Analysis of multiple ergodic averages

Summer School, Kopp *

September 2023

Organizers:

Asgar Jamneshan, Koç University Istanbul

Rajula Srivastava, Universität Bonn

Christoph Thiele, Universität Bonn

*supported by Hausdorff Center for Mathematics, Bonn

Contents

1 Boxes, extended boxes and sets of positive upper density in the Euclidean space

After P. Durcik and V. Kovač [DK]

A summary written by Ethan Ackelsberg

Abstract

We present an argument of Durcik and Kovač [DK] showing that sets of positive upper Banach density in sufficiently high dimension contain congruent copies of all large dilates of the 2^n vertices of an n-dimensional rectangular box, as well as the extension of a box by n vertices completing 3-term arithmetic progressions.

1.1 Introduction

Euclidean Ramsey theory is concerned with finding congruent copies of geometric configurations in large subsets of Euclidean space. The relevant notion of largeness for this paper is as follows: the *upper Banach density* of a measurable subset $A \subseteq \mathbb{R}^d$ is the quantity

$$\overline{\delta}(A) := \limsup_{N\to\infty} \sup_{\mathbf{x}\in\mathbb{R}^d} \frac{\big|A \cap \big(\mathbf{x} + [0,N]^d\big)\big|}{N^d},$$

where $|\cdot|$ denotes the d-dimensional Lebesgue measure on $\mathbb{R}^d$.

The first result states that sets of positive upper Banach density in sufficiently high dimension contain congruent copies of all large dilates of the 2^n vertices of any given n-dimensional rectangular box.

Theorem 1 ([DK], Theorem 1)**.** *Let $a_1, \ldots, a_n > 0$. For any natural numbers $d_1, \ldots, d_n \geq 5$ and any measurable set $A \subseteq \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}$ with $\overline{\delta}(A) > 0$, there exists $\lambda_0 > 0$ such that for any $\lambda \geq \lambda_0$, the set A contains a box*

$$B(x_1, \ldots, x_n; s_1, \ldots, s_n) := \big\{(x_1 + \epsilon_1 s_1, \ldots, x_n + \epsilon_n s_n) : \epsilon_1, \ldots, \epsilon_n \in \{0,1\}\big\}$$

with $x_j, s_j \in \mathbb{R}^{d_j}$ and $\|s_j\|_{\ell^2} = \lambda a_j$.

Extending the boxes appearing in Theorem 1 introduces new difficulties: Bourgain [B] constructed a set $A \subseteq \mathbb{R}$ with $\overline{\delta}(A) > 0$ for which there is an unbounded sequence $s_m \to \infty$ such that A does not contain any 3-term arithmetic progressions with common difference s_m. However, in higher dimensions, Cook, Magyar, and Pramanik [CMP] proved a density theorem for 3-term arithmetic progressions when the size of the common difference is measured in the ℓ^p norm for $1 < p < \infty$, $p \neq 2$.

The second main result of this paper is a common generalization of Theorem 1 and the aforementioned result from [CMP] about 3-term arithmetic progressions.

Theorem 2 ([DK], Theorem 2)**.** *Let $a_1, \ldots, a_n > 0$, and let $1 < p < \infty$, $p \neq 2$. There exists a threshold d_0 such that for any natural numbers $d_1, \ldots, d_n \geq d_0$ and any measurable set $A \subseteq \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}$ with $\overline{\delta}(A) > 0$, there exists $\lambda_0 > 0$ such that for any $\lambda \geq \lambda_0$, the set A contains a 3AP-extended box*

$$B_{3AP}(\mathbf{x}, \mathbf{s}) := B(\mathbf{x}, \mathbf{s}) \cup \{(x_1 + 2s_1, \ldots, x_n), \ldots, (x_1, x_2, \ldots, x_n + 2s_n)\}$$

with $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{s} = (s_1, \ldots, s_n) \in \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}$ and $\|s_j\|_{\ell^p} = \lambda a_j$.

1.2 Strategy of the proof

Rather than proving Theorem 2 directly, we deduce it from a result about higher-dimensional configurations and then project. Namely, we will prove a similar statement for *corner-extended boxes*

$$\begin{aligned} &B_{\llcorner}(x_1, \ldots, x_n; y_1, \ldots, y_n; s_1, \ldots, s_n) \\ &\qquad := \{x_1 + \epsilon_1 s_1, \ldots x_n + \epsilon_n s_n, y_1, \ldots, y_n) : \epsilon_1, \ldots, \epsilon_n \in \{0, 1\}\} \\ &\qquad\quad \cup \{(x_1, \ldots, x_n, y_1 + s_1, \ldots, y_n), \ldots, (x_1, \ldots, x_n, y_1, \ldots, y_n + s_n)\} \end{aligned}$$

with $x_j, y_j, s_j \in \mathbb{R}^{d_j}$, $s_j \neq \mathbf{0}$. Note that the projection $(x_j, y_j) \mapsto y_j - x_j$ sends the corner-extended box $B_{\llcorner}(\mathbf{x}, \mathbf{y}, \mathbf{s}) \subseteq \left(\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}\right)^2$ to the 3AP-extended box $B_{3AP}(\mathbf{y} - \mathbf{x}, \mathbf{s}) \subseteq \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}$. Theorem 2 therefore follows from

Theorem 3 ([DK], Theorem 3)**.** *Let $a_1, \ldots, a_n > 0$, and let $1 < p < \infty$, $p \neq 2$. There exists a threshold d_0 such that for any natural numbers $d_1, \ldots, d_n \geq d_0$ and any measurable set $A \subseteq \left(\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}\right)^2$ with $\overline{\delta}(A) > 0$, there exists $\lambda_0 > 0$ such that for any $\lambda \geq \lambda_0$, the set A contains a corner-extended box $B_{\llcorner}(\mathbf{x}, \mathbf{y}, \mathbf{s})$ for some $\mathbf{x}, \mathbf{y}, \mathbf{s} \in \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}$ with $\|s_j\|_{\ell^p} = \lambda a_j$.*

In order to prove the existence of a configuration in the subset A, we will establish positivity of an associated average that provides a normalized count of configurations in A. To that end, define the measure $\sigma^{d,p}$ on $\mathbb{R}^d$ by $\sigma^{d,p}(s) = \delta(1 - \|s\|_{\ell^p}^p)$, where δ is the Dirac δ distribution. The dilated measure $\sigma_\lambda^{d,p}$, defined by $\sigma_\lambda^{d,p}(A) = \sigma^{d,p}(\lambda^{-1}A)$, is supported on the surface $\{s \in \mathbb{R}^d : \|s\|_{\ell^p} = \lambda\}$.

Fix $d_1, \ldots, d_n$, and let $D = d_1 + \cdots + d_n$. Let $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n_{>0}$ be fixed. For $\lambda > 0$, we let $\sigma_{\lambda\mathbf{a}}^p$ denote the product measure $\sigma_{\lambda a_1}^{d_1,p} \times \cdots \times \sigma_{\lambda a_n}^{d_n,p}$ on $\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n} \cong \mathbb{R}^D$. We then define the *pattern-counting forms* at a scale $\lambda > 0$ by

$$\mathcal{N}_\lambda^p(f) := \int_{\mathbb{R}^{2D}} \prod_{z \in B(\mathbf{x},\mathbf{s})} f(z)\, d\sigma_{\lambda\mathbf{a}}^p(\mathbf{s})\, d\mathbf{x}$$

and

$$\widetilde{\mathcal{N}}_\lambda^p(f) := \int_{\mathbb{R}^{3D}} \prod_{z \in B_\llcorner(\mathbf{x},\mathbf{y},\mathbf{z})} f(z)\, d\sigma_{\lambda\mathbf{a}}^p(\mathbf{s})\, d\mathbf{x}\, d\mathbf{y}.$$

The quantities $\mathcal{N}_\lambda^p$ and $\widetilde{\mathcal{N}}_\lambda^p$ are normalized counts of boxes and corner-extended boxes at the scale λ, weighted by the function f. In particular, if $\mathcal{N}_\lambda^p(\mathbb{1}_A) > 0$, then A contains a box $B(\mathbf{x}, \mathbf{s})$ for some $\mathbf{x}, \mathbf{s} \in \mathbb{R}^D$ with $\|s_j\|_{\ell^p} = \lambda a_j$. Similarly, if $\widetilde{\mathcal{N}}_\lambda^p(\mathbb{1}_A) > 0$, then A contains a corner-extended box $B(\mathbf{x}, \mathbf{y}, \mathbf{s})$ for some $\mathbf{x}, \mathbf{y}, \mathbf{s} \in \mathbb{R}^D$ with $\|s_j\|_{\ell^p} = \lambda a_j$.

We will also work with smoothed approximations of the pattern-counting forms, defined by

$$\mathcal{M}_\lambda^{p,\varepsilon}(f) := \int_{\mathbb{R}^{2D}} \prod_{z \in B(\mathbf{x},\mathbf{s})} f(z)\, \omega_{\lambda\mathbf{a}}^{p,\varepsilon}(\mathbf{s})\, d\mathbf{s}\, d\mathbf{x}$$

and

$$\widetilde{\mathcal{M}}_\lambda^{p,\varepsilon}(f) := \int_{\mathbb{R}^{3D}} \prod_{z \in B_\llcorner(\mathbf{x},\mathbf{y},\mathbf{z})} f(z)\, \omega_{\lambda\mathbf{a}}^{p,\varepsilon}(\mathbf{s})\, d\mathbf{s}\, d\mathbf{x}\, d\mathbf{y},$$

where $\omega_{\lambda\mathbf{a}}^{p,\varepsilon}$ is a smooth bump function supported in an ε-neighborhood of the surface $\big\{\mathbf{s} \in \mathbb{R}^D : \|s_j\|_{\ell^p} = \lambda a_j \text{ for } j \in \{1, \ldots, n\}\big\}$. A precise definition for $\omega_{\lambda\mathbf{a}}^{p,\varepsilon}$ can be found in [DK, Section 2]. Finally, let

$$\mathcal{E}_\lambda^{p,\varepsilon}(f) := \mathcal{M}_\lambda^{p,\varepsilon}(f) - b(p,\varepsilon)\mathcal{M}_\lambda^{p,1}(f)$$

and

$$\widetilde{\mathcal{E}}_{\lambda}^{p,\varepsilon}(f) := \widetilde{\mathcal{M}}_{\lambda}^{p,\varepsilon}(f) - b(p,\varepsilon)\widetilde{\mathcal{M}}_{\lambda}^{p,1}(f),$$

where

$$b(p,\varepsilon) := \frac{\int_{\mathbb{R}^D} \omega^{p,\varepsilon}(\mathbf{s})\ d\mathbf{s}}{\int_{\mathbb{R}^D} \omega^{p,1}(\mathbf{s})\ d\mathbf{s}}.$$

We now sketch the proof of Theorem 1. The proof strategy for Theorem 3 is completely analogous.

Proof of Theorem 1 (sketch). A standard argument by contradiction shows that it suffices to prove the following: for any sufficiently large $N \geq N(\delta)$, any measurable subset $A \subseteq [0,N]^D$ with $|A| \geq \delta N^D$, and any sequence of scales $\lambda_1 < \lambda_2 < \cdots < \lambda_M \leq N$ with $\lambda_{m+1} \geq 2\lambda_m$ and $M \geq M(\delta)$, there exists $m \in \{1,\ldots,M\}$ such that $\mathcal{N}_{\lambda_m}^2(\mathbb{1}_A) > 0$.

We may write

$$\mathcal{N}_{\lambda_m}^2(\mathbb{1}_A) = b(2,\varepsilon)\mathcal{M}_{\lambda_m}^{2,1}(\mathbb{1}_A) + \big(\mathcal{N}_{\lambda_m}^2(\mathbb{1}_A) - \mathcal{M}_{\lambda_m}^{2,\varepsilon}(\mathbb{1}_A)\big) + \mathcal{E}_{\lambda_m}^{2,\varepsilon}(f).$$

The main term $b(2,\varepsilon)\mathcal{M}_{\lambda_m}^{2,1}(\mathbb{1}_A)$ is large ($\geq C(D,\delta)N^D$) by an application of the multidimensional Szemerédi theorem of Furstenberg and Katznelson [FK] and an estimate on $b(2,\varepsilon)$ from [CMP]. The first error term $\mathcal{N}_{\lambda_m}^2(\mathbb{1}_A) - \mathcal{M}_{\lambda_m}^{2,\varepsilon}(\mathbb{1}_A)$ tends to zero as $\varepsilon \to 0$, since $\mathcal{M}$ is a smoothed approximation of $\mathcal{N}$. The remaining error term $\mathcal{E}_{\lambda_m}^{2,\varepsilon}(f)$ is small for all large enough values of m as a consequence of a singular Brascamp–Lieb inequality due to Durcik and Thiele [DT]. Hence, taking ε sufficiently small and m sufficiently large, $\mathcal{N}_{\lambda_m}^2(\mathbb{1}_A) > 0$ as desired.

