transform method is well adapted . With this done the proof of our result will become smooth making use of the interpolation between L^{p_0} and L^2 results obtained in the previous sections . Indeed we have:

$$\sup_k |\, f * K_{2^k}| \leq \sum_{s'} \sup_{k \geq 4^{s'}} \left| \mathcal{F}^{-1}[(\Omega_{k,s'} - \Omega_{k,s'-1})\mathcal{F}f] \right| +$$

$$\sum_s \sup_{k \in \mathcal{K}_s} \left| (f * K_{2^k}) - \mathcal{F}^{-1}[\Omega_{k,s}\mathcal{F}f] \right|$$

Now from the facts mentioned before we have:

$$\left\| \sup_{k \geq 4^{s'}} \left| \mathcal{F}^{-1}[(\Omega_{k,s'} - \Omega_{k,s'-1})\mathcal{F}f] \right| \right\|_{p_0} \lesssim \|f\|_{p_0}$$

$$\left\| \sup_{k \in \mathcal{K}_s} \left| (f * K_{2^k}) - \mathcal{F}^{-1}[\Omega_{k,s}\mathcal{F}f] \right| \right\|_{p_0} \lesssim s \, \|f\|_{p_0}$$

As anounced our aim now is to interpolate the above relations with some better L^2 - estimates ; these last ones are obtained as follows:

$$\left\| \sup_{k \geq 4^{s'}} \left| (f * K_{2^k}) - \mathcal{F}^{-1}[\Omega_{k,s}\mathcal{F}f] \right| \right\|_2 \leq C \sum_{k \geq 4^{s'}} 2^{-k\delta_1} +$$

$$\left\| \sup_{k \geq 4^{s'}} \left| \sum_{0 \leq r \leq s'} \mathcal{F}^{-1}[\psi_{r,2^k}\mathcal{F}f] - \mathcal{F}^{-1}[\Omega_{k,s}\mathcal{F}f] \right| \right\|_2 + \sum_{r > s'} \left\| \sup_k \left| \mathcal{F}^{-1}[\psi_{r,2^k}\mathcal{F}f] \right| \right\|_2$$

where here we have used the lemma 1 . Now only the second term need attention :

$$\Omega_{k',s'} - \sum_{r \leq s'} \psi_{r,2^k} \; =$$

$$\sum_{r \leq s'} \sum_{\theta \in \mathcal{R}_r} S(\theta) w_{2^k}(\alpha - \theta)[\zeta\left(\mathcal{Q}^2_{s'}(\alpha - \theta)\right) - \zeta\left(10^r(\alpha - \theta)\right)]$$

$$+ \sum_{\substack{q | \mathcal{Q}_{s'} \\ q \geq 2^{s'+1}}} \sum_{\substack{1 \leq a \leq q \\ (a,q)=1}} S\left(\frac{a}{q}\right) w_{2^k}\left(\alpha - \frac{a}{q}\right) \zeta\left(\mathcal{Q}^2_{s'}\left(\alpha - \frac{a}{q}\right)\right) = (1) + (2)$$

For (1) we have an l^∞ estimate ($k \geq 4^{s'}$) :

$$(1) \lesssim 4^{s'} \sup_{|\beta|>\mathcal{Q}_{s'}^{-2}} |w_{2^k}(\beta)| \lesssim 2^{-k/2}$$

For the second sum one may prove (using the same arguments as those presented in section 2) that the maximal function associated to (2) is bounded in the l^2 norm by $C2^{-\delta' s'}$.Hence we obtain that:

$$\left\| \sup_{k\in\mathcal{K}_s} \left|(f * K_{2^k}) - \mathcal{F}^{-1}[\Omega_{k,s}\mathcal{F}f]\right| \right\|_2 \lesssim 2^{-\delta' s'} \|f\|_2$$

and by substraction :

$$\left\| \sup_{k\geq 4^{s'}} \left|\mathcal{F}^{-1}[(\Omega_{k,s'} - \Omega_{k,s'-1})\mathcal{F}f]\right| \right\|_2 \lesssim 2^{-\delta' s'} \|f\|_2$$

Now interpolating between p_0 and 2 we obtain for $p_0 < p < 2$ and some $\delta_p > 0$:

$$\left\| \sup_k |f * K_{2^k}| \right\|_p \leq \sum_{s'} \left\| \sup_{k\geq 4^{s'}} \left|\mathcal{F}^{-1}[(\Omega_{k,s'} - \Omega_{k,s'-1})\mathcal{F}f]\right| \right\|_p +$$

$$\sum_s \left\| \sup_{k\in\mathcal{K}_s} \left|(f * K_{2^k}) - \mathcal{F}^{-1}[\Omega_{k,s}\mathcal{F}f]\right| \right\|_p \lesssim_p \sum_{s'} 2^{-\delta_p s'} + \sum_s 2^{-\delta_p s} < C$$

ending our proof .

7.5 The outlines of the Theorem 2

In this section we need to extend the a.s.converges of $A_N f$ for polynomials with real coefficients (this thing will be done only in the L^∞ - case).
Remember that now:

$$A_N f = \frac{1}{N} \sum_{n=1}^{N} f \circ T^{\lfloor p(n) \rfloor}$$

where $P(x) = b_0 + b_1 x + ... + b_d x^d$ ($b_d > 0$ & $d \geq 1$) is a polynomial with real coefficinets and and $\lfloor l \rfloor$ stands for the entire part of l.
Now we observe that since we can suppose that at least one of $b_1, b_2, ...b_d$

is not rational (otherwise we can reduce our problem at the one dicussed before) the sequence:

$$\{\, p(n) - \lfloor p(n) \rfloor \mid n = 1, 2, ...\}$$

is uniformly distributed in $[0, 1]$. Now fixing $\epsilon > 0$ small we define the affine continuous function $\tau = \tau_\epsilon$ on R by $\tau = 1$ on $[\epsilon, 1 - \epsilon]$ and $\tau = 0$ outside of $[0, 1]$.With this done we set:

$$\tilde{A}_N f = \frac{1}{N} \sum_{n=1}^{N} \sum_{m \in Z} \tau(p(n) - m)\, T^m\, f$$

and from the uniform distribution property we have the pointwise estimation (N large enough) :

$$|A_N f - \tilde{A}_N f| \leq \frac{\|f\|_\infty}{N} \#\{\, 1 \leq n \leq N \mid dist(p(n), Z) < \epsilon\} \leq 3\epsilon \, \|f\|_\infty$$

Thus will be suffice to show the a.s. convergence for $\tilde{A}_N f$ with ϵ fixed . Transferring this problem for the Shift model (Z, S) as before , we realize that the Fourier tranform of the kernel of $\tilde{A}_N$ is given by:

$$\mathcal{F}[K_N](\alpha) = \sum_{k \in Z} \hat{\tau}(k - \alpha) \left\{ \frac{1}{N} \sum_{n \leq N} e^{2\pi i (k-\alpha) p(n)} \right\}$$

Now this expression is very similar with the one appearing in the section 2 and so using the same techniques as before one can show that:

$$\left\| \sup |\tilde{A}_N f| \right\|_2 \lesssim_\epsilon \|f\|_2$$

finishing our argument .

References

[1] Bourgain, J., *Pointwise Ergodic Theorems for arithmetic sets.* Inst. Hautes Études Sci. Publ. Math., **69** (1989), pp. 5-45.

[2] Bourgain, J *On the Pointwise Ergodic Theorem on L^p for arithmetic sets.* Israel J. Math, 61 (1) (1988), 73-84.

Victor Lie ,Ucla
email: **vilie@math.ucla.edu**

8 The Ergodic Theoretical Proof of Szemerédi's Theorem II

after H. Furstenberg, Y. Katnelson, and D. Ornstein [1] A summary written by Anne E. McCarthy

Abstract

We complete the proof of H. Furstenberg's multiple recurrence theorem, via an analysis of factors of a measure-preserving system. We prove the existance of a maximal factor satisfying property SZ, and that any proper SZ subfactor has a SZ extension.

8.1 Introduction

It is has been established in the preceeding expostion that Szemerédi's Theorem is equivalent to a statement about multiple recurrence for measure-preserving systems (*m.p.s.*) We will complete the proof of this ergodic theory result:

Theorem 1. *Let $(X, \mathcal{B}, \mu, T)$ be a m.p.s., and $A \in \mathcal{B}$ be set with $\mu(A) > 0$. Then for any positive integer k there exists $n \geq 1$, such that*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) > 0.$$

The validity of this theorem has been established in the two prototypical examples of weak mixing systems and compact systems. We have also seen that any *m.p.s.* that is not weak mixing has a compact factor. Therefore, it is now known that theorem 1 holds for a factor of any *m.p.s.* In this exposition we extend to the full generality of the theorem.

Definition 2. *We say that a m.p.s. satisfies property SZ if the conclusion of 1 holds.*

We proceed with the proof as follows. For a fixed *m.p.s.*, $(X, \mathcal{B}, \mu, T)$, we consider the family $\mathcal{F}$ of all factors $(X, \mathcal{B}_1, \mu, T)$ that satisfy property SZ, ordered by inclusion. We will show

1. The family $\mathcal{F}$ contains a maximal element.

2. No proper factor of $(X, \mathcal{B}, \mu, T)$ can be maximal.

These two facts taken together imply that the maximal element must be $(X, \mathcal{B}, \mu, T)$ itself. Therefore, the sytem satisfies property SZ.

The proof of item 1. is fairly straight forward, and will be proved in 8.3. To prove item 2. we will need to define the notions of a *relatively weak mixing* extension and *relatively compact* extensions. Using these notions, we proceed much like we did in showing that theorem 1 holds for some factor. We prove that a relatively weak mixing extension is "relative weak mixing of all orders," and hence SZ. We also show that a relatively compact extension of a SZ system is also SZ. To conclude, we show that if an extension $(X, \mathcal{B}, \mu, T)$ of a SZ system $(Y, \mathcal{D}, \nu, S)$ is not realtively weak mixing, there must be an intermediate extension that is relatively compact with respect to $(Y, \mathcal{D}, \nu, S)$. This final step is analogous to showing that a system that is not weak mixing has a compact factor.

8.2 Factors

We begin by establishing intuition for a factor of a *m.p.s*, as well as some useful properties.

8.2.1 skew products

Let $(Y, \mathcal{D}, \nu, S)$ be a *m.p.s.*, and $(Z, \mathcal{E}, \theta)$ be a measure space. Let $y \mapsto \sigma(y)$ be a function from Y into measureable transformations on Z, such that $(y, z) \mapsto \sigma(y)(z)$ is a measureable. We define the skew product of Y with Z as follows. Let $X = Y \times Z$, $\mathcal{B} = \mathcal{D} \times \mathcal{E}$, and $\mu = \nu \times \theta$, and note that

$$T(y, z) = (S(y), \sigma(y)(z)),$$

preserves measure on $(X, \mathcal{B}, \mu)$. Let $\pi : X \to Y$ be the projection $\pi(y, z) = y$. Setting $\mathcal{B}_1 = \pi^{-1}\mathcal{D}$, we can identify $(Y, \mathcal{D}, \nu, S)$ with the factor system $(X, \mathcal{B}_1, \mu, T)$, becuase they are isomorphic as *m.p.s*. Furthermore, a theorem of Rohklin states that given a *m.p.s.*, any factor can be thought of as existing within such a skew product.

Although we do not require the full strength of Rohklin's theorem, we will discuss one useful consequence. Because X is a product space, we are able to disintegrate the measure μ. By Fubini's theorem, for any set $A \in \mathcal{B}$,

we have

$$\mu(A) = \int \mu_y(A) d\nu(y)$$

Where $\mu_y = (\delta_y \times \theta)$, is a measure supported on the fiber $\pi^{-1}(y)$. We also note that our measure preserving transformations respect this measure in the sense that $\mu_y(T^{-1}A) = \mu_{S_y}(A)$.

8.2.2 projecting functions

We will denote $\mathfrak{X} = (X, \mathcal{B}, \mu)$, and $\mathfrak{Y} = (Y, \mathcal{D}, \nu)$, where $\mathfrak{Y} \simeq (X, \mathcal{B}_1, \mu)$ will always be thought of as a factor of $\mathfrak{X}$ in the above sense. Given a function $g \in L^2(\mathfrak{X})$, we define the $L^2(\mathfrak{Y})$ function, g conditioned on $\mathfrak{Y}$, via

$$E(g|\mathfrak{Y})(y) = \int_{\pi^{-1}(y)} g \; d\mu_y$$

Alternatively, $L^2(X, \mathcal{B}_1, \mu)$ is a closed subspace of $L^2\mathfrak{X}$. It is equivalent to define $E(g|\mathfrak{Y})$ as the orthogonal projection of $g \in L^2(\mathfrak{X})$ to $E(g|\mathfrak{Y}) \in L^2(X, B_1, \mu)$. Viewing $L^2(X, \mathcal{B}_1, \mu) \subset L^2(\mathfrak{X})$, we can also view functions on $\mathfrak{Y}$ as functions on $\mathfrak{X}$ as well. We note some useful properties of projected functions:

$$E(gf|\mathfrak{Y}) = g \; E(f|\mathfrak{Y}) \text{ if } g \text{ is } \mathcal{B}_1 \text{measureable}$$
$$SE(f|\mathfrak{Y}) = E(Tf|\mathfrak{Y})$$

8.2.3 fiber squares (or relative squares)

Using the skew-product structure $X = Y \times Z$, with assignemt $\sigma(y)$, we define a new space $X \times_Y X$ as follows. Let $\tilde{X} = Y \times Z \times Z$, $\tilde{\mathcal{B}} = \mathcal{D} \times \mathcal{E} \times \mathcal{E}$, and $\tilde{\mu} = \nu \times \theta \times \theta$. We define the trasformation $\tilde{T}(y, z, z') = (S(y), \sigma(y)(z), \sigma(y)(z'))$.