The full argument can be found in [DK, Section 2], where the main term is handled in [DK, Proposition 4], the first error term in [DK, Proposition 5], and the second error term in [DK, Proposition 6]. □

1.3 Key estimate

The most technically demanding part of the proof outlined above is the estimate of the second error term $\mathcal{E}_{\lambda_m}^{2,\varepsilon}(f)$. We use the notation $A(N) \ll_P B(N)$ to denote an inequality of the form $A(N) \leq CB(N)$, where C is a constant depending on the parameters P.

Proposition 4 ([DK], Proposition 6)**.** *Let $1 < p < \infty$, $0 < \varepsilon < \frac{1}{10D}$, and $0 < \lambda_1 < \cdots < \lambda_M$ with $\lambda_{m+1} \geq 2\lambda_m$.*

1. If $f : \mathbb{R}^D \to [0,1]$ is a measurable function supported on $[0,N]^D$, then

$$\sum_{m=1}^{M} \left|\mathcal{E}^{p,\varepsilon}_{\lambda_m}(f)\right| \ll_{D,\varepsilon} N^D.$$

2. If $f : \mathbb{R}^{2D} \to [0,1]$ is a measurable function supported on $[0,N]^{2D}$, then

$$\left(\sum_{m=1}^{M} \left|\widetilde{\mathcal{E}}^{p,\varepsilon}_{\lambda_m}(f)\right|^2\right)^{1/2} \ll_{D,\varepsilon} N^{2D}.$$

Part 1 of Proposition 4 is deduced from the following singular Brascamp–Lieb inequality (Part 2 follows from a similar inequality corresponding to corner-extended boxes):

Theorem 5 ([DK], Theorem 10(a))**.** *Let $K : \mathbb{R}^D \to \mathbb{C}$ be a bounded compactly supported function satisfying the symbol estimates*

$$\left|\partial^\kappa \widehat{K}(\xi)\right| \leq C_\kappa \|\xi\|_{\ell^2}^{-|\kappa|} \tag{1}$$

for any multi-index κ. Then

$$\left|\int_{\mathbb{R}^{2D}} K(\mathbf{s}) \prod_{\epsilon\in\{0,1\}^n} F_\epsilon(\mathbf{x}+\epsilon\mathbf{s})\ d\mathbf{s}\ d\mathbf{x}\right| \ll_{(C_\kappa)_\kappa, d_1,\ldots,d_n} \prod_{\epsilon\in\{0,1\}^n} \|F_\epsilon\|_{L^{2^n}},$$

where $\epsilon\mathbf{s}$ denotes the point $(\epsilon_1 s_1, \ldots, \epsilon_n s_n) \in \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}$.

In the special case $d_1 = \cdots = d_n$, Theorem 5 follows from [DT, Theorem 1], and the general case can be reduced to this one (see [DK, Section 4] for details).

We end with a sketch of the proof of Proposition 4 using Theorem 5.

Proof of Proposition 4 (sketch). For each $m \in \{1, \ldots, M\}$, expand

$$\mathcal{E}^{p,\varepsilon}_{\lambda_m}(f) = \int_{\mathbb{R}^{2D}} \prod_{z\in B(\mathbf{x},\mathbf{s})} f(z)\ \left(\omega^{p,\varepsilon}_{\lambda_m\mathbf{a}}(\mathbf{s}) - b(p,\varepsilon)\omega^{p,1}_{\lambda_m\mathbf{a}}(\mathbf{s})\right)\ d\mathbf{s}\ d\mathbf{x}.$$

The function $\omega^{p,\varepsilon}_{\lambda_m\mathbf{a}}(\mathbf{s}) - b(p,\varepsilon)\omega^{p,1}_{\lambda_m\mathbf{a}}(\mathbf{s})$ may be written in the form

$$\sum_{i=1}^{n}\prod_{j=1}^{n} \varphi^{i,j}_{\lambda_m}(s_j)$$

for some C^1 functions $\varphi_{\lambda_m}^{i,j}$ with $\int \varphi_{\lambda_m}^{i,i} = 0$ and $\varphi_{\lambda_m}^{i,j} \geq 0$ for $i \neq j$.

Let $\alpha_m \in \{-1, 1\}$ so that

$$\left|\mathcal{E}_{\lambda_m}^{p,\varepsilon}\right| = \alpha_m \mathcal{E}_{\lambda_m}^{p,\varepsilon}.$$

Then

$$\sum_{m=1}^{M} \left|\mathcal{E}_{\lambda_m}^{p,\varepsilon}\right| = \sum_{i=1}^{n} \int_{\mathbb{R}^{2D}} K_i(\mathbf{s}) \prod_{z \in B(\mathbf{x},\mathbf{s})} f(z)\, d\mathbf{s}\, d\mathbf{x}$$

for

$$K_i(\mathbf{s}) = \sum_{m=1}^{M} \alpha_m \varphi_{\lambda_m}^{i,1}(s_1) \ldots \varphi_{\lambda_m}^{i,n}(s_n).$$

One can check that the kernels K_i satisfy the estimates (1), so we conclude

$$\sum_{m=1}^{M} \left|\mathcal{E}_{\lambda_m}^{p,\varepsilon}\right| \ll \|f\|_{L^{2^n}}^{2^n} \leq N^D.$$

□

References

[B] Bourgain, J., *A Szemerédi type theorem for sets of positive density in* $\mathbb{R}^k$, Israel J. Math. **54** (1986), no. 3, 307–316.

[CMP] Cook, B., Magyar, A., and Pramanik, M., *A Roth type theorem for dense subsets of* $\mathbb{R}^d$, Bull. London Math. Soc. **49** (2017), no. 4, 676–689.

[DK] Durcik, P. and Kovač, V., *Boxes, extended boxes and sets of positive upper density in the Euclidean space* , Math. Proc. Cambridge Philos. Soc. **171** (2021), no. 3, 481–501.

[DT] Durcik, P. and Thiele, C., *Singular Brascamp-Lieb inequalities with cubical structure*, Bull. Lond. Math. Soc. **52** (2020), no. 2, 283–298.

[FK] Furstenberg, H. and Katznelson, Y., *An ergodic Szemerédi theorem for commuting transformations*, J. Anal. Math. **38** (1978), no. 1, 275–291.

Ethan Ackelsberg, IAS/EPFL
email: `eackelsberg@ias.edu`

2 Quantitative bounds in the nonlinear Roth theorem, part II

After S. Peluse and S. Prendiville [PP 22]

A summary written by Seljon Akhmedli

Abstract

We summarize a quantitatively effective version of the polynomial Szemerédi theorem for nonlinear Roth configurations of the form $x, x+y, x+y^2$.

2.1 Introduction

One of the classical theorems in combinatorics is of Szemerédi's which says a set with positive density will contain arbitrarily long arithmetic progressions. In a similar spirit, there are other types of configurations (besides arithmetic progressions) which arise in dense sets such as polynomial progressions. Indeed, the polynomial Szemerédi theorem says if $p_1, ..., p_n \in \mathbb{Z}[y]$ with $p_i(0) = 0$ for all $1 \leq i \leq n$, then any set $S \subset \mathbb{N}$ with positive upper density must contain a nontrivial progression of the form $x, x+p_1(y), ..., x+p_n(y)$. Another equivalent form of saying this is that the size of the largest subset of $\{1, 2, ..., N\}$ lacking the nontrivial polynomial progression is $o(N)$. Gowers [G 01] produces bounds on the density in Szemerédi's theorem, namely, for all $k \in \mathbb{N}$, there exists $c(k) > 0$ s.t. any subset of $\{1, 2, ..., N\}$ of density at least $(\log\log N)^{-c}$ contains an arithmetic progression of length k. In [Pre 17], quantitative bounds are given for the polynomial Szemerédi theorem but in the homogeneous case, namely, when the polynomials $p_i(y)$, $1 \leq i \leq n$, are all of the same degree. The main result in [PP 22] given below is the first quantitative version of the theorem for polynomial progressions of length three and of differing degrees over $\mathbb{Z}$.

Theorem 1. *There exists $c > 0$ such that if $A \subset \{1, 2, ..., N\}$ does not contain a nontrivial progression of the form*

$$x, x+y, x+y^2,$$

then

$$|A| \ll N(\log\log N)^{-c}.$$

We note that having a zero constant term in the polynomial progressions is necessary; for example, within the even numbers, we already cannot find a configuration of the form $x, x+y+1, x+y^2$ by parity reasons. Bourgain and Chang gave quantitative bounds on the nonlinear Roth configuration $x, x+y, x+y^2$ over $\mathbb{F}_p$ [BC 17]. However, there are difficulties in adapting their method to $\mathbb{Z}$. Peluse [Pel 19] shows that under certain conditions on the characteristic of $\mathbb{F}_p$, there exists $c > 0$ such that any subset of $\mathbb{F}_p$ of size at least p^{1-c} will contain a nontrivial polynomial progression

$$x, x+p_1(y), ..., x+p_m(y)$$

where $p_i(0) = 0$ for all $1 \leq i \leq m$. Peluse and Prendiville are able to utilize an important idea from [Pel 19], that is, if one can control the nonlinear Roth progressions by a Gowers U^s norm, then we can descend in s to obtain control by a U^1-seminorm through the degree lowering method. In the setting of finite fields, the so-called PET induction scheme of Bergelson and Leibman controls these configurations by a global U^s norm. Because of difficulties which arise in the integer setting, the PET induction only reduces to working with an average of constrained U^1-seminorms and we instead get localized U^1 norm control. The authors in [PP 22] then show these averages are controlled by a global Gowers U^s norm.

2.2 Control by a global Gowers norm

Definition 1. Let $f : \mathbb{Z} \to \mathbb{C}$ and $\Delta_h f : \mathbb{Z} \to \mathbb{C}$ be the difference function of f given by

$$\Delta_h f(x) = f(x)\overline{f(x+h)}.$$

Then the *Gowers U^s-norm* of f is given by

$$||f||_{U^s} := \Big(\sum_{x, h_1, \ldots, h_s} \Delta_{h_1, \ldots, h_s} f(x) \Big)^{\frac{1}{2^s}}.$$

If $S \subset \mathbb{Z}$, then the localized Gowers U^s-norm is $||f||_{U^s(S)} := ||f 1_S||_{U^s}$.

Definition 2. The *counting operator* on the functions $f_i : \mathbb{Z} \to \mathbb{C}$ by

$$\Lambda_q(f_0, f_1, f_2) := \sum_{x \in \mathbb{Z}} \sum_{y \in \mathbb{N}} f_0(x) f_1(x+y) f_2(x+qy^2).$$

Notice when $f_i = 1_{[N]}$ for $i = 0, 1, 2$, the counting operator $\Lambda_q(f_0, f_1, f_2)$ counts the number of nonlinear Roth progressions in $[N]$. The theorem below provides control of this count by a Gowers U^5-norm.

Theorem 2. *Let $g_0, g_1, f : \mathbb{Z} \to \mathbb{C}$ be 1-bounded functions, each supported in $[N] := \{1, 2, ..., N\}$. Suppose that*

$$\Big|\sum_{x\in\mathbb{Z}}\sum_{y\in\mathbb{N}} g_0(x)g_1(x+y)f(x+qy^2)\Big| \geq \delta \sum_{x\in\mathbb{Z}}\sum_{y\in\mathbb{N}} 1_{[N]}(x)1_{[N]}(x+y)1_{[N]}(x+qy^2).$$

Then either $N \ll q$ or

$$\sum_{u\in[q]} ||f||^{2^5}_{U^5(u+q\mathbb{Z})} \gg \delta^{O(1)} \sum_{u\in[q]} ||1_{[N]}||^{2^5}_{U^5(u+q\mathbb{Z})}.$$

In the proof of the above theorem, there are multiple technical lemmas used. We will highlight some of those main ingredients and tools in the following subsections.

2.3 PET Induction and van der Corput

The polynomial exhaustion technique (PET) was first developed by Bergelson and Leibman [BL 96] in the proof of the classical polynomial van der Waerden theorem. In [PP 22], the tool is used to replace working with univariate polynomials like y^2 in our progression to instead working with bilinear forms such as ah. Notice the polynomial y^2 has a sparse image compared to ah and this creates difficulties in trying to obtain control by some U^s norm. The van der Corput method in some sense takes a large nonlinear average and bounds it by a large linear average. By applying Cauchy-Schwarz, van der Corput, and a change of variables, [PP 22, Lemma 3.2] shows how difference functions control linear progressions like $x, x + ay, a + by, x + (a + b)y$, where $a, b \in \mathbb{Z}$. The next step is linearization where the counting operator $\Lambda_q(f_0, f_1, f_2)$ will be controlled by difference functions of bilinear forms. We give this lemma below.