Given the projection $\pi : X \to Y$, we can also identify $\tilde{X} = \cup_{y \in Y} \pi^{-1}(y) \times \pi^{-1}(y) \subset X \times X$. Also $\tilde{\mu} = \int \tilde{\mu}_y d\nu(y)$, where $\tilde{\mu}_y = \mu_y \times \mu_y$, and $\tilde{\mathcal{B}}$ is the restriction of $\mathcal{B} \times \mathcal{B}$ to $\tilde{X}$. We denote this system $\tilde{X} = X \times_Y X$. In the case of relative weak mixing and relative compactness with respect to a factor, we replace the role of $X \times X$ with $X \times_Y X$. In this setting, the following identity will be helpful:

$$\int_{X \times_Y X} f(y, z) f(y, z') d\tilde{\mu}(y, z, z') = \int_Y E(f|\mathfrak{Y})^2 d\nu(y)$$

8.3 Maximal SZ Factors

Definition 3. *A factor $(X, \mathcal{B}_1, \mu, T)$ is said to be Szemeredi (SZ) if*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) > 0$$

for all $k \in \mathbb{Z}^+$, and for all $A \in \mathcal{B}_1$ with $\mu(A) > 0$.

Recall that for every *m.p.s.* $(X, \mathcal{B}, \mu, T)$, there exists a factor $(X, \mathcal{B}_1, \mu, T)$ that is SZ. Let $\mathcal{F}$ denote the family of all factors of $(X, \mathcal{B}, \mu, T)$ that are SZ, ordered by inclusion. We will show that $\mathcal{F}$ has a maximal element. For a totally ordered set of factors $\mathcal{B}_\alpha$, the σ-algebra $\sup_\alpha \mathcal{B}_\alpha$ is characterized by those $A \in \mathcal{B}$ that satisfy the condition: given any $\varepsilon > 0$, there exists $A_0 \in \cup_\alpha \mathcal{B}_\alpha$ such that $\mu(\triangle(A, A_0)) < \varepsilon$.

Theorem 4. *Let $\{\mathcal{B}_\alpha\}$ be a totally ordered family of factors, and suppose $\{\mathcal{B}_\alpha\} \subset \mathcal{F}$. Then $\sup_\alpha \mathcal{B}_\alpha$ is SZ.*

Proof. Fix $k > 0$ and $A \in \sup_\alpha \mathcal{B}_\alpha$ with $\mu(A) > 0$. Set $\eta = \frac{1}{2(k+1)}$ and $A_0' \in \mathcal{B}_{\alpha_0}$ such that $\mu(\triangle(A, A_0')) < \frac{\eta}{4}\mu(A)$.

Consider the factor $(X, \mathcal{B}_{\alpha_0}, \mu, T) \simeq (Y, \mathcal{D}_0, \nu, T_0)$, and let $A_0'' = \pi(A_0') \in \mathcal{D}_0$. A simple calculation shows that $\nu(\{y \in A_0'' s.t. \mu_y(A) < 1 - \eta\}) < \frac{1}{4}\mu(A)$.

We now define the set $A_0 = \{y \in A_0'' s.t. \mu_y(A) > 1 - \eta$. By the previous comment, we can conclude that $\nu(A_0) > \frac{1}{2}\mu(A)$.

Because we $\mathcal{B}_{\alpha_0} \in \mathcal{F}$, the system $(Y, \mathcal{D}_0, \nu, T_0)$ is SZ. Therefore,

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \nu(A_0 \cap T^{-j}A_0 \cap \cdots \cap T^{-kj}A_0) = a > 0.$$

We can show that for $y \in A_0 \cap T^{-j}A_0 \cap \cdots \cap T^{-kj}A_0$, we have the inequality,

$$\mu_y(A \cap T^{-j}A \cap \cdots \cap T^{-kj}A) > \frac{1}{2}.$$

Since for any $B \in \mathcal{B}$ we have that $\mu(B) = \int \mu_y(B) d\nu(y)$, we can use the above inequality to bound the measure of $(\bigcap T^{-jl}A)$ below by $\frac{1}{2}\nu(\bigcap T^{-jl}A_0) > a/2 > 0$. $\square$

We now apply Zorn's Lemma to $\mathcal{F}$, to conclude that there is a maximal factor that is SZ.

8.4 Relative Weak Mixing

In previous work, we have established that if a system is weak mixing it is SZ. We will now define the notion of an extension $(x, \mathcal{B}, \mu, T)$ being weakly mixing *relative* to a factor $(Y, \mathcal{D}, \nu, S)$. The main goal is to show that if an extension $(X, \mathcal{B}, \mu, T)$ is weakly mixing *relative* to a SZ factor $(Y, \mathcal{B}, \mu, T)$, then the extension is SZ as well.

Let us briefly highlight a few facts about weak mixing systems. We say that $(X, \mathcal{B}, \mu, T)$ is weak mixing if the product $T \times T$ is ergodic. This is equivalent to the statement that for all $f, g \in L^\infty(X, \mathcal{B}, \mu)$,

$$\frac{1}{N} \sum \left(\int f T^n g d\mu - \int f d\mu \int g d\mu \right)^2 \to 0.$$

We say that a system $(X, \mathcal{B}, \mu, T)$ is weak mixing relative to a factor $(Y, \mathcal{D}, \nu, S)$ if the system $(X \times_Y X, \tilde{\mathcal{B}}, \tilde{\mu}, \tilde{T})$ is ergodic. For a weakly mixing system, we have that for all $f, g \in L^\infty(X, \mathcal{B}, \mu)$,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int \left(E(f T^n g | \mathfrak{Y}) - E(f | \mathfrak{Y}) S^n E(g | \mathfrak{Y}) \right)^2 d\nu = 0.$$

In the case where f and g are characteristic functions of sets A and B, the previous expression can be rewritten as

$$\int \frac{1}{N} \sum_{n=1}^{N} |\mu_y(A \cap T^{-n} B) - \mu_y(A) \mu_Y(T^{-n} B)|^2 d\nu \to 0.$$

This says that for a relative weak mixing extension, for most n and $y \in Y$, the sets A and $T^{-n}B$ are independent with respect to μ_y. Interpret this as 'relative weak mixing gives weak mixing on the fibers.' The below theorem can be explained intuitively bt saying that if the base action of $(Y, \mathcal{D}, \nu, S)$ is SZ, and the fibers are weak mixing, then the whole system is SZ.

Theorem 5. *Let $(X, \mathcal{B}, \mu, T)$ be a relative weak mixing extension of the system $(Y, \mathcal{D}, \nu, S)$. If the action of S on $\mathcal{D}$ is SZ, then so is that of T on X.*

The proof follows the structure of the proof that weak mixing systems are SZ. We show that relative weak mixing gives a 'relative weak mixing of all orders.' Much like the weak mixing case, this statement is established through an inductive argument using two inequalities: one the condition for realtive weak mixing of order k, and another giving strong convergence in L^2, however we now consider functions projected onto $L^2(Y, \mathcal{D}, \nu)$.

8.5 Compact Extensions and Existence

A function $f \in L^2(X, \mathcal{B}, \mu)$ *almost periodic relative to* $\mathfrak{Y}$ (AP) if for all $\delta > 0$ there exist $g_1, g_2, \ldots g_n \in L^2(X, \mathcal{B}, \mu)$ such that for all $j \in \mathbb{Z}$ we have $\inf_{1 \leq s \leq n} ||T^j f - g_s||_{L^2(\mu_y)} < \delta$, for all $y \in Y$. The extension $(X, \mathcal{B}, \mu, T)$ is a *compact extension* of $(y, \mathcal{D}, \nu, S)$ if the set of AP functions is dense in $L^2(X, \mathcal{B}, \mu)$.

Theorem 6. *Let $(X, \mathcal{B}, \mu, T)$ be a compact extension of $(Y, \mathcal{D}, \nu, S)$. If the system $(Y, \mathcal{D}, \nu, S)$ is SZ then so is $(X, \mathcal{B}, \mu, T)$.*

Some elements of this proof: Fix a set $A \in \mathcal{B}$. Since AP functions are dense, we are able to slightly modify the set A so that $f = 1_A$ is an AP function. We then remove from A the fibers $\pi^{-1}(y)$ for which $\mu_y(A) < \frac{\mu(A)}{2}$. Let A_1 be the set of $y \in Y$, for which $\mu_y(A) \geq \frac{\mu(A)}{2}$. This set A_1 has $\nu(A) > \frac{\mu(A)}{2}$. For $y \in A_1$ we consider the sequence of vectors $\{\mathcal{L}(k, f, y) = (f, T^n f, \cdots, T^{kn} f)\}_{n \in \mathbb{Z}} \in \bigoplus_{l=0}^{k} L^2(\mu_y)$. The fact that f is AP (all $T^j f$ close to $g_i, 1 \leq i \leq s$), tells us that this sequence is totally bounded. This boundedness allows us to find maximal ε-separated sets. We are then able to prove that the system is SZ using ideas similar to those used in proving that a compact system is SZ.

The proof of Szemeredi's theorem can now be concluded by establishing that any proper SZ factor has either a compact extension, or an extension that is relatively weak mixing. For then we can conclude that the maximal SZ factor must be the original system itself. This result is proved in the last section, and is similar to the proof that a system that is not weak mixing has a compact factor.

References

[1] Furstenberg, H., Katznelson, Y., and Ornstein, D., *The Ergodic Theoretic Proof of Szemerédi's Theorem* Comment. Math. Helv. **62** (1987), no. 1, 18–37;

Anne E. McCarthy, Temple University
email: **aemc@math.temple.edu**

9 Multiple recurrence and Szemerédi's theorem

after H. Furstenberg, Y. Katznelson, and D. Ornstein [2]
A summary written by Richard Oberlin

Abstract

In [2] it is shown that Szemerédi's theorem concerning arithmetic progressions is equivalent to a statement about multiple recurrence of measure-preserving transformations, and an ergodic theoretic proof of this statement is given. We summarize the first five sections of [2].

9.1 Introduction

Let $(X, \mathcal{B}, \mu)$ be a probability measure space, and let T be an invertible measure-preserving transformation on $(X, \mathcal{B}, \mu)$. Recurrence is the notion that the orbits of T should return close to their initial position. This was originally shown to occur by Poincaré:

Theorem 1. *Let $A \in \mathcal{B}$ be a set of positive measure. Then for some $n > 0$*

$$\mu(A \cap T^{-n}A) > 0. \tag{1}$$

The proof of the Poincaré recurrence theorem is simple. Suppose that (1) does not hold for any n. Then, since T is measure preserving, $\mu(T^{-j}A \cap T^{-i}A) = \mu(T^{-(j-i)}A \cap A) = 0$ for every $j > i$. Thus, $\bigcup_{j=0}^{\infty} T^{-j}A$ is an essentially disjoint union and has measure $\sum_{j=0}^{\infty} \mu(A) = \infty$. This contradicts the fact that μ is a probability measure.

The following "multiple recurrence" theorem of Furstenberg generalizes Theorem 1, showing that for every k the transformations $T, T^2, \ldots, T^{k-1}$ must exhibit simultaneous recurrence.

Theorem 2. *Let $(X, \mathcal{B}, \mu)$ be a probability measure space and let T be an invertible measure-preserving transformation on $(X, \mathcal{B}, \mu)$. Suppose $A \in \mathcal{B}$ has positive measure. Then for every $k > 0$ there exists an $n > 0$ such that*

$$\mu\left(\bigcap_{j=0}^{k-1} T^{-jn}A\right) > 0.$$

The proof of Theorem 2 is decidedly more subtle than that of Theorem 1. Indeed, Furstenberg observed that Theorem 2 is equivalent to Szemerédi's theorem, a deep result in combinatorial number theory.

Let $\Lambda \subset \mathbb{Z}$. We define the upper-density of Λ

$$UD(\Lambda) := \limsup_{N\to\infty} \sup_{b-a=N} \frac{|\Lambda \cap [a,b)|}{N},$$

where $|\cdot|$ denotes cardinality. Szemerédi's theorem, proven by Szemerédi in 1975, answered a forty year old conjecture of Erdös and Turán.