Lemma 1. *Let $f_i : \mathbb{Z} \to \mathbb{C}$ be such that $|f_i| \leq 1$ and with support in $[N]$. Then for any $1 \leq H \leq M$ we have*

$$\Big|\frac{1}{NM}\Lambda_q(f_0, f_1, f_2)\Big|^{32} \ll \sum_{a,b,h} \mu_M(a)\mu_M(b)\mu_H(h)\mathbb{E}_{x\in[N]}\Delta_{2q(a+b)h_1,2qbh_2,2qah_3} f_2(x)$$

where $M := \lfloor \sqrt{N/q} \rfloor$.

Thus the largeness of the counting operator implies the largeness of

$$(1) \qquad \sum_{a,b\in[N^{\frac{1}{2}}]} \sum_{h_1,h_2,h_3\in[N^{\frac{1}{2}}]} \sum_x \Delta_{ah_1,bh_2,(a+b)h_3} f_2(x),$$

and with an inverse theorem and concatenation result we will be able to further show largeness of $||f_2||_{U^5}$.

2.4 An Inverse Theorem

Lemma 2. *Let a, b be positive integers and coprime. Suppose that $f : \mathbb{Z} \to \mathbb{C}$ is 1-bounded with support in the interval $[N]$ and satisfies*

$$(2) \quad \sum_{h,x} \mu_H(h) \Delta_{ah_1,bh_2} f(x) \geq \delta N.$$

Then there exists 1-bounded functions $g, h : \mathbb{Z} \to \mathbb{C}$ such that g is a-periodic, h is almost b-periodic and furthermore

$$\Big| \sum_x f(x)g(x)h(x) \Big| \geq \delta \lfloor H \rfloor^2 - 2\Big(\frac{H}{a} + \frac{Hb}{N}\Big) \lfloor H \rfloor^2.$$

The proof of the above reduces to relating (2) to a box norm. In fact, understanding the largeness of (2) comes down to understanding the largeness of two dimensional Gowers box norms in directions 'a' and 'b'. Indeed, the largeness of this norm will imply a correlation of f with a product of periodic and almost periodic 1-bounded functions.

2.5 Quantitative Concatenation

Notice in (1), the coefficients of the h_i in the difference function of f_2 have some linear dependence. The purpose of the concatenation theorems are to work with this type of dependence. Roughly speaking, if one can understand the largeness of a function in two different directions, then one can understand the largeness in both directions jointly. There are a few technical lemmas in the final step to obtain control by a global U^5 norm. The essence of [PP 22, Lemma 5.3] is if $\sum_{a,h\in[N^{\frac{1}{2}}]} \sum_x \Delta_{ah} f(x)$ is large, then

$\sum_{k\in(-N,N)}\sum_x \Delta_k f(x)$ is large as well. So on average the behavior of these functions is relatively the same. Similarly, the motivation behind [PP 22, Lemma 5.4] is that $\Delta_{ah} f$ behaves like fg_a on average, where g_a is an a-periodic function. The proof of Theorem 2 uses all of the above tools and methods but there is still more work needed to obtain U^5 control for the nonlinear Roth configurations.

References

[BC 17] J. Bourgain and M.-C. Chang., *Nonlinear Roth type theorems in finite fields.* Israel J. Math., 221(2):853–867, 2017.

[BL 96] V. Bergelson and A. Leibman. *Polynomial extensions of van der Waerden's and Szemerédi's theorems.* J. Amer. Math. Soc., 9(3):725–753, 1996.

[G 01] W. T. Gowers. *A new proof of Szemerédi's theorem.* Geom. Funct. Anal., 11(3):465–588, 2001.

[Pre 17] S. Prendiville. *Quantitative bounds in the polynomial Szemerédi theorem: the homogeneous case.* Discrete Anal., pages 34, Paper No. 5, 2017.

[Pel 19] S. Peluse. *On the polynomial Szemerédi theorem in finite fields.* Duke Math. J., 168(5):749–774, 04 2019.

[PP 22] S. Peluse and S. Prendiville. *Quantitative bounds in the nonlinear Roth theorem.* arXiv:1903.02592.

Seljon Akhmedli, Northwestern University
email: `seljonakhmedli2026@u.northwestern.edu`

3 Singular Brascamp-Lieb inequalities with cubical structure

After P. Durcik and C. Thiele [DuTh]

A summary written by Michel Alexis

Abstract

In [DuTh], Durcik-Thiele show that for a collection of linear transformations possessing "cubical symmetry", the associated singular Brascamp-Lieb inequalities hold if and only if the standard dimensional Brascamp-Lieb criterion holds across all linear subspaces V of the ambient space. Because this criterion is only known to be necessary for singular Brascamp-Lieb inequalities, this result of Durcik-Thiele strengthens the conjecture that a general singular Brascamp-Lieb inequality holds if and only if the same Brascamp-Lieb criterion holds across all linear subspaces V.

3.1 Introduction

Fix surjective linear maps $\Pi_i : \mathbb{R}^m \to \mathbb{R}^{k_i}$ and exponents $p_i \in [1,\infty]$ for $i = 1, \dots, n$. One may ask whether there exists a constant C for which the following multilinear inequality holds,

$$\left| \int_{\mathbb{R}^m} \left(\prod_{i=1}^n F_i \left(\Pi_i x\right) \right) dx \right| \leq C \prod_{i=1}^n \|F_i\|_{p_i} . \tag{1}$$

This sort of inequality is known as a Brascamp-Lieb inequality. In [BeCaChTa], it was shown that inequality (1) holds if and only if for every linear subspace $V \subset \mathbb{R}^m$ we have

$$\dim(V) \leq \sum_{i=1}^n \frac{1}{p_i} \dim(\Pi_i V) , \tag{2}$$

with equality in (2) when $V = \mathbb{R}^m$. To communicate to the reader why these are inequalities of interest to analysts, consider for instance the case that Π_i equals the identity on $\mathbb{R}^m$: then the Brascamp-Lieb inequality (1) reduces to Hölder's inequality for the n functions $\{F_i\}_{1\leq i\leq n}$, and (2) reduces to the classical condition $\sum_{i=1}^{n} \frac{1}{p_i} = 1$. And for instance when $n = 2$, if we consider the linear surjective maps $\Pi_1 : \mathbb{R}^4 \to \mathbb{R}^2$ and $\Pi_2 : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\Pi_1(x, y) = x - y\,, \qquad \Pi(y) = y\,, \qquad x, y \in \mathbb{R}^2\,,$$

then (1) yields a particular case of Young's inequality

$$\left| \int \int F_1(x - y) F_2(y) dy dx \right| \leq C \left\| F_1 \right\|_{p_1} \left\| F_2 \right\|_{p_2} .$$

and (2) reduces to the corresponding exponents $\frac{1}{p_2} = \frac{1}{p_2} = 1$.

In their paper [DuTh], Durcik-Thiele consider singular Brascamp-Lieb inequalities, i.e. inequalities of the form

$$\left| \int_{\mathbb{R}^m} \left(\prod_{i=1}^{n} F_i \left(\Pi_i x \right) \right) K \left(\Pi x \right) dx \right| \leq C \prod_{i=1}^{n} \left\| F_i \right\|_{p_i}\,, \tag{3}$$

where in addition we consider $\Pi : \mathbb{R}^m \to \mathbb{R}^k$ a surjective linear map and all Calderón-Zygmund kernels K, i.e. all kernels K satisfying a Mikhlin multiplier condition

$$\left| \partial^{\alpha} \widehat{K} \left(\xi \right) \right| \leq \left| \xi \right|^{-|\alpha|} \tag{4}$$

for all α up to a certain threshhold. Durcik-Thiele note that a necessary condition for (3) to hold for all kernels K as above is that for every linear subspace $V \subset \ker \Pi$, we have

$$\dim \left(V \right) \leq \sum_{i=1}^{n} \frac{1}{p_i} \dim \left(\Pi_i \Big|_{\ker \Pi} V \right)\,, \tag{5}$$

with equality in (5) when $V = \mathbb{R}^m$. However it is not known if (5) is sufficient for (3) to hold; moreover, no general necessary and sufficient condition is known for singular Brascamp-Lieb inequalities, and most work on them has been done case-by-case (see for instance [LaTh] for a drastically approach from Durcik-Thiele [DuTh]).

However, under the assumption of some additional symmetry, dubbed "cubical structure," on the linear maps Π_i and particular choice of the coefficient p_i, Durcik-Thiele are able to verify (5) is sufficient for (3). This work of Durcik-Thiele lends credence to the idea that (3) if and only if (5) is the "right theorem" one should be aiming for.

More precisely, to the cube $Q \equiv [0,1]^n$ in $\mathbb{R}^m$, we may associate each vertex to a function $j : \{1, \ldots, m\} \to \{0, 1\}$. In an abuse of language, we let the cube Q denote the set of all such functions j. And finally, given $x \in \mathbb{R}^{2m}$, we will write it as $x = (x_1^0, x_2^0, \ldots, x_m^0, x_1^1, x_2^1, \ldots, x_m^1)$.

Theorem 1 ([DuTh, Theorem 1])**.** *Let $\Pi_i : \mathbb{R}^{2m} \to \mathbb{R}^m$ be given by $\Pi_i x = (x_1^{j(1)}, x_2^{j(2)}, \ldots, x_m^{j(m)})$. Then* (5) *holds across all subspaces V of $\ker \Pi$, with equality if $V = \ker \Pi$, if and only if* (3) *holds.*

3.2 Broad picture of the proof: an induction using symmetry and Gaussians

Durcik-Thiele prove Theorem 1 using a clever induction to leverage the symmetry and cubical structure of their setup. First, without loss of generality they may assume that their singular Brascamp-Lieb integral is always of the form

$$\Lambda (K, A) \equiv \int\limits_{\mathbb{R}^{2m}} \left(\prod_{i=1}^{n} F_j (\Pi_j x) \right) K ((I A) x) \, dx \,, \tag{6}$$

where A is some $m \times m$ matrix satisfying some non-degeneracy condition (see e.g. (7) below, where $\epsilon = \epsilon(A)$). Then Durcik-Thiele spend most of their paper proving the following lemma, by induction. In what follows, let $g(x) \equiv e^{-\pi |x|^2}$ denote the standard Gaussian, and given a function j in the cube Q, define the reflection $j * i$ to be the element in the cube Q such that $j * i(a) = j(a)$ if $a \neq i$, and $j * i(i) = 1 - j(i)$.

Lemma 2. *Let $m \geq 1$, $0 \leq l \leq m$ and let $0 < \epsilon < 1$. Furthermore, let A be an $m \times m$ matrix that is such that*

$$\left| \det (I \ A) \Pi_j^T \right| > \epsilon \quad \text{and} \quad \|A\|_{HS} \leq \epsilon^{-1} \,, \tag{7}$$

for all $1 \leq j \leq m$, and assume the first l rows of A are equal to the first l rows of $-I$. Finally, let $(F_j)_{j \in Q}$ be a tuple of functions with the symmetry

$$F_j = F_{j*i} \quad \text{and} \quad \|F_j\|_{2^m} = 1 \,.$$

1. *If K is a kernel satisfying the Mihklin multiplier condition* (4) *for all α up to a certain threshold, and $\hat{K}$ satisfies the vanishing condition*

$$\hat{K}\left(\xi_1, \ldots, \xi_l, 0, \ldots, 0\right) \equiv 0\,,$$

   then

$$\left|\Lambda\left(K, A\right)\right| \lesssim_{m,l,\epsilon} 1.$$

2. *Let $l < i \leq m$, let $u \in \mathbb{R}^m$ and for each $t \in (0, \infty)$ let $c_t \in L^\infty(\mathbb{R}^m)$ with $|c|_\infty \leq 1$. If K is the kernel defined by*

$$\hat{K}\left(\xi\right) \equiv \int_0^\infty c_t(u)\widehat{\partial_i \partial_{i+l} g}\left(\left(I \;\; A\right)^T \left(t\xi\right)\right) e^{2\pi i u t \xi} \frac{dt}{t}\,,$$

   then

$$\left|\Lambda\left(K, A\right)\right| \lesssim_{m,l,\epsilon} \left(1 + \|u\|\right)^{2d(m-1)}\,.$$

Lemma 2 has two statements, (1) and (2). As Durcik-Thiele note, when $l = 0$, statement (1) yields Theorem 1. And when $l = m$, statement (1) is trivial because the kernel $K \equiv 0$. As such, the authors proceed by induction on l to get from statement (1) with $l = m$ to statement (1) with $l = 0$. More precisely, they show that statement (1) with $l = m$ implies statement (2) with $l = m - 1$, then they show that (2) with $l = m - 1$ implies (1) with $l = m - 1$, then that (1) with $l = m - 1$ implies (2) with $l = m - 2$, and so on ...; see e.g. Figure 3.2 for further details.