Theorem 3. *Suppose that $UD(\Lambda) > 0$. Then Λ contains arbitrarily long arithmetic progressions.*

Furstenberg's observation in 1976 that Theorem 3 is equivalent to Theorem 2 thus proved Theorem 2. However, Furstenberg went much further by giving a direct ergodic theoretic proof of Theorem 2, thereby giving a new proof of Szemerédi's theorem. The method of Furstenberg's proof has turned out to be flexible enough to give certain generalizations of Theorem 3 which seem to be inaccessible by Szemerédi's method, and Szemerédi's method gives quantitative estimates related to Theorem 3 which seem to be inaccessible by Furstenberg's method.

9.2 Theorem 2 implies Theorem 3

Here, we focus on the more difficult half of the equivalence of Theorems 2 and 3.

Let $\Lambda \subset \mathbb{Z}$ with $UD(\Lambda) > 0$, and let $k > 0$. We need to find $a \in \mathbb{Z}$ and $b \in \mathbb{Z} \setminus \{0\}$ such that

$$\{a + bj\}_{j=0}^{k-1} \subset \Lambda. \tag{2}$$

Consider the metric on $2^{\mathbb{Z}}$, say,

$$d(x,y) := \sum_{j\in\mathbb{Z}\,:\,x_j\neq y_j} 2^{-|j|} \tag{3}$$

and let T be the shift on $2^{\mathbb{Z}}$, $(Tx)_j := x_{j+1}$. Let $\overline{\omega}$ be the characteristic function of Λ, and let X be the closure in $2^{\mathbb{Z}}$ of $\{T^n\overline{\omega} : n \in \mathbb{Z}\}$. Note that X is compact (since $2^{\mathbb{Z}}$ is compact) and T-invariant.

We define the open subset of X, $A := \{\omega \in X : \omega_0 = 1\}$. Suppose that for some $n > 0$:

$$\bigcap_{j=0}^{k-1} T^{-jn} A \neq \emptyset. \tag{4}$$

Every element of the intersection (4) contains the arithmetic progression $\{jn\}_{j=0}^{k-1}$. Since each $T^{-jn}A$ is open, the intersection (4) is open, and thus by density contains some translate of $\overline{\omega}$. Hence, we obtain (2).

To see (4) from Theorem 2, it remains to find a T-invariant probability measure μ on X such that $\mu(A) > 0$. Such a measure must exist based on the assumption that $UD(\Lambda) > 0$. Indeed, choose $\epsilon > 0$ and sequences $\{a_j\}_{j=1}^{\infty}$ and $\{b_j\}_{j=1}^{\infty}$ such that

$$\lim_{j\to\infty} (b_j - a_j) = \infty \tag{5}$$

and each

$$\frac{|\Lambda \cap [a_j, b_j)|}{b_j - a_j} > \epsilon. \tag{6}$$

Define the sequence of Borel probability measures on X

$$\mu_j(E) := \frac{|\{l \in [a_j, b_j) : T^l \overline{\omega} \in E\}|}{b_j - a_j}.$$

By the Banach-Alaoglu theorem, the closed unit ball in $C(X)^*$ is weak*-compact. Additionally, since X is a compact metric space, $C(X)$ is separable and hence the closed unit ball in $C(X)^*$ with the weak*-topology is metrizable. Thus, some subsequence of $\{\mu_j\}_{j=1}^{\infty}$ converges to a probability measure μ on X (Alternately, one may use a simple diagonalization argument in conjunction with the separability of $C(X)$.). From (6), we see that each $\mu_j(A) > \epsilon$ and hence $\mu(A) \geq \epsilon$. From (5), we see that μ is T-invariant.

9.3 Beginning the proof of Theorem 2

Associated to each measure-preserving system $(X, \mathcal{B}, \mu, T)$ is the unitary operator U on $L^2(X, \mathcal{B}, \mu)$ given by $Uf(x) := f(T(x))$. We first consider Theorem 2 under the assumption that the spectral behavor of U lies on either of two extreme ends, and later we treat the cases in between.

9.3.1 Weak-mixing systems

For certain types of measure-preserving systems, we may strengthen the conclusion of Theorem 1. If $(X, \mathcal{B}, \mu, T)$ is ergodic, it follows from a weak form of the mean ergodic theorem that for any $A_0, A_1 \in \mathcal{B}$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu(A_0 \cap T^{-n} A_1) = \mu(A_0)\mu(A_1). \tag{7}$$

In fact it is easily seen that this only holds when $(X, \mathcal{B}, \mu, T)$ is ergodic.

If we make the stronger assumption that the product system $(X \times X, \mathcal{B} \times \mathcal{B}, \mu \times \mu, T \times T)$ is ergodic, then $(X, \mathcal{B}, \mu, T)$ is said to be *weak-mixing*, and we find that for every $A_0, A_1 \in \mathcal{B}$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |\mu(A_0 \cap T^{-n} A_1) - \mu(A_0)\mu(A_1)| = 0. \tag{8}$$

Again, if (8) holds for every $A_0, A_1 \in \mathcal{B}$, then $(X, \mathcal{B}, \mu, T)$ must be weak-mixing. Additionally, it is well known that $(X, \mathcal{B}, \mu, T)$ is weak-mixing if and only if the unitary operator induced by T has absolutely continuous spectrum except for the eigenspace of constant functions (see [3]).

A measure-preserving system $(X, \mathcal{B}, \mu, T)$ is said to be "weak-mixing of all orders" if for every $k > 0$ and $A_0, \ldots, A_{k-1} \in \mathcal{B}$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left| \mu\left(\bigcap_{j=0}^{k-1} T^{-jn} A_j \right) - \prod_{j=0}^{k-1} \mu(A_j) \right| = 0. \tag{9}$$

Theorem 4. *Every weak-mixing system is weak-mixing of all orders.*

Thus, for weak-mixing systems we obtain a stronger form of Theorem 2.

Let $X = 2^{\mathbb{Z}}$, $\mathcal{B}$ the Borel σ-algebra with respect to (3), μ the product over $\mathbb{Z}$ of the probability measure on $\{0, 1\}$: $\tilde{\mu}(\{0\}) = \tilde{\mu}(\{1\}) = \frac{1}{2}$, and let T be the shift as in Section 9.2. This system is a Bernoulli system. Bernoulli systems are weak-mixing and one may verify directly that they are weak-mixing of all orders.

9.3.2 Compact systems

If T^p is the identity for some $p > 0$ then T is said to be periodic and Theorem 2 is a triviality. If T is "almost periodic" for every $f \in L^2(X, B, \mu)$, i.e. if

the set of orbits of f, $\{f, Tf, T^2f, \ldots\}$, has compact closure in L^2, then the system $(X, \mathcal{B}, \mu, T)$ is said to be *compact*.

A system is compact if and only if the unitary operator induced by T has discrete spectrum, and it is in this way that compact systems are the opposite of weak-mixing systems.

Theorem 5. *Suppose $(X, \mathcal{B}, \mu, T)$ is compact. Then for every $k > 0$ and every set $A \in \mathcal{B}$ of positive measure*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mu\left(\bigcap_{j=0}^{k-1} T^{-jn} A\right) > 0.$$

Thus, we obtain a stronger (though not as strong as (9)) version of Theorem 2 for compact systems (and in fact this is the version that will be proven for general systems).

The class of compact systems is exemplified by the system of irrational rotations of the circle. Let X be the circle, represented by the reals modulo the integers $\mathbb{R}/\mathbb{Z}$, with Lebesgue measure μ and the collection of Borel subsets of X, $\mathcal{B}$. We set $T(x) := x + \alpha$ where $\alpha \in \mathbb{R}$ is fixed (and irrational if one wants to avoid a trivial periodic system).

Let $A \in \mathcal{B}$ be a set of positive measure, and $k > 0$. Since the convolution $\chi_A * \chi_{-A}$ is continuous, we may find $\delta > 0$ such that $\mu(A \cap (A + x)) > \mu(A) - \frac{\mu(A)}{2k}$ when $|x| < \delta$. In particular, this implies that

$$\mu\left(\bigcap_{j=0}^{k-1} (A + jx)\right) > 0 \tag{10}$$

when $|x| < \delta$.

It is well known (see for example [4], page 302) that for any positive integer n, we may find a rational number $\frac{a}{b}$ with $b \leq n$ and

$$\left|\alpha - \frac{a}{b}\right| \leq \frac{1}{b(n+1)}.$$

Choosing $\frac{1}{n+1} < \delta$, we see that T^{-b} is the rotation $y \to y + x$ where $|x| < \delta$. Hence, from (10) we obtain

$$\mu\left(\bigcap_{j=0}^{k-1} T^{-jb} A\right) > 0,$$

and thus Theorem 2 holds for $(X, \mathcal{B}, \mu, T)$.

9.3.3 Weak-mixing and compact factors

Given a measure-preserving system $(X, \mathcal{B}, \mu, T)$, suppose $\mathcal{B}_1 \subset \mathcal{B}$ is a nontrivial T-invariant σ-algebra. Then we may form the new measure-preserving system $(X, \mathcal{B}_1, \mu, T)$. Such a system is referred to as a *factor* of $(X, \mathcal{B}, \mu, T)$. Although we are still far from proving Theorem 2 for arbitrary measure-preserving systems, the following theorem, when combined with Theorems 4 and 5, shows that Theorem 2 at least holds for a factor of every measure-preserving system.

Theorem 6. *A system $(X, \mathcal{B}, \mu, T)$ is weak-mixing if and only if it has no nontrivial compact factors*

Sketch of proof. Suppose $(X, \mathcal{B}, \mu, T)$ is not weak-mixing. Then, there is a nonconstant $\mathcal{B}$-measurable eigenfunction f of T with eigenvalue λ. Since T is unitary, $|\lambda| = 1$. The orbit closure $\overline{\{T^n f\}}_{n=0}^{\infty}$ is isomorphic to $\overline{\{\lambda^n\}}_{n=0}^{\infty}$, and hence is compact. One may verify that if g is $\mathcal{B}_0$-measurable, where $\mathcal{B}_0$ is the smallest σ-algebra such that f is $\mathcal{B}_0$-measurable, then $\overline{\{T^n g\}}_{n=0}^{\infty}$ is compact. Letting $\mathcal{B}_1$ be the smallest σ-algebra containing $\bigcup_{n=0}^{\infty} T^{-n}\mathcal{B}_0$, we see that $\mathcal{B}_1$ is T-invariant and $(X, \mathcal{B}_1, \mu, T)$ is a compact system.

If $(X, \mathcal{B}, \mu, T)$ has a compact factor, then we may choose $A \in \mathcal{B}$ with $0 < \mu(A) < 1$ such that $\overline{\{T^n \chi_A\}}_{n=0}^{\infty}$ has compact closure. This implies that $\|\chi_A - \chi_{T^{-j}A}\|_{L^2}$ may be taken arbitrarily small for a set of j's with positive density. Thus we may not have (8) with $A_0 = A_1 = A$. $\square$

References

[1] Furstenberg, H., *Recurrence in Ergodic Theory and Combinatorial Number Theory.* Princeton University Press, 1981.

[2] Furstenberg, H., Katznelson, Y., and Ornstein, D., *The ergodic theoretical proof of Szemeredi's theorem.* Bull. Amer. Math. Soc. **7** (1982), pp. 527-552.

[3] Halmos, P.R., *Ergodic Theory.* Chelsea Publishing Company, 1956.

[4] Montgomery, H.L., Niven, I., Zuckerman, H.S., *An introduction to the theory of numbers, Fifth Edition.* John Wiley & Sons Inc., 1991.

Richard Oberlin, University of Wisconsin Madison
email: **oberlin@math.wisc.edu**

10 Entropy of Convolutions on the Circle, I

after E. Lindenstrauss, D. Meiri and Y. Peres [5]
A summary written by Robert C. Rhoades

Abstract

We begin with a brief introduction to the main theorems of [5]. Our discussion focuses on the connections of this work with previous works of Furstenberg, Lyons, Johnson, Rudolph, and Host. The starting point is Furstenberg's paper [1].

10.1 Furstenberg's Conjecture

Lindenstrauss, Meiri, and Peres write

> Let $p \geq 2$ be any integer (p need not be prime), and $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ the 1-torus. Denote by σ_p the p-to-one map $x \mapsto px \pmod 1$. The pair $(\mathbb{T}, \sigma_p)$ is a dynamical system that has additional structure: $\mathbb{T}$ is a commutative group (with the group operation being addition mod 1), and σ_p is an endomorphism of it. Even in such a simple system, the interaction between the dynamics and the algebraic structure of $\mathbb{T}$ can be quite subtle; the present work continues the study of this interaction, inspired by the fundamental work of Furstenberg [1].

Furstenberg [1] gave the following very general conjecture about the connection between the group theoretic structure and the dynamics of the system.

Conjecture 2 (Furstenberg's Conjecture, [1])**.** *The only ergodic invariant measures for the semi-group of circle endomorphisms generated by σ_p and σ_q for p and q multiplicatively independent (i.e. $\log p/\log q \notin \mathbb{Q}$, that is p and q are not integral powers of a single integer) are Lebesque measure, and atomic measures concentrated on periodic orbits.*

This problem was out of the reach of Furstenberg, however he was able to prove the corresponding topological result. We say a set $X \subset \mathbb{T}$ is n-invariant if for all $x \in X$, $nx \pmod 1 \in X$. Then the corresponding topological result is:

Theorem 1. *Any infinite closed set in $\mathbb{T}$, invariant under multiplication by two multiplicatively independent integers p and q must be all of $\mathbb{T}$.*

The original conjecture has had a long history and as stated above is still open. The strongest result is due to Johnson [2], who assume the additional condition that μ has positive entropy.

A similar problem that has been studied more recently, and is perhaps more general, is the one where the measure μ is assumed to be σ_p-invariant and we study the action of σ_{c_n} on μ for some sequence $\{c_n\}$. This problem is perhaps most closely related to the conjecture of Furstenberg in the case where $c_n = q^n$. This particular case was studied by Host [3].

One variation of the problem, which seems to be the simplest to consider is to investigate

$$\frac{1}{N}\sum_{n=0}^{N-1} \sigma_{c_n}\mu.$$

In the paper of Host [3], the following theorem is obtained

Theorem 2 (Host, [3])**.** *Let $p, q > 1$ and relatively prime integers. Let μ be a Borel measure on the torus $\mathbb{T}$, invariant under σ_p, ergodic with positive entropy, then*

$$\frac{1}{N}\sum_{n=0}^{N-1} \sigma_{q^n}\mu \to \mu$$

in the weak topology.*

This yields the following definition: Let $\{c_n\}$ be a sequence of integers, then a measure μ of $\mathbb{T}$ is $\{c_n\}$-generic if

$$\frac{1}{N}\sum_{n=0}^{N-1} \sigma_{c_n}\mu \to \lambda$$

in the weak* topology.

In fact, Host proves a much stronger result. Define a sequence measure μ of $\mathbb{T}$ to be $\{c_n\}$-normal if $\{c_n x \pmod 1\}$ is uniformly distributed for μ-almost every $x \in \mathbb{T}$. Then Host proves the following

Theorem 3 (Host, [3])**.** *Let $p, q > 1$ be relatively prime integers. If μ is a Borel measure on the torus, invariant under σ_p, and ergodic with positive entropy, then μ is $\{q^n\}$-normal.*

By Weyl's criterion this result is equivalent to the fact that for all $a \in \mathbb{Z} \setminus \{0\}$,

$$\frac{1}{N}\sum_{n=0}^{N-1} e(aq^n x) \to 0 \ \mu - \text{a.e.},$$

where $e(x) = e^{2\pi i x}$. In turn, this condition implies that μ is $\{q^n\}$-generic.

To show how we can then deduce that $\mu = \lambda$, when $\sigma_q \mu = \mu$, and thus obtain the original desired result of Furstenberg, we apply the following lemma due to Johnson and Rudolph:

Lemma 4 (Johnson and Rudolph [4])**.** *Suppose that $\nu, \nu_1, \nu_2, \cdots$ are invariant measures under σ_p, and that ν is ergodic. Suppose also that*

$$\frac{1}{N}\sum_{n=1}^{N} \nu_n \to \nu$$

in the weak topology. Then there exists a zero-density set $J \subset \mathbb{N}$ such that $\nu_n \to \nu$ (weakly) as $n \to \infty$ for $n \notin J$.*

This line of reasoning shows how Furstenberg's original conjecture is the weakest in a string of results about the interaction between algebraic structure of $\mathbb{T}$ and the dynamics on $\mathbb{T}$. From here on out we will focus our attention not on Frustenberg's conjecture but on the other results that are more general and stronger.

10.2 Uniform Distribution and $\{c_n\}$-genericity

We have seen that via Weyl's criterion we can move from a uniform distribution result about a measure μ and a sequence $\{c_n\}$ to a result about the "on average" behavior of $\sigma_{c_n}\mu$. This approach was originally due to Host [3]. His approach leads naturally to the definition of the p-adic collision exponent: Given an integer-sequence $\{c_n\}$ and an integer $p > 1$, we define the p-adic collision exponent of the sequence as

$$\Gamma_p(\{e_n\}) = \limsup_{n\to\infty} \frac{\log |\{0 \le k, \ell < p^n : c_k \equiv c_\ell \pmod{p^n}\}|}{n \log p}.$$

Remark 5. *Since $k = \ell$ is allowed we know that we will always have $1 \le \Gamma_p(\{c_n\}) \le 2$.*

Meiri greatly generalized Host's result by proving the following theorem

Theorem 6 (Meiri [6], Proposition 2.2 [5]). *Fix an integer $p > 1$ and a sequence $\{c_k\}$ with p-adic collision exponent < 2. Then there exists a constant $h_0 < \log p$ such that every p-invariant and ergodic measure μ with $h(\mu, \sigma_p) > h_0$ is $\{c_n\}$-normal. In particular, μ is $\{c_n\}$-generic.*

The condition on the p-adic collision exponent does not seem too strict. However, in contrast, the positive entropy condition seems to play an important role in this theorem. Having a large entropy means that the measure is not structured. (Is the positive entropy here analogous to the positive upper-density in Szemeredi's theorem?)

10.3 The Role of Entropy

The role of positive entropy seems to be an important one. The big idea from the paper [5] is to take a measure that does not have large enough entropy and some how create a new measure with greater entropy which will have large enough entropy to have the desired $\{c_n\}$-generic property. Then show that the new measure having this property implies that the old one had this property. This is the basic idea of the bootstrap.

The idea is that convolution will help us create new measures that have increased entropy. Theorem 1 of [5] suggests that convolution may be a helpful tool in this problem.

Theorem 7 (Theorem 1.1 [5]). *Let $\{\mu_i\}$ be a sequence of p-invariant and ergodic measures on $\mathbb{T}$ whose normalized base-p entropies $h_i = h(\mu_i, \sigma_p)/\log p$ satisfy*

$$\sum_{i=1}^{\infty} \frac{h_i}{|\log h_i|} = \infty.$$

Then

$$h(\mu_1 * \cdots * \mu_n, \sigma_p) \to \log p$$

monotonically as $n \to \infty$. In particular, $\mu_1 * \cdots * \mu_n \to \lambda$ weak and in the $\overline{d}$ metric (with respect to the base-p partition).*

Convolution also has the following property

Proposition 8 (Proposition 2.4 [5]). *An invariant measure μ is $\{c_n\}$-generic if and only if $\mu * \mu$ is $\{c_n\}$-generic.*

Remark 9. *This proposition deals with $\{c_n\}$-generic not $\{c_n\}$-normal. This is the reason why Theorem 1.4 of [5] is about $\{c_n\}$-generic and not $\{c_n\}$-normal.*

With all this in mind we state the Bootstrap Lemma:

Theorem 10 (The Bootstrap Lemma, Theorem 1.7 [5])**.** *Suppose that $\mathcal{C}$ is a class of p-invariant measures on $\mathbb{T}$ with the following properties:*

1. *If μ is p-invariant and ergodic and $\mu * \mu \in \mathcal{C}$, then $\mu \in \mathcal{C}$.*
2. *If μ is p-invariant and almost every ergodic component of μ is in $\mathcal{C}$, then $\mu \in \mathcal{C}$.*
3. *There exists some constant $h_0 < \log p$ such that every p-invariant and ergodic measure μ with $h(\mu, \sigma_p) > h_0$ is in $\mathcal{C}$.*

Then $\mathcal{C}$ contains all p-invariant ergodic measures with positive entropy.

Now we can get back to proving the type of $\{c_n\}$-generic results that we are looking for:

Theorem 11 (Theorem 1.4 [5])**.** *Let $\{c_n\}$ be a sequence with p-adic collision exponent < 2, for some $p > 1$. Then any p-invariant ergodic measure μ on $\mathbb{T}$ with positive entropy is $\{c_n\}$-generic.*

Proof. The proof follows from the Bootstapping Lemma with $\mathcal{C}$ being the class of $\{c_n\}$-generic measures where $\{c_n\}$ has p-adic collision exponent less than 2. We apply Theorems 6 and 8 to check the first and third conditions in the Bootstrapping Lemma. All that remains is verify that if μ is p-invariant and almost every ergodic component of μ is $\{c_n\}$-generic then μ is $\{c_n\}$-generic. □

In fact, with a modified version of the Bootstrap Lemma it is possible to give a strengthened version of this theorem. For $k \in \mathbb{Z} \setminus \{0\}$, define

$$g_N^{(k)} = \frac{1}{N} \sum_{n=0}^{N-1} e(kc_n x).$$

With these functions $\{c_n\}$-generic is equivalent to

$$\int g_N^{(k)}(x) d\mu \to 0.$$

The following slightly stronger version of Theorem 11 is obtained as well:

Theorem 12 (Theorem 1.5 [5]). *Under the same conditions as Theorem 11, μ is $\{c_n\}$-normal in probability. Equivalently*

$$\int |g_N^{(k)}(x)|^2 d\mu \to 0,$$

as $N \to \infty$ for $k \neq 0$.

This finishes the summary of the relation of the results in [5] to the conjecture of Furstenberg as well as results of Lyons, Johnson, Ruldolph and Host. In the talks I will give on this paper I will present some background on topics such as Hausdorff dimension, Entropy, and their relationship with one-another. I will also give in detail the proofs of Theorems 10 and 11. Time permitting I will finish with some discussion of examples and open questions related to [5].

References

[1] H. Furstenberg, *Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation*, Math. Systems Theory **1** (1967), 1-49.

[2] A. Johnson, *Measures on the circle invariant under multiplication by a nonlacunary subsemigroup of the integers*, Israel J. Math. **77** (1992), 211 - 240.

[3] B. Host, *Nombres normaux, entropie, translations*, Isreal J. Math. **91** (1995), 419-428.

[4] A. Johnson and D. Rudolph, *Convergence under $\times_p$ of $\times_q$ invariant measures on the circle*, Adv. in Math. **115** (1995), 117-140.

[5] E. Lindenstrauss, D. Meiri and Y. Peres, *Entropy of convolutions on the circle* Ann. of Math. **149** (1999), 871-904.

[6] D. Meiri, *Entropy and uniform distribution of orbits in $\mathbb{T}^d$*, Israel J. Math. **105** (1998), 155 - 183.

Robert C. Rhoades, University of Wisconsin
email: **rob.rhoades@gmail.com**

11 Entropy of convolutions on the circle

after Elon Lindenstrauss, David Meiri and Yuval Peres [3]
A summary written by Shuanglin Shao

Abstract

Given ergodic p-invariant measures $\{\mu_i\}$ on the torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, we give a sharp condition on their entropies, guanranteeing that the entropy of the convolution $\mu_1 * \cdots * \mu_n$ converges to $\log p$. we also prove a variant of this result for joinings of full entropy on $\mathbb{T}^{\mathbb{N}}$. We also obtain the following corollary concerning the Hausdorff dimension of sum sets: For any sequence $\{S_i\}$ of p-invariant closed subsets of $\mathbb{T}$, if $\sum \dim_H(S_i)/|\log \dim_H(S_i)| = \infty$, then $\dim_H(S_1 + \cdots + S_n) \to 1$

11.1 Introduction

Let $p \geq 2$ be any integer(p need not to be prime), and $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ the 1-torus. Denote by σ_p the p-to-one map $x \mapsto px(\bmod 1)$. The pair $(\mathbb{T}, \sigma_p)$ is a dynamical system that has additional structure: $\mathbb{T}$ is a commutative group(with the group operation being addition mod 1), and σ_p is an endormorphism of it. Inspired by the fundamental work of Furstenberg([2]), the present work continues the study of the intersection between the dynamics and the algebraic structure of $\mathbb{T}$. Say that a measure μ on $\mathbb{T}$ is p-invariant if $\sigma_p\mu = \mu$, where for every set $A \subset \mathbb{T}$

$$(\sigma_p\mu)(A) = \mu(\sigma_p^{-1}A)$$

(All measures we consider are Borel probability measure). Lebegue measure on $\mathbb{T}$, denoted by λ, has entropy $\log p$ with respect to the σ_p, and is the unique p-invariant measure of maximal entropy. Given two p-invariant measures μ and ν, the group structure of $\mathbb{T}$ naturally yields another p-invariant measure-the convolution $\mu * \nu$. Our main results, Theorem 1 and Theorem 4, concern the entropy growth for convolutions of p-invariant measures and their ergodic components. These results have applications to the Hausdorff dimension of sum sets and to genericity of the orbits of measures with positive entropy under multiplication by certain integer sequences.

Theorem 1 (The Convolution theorem)**.** *Let $\{\mu_i\}$ be a sequence of p-invariant and ergodic measures on $\mathbb{T}$ whose normalized base-p entropies $h_i = h(\mu_i, \sigma_p)/\log p$ satisfy*

$$\sum_{i=1}^{\infty} \frac{h_i}{|\log h_i|} = \infty. \tag{1}$$

Then

$$h(\mu_1 * \cdots * \mu_n, \sigma_p) \to \log p, \quad \textit{monotonically}, \quad \textit{as} \quad n \to \infty.$$

In particular, $\mu_1 * \cdots * \mu_n \to \lambda$ weak*.