$l =$	0		1		2	...	m-1		m
Lemma 2 part (1)	**Theorem 1**		x		x	...	x		**Trivial**
	$\uparrow$	$\swarrow$	$\uparrow$	$\swarrow$	$\uparrow$	...	$\uparrow$	$\swarrow$	
Lemma 2 part (2)	x		x		x	...	x		(Ignore)

Figure 1: Implications of (1) and (2) in induction argument

The argument showing (1) $\Leftarrow$ (2) for the same l is more straightforward of the two arguments, as similar ideas have been applied elsewherein harmonic analysis. Namely the authors do a cone decomposition of the kernel K, which will preserve all of its Mihklin multiplier conditions. They choose the cone small enough so that the conic decompositions will also preserve nondegeneracy of the newly resulting A. Of course, the authors also leverage the vanishing condition for K. While reading this part of the argument, one might wonder though, why did the authors chose to use particular Gaussians in their conic decomposition?

The answer lies in the argument showing

$$\{\text{statement (2) for } l\} \Leftarrow \{\text{statement (1) for } l-1\}.$$

To prove this portion of the argument, the authors make use of the special of Gaussians. Namely, they first note that in a special case of (2), the form $\Lambda(K, A)$ is simply an integral of a product of Gaussians times squares of F_j's, using the symmetry inherent to the problem. Thus the authors need not worry about cancellation, and they need only show a sum of special cases of (2) is under control. They do this using the extra symmetries resulting from the sum, which then allow them to use general estimates involving Gaussians. (After all, this is generally how one estimates integrals of Gaussians in the first place, using symmetry). The authors then essentially demonstrate that this "special" case is in fact the only case that matters, and they show one can always reduce down to those considerations, using ubiquitously along the way the special multiplicative properties of Gaussians and the cubical symmetries in their setup the problem.

References

[BeCaChTa] Bennett, J. Carbery, A., Christ, F.M., and Tao, T. *The Brascamp-Lieb inequalities: finiteness, structure and extremals.* Geom. Func. Anal. **17** (2008), no. 5, 1343-1415.

[DuTh] Durcik, P. and Thiele, C., *Singular Brascamp-Lieb Inequalities with Cubical Structure,* `arXiv: 1809.08688v2`

[LaTh] Lacey, M. and Thiele, C., L^p *estimates for the bilinear Hilbert transform*, Proc. Nat. Acad. Sci. U.S.A., **94** (1997), no. 1, 33-35

Michel Alexis, University of Bonn
email: `micalexis.math@gmail.com`

4 On the polynomial Szemerédi theorem in finite fields I

After S. Peluse [Pe]

A summary written by Lars Becker

Abstract

Let $P_1, \dots, P_m \in \mathbb{Z}[y]$ be any linearly independent polynomials with zero constant term. We show that there exists $\gamma > 0$ such that any subset of $\mathbb{F}_q$ of size at least $q^{1-\gamma}$ contains a nontrivial polynomial progression $x, x + P_1(y), \dots, x + P_m(y)$, provided the characteristic of $\mathbb{F}_q$ is large enough.

4.1 Introduction

Let $P_1, \dots, P_m \in \mathbb{Z}[y]$ and let S be either $[N] = \{1, \dots, N\}$ or $\mathbb{F}_q$. We denote by $r_{P_1,\dots,P_m}(S)$ the size of the largest subset of S containing no polynomial progression $x, x + P_1(y), \dots, x + P_m(y)$ with $y \neq 0$. To avoid simple congruence obstructions to the existence of polynomial progressions in large subsets, we will always assume that $P_1(0) = \dots = P_m(0) = 0$, and we call the space of such polynomials $\mathbb{Z}[y]_0$. In this notation, Szemerédi's theorem states that

$$r_{y,2y,\dots,(k-1)y}([N]) = o_k(N)\,,$$

and Gower's proved the quantitative bound

$$r_{y,2y,\dots,(k-1)y}([N]) \lesssim_k \frac{N}{(\log\log N)^{c_k}}\,.$$

For general polynomials it is known by work of Bergelson and Leibman [BL] that

$$r_{P_1,\dots,P_m}([N]) = o_{P_1,\dots,P_m}(N)\,.$$

This talk is about a quantitative version of their result, but for finite fields $\mathbb{F}_q$ instead of for $[N]$.

Theorem 1. *Let $P_1, \dots, P_m \in \mathbb{Z}[y]_0$ be linearly independent over $\mathbb{Q}$. There exist $c, \gamma > 0$ such that if the characteristic of $\mathbb{F}_q$ is at least c, then*

$$r_{P_1,\dots,P_m}(\mathbb{F}_q) \lesssim_{P_1,\dots,P_m} q^{1-\gamma}\,,$$

and, more precisely,

$$\#\{(x,y) \in \mathbb{F}_q^2 \,:\, x, x+P_1(y), \dots, x+P_m(y) \in A\} \\ = \frac{|A|^{m+1}}{q^{m-1}} + O_{P_1,\dots,P_m}(q^{2-(m+1)\gamma})\,. \quad (1)$$

Let $m_1 \geq 1$, $m_2 \geq 0$, $P_1, \dots, P_{m_1}, Q_1, \dots, Q_{m_2} \in \mathbb{Z}[y]$. For every $F = (f_0, \dots, f_{m_1})$, $G = (g_0, \dots, g_{m_2})$ with $f_i, g_i : \mathbb{F}_q \to \mathbb{C}$ define

$$\Lambda_{P_1,\dots,P_{m_1}}^{Q_1,\dots,Q_{m_2}}(F;G) := \mathbb{E}_{x,y}[f_0(x) \prod_{j=1}^{m_1} f_j(x+P_j(y)) \prod_{j=1}^{m_2} g_j(Q_j(y))]\,,$$

where $\mathbb{E}$ denotes expectation with respect to the uniform probability measure on $\mathbb{F}_q$. Theorem 1 is an easy consequence of the following theorem with $m_2 = 0$, $m_1 = m$ and $f_0 = \dots = f_m = \mathbf{1}_A$.

Theorem 2. *Let $m_1 \geq 1$, $m_2 \geq 0$ and let $P_1, \dots, P_{m_1}, Q_1, \dots, Q_{m_2} \in \mathbb{Z}[y]_0$ be linearly independent over $\mathbb{Q}$. There exist $c, \gamma > 0$ such that if the characteristic of $\mathbb{F}_q$ is at least c, then*

$$\Lambda_{P_1,\dots,P_{m_1}}^{Q_1,\dots,Q_{m_2}}(F, \Psi) = \mathbf{1}_{\Psi \equiv 1} \prod_{j=0}^{m_1} \mathbb{E}[f_j] + O_{P_1,\dots,P_{m_1},Q_1,\dots,Q_{m_2}}(q^{-\gamma})\,,$$

whenever $F = (f_0, \dots, f_{m_1})$ is 1-bounded and $\Psi \in (\widehat{\mathbb{F}}_q)^{m_2}$.

4.2 Preliminaries

4.2.1 Upper bounds in terms of some U^s-norm

The first ingredient in the proof of Theorem 2 is the following bound for $\Lambda_{P_1,\dots,P_{m_1}}^{Q_1,\dots,Q_{m_2}}$ in terms of some U^s norm (for possibly very large s) of the f_i.

Proposition 3. *Let $P_1, \dots, P_{m_1}$, $Q_1, \dots, Q_{m_2} \in \mathbb{Z}[y]_0$. There exists $1 \geq \beta > 0$ and $s \in \mathbb{N}$ such that*

$$|\Lambda_{P_1,\dots,P_{m_1}}^{Q_1,\dots,Q_{m_2}}(F;\Psi)| \leq \min_j \|f_j\|_{U^s}^{\beta} + O_{P_1,\dots,P_{m_1},Q_1,\dots,Q_{m_2}}(q^{-\beta})$$

for all 1-bounded $F = (f_0, \dots, f_{m_1})$ and $\Psi \in (\hat{\mathbb{F}}_q)^{m_2}$.

This proposition is a slight generalization of a bound that already occurs in [Pr], and the proof is based on arguments from [BL].

4.2.2 Decomposing functions

Given a norm $\|\cdot\|$ on $\mathbb{R}^n$, we denote by

$$\|\phi\|^* := \sup\{\langle \phi, x\rangle \ : \ \|x\| \leq 1\}$$

the dual norm. The following proposition allow us to decompose a function into pieces, each of which is of controlled size in a certain given norm.

Proposition 4. *Let $\|\cdot\|$ be any norm on the $\mathbb{C}$-vector space of complex valued functions on $\mathbb{F}_q$, and let $\delta_1, \delta_2, \delta_3, \delta_4 > 0$. Suppose that $f : \mathbb{F}_q \to \mathbb{C}$ with $\|f\|_2 \leq 1$. If $q^{\delta_2-\delta_3} + q^{\delta_4-\delta_1} \leq 1/2$, then there exist $f_a, f_b, f_c : \mathbb{F}_q \to \mathbb{C}$ such that*

$$f = f_a + f_b + f_c$$

and $\|f_a\|^* \leq q^{\delta_1}$, $\|f_b\|_{L^1} \leq q^{-\delta_2}$, $\|f_c\|_{L^\infty} \leq q^{\delta_3}$ and $\|f_c\| \leq q^{-\delta_4}$.

The proof is based on the finite-dimensional Hahn-Banach theorem and simple properties of $\|\cdot\|_{L^1}, \|\cdot\|_{L^2}$ and $\|\cdot\|_{L^\infty}$.

4.3 The induction scheme

4.3.1 Overview of the argument

Theorem 2 is proven by induction on m_1. More specifically, denote by $\mathrm{E}(s)$ the statement that there exists $1 \geq \beta > 0$ such that the estimate

$$|\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}(F;\Psi)| \leq \min_j \|f_j\|^{\beta}_{U^s} + O(q^{-\beta}) \tag{E(s)}$$

holds for all 1-bounded F. Then the key step is to show that the case $(m_1 - 1, m_2 + 1)$ of Theorem 2 together with $\mathrm{E}(s)$ implies $\mathrm{E}(s-1)$. Since $\mathrm{E}(s)$ holds for some s by Proposition 3, we can iteratively apply this to deduce $\mathrm{E}(1)$, which is

$$|\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}(F;\Psi)| \leq \min_j \|f_j\|^{\beta}_{U^1} + O(q^{-\beta})\,. \tag{E(1)}$$

To deduce from E(1) the case (m_1, m_2) of Theorem 2, write $f'_{m_1} = f_{m_1} - \mathbb{E}[f_{m_1}]$. Then

$$\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}(F,\Psi) = \Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1-1}}(f_0,\ldots,f_{m_1-1};\Psi)\mathbb{E}[f_{m_1}] + 2\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}(f_1,\ldots,f_{m_1-1},\frac{1}{2}f'_{m_1};\Psi)$$

By the case (m_1-1, m_2) of Theorem 2 and 1-boundedness of f_{m_1}, the first term is

$$\mathbf{1}_{\Psi\equiv 1}\prod_{j=1}^{m_1}\mathbb{E}[f_j] + O(q^{-\gamma})\,.$$

By E(1), the second term is bounded by

$$2\min_j \|f_j\|^{\beta}_{U^1} + O(q^{-\beta}) = O(q^{-\beta})\,,$$

since $\|f'_{m_1}\|_{U^1} = |\mathbb{E}[f'_{m_1}]| = 0$. This completes the induction step.

4.3.2 Base case of the induction

We need the following lemma, which is a consequence of the Weil bound.

Lemma 5. *Let $P_1,\ldots,P_m \in \mathbb{Z}[y]_0$ be linearly independent over $\mathbb{Q}$. There exists $c > 0$ such that if the characteristic of $\mathbb{F}_q$ is at least c and $\psi_1,\ldots,\psi_m \in \hat{\mathbb{F}}_q$ are not all trivial, then*

$$\mathbb{E}_y\prod_{j=1}^{m}\psi_j(P_j(y)) \lesssim_{P_1,\ldots,P_m} q^{-1/2}\,.$$

Now we can prove the $m_1 = 1$ case of Theorem 1:

Lemma 6. *Let $m_2 \geq 0$ and $P_1, Q_1,\ldots,Q_{m_2} \in \mathbb{Z}[y]_0$ be linearly independent over $\mathbb{Q}$. There exists $c > 0$ such that if the characteristic of $\mathbb{F}_q$ is at least c, then*

$$|\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1}(F;\Psi) - \mathbf{1}_{\Psi\equiv 1}\prod_{j=0}^{1}\mathbb{E}_y[f_i(y)]| \lesssim_{P_1,Q_1,\ldots,Q_{m_2}} q^{-1/2}\,,$$

whenever $F = (f_0, f_1)$ is 1-bounded and $\Psi \in (\hat{\mathbb{F}}_q)^{m_2}$.