It is relatively easy to see that under hypotheses of theorem, $\mu_1 * \cdots * \mu_n \to \lambda$ weak*. This means that

$$\int f(x) d\mu_1 * \cdots * \mu_n \to \int f(x) d\lambda$$

for all continuous f, and gives very little information on the dynamics of $\mu_1 * \cdots * \mu_n$. The entropy condition in Theorem 1 is sharp: if $\{h_i\}$ is a sequence of numbers in the range $(0, 1)$ with $\sum h_i/|\log h_i| < \infty$, then there exists a sequence of p-invariant ergodic measures $\{\mu_i\}$, such that $h_i = h(\mu_i, \sigma_p)/\log p$, yet $\mu_1 * \cdots * \mu_n$ doesn't converge to Lebesgue measure λ even in the weak* topology. The convolution theorem has implications for Hausdorff dimension of sum sets:

Corollary 2. *Let $\{S_i\}$ be a sequence of p-invariant closed subsets of $\mathbb{T}$, and suppose that*

$$\sum_{i=1}^{\infty} \frac{\dim_H(S_i)}{|\log \dim_H(S_i)|} = \infty$$

Then $\dim_H(S_1 + \cdots + S_n) \to 1$.

Is the dimension condition of Theorem 2 sharp as well? specifically, given a sequence of numbers $0 < d_i < 1$ such that $\sum d_i/|\log d_i| < \infty$, can one always find p-invariant closed subsets $S_i \subset \mathbb{T}$ with $\dim_H S_i = d_i$ and $\lim \dim_H(S_1 + \cdots + S_n) \to < 1$? Currently we can construct sets satisfying the desired conclusion, but only when $\sum d_i/|\log d_i|$ is small enough.

Theorem 3. *Let $\{\mu_i\}_{i=1}^{\infty}$ be a sequence of p-invariant and ergodic measures on $\mathbb{T}$ such that* $\inf_i h(\mu_i, \sigma_p) > 0$. *Suppose that $\tilde{\mu}$ is a joining of full entropy of $\{\mu_i\}$. Define $\Theta^n : \mathbb{T}^{\mathbb{N}} \to \mathbb{T}$ by $\Theta^n(x) = x_1 + \cdots + x_n (\mathrm{mod}\, 1)$. Then*

$$h(\Theta^n \tilde{\mu}, \sigma_p) \to \log p, \quad \textit{monotonically}, \quad \textit{as} \quad n \to \infty.$$

Theorem 3 is not valid under the weaker entropy assumptions of Theorem 1. Indeed, it is possible to find a joining of full entropy $\tilde{\mu}$ with entropies satisfying (1), such that $\Theta^n \tilde{\mu}$ doesn't even converge weak* to λ.

Theorem 4. *Let $\{\mu_i\}_{i=1}^{\infty}$ be a sequence of p-invariant and ergodic measures on $\mathbb{T}$ such that $\inf_i h(\mu_i, \sigma_p) > 0$. Let $\{S_i\}$ be a sequence of Borel sets of $\mathbb{T}$, and suppose that $\mu_i(S_i) > 0$ for all $i \geq 1$. Then $\dim_H(S_1 + \cdots + S_n) \to 1$.*

This summary is organized as follows and we will list the main theorems in each section for convenience. In Section 11.2, we use the connection between entropy of measures and Hausdorff dimension, to derive Theorem 2 and Theorem 4 from the results about convolutions of measures. In Section 11.3-11.5, we prove our main results, Theorem 1 and Theorem 4. Section 11.3 contains results about finite cyclic groups which are crucial to the proof of the convolution theorems. Lemma 7 and Lemma 8 study convolutions of measures on a finite cyclic group and contain one key idea in the proof, namely that the convolution of a sequence of measures on a finite cyclic group of order N(we shall use $N = p^k$) tends to be invariant under a subgroup that will typically rather large (in the cases we will be interested in, this subgroup will be of order approximately $p^{\alpha k}$ for some $0 < \alpha < 1$). Lemma 9 shows that if a measure on $\mathbb{Z}/p^k\mathbb{Z}$ is almost invariant under a subgroup of size $p^{\alpha k}$, the distribution of the αk high order digits is nearly uniform. In section 11.4 we begin to show how convolutions of measures on $\mathbb{Z}/p^k\mathbb{Z}$ relate to convolutions of measures on $\mathbb{T}$, where we get measures on $\mathbb{Z}/p^k\mathbb{Z}$ from measures on $\mathbb{T}$ by considering the conditional distributional of the first k digits in the base p expansion of $x \in \mathbb{T}$, given the rest digits. In section 11.5 we continue this approach and prove Theorem 1. The basic argument is that if that if the entropy of $mu_1 * \cdots * \mu_n$ is almost

$$\sup_{N \in \mathbb{N}} h(\mu_1 * \cdots * \mu_N, \sigma_p),$$

then for any $k \geq 1$ the distribution of the first k digits of x given the rest of the digits (chosen according to $\mu_1 * \cdots * \mu_n$) must be nearly invariant under a subgroup $G \in \mathbb{Z}/p^k\mathbb{Z}$ of size $p^{\alpha k}$-for if it was not, then the entropy of $mu_1 * \cdots * \mu_n$ can significantly increase by further convolutions. Using Lemma 9, this implies that the first αk digits of x are distributed nearly uniformly. Since k is arbitrary, we it follows that $h(\mu_1 * \cdots * \mu_n, \sigma_p) \backsimeq \log p$.

11.2 Dimension of sum sets

For any any measure μ, let

$$\dim \mu = \inf\{\dim_H S| \quad \text{S is a Borel set with} \quad \mu(S) = 1\}. \tag{2}$$

Define the *lower dimension* of μ

$$\underline{\dim}\mu = \inf\{\dim_H S| \quad \text{S is a Borel set with} \quad \mu(S) > 0\}. \tag{3}$$

Lemma 5 (Billingsley)**.** *Let μ be a positive finite measure on $\mathbb{T}$. Assume $K \subset \mathbb{T}$ is a Borel set satisfying $\mu(K) > 0$ and*

$$K \subset \{x \in \mathbb{T} : \liminf_{\epsilon \to 0} \frac{\log \mu[B_\epsilon(x)]}{\log \epsilon} \leq \gamma\}.$$

Then $\dim_H K \leq \gamma$. If the $\liminf$ is γ a. e. , then $\dim_H K = \gamma$.

In [4], Meiri and Peres proved that $\dim \mu_\theta = \text{ess-sup}_\theta \dim \mu_\theta$ in a more general context. We have the equivalent statement for lower dimension.

Theorem 6. *Let μ be a p-invariant measure on $\mathbb{T}$, and denote its ergodic decomposition by $\mu = \int \nu_\theta d\theta$. Then $\underline{\dim}\mu_\theta = \text{ess-inf}_\theta \dim \mu_\theta$.*

11.3 Uniform distribution on subgroups

The following lemmas are the key to prove Theorem 1.

Lemma 7. *Let $\{X_n\}$ be an infinite sequence of independent random variables with values in $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$, for some fixed integer $N > 1$. Suppose that for some non-zero $g \in \mathbb{Z}_N$,*

$$\sum_{j=1}^{\infty} \sum_{x=0}^{N-1} \min\{\mathbb{P}(X_j = x), \mathbb{P}(X_j = x + g)\} = \infty. \tag{4}$$

Let $S_n = X_1 + \cdots + X_n \pmod N$. Then for any $x \in \mathbb{Z}_N$,

$$\lim_{n \to \infty} (\mathbb{P}(S_n = x + g) - \mathbb{P}(S_n = x)) = 0. \tag{5}$$

Lemma 8. *Let $\{X_n\}$ be an infinite sequence of independent random variables with values in $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$, for some fixed integer $N > 1$. Suppose that there exists a subgroup $G \subset \mathbb{Z}_N$, generated by $g_1, \ldots, g_r$, such that (4) holds for $g = g_1, \ldots, g_r$. let $S_n = X_1 + \cdots + X_n \pmod N$, and let $S_n \bmod G$ denote the projection of S_n to $\mathbb{Z}_N/G$. Then*

$$\mathbb{E}H(S_n|S_n \bmod G) \to \log|G|.$$

Lemma 9. *Let Y be an $\mathbb{Z}_{p^k}$-valued random variable. Let G be a subgroup of $\mathbb{Z}_{p^k}$, and suppose that n satisfies $p^n \geq |G|$. Then*

$$H(\pi_n(Y)) \geq H(Y|Y \bmod G) = H(Y) - H(Y \bmod G).$$

11.4 Entropy and subgroups

Proposition 10. *Let μ, ν be two p-invariant measures, and G a subgroup of $\mathbb{Z}/p^k\mathbb{Z}$ for some $k \in \mathbb{N}$. Then*

$$\mathbb{E}H_{\mu*\nu}(x_{1\ldots k} \bmod G|x_{k+1\ldots\infty}) \geq \mathbb{E}H_{\mu}(x_{1\ldots k} \bmod G|x_{k+1\ldots\infty})$$

Corollary 11. *let $\{\mu_i\}$ be a sequence of p-invariant measures, and denote $\tilde{\mu} = \Pi\mu_i$. Suppose that G is a subgroup of $\mathbb{Z}/p^k\mathbb{Z}$ for some $k \in \mathbb{N}$. Then*

$$\mathbb{E}H_{\tilde{\mu}}(\theta^n_{1\ldots k} \bmod G|\theta^n_{k+1\ldots\infty})$$

is monotone nondecreasing in n.

Lemma 12. *Let μ be a measure on $\mathbb{T}$, and suppose that $G \subset \mathbb{Z}/p^k\mathbb{Z}$ is a group of size $\geq p^l$. Then*

$$H(\alpha_{1\ldots l}) \geq (l-1)\log p - \log|G| + \int H(\alpha_{1\ldots k}|\alpha_{1\ldots k} \bmod G \vee \alpha_{k+1\ldots\infty})d\mu$$

11.5 The convolution theorem

Lemma 13. *Let $G \subset \mathbb{Z}_{p^k}$, and suppose that μ and ν are non-atomic measures on $\mathbb{T}$. Then*

$$\mathbb{E}I_k(\mu * \nu) - \mathbb{E}I_k(\mu) = \mathbb{E}I_G(\mu * \nu) - \mathbb{E}I_G(\mu).$$

Lemma 14. *Let $\{\mu_n\}$ be a sequence of probability measures on $\mathbb{T}$, and form the product measure $\tilde{\mu} = \Pi\mu_i$. Suppose that for some fixed number k,*

$$\sum_{i=1}^{\infty} \mathbb{E}\psi^{-1}(H_{\mu_i}(x_k^i | t_{k+1\ldots\infty}^i)) = \infty. \tag{6}$$

Then for $\tilde{\mu}$-almost every $t \in \mathbb{T}^{\mathbb{N}}$ there exists a group $G_k(t) \subset \mathbb{Z}^{p^k}$ such that

$$\mathbb{E}I_{G_k(t)}(\mu_1 * \cdots * \mu_n) \to \mathbb{E}\log|G_k(t)|. \tag{7}$$

Furthermore, $|G_k(t)| \geq p_*^k$ a. e. , where p_* is the smallest prime factor of p.

Lemma 15. *Under the assumptions of Lemma 14, define*

$$h_k = \sup_n \frac{1}{k}\mathbb{E}H_{\mu_1 * \cdots * \mu_n}(x_{1\ldots k} | x_{k+1\ldots\infty}). \tag{8}$$

For any m, if $\frac{1}{k}\mathbb{E}H_{\mu_1 * \cdots * \mu_m}(x_{1\ldots k} | x_{k+1\ldots\infty}) > h_k - \epsilon$, then

$$H_{\mu_1 * \cdots * \mu_m}(t_{1\ldots l}) \geq (l-1)\log p - (k+1)\epsilon, \tag{9}$$

Where $l = l(k) = \llcorner k\frac{\log p_*}{\log p}\lrcorner$.

References

[1] . Billingsley, *Ergodic Theorey and Information*, (1965), Wiley, New York.

[2] H. Furstenberg, *Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation*, Math. Systems Theory **1** (1967), 1-49.

[3] E. Lindenstrauss, D. Meiri and Y. Peres *Entropy of convolutions on the circle*, Ann. of Math. **149** (1999), 871-904.

[4] D. Meiri and Y. Peres, *Bi-invariant sets and measures have integer hausdorff dimension*, Ergodic Theory and Dynamical Systems, **19** (1999), no. 2, 523–534.

Shuanglin Shao, UCLA
email: **slshao@math. ucla. edu**

12 Bourgain's Entropy Estimates

after J. Bourgain [2]
A summary written by John T. Workman

Abstract

It is shown that if certain sequences of operators (S_n) converge almost surely on some L^p, then there is a uniform estimate on the L^2-entropy of $\{S_n f : n \in \mathbb{N}\}$ for all f in the L^2-unit ball. This shows that there are bounded functions for which Bellow's averages do not converge almost surely.

12.1 Preliminaries

Let (X, μ) be a probability space. Let $(T_j)_{j\in\mathbb{N}}$ be a sequence of linear operators from the X-measurable functions to the X-measurable functions satisfying: $T_j : L^1(\mu) \to L^1(\mu)$ are bounded; $T_j : L^2(\mu) \to L^2(\mu)$ are isometries; T_j are positive, i.e., if $f \geq 0$ a.s.$[\mu]$ then $T_j(f) \geq 0$ a.s.$[\mu]$; and $T_j(1) = 1$.