Proof. Write $f_1' = f_1 - \mathbb{E}[f_1]$ and $F' = (f_0, f_1')$. Then

$$\begin{aligned}\Lambda_{P_1}^{Q_1,\dots,Q_{m_2}}(F;\Psi) &= \Lambda_{P_1}^{Q_1,\dots,Q_{m_2}}(F';\Psi) + \mathbb{E}[f_1]\mathbb{E}_{x,y} f_0(x)\prod_{j=1}^{m_2}\psi_j(Q_j(y)) \\ &= \Lambda_{P_1}^{Q_1,\dots,Q_{m_2}}(F';\Psi) + \mathbb{E}[f_1]\mathbb{E}[f_0]\mathbb{E}_y\prod_{j=1}^{m_2}\psi_j(Q_j(y))\,.\end{aligned}$$

The y-expectation term equals 1 if all $\psi_j = 1$, and else it is $\lesssim q^{-1/2}$ if the characteristic of $\mathbb{F}_q$ is large enough, by Lemma 5. Since f_0, f_1 are 1-bounded, it follows that

$$\Lambda_{P_1}^{Q_1,\dots,Q_{m_2}}(F;\Psi) = \Lambda_{P_1}^{Q_1,\dots,Q_{m_2}}(F';\Psi) + \mathbf{1}_{\Psi\equiv 1}\mathbb{E}[f_0]\mathbb{E}[f_1] + O(q^{-1/2})\,.$$

The first term equals, by Fourier inversion and since $\mathbb{E}[f_1'] = \hat{f}_1'(1) = 0$:

$$\begin{aligned}&\sum_{\eta_0,\eta_1\in\hat{\mathbb{F}}_q}\hat{f}_0(\eta_0)\hat{f}_1'(\eta_1)\mathbb{E}_x[\eta_0(x)\eta_1(x)]\mathbb{E}_y[\eta_1(P_1(y))\prod_{j=1}^{m_2}\psi_j(Q_j(y))] \\ &= \sum_{\eta\neq 0}\hat{f}_0(\eta)\hat{f}_1'(\bar{\eta})\mathbb{E}_y[\eta(P_1(y))\prod_{j=1}^{m_2}\psi_j(Q_j(y))]\,.\end{aligned}$$

By Lemma 5 the y-expectation is $\lesssim q^{-1/2}$ if the characteristic of $\mathbb{F}_q$ is sufficiently large. By Cauchy-Schwarz, Plancherel and the 1-boundedness of f_0 and f_1, it follows that the whole expression is $\lesssim q^{-1/2}$, which completes the proof. □

References

[Pe] Peluse, S., *On the polynomial Szemerédi theorem in finite fields.* Duke Mathematical Journal, 168(5), 749-774.

[Pr] Prendiville, S., *Quantitative bounds in the polynomial Szemer ́edi theorem: the homogeneous case.* Discrete Anal., (5), 2017.

[BL] Bergelson, V. and Leibman, A., *Polynomial extensions of van der Waerden's and Szemerédi's theorems.* J. Amer. Math. Soc., 9(3):725–753, 1996.

Lars Becker, Universität Bonn
email: becker@math.uni-bonn.de

5 On the norm convergence of non-conventional ergodic averages

After T. Austin [A]

A summary written by Noa Bihlmaier

Abstract

We show the norm convergence of ergodic averages of the form

$$\frac{1}{2N+1} \sum_{n=-N}^{N} \prod_{i=1}^{d} f_i \circ T_i^n,$$

following the proof by Austin in [A], which proceeds by building suitable structured extensions of the initial system.

5.1 Statement of the theorem

Given a probability preserving transformation T on a standard measure space (X, Σ, μ), the multiple ergodic averages

$$\frac{1}{2N+1} \sum_{n=-N}^{N} \prod_{i=1}^{d} f_i \circ T^{ni}$$

are of central interest in ergodic theory. Most importantly, Host and Kra in [HK] and independently Ziegler in [Z] showed the $L^2(\mu)$–convergence of these averages for all $f_i \in L^\infty(\mu)$. Soon after, Tao in [T] generalized this situation to averaging d commuting transformations T_i rather than powers of a single operator T by converting the problem into a finitary problem. We prove the same result as Tao, with the proof of Austin [A].

Given a standard Borel measure space (X, Σ, μ) together with invertible commuting measure preserving continuous transformations $T_i \colon X \to X$ (for $i = 1, \dots, d$), we prove the following ergodic theorem which is a special case of [A, Thm 1.1].

Theorem 1. *For any choice of measurable functions $f_1, \dots, f_d$ in $L^\infty(\mu)$ the ergodic averages*

$$\frac{1}{2N+1} \sum_{n=-N}^{N} \prod_{i=1}^{d} f_i \circ T_i^n \qquad (\star)$$

converge in L^2–norm as N tends to infinity.

5.2 The proof

We prove this by induction on d. The induction start $d = 1$ is simply the von Neumann ergodic theorem. In every step $d-1 \to d$ we prove the convergence by finding a suitable "pleasant" extension of our initial system for which the convergence of the multiple ergodic averages can be proven.

5.2.1 Part 1: Reduction to pleasant systems

The main idea of this part is that in order to prove the convergence of multiple ergodic averages, it might suffice to prove convergence for an easier set of functions which still encodes all the information of the convergence of the averages on the original system. As we want to prove the statement by induction, we restrict only the first function to a smaller σ-subalgebra. This leads us to the notion of *pleasant systems.*

Definition 2. *We denote by Σ^{T_i} the invariant factor of T_i, i.e. the subalgebra of Σ consisting of all $A \in \Sigma$ with $\mu(A \Delta T_i(A)) = 0$. Further we denote by $\Sigma^{T_i = T_j}$ the invariant factor of $T_i \circ T_j^{-1}$. Now we call a system $T = (T_i)_{i=1,\ldots,d}$ pleasant if restricting the first coordinate to the σ–subalgebra*

$$\Xi := \Sigma^{T_1} \vee \bigvee_{i=2}^{d} \Sigma^{T_i = T_1}$$

yields a characteristic factor, i.e. for any functions $f_1, \ldots, f_d \in L^\infty(\mu)$ we obtain

$$\frac{1}{2N+1} \sum_{n=-N}^{N} \prod_{i=1}^{d} f_i \circ T_i^n - \frac{1}{2N+1} \sum_{n=-N}^{N} E_\mu[f_1 \mid \Xi] \circ T_1^n \cdot \prod_{i=2}^{d} f_i \circ T_i^n \to 0$$

in $L^2(\mu)$ as $N \to \infty$.

For such pleasant systems we are able to complete the induction step, noting that in the induction hypothesis we do not only assume the convergence for pleasant systems but rather for all systems.

Proposition 3. *If $T = (T_i)_{i=1,\ldots,d}$ is a pleasant system and Theorem 1 is true for all systems of $d-1$ commuting actions, then it also holds for (X, Σ, μ, T).*

Proof. In order to prove this, by pleasantness of the system we can first assume f_1 to be $\Sigma^{T_1} \vee \bigvee_{i=2}^d \Sigma^{T_i=T_1}$–measurable. Then we note that we can replace f_1 by a sequence $(f_1^{(m)})$ in $L^\infty(\mu)$ converging to f_1 in $L^2(\mu)$. As f_1 is $\Sigma^{T_1} \vee \bigvee_{i=2}^d \Sigma^{T_i=T_1}$–measurable we can assume the approximating sequence to consist of finite sums of products $g_1 \cdots g_d$, where $g_1 \in L^\infty(\mu\ |_{\sigma^{T_1}})$ and $g_i \in L^\infty(\mu\ |_{\Sigma^{T_1=T_i}})$.
Thus it suffices to prove the convergence of averages of the form

$$\frac{1}{2N+1} \sum_{n=-N}^{N} ((g_1 \cdot g_2 \cdots g_d) \circ T_1^n) \cdot \prod_{i=2}^{d} f_i \circ T_i^n.$$

Since $g_1 \in L^\infty(\mu\ |_{\Sigma^{T_1}})$, we obtain $g_1 \circ T_1^n = g_1$ and similarly $g_i \circ T_1^n = g_i \circ T_i^n$ for all i. This simplifies the above term to

$$g_1 \cdot \frac{1}{2N+1} \sum_{n=-N}^{N} \prod_{i=2}^{d} (g_i \cdot f_i) \circ T_i^n,$$

but this converges by induction hypothesis. $\square$

5.2.2 Part 2: Constructing pleasant extensions

Knowing the convergence of $(\star)$ for all pleasant systems reduces the problem to finding "enough" pleasant systems to obtain the convergence of $(\star)$ for all systems, i.e. the following proposition together with the above Proposition 3 clearly imply the induction step and hence the desired statement.

Proposition 4. *Every system $T = (T_i)_{i=1,\dots,d}$ has a pleasant extension, i.e., there exists a pleasant system $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T})$ and a factor map Ψ intertwining the actions.*

In order to construct such a pleasant extension we need to iteratively pass to an extension which controls the previous averages better. This step is done via the *Furstenberg self-joining*.

Definition 5. *For any action $T = (T_i)_{i=1,\dots,d}$ on (X, Σ, μ) we define the Furstenberg self-joining of (X, Σ, μ) corresponding to T as $(X^d, \Sigma^{\otimes d}, \mu^{*d})$, where μ^{*d} is defined via*

$$\mu^{*d}(A_1 \times \cdots \times A_d) := \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} \int_X \prod_{i=1}^{d} 1_{A_i} \circ T_i^n d\mu.$$

Note that the well-definedness of the above limit follows by the induction hypothesis, and that the Furstenberg self-joining is invariant under the actions of $T_i \times T_i \times \cdots \times T_i$ and of $T_1 \times T_2 \times \cdots \times T_d$. Using these actions, we are now ready to construct the pleasant extension starting with an initial system (X, Σ, μ, T).

Construction of a pleasant extension.

1) First we iteratively construct actions on Furstenberg self-joinings.
Define $(X^{(0)}, \Sigma^{(0)}, \mu^{(0)}, T^{(0)}) = (X, \Sigma, \mu, T)$ and now build extensions

$$\psi^{(m)} \colon (X^{(m)}, \Sigma^{(m)}, \mu^{(m)}, T^{(m)}) \to (X^{(m-1)}, \Sigma^{(m-1)}), \mu^{(m-1)}, T^{(m-1)})$$

iteratively for all $m \in \mathbb{N}$ by setting $(X^{(m)}, \Sigma^{(m)}, \mu^{(m)})$ as the Furstenberg self-joining of $(X^{(m-1)}, \Sigma^{(m-1)}, \mu^{(m-1)}, T^{(m-1)})$ and $T^{(m)} = (T_i^{(m)})_{i=1,\dots,d}$ via

$$\begin{aligned}
T_1^{(m)} &\coloneqq T_1^{(m-1)} \times T_2^{(m-1)} \times \cdots \times T_d^{(m-1)} \\
T_2^{(m)} &\coloneqq T_2^{(m-1)} \times T_2^{(m-1)} \times \cdots \times T_2^{(m-1)} \\
&\vdots \\
T_d^{(m)} &\coloneqq T_d^{(m-1)} \times T_d^{(m-1)} \times \cdots \times T_d^{(m-1)}.
\end{aligned}$$

The projection $\psi^{(m)}$ is now given by the projection onto the first coordinate of $X^{(m)}$.

2) Having constructed this projective system of measure preserving actions, we want to pass to a limit as each step "controls" the previous one.
Thus we define the desired extension $(\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T})$ as the inverse (projective) limit of this system of extensions, i.e.,

$$(\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T}) \coloneqq \varprojlim (X^{(m)}, \Sigma^{(m)}, \mu^{(m)}, T^{(m)})$$

equipped with the corresponding factor map

$$\Psi \colon (\tilde{X}, \tilde{\Sigma}, \tilde{\mu}, \tilde{T}) \to (X^{(0)}, \Sigma^{(0)}, \mu^{(0)}, T^{(0)}) = (X, \Sigma, \mu, T).$$

Proof of pleasantness.

To see that this extension is indeed pleasant, first we show that it suffices to prove that for any $\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_d \in L^\infty(\tilde{\mu})$ and $\tilde{g} \in L^\infty(\mu^{*d} \mid_{(\Sigma^{\otimes d})^{\tilde{T}_1 \times \tilde{T}_2 \times \cdots \times \tilde{T}_d}})$ we have

$$\int_{\tilde{X}^d} \tilde{f}_1 \circ \tilde{\pi}_1 \cdot \left(\prod_{i=2}^d \tilde{f}_i \circ \tilde{\pi}_i \right) \cdot \tilde{g} d\tilde{\mu}^{*d} = \int_{\tilde{X}^d} E_{\tilde{\mu}}[\tilde{f}_1 \mid \Xi] \circ \tilde{\pi}_1 \cdot \left(\prod_{i=2}^d \tilde{f}_i \circ \tilde{\pi}_i \right) \cdot \tilde{g} d\tilde{\mu}^{*d},$$

where

$$\Xi := \tilde{\Sigma}^{\tilde{T}_1} \vee \tilde{\Sigma}^{\tilde{T}_2=\tilde{T}_1} \vee \cdots \vee \tilde{\Sigma}^{\tilde{T}_d=\tilde{T}_1}$$

and the $\tilde{\pi}_i$ are the corresponding coordinate projections. This is done by showing that the Furstenberg self-joinings control our averages, i.e. we prove that if $f_1 \in L^\infty(\mu)$ fulfills

$$\int_X^d f_1 \circ \pi_1 \cdot \left(\prod_{i=2}^d f_i \circ \pi_i \right) \cdot g d\mu^{*d} = 0$$

for every $f_2, \dots, f_d \in L^\infty(\mu)$ and $g \in L^\infty(\mu^{*d} \mid_{(\Sigma^{\otimes d})^{T_1 \times T_2 \times \cdots \times T_d}})$, then

$$\frac{1}{2N+1} \sum_{n=-N}^{N} \prod_{i=1}^{d} f_i \circ T_i^n \to 0$$

for every $f_2, \dots, f_d \in L^\infty(\mu)$.

Second we prove that the invariant factors behave nicely under projective limits and hence we can obtain a good description of Ξ in terms of the σ–algebras

$$\Xi^{(m)} := (\Sigma^{(m)})^{T_1^{(m)}} \vee (\Sigma^{(m)})^{T_2^{(m)}=T_1^{(m)}} \vee \cdots \vee (\Sigma^{(m)})^{T_d^{(m)}=T_1^{(m)}}.$$

This allows us to translate our problem onto some finite level of the projective system. There the statement can be proven with some calculations.

□

References

[A] Austin, T., *On the norm convergence of nonconventional ergodic averages.* Ergodic Theory Dynam. Systems, 30(2) (2010), 321–338.

[HK] Host, B. and Kra, B., *Nonconventional ergodic averages and nilmanifolds.* Ann. Math., 161(1) (2005), 397–488.

[Z] Ziegler, T. *Universal characteristic factors and Furstenberg averages.* J. Amer. Math. Soc., 20(1) (2007), 35–97.

[T] Tao, T. *Norm convergence of multiple ergodic averages for commuting transformations.* Ergodic Theory Dynam. Systems, 28 (2008), 657–688.

Noa Bihlmaier, Tübingen
email: nobi@fa.uni-tuebingen.de

6 Using Quadratic Fourier Analysis to Find 4-term Arithmetic Progressions

After B. Green [G]

A summary written by Bora Çalım and Nihan Tanısalı

Abstract

We give an inverse theorem for the Gowers U^3 norm on $\mathbb{F}_5^n$ and use it to prove the existence of many (proportional to the density of the set) 4 term arithmetic progressions with the same step size in subsets of $\mathbb{F}_5^n$.

6.1 Introduction

Let $1 \geq \alpha > 0$ be a real number. We aim to show the existence of 4 term arithmetic progressions in subsets $A \subset \mathbb{F}_5^n$ with density α for large enough n. Throughout the summary, G will denote $\mathbb{F}_5^n$, and $N = |G|$.

Theorem 1. *Let $\alpha, \epsilon > 0$ be real numbers. Then there is an $n_0 = n_0(\alpha, \epsilon)$ with the following property. Suppose that $n > n_0(\alpha, \epsilon)$, and that $A \subseteq G$ is a set with density α. Then there is some $d \neq 0$ such that A contains at least $(\alpha^4 - \epsilon)\, N$ four-term arithmetic progressions with common difference d.*

Instead of working with the set $A \subset \mathbb{G}$, we will consider its characteristic function $1_A : \mathbb{G} \to \{0, 1\}$. The averages, the Fourier transform, and the Gowers uniformity norm of functions carry information about the number of arithmetic progressions in A. However, the techniques used to prove the existence of 3-APs cannot be directly generalized to 4-APs. We summarize these differences and introduce the required notions.

Definition 2 (Λ_3, Λ_4)**.** *For $f_i : G \to [-1, 1]$ we define $\Lambda_3(f_1, f_2, f_3) = \mathbb{E}_{x,d} f_1(x) f_2(x + d) f_3(x + 2d)$, and $\Lambda_4(f_1, f_2, f_3, f_4)$ analogously.*

Definition 3 (Gowers norms)**.** *The Gowers uniformity norm of $f : G \to \mathbb{R}$ for integer $d \geq 2$ is defined as follows*

$$\|f\|_{U^d}^{2^d} := \sum_{x,h_1,\ldots,h_d} \prod_{\omega_1,\ldots,\omega_d \in \{0,1\}} f(x + h_1 w_1 + \ldots + h_d \omega_d).$$

In $k = 3$ case

- The operator Λ_3 is controlled by the Gowers U^2-norm. Specifically for any three functions $f_1, f_2, f_3 : G \to [-1, 1]$ we have

$$|\Lambda_3 (f_1, f_2, f_3)| \leqslant \inf_{i=1,2,3} \|f_i\|_{U^2} .$$

- (Gowers inverse theorem) If the Gowers U^2-norm of a function $f : G \to [-1, 1]$ is large, f must have a large Fourier coefficient:

$$\|f\|_{U^2} \geqslant \delta \quad \Rightarrow \quad \|\widehat{f}\|_\infty \geqslant \delta^2.$$

The first item is directly generalized, while the second item is not. The following proposition and example illustrate this.

Proposition 4. *Let $f_1, \ldots, f_4 : G \to [-1, 1]$ be any four functions. Then we have*

$$|\Lambda_4 (f_1, \ldots, f_4)| \leqslant \inf_{i=1,\ldots,4} \|f_i\|_{U^3} .$$

Example 5. *There is a function $f : G \to \mathbb{C}$ with $\|f\|_\infty \leqslant 1$ such that $\|f\|_{U^3} = 1$, but such that $\|\widehat{f}\|_\infty \leqslant N^{-1/2}$. Namely $f = w^{x^T x}$.*

Instead, we find that f has significant correlation with a quadratic phase:

Theorem 6. *Suppose that $f : G \to [-1, 1]$ is a function for which $\|f\|_{U^3} \geqslant \delta$. Then there is a matrix $M \in \mathfrak{M}_n (\mathbb{F}_5)$ and a vector $r \in \mathbb{F}_5^n$ so that*

$$\left| \mathbb{E}_{x \in G} f(x) \omega^{x^T M x + r^T x} \right| \gg_\delta 1.$$

6.2 Proof of Theorem 6

There are 3 steps in proving a function f with large U^3 norm correlates with a quadratic phase $w^{x^T M x + r^T x}$. Throughout, $|G| \gg_\delta 1$ whenever needed, $f : G \to [-1, 1]$, $\|f\|_{U^3} \geqslant \delta$, M denotes an $n \times n$ matrix with entries from $\mathbb{F}_5$, b denotes a vector in $\mathbb{F}_5^n$, and $\Delta(f; h)(x) = f(x)f(x - h)$ is a "multiplicative derivative".
The first step is to show that the derivative of f obeys a "weak linearity" property: There is a function $\phi : G \to \widehat{G}$ and $S \subseteq G$ with $|S| \gg_\delta |G|$ such that

1. $|\Delta(f;h)^\wedge(\phi(h))| \gg_\delta 1$ for all $h \in S$

2. There are $\gg_\delta |G|^3$ quadruples $(s_1, s_2, s_3, s_4) \in S^4$ such that $s_1 + s_2 = s_3 + s_4$ and $\phi(s_1) + \phi(s_2) = \phi(s_3) + \phi(s_4)$.

The second step is to show that this weak linearity property implies a stronger linearity property: If $\phi : G \to \widehat{G}$, $S \subseteq G$ satisfy the conclusions 1 and 2 of the previous step, then there is some linear function $\psi(x) = Mx + b$ such that $\psi(x) = \phi(x)$ for $\gg_\delta |G|$ values of $x \in S$. We give a sketch of the proof of this step.
Consider $\Gamma = \{(h, \phi(h)) : h \in S\}$. By conclusion 2 of the first step, we can use the Balog-Szemerédi-Gowers theorem to find some $\Gamma' \subseteq \Gamma$ such that $|\Gamma'| \gg_\delta |\Gamma| \gg_\delta |G|$ and $|\Gamma' + \Gamma'| \ll_\delta |\Gamma'|$. Identifying $G \times \widehat{G}$ with $\mathbb{F}_5^{2n}$, by Freiman's theorem, we can find a subspace $H \subseteq \mathbb{F}_5^{2n}$ containing Γ' such that $|H| \ll_\delta |\Gamma'| \ll_\delta |G|$.
Consider the canonical projection $\pi : H \to G$ to the first factor, and let $S' = \pi(\Gamma')$, so that $|\pi(H)| \geq |S'| \gg_\delta |G|$. By the rank-nullity theorem, it follows that $\dim \ker(\pi) \ll_\delta 1$. Let $H' = (\ker(\pi))^\perp$, so that $H = \bigcup_{x \in \ker(\pi)} (H' + x)$, where the union is disjoint and taken over $\ll_\delta 1$ elements. Observe that π is injective on each of the cosets in the union. By the pigeonhole principle, there is some x such that $|(x + H') \cap \Gamma'| \gg_\delta |\Gamma'| \gg_\delta |G|$. Let $\Gamma'' = (x + H') \cap \Gamma'$ and $S'' = \pi(\Gamma'')$, $V = \pi(x + H')$. Then $\psi : V \to \widehat{G}$ given by the composition of π^{-1} and the canonical projection to the second factor is an affine map, so $\psi(x) = Mx + b$ for some M, b. It can be seen that $\psi(x) = \phi(x)$ for all $x \in S''$, so the proof is complete.
Combining the two steps, we can find some M, b such that

$$\mathbb{E}_h \left|\Delta(f;h)^\wedge(Mh + b)\right|^2 \gg_\delta 1.$$

It turns out that a Matrix M satisfying the above bound is approximately symmetric in a precise sense: If

$$\mathbb{E}_h \left|\Delta(f;h)^\wedge(Mh + b)\right|^2 \gg_\delta 1,$$

Then $\text{rank}(M) \ll_\delta 1$.
From this we can recover a fully symmetric matrix M', which gives theorem 6.

6.3 Arithmetic Regularity for U^3

In this section, the main objective is to decompose a function $f : \mathbb{G} \to [-1, 1]$ into three parts. The first one, $\mathbb{E}(f \mid \mathcal{B})$, is constant on certain sets, the second one is the error term in the sense of having a small L_2 norm, and the third has a small U^3 norm.

Definition 7 (Factors, Conditional Expectation, Rank of a Quadratic Factor)**.** *Let $\phi_1, \ldots, \phi_k : G \to G$ be any functions. The σ-algebra, $\mathcal{B}$, generated by the sets (atoms) of the form $\{x \in G \mid \phi_1(x) = c_1, \ldots, \phi_k(x) = c_k\}$ are called a factor. The conditional expectation of f is defined as*

$$\mathbb{E}(f \mid \mathcal{B})(x) := \mathbb{E}_{x \in \mathcal{B}(x)} f(x)$$

where $\mathcal{B}(x)$ is the atom of $\mathcal{B}$ containing x. If all the functions $\phi_i(x)$ $i \leq k$ are of the form $r_i^T x$ for some $r_i \in G$ the factor $\mathcal{B}$ generated by $\phi_i, i \leq k$ is called a linear factor of complexity at most k.
Let $i \leq d_i, r_i \in G$ and $M_j, j \leq d_2$ be symmetric matrices in $\mathcal{M}_n(G)$. Let $\mathcal{B}_1$ be the factor generated by the linear functions $\phi_i(x) = r_i^T x$; and $\mathcal{B}_2$ be the factor generated by $\phi_i(x) = r_i^T x, i \leq d_1$ and $\psi_j(x) = x^T M_j x$, $j \leq d_1$. $\mathcal{B}_2$ is a refinement of $\mathcal{B}_1$. $(\mathcal{B}_1, \mathcal{B}_2)$ is called a factor of complexity (d_1, d_2). We say that $(\mathcal{B}_1, \mathcal{B}_2)$ has rank at least r if for all nontrivial linear combinations of $M_1, \ldots, M_{d_2}$ has rank at least r.

With the following lemma, we write any function $f : G \to [-1, 1]$ as a sum of a measurable function with respect to a quadratic factor and two error terms that are small, respectively, in L^2 and U^3. The strength of the lemma is to make $\|f_3\|_{U^3}$ arbitrarily small by choosing a suitable growth function ω_2 with the cost of making the complexity higher.