Also, assume the T_j satisfy a mean ergodic condition. Namely, for every $1 \leq p < \infty$

$$\frac{1}{J}\sum_{j=1}^{J} T_j f \to \int_X f(x)\,\mu(dx) \quad \text{in } L^p(\mu)\text{-norm, for all } f \in L^p(\mu).$$

Let $(S_n)_{n\in\mathbb{N}}$ be another sequence of linear operators satisfying $S_n : L^2(\mu) \to L^2(\mu)$ with $\|S_n(f)\|_{L^2(\mu)} \leq \|f\|_{L^2(\mu)}$, and $T_j S_n = S_n T_j$ for all j, n. With these hypotheses, we gain the following results of Bellow and Jones [1].

Lemma 1. *Suppose that for some $1 \leq p < \infty$, $S_n f$ converges a.s.$[\mu]$ for all $f \in L^p(\mu)$. Then, there is a finite-valued function $C(\epsilon)$, for $\epsilon > 0$, so that*

$$\mu\Big\{\sup_n |S_n f| \leq C(\epsilon)\Big\} > 1 - \epsilon$$

for every $f \in L^\infty(\mu)$, $\|f\|_{L^p(\mu)} \leq 1$.

Lemma 2. *Suppose $S_n f$ converges a.s.$[\mu]$ for all $f \in L^\infty(\mu)$. Then, there is a finite-valued function $\delta(\epsilon)$, for $\epsilon > 0$, so that $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ and*

$$\int_X \sup_n |S_n f|\, d\mu < \delta(\epsilon)$$

whenever $\|f\|_{L^\infty(\mu)} \leq 1$ and $\|f\|_{L^1(\mu)} < \epsilon$.

The theory of normal random variables and Gaussian processes plays a pivotal role in the proofs of both entropy results. Let (Ω, P) be another probability space. Recall, a random variable g on Ω is normal (or Gaussian) with mean m and variance σ^2 if it has the distribution function

$$f(\omega) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\omega - m)^2}{2\sigma^2}\right).$$

A normal random variable with mean 0 and variance 1 is called a standard normal random variable.

Let T be a countable indexing set. A stochastic process $(G_t : t \in T)$ is called a Gaussian process if all finite linear combinations $\sum_t a_t G_t$ are normal random variables, and each G_t has mean 0. We define a pseudo-metric on T by $d_G(s,t) = \|G_s - G_t\|_{L^2(P)}$. Denote the entropy number of T by $N(T, d_G, \delta)$, the minimal number of δ-balls, under d_G, needed to cover T. The following fundamental result is Sudakov's inequality.

Theorem 3. *There is a constant $R > 0$ such that if $(G_t : t \in T)$ is a Gaussian process, then*

$$\sup_{\delta > 0} \delta \sqrt{\log N(T, d_G, \delta)} \leq R \left\| \sup_{t \in T} |G_t| \right\|_{L^1(P)}.$$

12.2 The First Entropy Result

For $\delta > 0$ and $f \in L^2(\mu)$, let $N_f(\delta)$ be the δ-entropy number of the set $\{S_n f : n \in \mathbb{N}\}$ in $L^2(\mu)$, i.e., the minimum number of δ-balls in $L^2(\mu)$ needed to cover $\{S_n f : n \in \mathbb{N}\}$. The following theorem is the first of Bourgain's entropy results.

Theorem 4. *Suppose that for some $1 \leq p < \infty$, $S_n f$ converges a.s.[μ] for all $f \in L^p(\mu)$. Then, there exists $C > 0$ such that $\delta(\log N_f(\delta))^{1/2} \leq C$ for all $\delta > 0$ and $\|f\|_{L^2(\mu)} \leq 1$.*

We give a brief idea of the proof. Let $(g_j)_{j \in \mathbb{N}}$ be a sequence of independent standard normal random variables on (Ω, P). As (X, μ) is a probability space, $L^r(\mu) \supset L^q(\mu)$ when $r < q$. So, assume without loss of generality that $p \geq 2$. Fix $f \in L^2(\mu)$, $\|f\|_{L^2(\mu)} \leq 1$. As $L^\infty(\mu)$ is dense in $L^2(\mu)$, we may also assume $f \in L^\infty(\mu)$.

For $M \in \mathbb{N}$, denote $\overline{M} = \{1, 2, \ldots, M\}$, and define a pseudo-metric on $\mathbb{N}$ by $d_f(n, n') = \|S_n f - S_{n'} f\|_{L^2(\mu)}$. Note, $\sup_M N(\overline{M}, d_f, \delta) = N_f(\delta)$. Fix M.

As $J^{-1} \sum T_j(f^2) \to \|f^2\|_{L^1(\mu)} = \|f\|^2_{L^2(\mu)} \leq 1$ in $L^{p/2}(\mu)$-norm (because $p \geq 2$), choose J so large that

$$\left\| \frac{1}{J} \sum T_j(f^2) \right\|_{L^{p/2}(\mu)} \leq 2. \tag{1}$$

For each pair $n, n' \leq M$, $J^{-1} \sum T_j((S_n f - S_{n'} f)^2) \to \|S_n f - S_{n'} f\|^2_{L^2(\mu)} = d_f(n, n')^2$ in $L^1(\mu)$-norm, and thus in probability. So, also choose J big enough that

$$X' = \left\{ \left(\frac{1}{J} \sum T_j\big((S_n f - S_{n'} f)^2\big) \right)^{1/2} \geq \frac{1}{2} d_f(n, n') \text{ for all } n, n' \in \overline{M} \right\}$$

has large probability. Define $F(x, \omega) = J^{-1/2} \sum_{j=1}^{J} g_j(\omega) T_j f(x)$.

Let $F^*(x, \omega) = \sup_{n \leq M} |S_n F(x, \omega)|$. From Lemma 1, (1), and some general probability results, there is a "universal" constant R' (independent of f, M, and J) such that

$$X'' = \left\{ x \in X : \int_\Omega F^*(x, \omega)\, P(d\omega) \leq R' \right\}$$

has large enough μ-probability, so that $\mu(X' \cap X'') > 0$. Fix $\overline{x} \in X' \cap X''$. Define a Gaussian process by $G_n(\omega) = S_n F(\overline{x}, \omega) = J^{-1/2} \sum g_j(\omega) T_j S_n f(\overline{x})$. By Sudakov's inequality and as $\overline{x} \in X''$,

$$\begin{aligned} \sup_{\delta > 0} \delta (\log N(\overline{M}, d_G, \delta))^{1/2} &\leq R \int_\Omega \sup_{n \in \overline{M}} |G_n(\omega)|\, P(d\omega) \\ &= R \int_\Omega F^*(\overline{x}, \omega)\, P(d\omega) \leq RR'. \end{aligned}$$

On the other hand, as $\overline{x} \in X'$,

$$d_G(n, n') = \left(\frac{1}{J} \sum T_j (S_n f - S_{n'} f)^2(\overline{x}) \right)^{1/2} \geq \frac{1}{2} d_f(n, n'),$$

where the equality follows from a simple result on normal random variables and the properties of (T_j). Thus, $N(\overline{M}, d_f, \delta) \leq N(\overline{M}, d_G, \delta/2)$ for all δ, so that $\delta(\log N(\overline{M}, d_f, \delta))^{1/2} \leq 2RR' =: C$. Taking the supremum over M, $\delta(\log N_f(\delta))^{1/2} \leq C$ for all $\delta > 0$.

12.3 The Second Entropy Result

Theorem 5. *Suppose $S_n f$ converges a.s.[μ] for all $f \in L^\infty(\mu)$. Then, there exists a finite-valued function $C(\delta)$, for $\delta > 0$, such that $N_f(\delta) \leq C(\delta)$ for all $\delta > 0$ and $\|f\|_2 \leq 1$.*

The proof is by contradiction. Suppose there is some $\delta > 0$ such that $N_f(\delta)$ is unbounded over the $L^2(\mu)$ unit ball. Fix $K \in \mathbb{N}$, $K > 1$. Then, there is some $f \in L^\infty$, $\|f\|_{L^2(\mu)} \leq 1$ such that $N_f(\delta) > K$. In particular, there is a subset $I \subset \mathbb{N}$ with $|I| = K$ (cardinality) and $\|S_n f - S_{n'} f\|_{L^2(\mu)} > \delta$ for all $n \neq n' \in I$.

Like in the previous proof, we must choose an appropriately large J, but we suppress those details. Define $F(x,\omega) = J^{-1/2} \sum_{j=1}^{J} g_j(\omega) T_j f(x)$. Write $F(x,\omega) = \varphi(x,\omega) + H(x,\omega)$ where

$$\varphi(x,\omega) = F(x,\omega)\chi_{\{|F(x,\omega)| \leq 6\sqrt{\log K}\}} \quad \text{and}$$
$$H(x,\omega) = F(x,\omega)\chi_{\{|F(x,\omega)| > 6\sqrt{\log K}\}}.$$

Much like the first proof, here we must show that three sets (in Ω) have sufficiently large probability, so that there is some $\overline{\omega}$ in all three sets. This is done by taking J large enough initially, an application of Sudakov's inequality, and more results in probability theory. This gives $\overline{\omega} \in \Omega$ such that

$$\int_X \sup_{n \in I} |S_n H(x,\overline{\omega})| \, \mu(dx) \leq c_1, \tag{2}$$

$$\int_X \sup_{n \in I} |S_n F(x,\overline{\omega})| \, \mu(dx) \geq c_2 (\log K)^{1/2}, \tag{3}$$

$$\int_X |\varphi(x,\overline{\omega})| \, \mu(dx) \leq c_3, \tag{4}$$

for universal constants c_1, c_2, c_3. By (2) and (3) we see

$$\int_X \sup_I |S_n \varphi(x,\overline{\omega})| \, \mu(dx) \geq c_2 (\log K)^{1/2} - c_1. \tag{5}$$

Define $\psi_K(x) = \frac{1}{6}(\log K)^{-1/2} \varphi(x,\overline{\omega})$. Simply by construction, we have $|\psi_K| \leq 1$. By (5), $\int_X \sup_I |S_n \psi_K(x)| \, \mu(dx) \geq c_2/6 - c_1(\log K)^{-1/2}/6$. Finally, by (4), we see $\|\psi_K(x)\|_{L^1(\mu)} \leq c_3(\log K)^{-1/2}/6$. As $K \in \mathbb{N}, K > 1$ was arbitrary, we can create a sequence ψ_K satisfying these three conditions. However, $c_3(\log K)^{-1/2}/6 \to 0$, while $c_2/6 - c_1(\log K)^{-1/2}/6 \geq c_2/12 > 0$ for large K. This contradicts Lemma 2, and completes the proof.

12.4 Application: Bellow's Averages

Let $\mathbf{T} = \mathbb{R}/\mathbb{Z}$. Then, a function $f : \mathbb{R} \to \mathbb{C}$ with period 1 can be viewed as a function on $\mathbf{T}$. Let m be Lebesgue measure, and consider the probability space $(\mathbf{T}, m)$. Let (a_j) be any non-zero sequence of real numbers which converge to 0. For functions $f : \mathbf{T} \to \mathbb{C}$, consider the operators

$$S_n f(x) = \frac{1}{n} \sum_{j=1}^{n} f(x + a_j).$$

Bellow asked whether $S_n f$ converges to f a.s.$[m]$ for all $f \in L^1(m)$. The answer to this turns out to be no. In fact, it is not even true for all $f \in L^\infty(m)$. We prove this using the second entropy criterion. First, we need the following result.

Lemma 6. *Let (a_j) be a sequence of non-zero real numbers converging to 0. Then, given any $r \in \mathbb{N}$, there exist integers $J_1 < J_2 < \ldots < J_r$ satisfying the following: if $\bar{\alpha} = (\alpha_1, \ldots, \alpha_r)$ is a vector of 0's and 1's, then there is an integer $n(\bar{\alpha})$ such that for all $1 \leq s \leq r$*

$$\left| 1 - J_s^{-1} \sum_{j \leq J_s} e^{2\pi i a_j n(\bar{\alpha})} \right| < \frac{1}{10} \quad \text{if } \alpha_s = 0,$$

$$\left| 1 - J_s^{-1} \sum_{j \leq J_s} e^{2\pi i a_j n(\bar{\alpha})} \right| > \frac{1}{2} \quad \text{if } \alpha_s = 1.$$

The proof of this lemma is somewhat technical, but it is straightforward and relies only on some basic complex arithmetic. We now proceed to the solution of Bellow's question.