Lemma 8. *Let $\delta > 0$ be a parameter, and let $\omega_1, \omega_2 : \mathbb{R}_+ \to \mathbb{R}_+$ be arbitrary growth functions (which may depend on δ). Let $n > n_0 (\delta, \omega_1, \omega_2)$ be sufficiently large, and let $f : G \to [1, 1]$ be a function. Let $\left(\mathcal{B}_1^{(0)}, \mathcal{B}_2^{(0)}\right)$ be a quadratic factor of complexity $\left(d_1^{(0)}, d_2^{(0)}\right)$. Then there is a quadratic factor $(\mathcal{B}_1, \mathcal{B}_2)$ with the following properties: $(\mathcal{B}_1, \mathcal{B}_2)$ refines $\left(\dot{\mathcal{B}}_1^{(0)}, \mathcal{B}_2^{(0)}\right)$; the complexity of $(\mathcal{B}_1, \mathcal{B}_2)$ is at most (d_1, d_2), where*

$$d_1, d_2 \leqslant C\left(\delta, \omega_1, \omega_2, d_1^{(0)}, d_2^{(0)}\right),$$

for some fixed function C; the rank of $(\mathcal{B}_1, \mathcal{B}_2)$ is at least $\omega_1(d_1+d_2)$; there is a decomposition $f = f_1 + f_2 + f_3$, where

$$f_1 := \mathbb{E}(f \mid \mathcal{B}_2),$$
$$\|f_2\|_2 \leqslant \delta,$$
$$\|f_3\|_{U^3} \leqslant 1/\omega_2(d_1+d_2).$$

6.4 Main Theorem

To understand $\mathcal{B}_2$ measurable functions, i.e., functions that are constant on the atoms of $\mathcal{B}_2$ with complexity (d_1, d_2), we study functions on the configuration space $\mathbb{F}_5^{d_1} \times \mathbb{F}_5^{d_2}$. We take $r_1, \dots, r_{d_1}$ linearly independent and define $\Gamma(x) := (r_1^T, \dots, r_{d_1}^T)$ and $\Phi(x) := (x^T M_1 x, \dots, x_{d_2}^T M_{d_2} x)$.

Proof of theorem 1. We apply theorem 8 to 1_A to obtain a decomposition $1_A = f_1 + f_2 + f_3$ such that the quadratic factor $(\mathcal{B}_1, \mathcal{B}_2)$ is with complexity (d_1, d_2) $d_i \leq d_0(\alpha, \epsilon)$ and the rank r is such that

$$r \geq 100(\log(1/\epsilon) + \log(1/\alpha) + d_1 + d_2).$$

The parameter δ and ω (which only depends on α and ϵ justifying the bound for d_0) will be specified afterwards. We define the $n - d_1$ dimensional space $H := \langle r_1, \dots, r_{d_1} \rangle^{\perp}$, and μ_H to be the normalised measure $\mu_H : 1_H / \mathbb{E} 1_H$. To prove the theorem, we show

$$\mathbb{E}_{x,d} 1_A(x) 1_A(x+d) 1_A(x+2d) 1_A(x+3d) \mu_H(d) > (\alpha^4 - \epsilon).$$

The left-hand side of the above expression splits into 81 parts after the substitution $1_A = f_1 + f_2 + f_3$.
Claim 1. The 65 terms containing f_2 has contribution $\leq \epsilon/200$.
Claim 2. The 65 terms containing f_3 has contribution $\leq \epsilon/200$.

Proof. Suppose that $g_1 = f_3$, the other cases are similar. We write the term as

$$\mathbb{E}_{x,d} g_1(x) g_2(x+d) g_3(x+2d) g_4(x+3d) \mu_H(d) \tag{1}$$

where g_2, g_3, g_4 are one of the f_1, f_2, f_3. We make the observation

$$1_H(d) = \sum_t 1_{t+H}(x) 1_{t+H}(x+2d)$$

where the sum is over all cosets of H in G. By proposition 4

$$\mathbb{E}_{x,d} g_1(x) g_2(x+d) 1_{t+H}(x+d) g_3(x+2d) 1_{t+H}(x+2d) g_4(x+3d) \le \|f_3\|_{U^3} \le 1/\omega_2(d_1+d_2).$$

Hence we bound (1) by $< 5^{2d_1}/\omega(d_1+d_2)$. Provided that $\omega(m) \ge 5^{m+4}/\epsilon$. □

Claim 3. As f is a $\mathcal{B}_2$ measurable function we define $\mathbf{f}_1 : \mathbb{F}_5^{d_1} \times \mathbb{F}_5^{d_2}$ such that $f_1(x) = \mathbf{f}_1(\Gamma(x), \phi(x))$ for all $x \in G$. Since the size of the factors are not equal, we have

$$\begin{aligned} &\mathbb{E}_{x,d} f_1(x) f_1(x+d) f_1(x+2d) f_1(x+3d) \mu_H(d) \\ &= \mathbb{E}_{\substack{a \in \mathbb{F}_5^{d_1}, b^{(1)}, \ldots, b^{(4)} \in \mathbb{F}_5^{d_2} \\ b^{(1)} - 3b^{(2)} + 3b^{(3)} - b^{(4)} = 0}} \mathbf{f}_1(a, b^{(1)}) \mathbf{f}_1(a, b^{(2)}) \mathbf{f}_1(a, b^{(3)}) \mathbf{f}_1(a, b^{(4)}) \\ &+ O(5^{2d_1 + 3d_2 - r/2}). \end{aligned}$$

The constraints on a and b is a result of two facts: $d \in H$ and $\Phi(x) - 3\Phi(x+d) + 3\Phi(x+2d) - \Phi(x+3d) = 0$.

$$\begin{aligned} (5^{-2d_1 - 3d_2} + O(5^{-r/2})) \sum_{a \in \mathbb{F}_5^n} \sum_{\substack{a \in \mathbb{F}_5^{d_1}, b^{(1)}, \ldots, b^{(4)} \in \mathbb{F}_5^{d_2} \\ b^{(1)} - 3b^{(2)} + 3b^{(3)} - b^{(4)} = 0}} & \mathbf{f}_1(a, b^{(1)}) \mathbf{f}_1(a, b^{(2)}) \\ & \times \mathbf{f}_1(a, b^{(3)}) \mathbf{f}_1(a, b^{(4)}) \\ \ge (5^{-2d_1 - 3d_2} + O(5^{-r/2})) (\mathbb{E}_{(a,b) \in \mathbb{F}_5^{d_1} \times \mathbb{F}_5^{d_2}} & \mathbf{f}_1(a, b))^4. \end{aligned}$$

The last line follows from two applications of Cauchy-Schwarz.
Claim 4. $\mathbb{E}_{(a,b) \in \mathbb{F}_5^{d_1} \times \mathbb{F}_5^{d_2}} \mathbf{f}_1(a, b) = \alpha(1 + O(5^{d_1 + d_2 - r/2}))$. This claim is a result of the fact that atoms are close in size. After some calculations, the theorem follows from these four claims.

□

References

[G] Green, B. *Montréal Notes on Quadratic Fourier Analysis.* Proceedings of the CRM-Clay Conference on Additive Combinatorics, Montréal 2006.

Bora Çalim, Nihan Tanisali, Koç
email: `bcalim21@ku.edu.tr`
email: `ntanisali16@ku.edu.tr`

7 Degree lowering for the polynomial Szemerédi theorem in finite fields

After S. Peluse [Pel19]

A summary written by Jonathan Chapman

Abstract

We give an exposition of a degree lowering argument used by Peluse to obtain quantitative bounds in the finite field polynomial Szemerédi theorem.

7.1 Introduction

In the previous summary, an overview was given of the recent work of Peluse on the polynomial Szemerédi theorem in finite fields. Recall that the goal is to obtain upper bounds on the size of an arbitrary $A \subseteq \mathbb{F}_q$ which does not contain a progression of the form $\{x, x+P_1(y), \ldots, x+P_m(y)\}$, where the P_i are integer polynomials with 0 constant term. Peluse's method involves studying properties of counting operators of the form

$$\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}(F;\Psi) := \mathbb{E}_{x,y} f_0(x) \prod_{i=1}^{m_1} f_i(x+P_i(y)) \prod_{j=1}^{m_2} \psi_j(Q_j(y)). \tag{1}$$

Specifically, Peluse's power-saving quantitative bounds for the finite field Szemerédi theorem [Pel19, Theorem 1.1] are a corollary of the following result.

Theorem 1 ([Pel19, Theorem 2.1])**.** *Let $m_1 \geq 1$ and $m_2 \geq 0$ and let $P_1, \ldots, P_{m_1}, Q_1, \ldots, Q_{m_2} \in \mathbb{Z}[y]_0$ be linearly independent over $\mathbb{Q}$. There exist $c, \gamma > 0$ such that if the characteristic of $\mathbb{F}_q$ is at least c, then*

$$\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}(F;\Psi) = 1_{\Psi=1} \prod_{i=0}^{m_1} \mathbb{E}_x f_i(x) + O_{P_1,\ldots,P_{m_1},Q_1,\ldots,Q_{m_2}}(q^{-\gamma}) \tag{2}$$

whenever $F = (f_0, \ldots, f_{m_1})$ is 1-bounded and $\Psi \in (\widehat{\mathbb{F}}_q)^{m_2}$.

The main purpose of this summary is to give a proof of this theorem using Peluse's degree lowering argument [Pel19, Lemma 4.1].

7.2 Discorrelation estimates and norm control

Our primary objective is to obtain an asymptotic formula (2) for the counting operators (1). For fixed Ψ, these counting operators are averages of functions f_i evaluated along polynomial progressions. Compare this with (2), where the main term is just a product of the averages of the f_i. As noted in [Kuc21], we can therefore think of (2) as demonstrating a kind of 'discorrelation' of the f_i. This motivates the following non-standard, but useful, definition.

Definition 2 (Discorrelation estimate)**.** *Let $m_1 \geqslant 1$ and $m_2 \geqslant 0$ and let $P_1, \ldots, P_{m_1}, Q_1, \ldots, Q_{m_2}$ be integer polynomials with 0 constant term. Let q be a prime power, and let $C, \gamma > 0$. We say that the counting operator $\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}$ defined by (1) satisfies a (C,γ)-*discorrelation estimate *if*

$$\left|\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}(F;\Psi) - 1_{\Psi=1}\prod_{i=0}^{m_1}\mathbb{E}_x f_i(x)\right| \leqslant Cq^{-\gamma}$$

holds for all 1-bounded $F = (f_0, \ldots, f_{m_1})$ and all $\Psi \in (\widehat{\mathbb{F}}_q)^{m_2}$.

Recall from the previous summary that the *Gowers U^s-norm*[1] is defined by the equation

$$\|f\|_{U^s}^{2^s} := \mathbb{E}_{x,h_1,\ldots,h_s\in\mathbb{F}_q}\Delta_{h_1,\ldots,h_s}f(x).$$

In additive combinatorics literature, one refers to a counting operator, such as (1), as being 'controlled' by the U^s-norm if the counting operator is small whenever one of the f_j has a small U^s-norm.[2] We formalise this by making the following (again, non-standard) definition.

Definition 3 (Norm control)**.** *Let $q, m_1, m_2, P_1, \ldots, P_{m_1}, Q_1, \ldots, Q_{m_2}$ be as in the previous definition. Let $b_1, b_2, b_3 > 0$. We say that the counting operator $\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}$ defined by (1) is* (b_1,b_2,b_3)-controlled by the U^s-norm *if*

$$\left|\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}(F;\Psi)\right| \leqslant b_1 \min_j \|f_j\|_{U^s}^{b_2} + b_3$$

holds for all 1-bounded $F = (f_0, \ldots, f_{m_1})$ and all $\Psi \in (\widehat{\mathbb{F}}_q)^{m_2}$.

[1] Note that these are norms for all $s \geqslant 2$, but for $s = 1$ this defines only a semi-norm.
[2] See, for example, [Pre17].

To demonstrate the utility of control by Gowers norms, we now prove Theorem 1 under the assumption of U^1-norm control.

Lemma 4. *Let $q, m_1, m_2, P_1, \ldots, P_{m_1}, Q_1, \ldots, Q_{m_2}$ be as above. Suppose that $\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}$ is (a, b, q^{-c})-controlled by the U^1-norm for some $a, b, c > 0$. If $\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1-1}}$ satisfies a (C, γ)-discorrelation estimate, then $\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}$ satisfies a $(C + 2, \min\{c, \gamma\})$-discorrelation estimate.*

Proof. Let $F = (f_0, \ldots, f_{m_1})$ be 1-bounded, and let $\Psi \in (\widehat{\mathbb{F}}_q)^{m_2}$. Write $h = f_{m_1} - \mathbb{E}_x f_{m_1}(x)$, whence $\|h\|_{U^1} = 0$ and $h/2$ is 1-bounded. Using our U^1 control hypothesis, we observe that

$$\begin{aligned} &|\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}(F; \Psi) - \mathbb{E}_x f_{m_1}(x) \Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1-1}}(f_0, \ldots, f_{m_1-1}; \Psi)| \\ &\qquad\qquad \leqslant 2|\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}(f_0, \ldots, f_{m_1-1}, h/2; \Psi)| \leqslant 2q^{-c}. \end{aligned}$$

The result now follows from our discorrelation estimate for $\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1-1}}$. $\square$

7.3 Degree lowering

In view of the previous lemma, if we can control the counting operators (1) by the U^1-norm, then we can prove Theorem 1 by induction on m_1. Unfortunately, as shown in [Pel19, Proposition 2.2] and [Pre17, §§3-5], one can usually only obtain control of (1) by a U^s-norm with s a very large number which depends on the degrees of the P_i and Q_j.
The key insight of Peluse was that it is possible to leverage discorrelation estimates for counting operators $\Lambda^{S_1,\ldots,S_{m_2+1}}_{R_1,\ldots,R_{m_1-1}}$ to improve U^s-control for $\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}$ to U^{s-1}-control. This strategy is known as *degree lowering* and has become a highly influential tool in additive combinatorics and beyond (for further examples, see the other summaries in these proceedings).
The degree lowering argument employed to prove Theorem 1 is encapsulated in the following lemma, which we have reworded using the definitions introduced in the previous section.