Theorem 7. *Let (a_j) be any sequence of real numbers which converge to 0 and $a_j \neq 0$ for all j. Then, there exists $f \in L^\infty(m)$ such that $S_n f$ does not converge a.s.$[m]$.*

Proof. For $b_j = (j-1)w$, where w is irrational, it follows from Birkhoff's erogdic theorem that the operators $T_j f(x) = f(x + b_j)$ satisfy the mean ergodic condition. That these S_n and T_j satisfy the other conditions laid out in Section 12.1 is immediate. By Theorem 5, it suffices to show

$$\sup\{N_f(\delta) : \|f\|_{L^2(m)} \leq 1\} = \infty$$

for some $\delta > 0$. In fact, we will do this with $\delta = 1/10$. Let $r \in \mathbb{N}$. By Lemma 6, choose integers $J_1 < \ldots < J_r$. Define a function f by

$$f(x) = 2^{-r/2} \sum_{\bar{\alpha} \in \{0,1\}^r} e^{2\pi i n(\bar{\alpha})x}.$$

Note, $\|f\|_2^2 = \int_0^1 f(x)\bar{f}(x)\,dx = 2^{-r}\sum_{\bar{\alpha}} 1 = 1$ by orthogonality. Also,

$$S_{J_s} f(x) = 2^{-r/2} \sum_{\bar{\alpha} \in \{0,1\}^r} \beta_{s,\bar{\alpha}} e^{2\pi i n(\bar{\alpha})x},$$

where

$$\beta_{s,\bar{\alpha}} = J_s^{-1} \sum_{j \le J_s} e^{2\pi i a_j n(\bar{\alpha})}.$$

Fix an $\bar{\alpha}$ and suppose $\alpha_s = 1$ and $\alpha_t = 0$. By Lemma 6, we have

$$|\beta_{s,\bar{\alpha}} - \beta_{t,\bar{\alpha}}| \ge |\beta_{s,\bar{\alpha}} - 1| - |1 - \beta_{t,\bar{\alpha}}| > \frac{1}{2} - \frac{1}{10} = \frac{2}{5}.$$

Hence, by orthogonality and above,

$$\begin{aligned} \|S_{J_s} f - S_{J_t} f\|_{L^2(m)} &= 2^{-r/2} \left(\sum_{\bar{\alpha}} |\beta_{s,\bar{\alpha}} - \beta_{t,\bar{\alpha}}|^2 \right)^{1/2} \\ &\ge 2^{-r/2} \left(\sum_{\bar{\alpha}:\alpha_s \ne \alpha_t} |\beta_{s,\bar{\alpha}} - \beta_{t,\bar{\alpha}}|^2 \right)^{1/2} > 1/5. \end{aligned}$$

As no two $S_{J_s} f$ could be contained in the same 1/10-ball in L^2, we see $N_f(1/10) \ge r$. As r is arbitrary, $\sup N_f(1/10) = \infty$. □

References

[1] Bellow, A. and Jones, R. L., *A Banach principle for L^∞*. Adv. Math. **120** (1996), 155–72;

[2] Bourgain, J., *Almost sure convergence and bounded entropy.* Israel J. Math. **63** (1988), no. 1, 79–97;

John Workman, Cornell University
email: **jworkman@math.cornell.edu**

13 Pointwise Ergodic Theorems for Arithmetic Sets, part 1

after J. Bourgain [1]
A summary written by Andrew Yingst

Abstract

We reduce a pointwise Ergodic theorem along polynomial iterates to certain inequalities. We also prove a lemma to be used in part 2.

13.1 Introduction

Let $(\Omega, \mathcal{B}, \mu, T)$ be a dynamical system. (For us, this is a probability measure space, with a measure-preserving automorphism.) We are interested in adaptations of Birkhoff's individual ergodic theorem which replace the usual ergodic means with means along some polynomial:

$$A_N f = \frac{1}{N} \sum_{n=1}^{N} f \circ T^{\lfloor p(n) \rfloor},$$

where p is a polynomial with real or integer coefficients. We follow Bourgain's arguments of the following two theorems. The first applies only to integer polynomials and includes a maximal inequality.

Theorem 1. *Let $(\Omega, \mathcal{B}, \mu, T)$ be a dynamical system and let p be a polynomial with integer coefficients. Given $r > 1$, there is a constant C so that*

$$\| \sup_{N \geq 1} |A_N f| \|_r \leq C \|f\|_r \tag{1}$$

holds for all $f \in L^r(\Omega, \mu)$. Furthermore, the sequence $A_N f$ converges pointwise almost everywhere. If T is weakly mixing and p is non-constant, then the limit is given by $\int f d\mu$.

In the second theorem, we generalize to all real polynomials.

Theorem 2. *Let $(\Omega, \mathcal{B}, \mu, T)$ be a dynamical system and let p be a polynomial with real coefficients. If f is any bounded measurable function on Ω, then $A_N f$ converges pointwise almost everywhere.*

13.2 Reduction to Inequalities.

The section following this one will argue the maximal inequalities to be used, and a certain lemma which follows from them. We begin by showing that Theorems 1 and 2 follow from these. Hence, for this subsection we assume that inequality 1 holds, as well as the following lemma.

Lemma 3. *Let $(\Omega, \mathcal{B}, \mu, T)$ be the shift map on $\mathbb{Z}$ endowed with counting measure, and let p be a polynomial with real coefficients. Fix $\epsilon > 0$, and let $Z_\epsilon = \{\lfloor (1+\epsilon)^n \rfloor : n \in \mathbb{N}\}$. Let (N_j) be a sequence of positive integers such that $N_{j+1} > 2N_j$ for all $j > 1$. For $j \geq 1$ and f a function on $\mathbb{Z}$, let*

$$\mathcal{M}_j f = \sup_{N_j \leq N \leq N_{j+1}, N \in Z_\epsilon} |A_N f - A_{N_j} f|.$$

Then there is a function $g : \mathbb{N} \to \mathbb{R}^+$ so that $\frac{g(J)}{J} \to 0$ as $J \to \infty$, and so that for all bounded, measurable f on Ω, we have

$$\sum_{j=1}^{J} \|\mathcal{M}_j f\|_2^2 \leq g(J) \|f\|_2^2.$$

We first show that Lemma 3 holds if $(\Omega, \mathcal{B}, \mu, T)$ is any dynamical system. (We do not state this identical lemma.)

Proof. Let f be a bounded measurable function on Ω, fix $J \geq 1$ and $x \in \Omega$, and let K be some large positive integer. Let ϕ be a function on $\mathbb{Z}$ by $\phi(k) = f(T^k(x))$ for $|k| \leq K$, and 0 otherwise. Using the shift map on $\mathbb{Z}$, we have that $A_N \phi(k) = \frac{1}{N} \sum_{n=1}^{N} \phi(\lfloor p(n) \rfloor + k) = A_N f(T^k(x))$, if $|k + \lfloor p(i) \rfloor| \leq K$ for all $i = 1 \ldots N$.

From this it follows that $\mathcal{M}_j \phi(k) = \mathcal{M}_j f(T^k(x))$, if $|k + \lfloor p(i) \rfloor| \leq K$ for all $i = 1 \ldots N_{j+1}$. Let $M = \max\{|p(i)| : i = 1 \ldots N_{J+1}\}$. Assume that K is so large that $K > M$. From Lemma 3, we have

$$\sum_{j=0}^{J} \sum_{n=-K+M}^{K-M} |\mathcal{M}_j \phi(n)|^2 \leq \sum_{j=0}^{J} \|\mathcal{M}_j \phi\|_2^2 \leq g(J) \|\phi\|_2^2 = g(J) \left(\sum_{n=-K}^{K} |\phi(n)|^2 \right).$$

For the values of n considered, ϕ and $\mathcal{M}_j \phi$ agree with an iterate of f:

$$\sum_{j=0}^{J} \sum_{n=-K+M}^{K-M} |\mathcal{M}_j f(T^n(x))|^2 \leq g(J) \left(\sum_{n=-K}^{K} |f(T^n(x))|^2 \right).$$

The above holds for all x, and we may integrate both sides with respect to x. Since T is invariant, these terms no longer depend on n. Dividing by $(2K+1-2M)$, we find:

$$\sum_{j=0}^{J} \|\mathcal{M}_j f\|^2 \leq g(J)\frac{2K+1}{2K+1-2M}\|f\|^2.$$

This holds for all large K, so letting K go to infinity gives the desired result. □

Next we show that the above lemma gives pointwise convergence of $A_N f$ when f is bounded.

Theorem 4. *Let $(\Omega, \mathcal{B}, \mu, T)$ be a dynamical system, and let p be a polynomial with real coefficients. If f is a bounded measurable function on Ω, then $A_n f$ converges pointwise almost everywhere.*

Proof. By way of contradiction, suppose f is a bounded function for which the above fails. We may assume $|f| < 1$. Let $L \leq \mu\{x \in \Omega : A_n f(x)$ doesnt converge$\}$, with $L \leq 1/2$. For $\epsilon > 0$ let E_ϵ denote the set of all points x so that ϵ witnesses that the sequence $(A_n f(x))$ is not a Cauchy sequence. Let $\hat{\epsilon} > 0$ be such that $\mu(E_{\hat{\epsilon}})$ has measure greater than $L/2$. Let $\epsilon = L\hat{\epsilon}/6$, and as above, let $Z_\epsilon = \{\lfloor (1+\epsilon)^n \rfloor : n \in \mathbb{N}$. We now inductively define a sequence (N_j) of positive integers with $N_{j+1} \geq 2N_j$.

Let $N_1 = 1$. Suppose that N_j has been defined. For $x \in E = E_{\hat{\epsilon}}$, there are arbitrarily large integers m and n so that $|A_m f(x) - A_n f(x)| > \hat{\epsilon}$. Hence there are arbitrarily large n so that $|A_n f(x) - A_{N_j} f(x)| > \hat{\epsilon}/2$. For such an n, let n' be the nearest element of Z_ϵ to n. Using the definition of Z_ϵ it can be verified that for large n, $|A_n f(x) - A_{n'} f(x)| \leq \frac{|n-n'|}{\min\{n,n'\}} \leq 2\epsilon$. Hence, $|A_{N_j} f(x) - A_{n'} f(x)| \geq \hat{\epsilon}/2 - 2\epsilon > \hat{\epsilon}/3$.

As before, let $\mathcal{M}_j f = \sup_{N_j \leq N \leq N_{j+1}, N \in Z_\epsilon} |A_N f - A_{N_j} f|$. By the above argument, as N_{j+1} becomes large, $\{x \in E : \mathcal{M}_j f(x) > \hat{\epsilon}/3\}$ increases to E. Therefore, we may choose $N_{j+1} > N_j$ so that $\mu\{x : \mathcal{M}_j f(x) > \hat{\epsilon}/3\} > L/2$. Thus the sequence N_j is defined, and we have $\|\mathcal{M}_j f(x)\|_2 \geq \sqrt{L/2}\hat{\epsilon}/3 > \epsilon$.

But by Lemma 3, we have $\sum_{j=1}^{J} \|\mathcal{M}_j f\|_2^2 \leq g(J)\|f\|_2^2$. In our case, this gives $J\epsilon^2 \leq g(J)\|f\|_2^2$, contradicting that $g(J)/J \to 0$. □

At this point, we have completed the proof of Theorem 2. To complete Theorem 1, we use the maximal inequality in a typical way:

Theorem 5. *Let $(\Omega, \mathcal{B}, \mu, T)$ be a dynamical system, and let p be a polynomial with integer coefficients. If $f \in L^r(\Omega, \mu)$, then $A_n f$ converges pointwise almost everywhere.*

Proof. Let $\delta, \epsilon > 0$. By the maximal inequality, (1) we may find a bounded function g with $\|f - g\|_r$ so small that $\|\sup_{n\geq 1} |A_n(f-g)|\|_r \leq \frac{\delta^{1/r}\epsilon}{3}$. We must then have $\sup_{n\geq 1} |A_n f - A_n g| \leq \epsilon/3$ except on a set of measure at most δ. But $A_n g(x)$ converges: for almost every $x \in \Omega$, there is $N(x)$ such that $m, n > N(x)$ implies $|A_m g(x) - A_n g(x)| < \epsilon/3$. Hence, for all x off a set of measure δ, there is $N(x)$ so that $m, n > N(x)$ implies $|A_n f(x) - A_m f(x)| < \epsilon$. Taking (δ_n) to be a summable sequence, and (ϵ_n) converging to zero, we find that $A_n f(x)$ is a Cauchy sequence except on a set of measure $\sum \delta_n$. $\square$

13.3 A Lemma

In this subsection, we outline Bourgain's argument of an inequality for certain Fourier multipliers. Let $\mathcal{F}$ denote the Fourier transform from $l^2(\mathbb{Z})$ to $L^2[0,1]$, or from $L^2(\mathbb{R})$ to $L^2(\mathbb{R})$. The goal of this section is to motivate the following:

Lemma 6. *Let $\lambda_1, \ldots, \lambda_K$ be points in $I = [0,1]$, with $d(\lambda_k, \lambda_{k'}) > 2 \cdot 2^{-s}$ for $k \neq k'$. For $j \geq 0$, let $R_j = \cup_{i=1}^K B(\lambda_i, 2^{-j})$. Then for any function f on $\mathbb{Z}$, the following inequality holds:*

$$\left\| \sup_{j\geq s} |\mathcal{F}^{-1}[\chi_{R_j} \mathcal{F} f]| \right\|_{l^2(\mathbb{Z})} \leq C(\log K)^2 \|f\|_{l^2(\mathbb{Z})},$$

where C is a constant independent of the choice of f or $\{\lambda_i\}$.

Note that the above supremum is over all those j so that the intervals of R_j are disjoint. Bourgain also generalizes this to take the supremum over all j, but in our applications, f is supported on some disjoint R_j, so this version suffices.

The argument begins by adapting a form of Doob's Oscillation Lemma for Martingales. We won't go through the argument, but do outline the adaptation. To avoid introducing terminology, we state only a special case of Doob's Oscillation Lemma. Given f a measurable function on $[0,1]$, let f_n denote $\sum_{I \in D_n} (\frac{1}{|I|} \int_I f) \chi_I$, where D_n is the collection of all dyadic intervals in $[0,1]$ of length 2^{-n}.

Lemma 7 (Doob's Oscillation Lemma)**.** *For $\lambda > 0$ and $x \in [0,1]$, let $N_\lambda(x)$ denote the number of jumps of length λ in the sequence $(f_n(x))$. Then for each $\lambda > 0$ and each measurable f, we have*

$$\|\lambda(N_\lambda)^{1/2}\|_2 \leq c\|f\|_2.$$

In our adaptation, we replace the scalar valued function f with a function from $f : \mathbb{R} \to \mathbb{R}^d$. We also replace the sequence $f_n(x)$ with the similar expression $(f * \phi_t)(x)$, where $\phi_t(y) = \frac{1}{t}\phi(\frac{y}{t})$. (Convolution with a vector-valued function is defined coordinate-wise.) Finally, we replace our gap-counting function $N_\lambda(x)$ with $M_\lambda(x)$, defined to be the number of balls of radius λ needed to cover the set $\{(f * \phi_t)(x) : t > 0\}$, taken to be zero if this set has diameter less than λ.

Lemma 8. *Let ϕ be a smooth real-valued function on $\mathbb{R}$ which vanishes at ∞, with $c_\phi = \int |x\phi'(x)| < \infty$. Then for every measurable $f : \mathbb{R} \to \mathbb{R}^d$ and every $s > 2$, we have*

$$\|\sup_{\lambda>0} \lambda(M_\lambda)^{1/s}\|_2 \leq c_\phi(s-2)^{-1}\|f\|_2.$$

From this, we easily obtain the following lemma.

Lemma 9. *With the notation above, there is a constant c'_ϕ depending only on ϕ, so that for $K \geq 1$ we have*

$$\|\int_0^\infty \min(K, M_\lambda(x))^{1/2} d\lambda\|_2 \leq c'_\phi (logK)^2 \|f\|_2.$$

Proof. Let $F(x) = \frac{1}{2}\sup_{t>0}\|f * \phi_t\|_2$. (The norm here is the vector norm.) Applying the Hardy-Littlewood Maximal Inequality coordinatewise, we get $\|F\|_2 \leq c_\phi\|f\|_2$. Note that if $\lambda \geq F(x)$, then $M_\lambda(x) = 0$. Hence, we may write

$$\int_0^\infty \min(K, M_\lambda(x))^{1/2} d\lambda \leq \int_0^{F(x)K^{-1/s}} K^{1/2} d\lambda + \int_{F(x)K^{-1/s}}^{F(x)} K^{\frac{1}{2}-\frac{1}{s}} M_\lambda(x)^{\frac{1}{s}} d\lambda$$

$$\leq K^{\frac{1}{2}-\frac{1}{s}} F(x) + K^{\frac{1}{2}-\frac{1}{s}} \frac{1}{s}\log(K) \cdot \sup_{\lambda>0} \lambda M_\lambda(x)^{1/s}.$$

Now we are done, by applying the previous lemma and taking s so that $\frac{1}{s} - \frac{1}{2} = \frac{1}{\log K}$. (This makes $K^{\frac{1}{2}-\frac{1}{s}} = e$ and $\frac{1}{s-2} = \frac{\log(K)-2}{4}$.) $\square$

We are now in a position to outline the proof of Lemma 6.

Proof. We begin with two reductions which we do not argue fully. First it is sufficient to argue the equivalent lemma for functions f on $\mathbb{R}$. (We still leave R_j a subset of $[0,1]$; we do not make it periodic.) The idea comes from noting that when f is a function on $\mathbb{Z}$ and $0 < \rho$ is small, the function $f_\rho = \rho^{-1}\sum_{n\in\mathbb{Z}} f(n)\chi_{[n,n+\rho]}$ on $\mathbb{R}$ has $\|f\|_{l^2(\mathbb{Z})} = \|f_\rho\|_{L^2(\mathbb{R})}$, and $\hat{f}_\rho \approx \hat{f}$ on $[0,1]$.

Second, we'd like to replace χ_{R_j} by a smooth function. Let $|\psi| \leq 1$ be some smooth function on $\mathbb{R}$ with $\psi = 1$ on $[-1/2, 1/2]$, and $\psi = 0$ off $(-1,1)$. Let $g_j(\lambda) = \sum_{k=1}^K \psi(2^j(\lambda - \lambda_k))$. The argument that it is sufficient to argue the lemma with g_j replacing χ_{R_j} comes from the observation that $\sum_{j\geq s} |g_j - \chi_{R_j}|$ is bounded.

We must now show the following inequality:

$$\left\| \sup_{j\geq s} |\mathcal{F}^{-1}[g_j \mathcal{F} f]| \right\|_{L^2(\mathbb{R})} \leq C(\log K)^2 \|f\|_{L^2(\mathbb{R})}.$$

Let $\phi = \mathcal{F}^{-1}\psi$ and again let $\phi_t(x) = \frac{1}{t}\phi(\frac{x}{t})$. Note that $\psi(2^j\lambda)\cdot\psi(2^{s-1}\lambda) = \psi(2^j\lambda)$. Taking the Fourier inverse of both sides of this gives $\phi_{2^j} * \phi_{2^{s-1}} = \phi_{2^j}$. Using this fact and usual rules of Fourier transforms, we may write the left-hand side of the above as:

$$\| \sup_{j\geq s} | \sum_{k=1}^K e^{2\pi i \lambda_k x} \cdot (f_k * \phi_{2^j})(x)) | \|_2, \tag{2}$$

where $f_k = [e^{-2\pi i \lambda_k x} f(x)] * [\phi_{2^{s-1}}(x)]$.

The Hardy-Littlewood Inequality gives that $\| \sup_{j\geq s} f_k * \phi_t \|_2$ is bounded by $c(\phi)\|f_k\|_2$. Using this we find an inequality of the form,

$$\| \sup_{j\geq s} | \sum_{k=1}^K e^{2\pi i \lambda_k x} \cdot (f_k * \phi_{2^j})(x)) | \|_2 \leq A(\sum_{k=1}^K \|f_k\|_2^2)^{1/2},$$

where $A = C\sqrt{K}$. Let $B = B(\phi, K)$ be the least value for which the above inequality holds. From Parseval's formula, we know $\sum_{k=1}^K \|f_k\|_2^2 = \sum_{k=1}^K \|\hat{f}_k\|_2^2 = \|\hat{f}\cdot g_{s-1}\|_2^2$ which is easily checked to be less than $2\|\hat{f}\|^2 = 2\|f\|^2$. Hence, we are done if we can show $B \leq C(\log K)^2$.

Because $\hat{f}_k$ is supported on the interval $[-2\cdot 2^{-s}, 2\cdot 2^{-s}]$, we can estimate that $\|(1-e^{2\pi i\lambda u})\hat{f}(\lambda)\|_2 \le \frac{1}{2}\|\hat{f}\|_2$ provided $|1-e^{2\pi i\lambda u}| < \frac{1}{2}$ on $[-2\cdot 2^{-s}, 2\cdot 2^{-s}]$, which will occur for example when $|u| < 2^s/100$. Taking $\mathcal{F}^{-1}$ gives

$$\|f_k - \sigma_u f_k\|_2 < \frac{1}{2}\|f_k\|_2, \text{ when } |u| < 2^s/100$$

where $\sigma_u g(x) = g(x+u)$. Writing $f_k = (f_k - \sigma_u f_k) + \sigma_u f_k$ and applying the triangle inequality gives

$$(2) \le \frac{B}{2}(\sum_{k=1}^{K} \|f_k\|_2^2)^{1/2} + \|\sup_{j\ge s}|\sum_{k=1}^{K} e^{2\pi i\lambda_k x}\cdot \sigma_u(f_k * \phi_{2^j})(x)\|_2.$$

Integrating this inequality over $u \in [0, 2^s/100]$ and changing variables lets us replace the last term with

$$10\sqrt{2^{-s}}\left\|\left(\int_0^{2^s} \sup_{t\ge 0}|\sum_{k=1}^{K} e^{2\pi i\lambda_k(x-u)}\cdot (f_k * \phi_t)(x))|^2\, du\right)^{1/2}\right\|_2.$$

If we can show the above is bounded by $C(\log K)^2(\sum_{k=1}^{K}\|f_k\|_2^2)^{1/2}$, we'll know that $C(\log K)^2 + \frac{B}{2}$ is a suitable choice of A in our desired inequality, and hence $C(\log K)^2 + \frac{B}{2} \ge B$, and so we'll be done.

Let F map into $\mathbb{R}^K$ by $F(x) = (f_1(x), \ldots, f_K(x))$, and consider the set $A_x = \{(F*\phi_t)(x) : t > 0\} = \{((f_1*\phi_t)(x), \ldots, (f_K*\phi_t)(x))\}$. As in Lemma 9, let $M_\lambda(x)$ denote the minimal number of λ balls needed to cover the set A_x, or zero if $\operatorname{diam}(A_x) < \lambda$. Fixing x, let q be the largest integer so that $M_{2^q}(x)$ is positive. For $j \le q$, let C_j be a collection of balls of radius 2^j covering A_x with $|C_j| = M_{2^j}(x)$. Next, for $j \le q$, define D_j by choosing a point of A_x from each ball in C_j. Finally, define B_j for $j \le q$ by taking $B_q = D_q$, and for $j < q$ let B_j consist of all points of the form $d - d'$ where $d \in D_j$ and d' is the nearest point of D_{j+1} to d.

Because of the covering properties of C_j, we know that if $j \le q$ and $y \in B_j$, then $\|y\| < 2\cdot 2^j$. (For $j = q$ we note that $\lim_{t\to\infty}(F*\phi_t)(x) = 0$, so we may assume $0 \in A_x$.) Further, we know that for each $a \in A_x$, we can write $a = \sum_{j\le q} y_j$ for some sequence $y_j \in B_j$. This sum holds coordinate-wise, and we may write $f_k * \phi_t(x) = \sum_{j\le q}\pi_k(y_j)$, leading us to the inequality

$$\sqrt{2^{-s}}\left(\int_0^{2^s} \sup_{t\ge 0}|\sum_{k=1}^{K} e^{2\pi i\lambda_k(x-u)}\cdot (f_k * \phi_t)(x)|^2 du\right)^{1/2}$$

$$\leq \sqrt{2^{-s}} \sum_{j \leq q} \left\| \max_{\vec{y} \in B_j} |\sum_{k=1}^{K} e^{2\pi i \lambda_k (x-u)} \cdot \pi_k(\vec{y})|^2 \right\|_{L^2[0,2^s],du} . \tag{3}$$

Now on the one hand, the interior of the maximum above is always less than $\sum_{k=1}^{K} |\pi_k(\vec{y})| \leq \sqrt{K}\|\vec{y}\|$. On the other hand, it also increases if we replace the max by a sum. We write

$$\max_{\vec{y} \in B_j} |\ldots| \leq \min\{\sqrt{K} \cdot 2 \cdot 2^j, [\sum_{\vec{y} \in B_j} |\sum_{k=1}^{K} e^{2\pi i \lambda_k (x-u)} \pi_k(\vec{y})|^2]^{1/2}\}.$$

For positive functions, the norm of the min is less than min of the norms, so we estimate the $L^2[0,2^s], du$ norm of each of the above. For the first, we get $2^{j+1}\sqrt{K2^s}$. For the second, the fact that $|\lambda_j - \lambda_k| > 2^{-s}$ for $j \neq k$ gives the following inequality, which we will not verify:

$$\|\sum_{k=1}^{K} a_k e^{2\pi i \lambda_k u}\|_{L^2[0,2^s],du} \leq C\sqrt{2^s}(\sum_{k=1}^{K} |a_k|^2)^{1/2}.$$

This gives that the $L^2[0,2^s], du$ norm of the second element of the above min is bounded by

$$\left(\sum_{\vec{y} \in B_j} C^2 2^s \|\vec{y}\|^2 \right)^{1/2} \leq \left(C^2 2^s |B_j| (2^{j+1})^2\right)^{1/2} = 2^{j+1}\sqrt{2^s}(M_{2^j}(x))^{1/2}$$

This gives us that (3) is bounded by

$$C \sum_{j \leq q} 2^{j+1} (\min\{K, M_{2^j}(x)\})^{1/2} \leq 2C \int_0^\infty (\min\{K, M_\lambda(x)\})^{1/2} d\lambda,$$

since $M_\lambda(x)$ decreases as λ increases. But now we are finished, by applying Lemma 9. □

References

[1] Bourgain, J., *Pointwise Ergodic Theorems for arithmetic sets.* Inst. Hautes Études Sci. Publ. Math., **69** (1989), pp. 5-45.

Andy Yingst, Univ. of S. Carolina Columbia
email: **andy.yingst@gmail.com**