Lemma 5 ([Pel19, Lemma 4.1])**.** *Let $P_1, \ldots, P_{m_1}, Q_1, \ldots, Q_{m_2} \in \mathbb{Z}[y]_0$ be linearly independent, for some Let $m_1 \geqslant 2$ and $m_2 \geqslant 0$. Suppose there exist $b_1, b_2, b_3, b_4, c_1, c_2, \gamma > 0$ and $s \in \mathbb{N}$ such that the following two conditions both hold whenever $\mathbb{F}_q$ has characteristic at least c_1.*

(I) *For all linearly independent* $R_1, \ldots, R_{m_1-1}, S_1, \ldots, S_{m_2+1} \in \mathbb{Z}[y]_0$*, the counting operator* $\Lambda^{S_1,\ldots,S_{m_2+1}}_{R_1,\ldots,R_{m_1-1}}$ *satisfies a* (c_2, γ)*-discorrelation estimate.*

(II) *The operator* $\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}$ *is* (b_1, b_2, b_3)*-controlled by the* U^s*-norm.*

Then there exist $c_1', c_2', \gamma' > 0$ *depending only on the* P_i *and* Q_j *such that, if* $\mathbb{F}_q$ *has characteristic at least* $\max(c_1', b_4)$*, then the following is true. For every* $\delta_1, \delta_2, \delta_3, \delta_4 > 0$ *satisfying* $q^{\delta_2-\delta_3} + q^{\delta_4-\delta_1} \leqslant 1/2$*, the counting operator* $\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}$ *is a* $(q^{\delta_1}, 2^{1-s}, \beta)$*-controlled by the* U^{s-1}*-norm, where*

$$\beta = q^{\delta_1}\left(\frac{c_2'}{q^{\gamma'}}\right)^{2^{2-2s}} + q^{-\delta_2} + q^{(1-b_2)\delta_3 - b_2\delta_4} b_1 + q^{\delta_3} b_3.$$

Remark 6. *Although the flexibility in the choice of* δ_i *is important in the proof of [Pel19, Theorem 2.1], we will not keep track too carefully of these parameters in our proofs. We similarly omit details regarding the* b_i *and* c_j*. The interested reader should consult [Pel19] for a more thorough account.*

Proof of Theorem 1. Iteratively applying Lemma 5 with an appropriate choice of parameters δ_i at each step (see [Pel19, Eqn. (19)]), we can control $\Lambda^{Q_1,\ldots,Q_{m_2}}_{P_1,\ldots,P_{m_1}}$ by the U^1-norm. The result then follows from Lemma 4. $\square$

Before concluding with the proof of Lemma 5, we require the following lemma, which relates the U^s-norm of a general counting operator with an average of U^2-norms of discrete partial derivatives. It is proved using multiple applications of the Cauchy-Schwarz inequality; we omit the details.

Lemma 7 ([Pel19, Lemma 5.1])**.** *Let* $f_1, \ldots, f_m : \mathbb{F}_q^2 \to \mathbb{C}$ *be* 1*-bounded. For every* $h_1, \ldots, h_t \in \mathbb{F}_q$*, define*

$$\Upsilon(x) := \mathbb{E}_y \prod_{i=1}^{m} f_i(x,y); \qquad \Upsilon_{h_1,\ldots,h_t}(x) := \mathbb{E}_y \prod_{i=1}^{m} \Delta^{(1)}_{h_1,\ldots,h_t} f_i(x,y).$$

Then $\|\Upsilon\|_{U^s}^{2^{2s-2}} \leqslant \mathbb{E}_{h_1,\ldots,h_{s-2}} \|\Upsilon_{h_1,\ldots,h_{s-2}}\|_{U^2}^4$ *for all* $s \geqslant 2$*.*

Proof of Lemma 5. Applying the regularity lemma [Pel19, Proposition 2.6], we can decompose

$$f_0 = f_a + f_b + f_c$$

for some $f_a, f_b, f_c : \mathbb{F}_q \to \mathbb{C}$ with $\|f_a\|_{U^s}^* \leqslant q^{\delta_1}$, $\|f_b\|_{L^1} \leqslant q^{-\delta_2}$, $\|f_c\|_{L^\infty} \leqslant q^{\delta_3}$, and $\|f_c\|_{U^s} \leqslant q^{-\delta_4}$. Writing $F_a = (f_a, f_1, \dots, f_{m_1})$, and similarly defining F_b and F_c, we have

$$\Lambda_{P_1,\dots,P_{m_1}}^{Q_1,\dots,Q_{m_2}}(F;\Psi) = \Lambda_{P_1,\dots,P_{m_1}}^{Q_1,\dots,Q_{m_2}}(F_a;\Psi) + \Lambda_{P_1,\dots,P_{m_1}}^{Q_1,\dots,Q_{m_2}}(F_b;\Psi) + \Lambda_{P_1,\dots,P_{m_1}}^{Q_1,\dots,Q_{m_2}}(F_c;\Psi).$$

The triangle inequality shows that the second term is at most $q^{-\delta_2}$. Using our U^s-norm control assumption (II), we have

$$|\Lambda_{P_1,\dots,P_{m_1}}^{Q_1,\dots,Q_{m_2}}(F_c;\Psi)| \leqslant q^{(1-b_2)\delta_3 - b_2\delta_4} b_1 + q^{\delta_3} b_3.$$

Introducing the auxilliary counting operator

$$\Upsilon(x) := \mathbb{E}_y \prod_{i=1}^{m_1} f_i(x + P_i(y)) \prod_{j=1}^{m_2} \psi_j(Q_j(y)),$$

we can bound the first term:

$$|\Lambda_{P_1,\dots,P_{m_1}}^{Q_1,\dots,Q_{m_2}}(F_a;\Psi)| = |\langle f_a, \Upsilon\rangle| \leqslant \|f_a\|_{U^s}^* \|\Upsilon\|_{U^s} \leqslant q^{\delta_1}\|\Upsilon\|_{U^s}.$$

It therefore remains to bound $\|\Upsilon\|_{U^s}$. Lemma 7 informs us that

$$\|\Upsilon\|_{U^s}^{2^{2s-2}} \leqslant \mathbb{E}_{h_1,\dots,h_{s-2}} \|\Upsilon_{h_1,\dots,h_{s-2}}\|_{U^2}^4. \tag{3}$$

Here, $\Upsilon_{h_1,\dots,h_{s-2}}$ is as defined in Lemma 7. In view of the well-known fact that $\|\psi\|_{U^2} = \|\hat{\psi}\|_{L^4}$ for any $\psi : \mathbb{F}_q \to \mathbb{C}$ (see [Pel19, Eqn. (6)] or [Pre17, Page 11]), we proceed to study the Fourier transform. For each $k \in \{1, \dots, m_1\}$ and $\phi_{m_2+1} \in \widehat{\mathbb{F}}_q$, we observe that

$$\widehat{\Upsilon_{h_1,\dots,h_{s-2}}}(\phi_{m_2+1}) = \Lambda_{R_1,\dots,R_{m_1-1}}^{S_1,\dots,S_{m_2+1}}(g_{0;k}, \dots, g_{m_1-1;k}; \phi_1, \dots, \phi_{m_2+1}),$$

where

$$R_i = R_{i;k} = \begin{cases} P_i - P_k & i \leqslant k-1 \\ P_{i+1} - P_k & i \geqslant k \end{cases}, \quad S_j = S_{j;k} = \begin{cases} Q_j & j \leqslant m_2 \\ P_k & j = m_2 + 1 \end{cases},$$

and

$$g_{i;k} = \begin{cases} \overline{\phi_{m_2+1}} \Delta_{h_1,\dots,h_{s-2}} f_k & i = 0 \\ \Delta_{h_1,\dots,h_{s-2}} f_i & i \leqslant k-1 \\ \Delta_{h_1,\dots,h_{s-2}} f_{i+1} & i \geqslant k \end{cases}.$$

We can therefore bound the Fourier coefficient using our discorrelation estimate assumption (I). For our choice of $g_{(i;k)}$'s this gives

$$|\widehat{\Upsilon_{h_1,\dots,h_{s-2}}}(\phi_{m_2+1})| \leqslant \min_{i\geqslant 1} |\mathbb{E}_z \Delta_{h_1,\dots,h_{s-2}} f_i(z)| + c_2 q^{-\gamma'}$$

for some $\gamma' > 0$. Incorporating this into (3) leads to the bound

$$|\Lambda_{P_1,\dots,P_{m_1}}^{Q_1,\dots,Q_{m_2}}(F;\Psi)| \leqslant q^{\delta_1} \min_{i\geqslant 1} \|f_i\|_{U^{s-1}}^{2^{1-s}} + \beta.$$

If we decompose f_1 instead of f_0 and follow this same argument, then we can extend the above minimum to cover $i = 0$ as well, completing the proof. □

References

[Kuc21] Kuca, B. *Further bounds in the polynomial Szemerédi theorem over finite fields*, Acta Arith. **198** (2021), no. 1, 77–108.

[Pel19] Peluse, S. *On the polynomial Szemerédi theorem in finite fields*, Duke Math. J. **168** (2019), no. 5, 749–774.

[Pre17] Prendiville, S. *Quantitative bounds in the polynomial Szemerédi theorem: the homogeneous case*, Discrete Anal. (2017), Paper No. 5, 34 pp.

Jonathan Chapman, University of Bristol
email: jonathan.chapman@bristol.ac.uk

8 A mean ergodic theorem for $\sum_{n=1}^{N} f(T^n)g(T^{n^2})$

After H. Furstenberg and B. Weiss [FW]

A summary written by Yoav Cohn

Abstract

We survey Furstenberg and Weiss's proof of L^2 convergence of non-conventional ergodic averages of the form $\sum_{n=1}^{N} f(T^n)g(T^{n^2})$

8.1 Introduction

Let $(X, \mathcal{B}, \mu)$ be a measure space, with $\mu(X) < \infty$, and let $T : X \to X$ be a measure preserving transformation of this space. Expressions of the form $\sum_{n=1}^{N} \prod_{j=1}^{\ell} f_j(T^{p_j(n)})$, Where $\{f_j\}_{j=1}^{\ell}$ are bounded functions, and $\{p_j\}_{j=1}^{\ell}$ are integer sequences are well - studied, and by now, are well understood in many cases (see, e.g [HK]). In this paper, we present Furstenberg and Weiss's result, regarding the specific case $\ell = 2,\ p_1(n) = n,\ p_2(n) = n^2$, that is:

Theorem 1. *For any measure-preserving system $(X, \mathcal{B}, \mu, T)$ and $f, g \in L^\infty(X)$, the averages $\sum_{n=1}^{N} f(T^n)g(T^{n^2})$ converge in $L^2(X)$.*

A central tool in the investigation of multiple ergodic averages is the construction of appropriate characteristic factors. This is true also for the case we will discuss here. Through a series of reductions, the authors reduce matters the case where the system is a group extension $Z \times_\rho S^1$ of the Kronecker factor, where $\rho : Z \to S^1$ is a cocycle, satisfying a functional equation to be specified (see Equation 1). Then, this functional equation is used to conclude the proof. In more detail, the main steps of the reduction are as follows:

topsep=0pt,1temsep=-1ex,p1rtopsep=1ex,p1rsep=1ex Showing that a characteristic factor for all schemes $\{rn, sn, tn\}$ where $t = r + s$ is a partial characteristic factor for $\{n, n^2\}$. (see [FW, Chapter 4] for the definition of a partial characteristic factor).

topsep=0pt,2temsep=-2ex,p2rtopsep=2ex,p2rsep=2ex Showing $\hat{Z}$ is a characteristic factor for the schemes $\{rn, sn, tn\}$.

topsep=0pt,3temsep=-3ex,p3rtopsep=3ex,p3rsep=3ex Reduction to the case where the system is **normal**. That is, $\hat{Z}$ is a group extension by some compact group G.

topsep=0pt,4temsep=-4ex,p4rtopsep=4ex,p4rsep=4ex Reduction to the case where G is abelian, and from there to S^1.

topsep=0pt,5temsep=-5ex,p5rtopsep=5ex,p5rsep=5ex Reduction to the case where the cocyle satisfies an equation of CL -type. (see Equation 1)

Our talk will focus on the last two steps of the reduction, and the conclusion of the proof.

8.2 Preliminaries

In this subsection we survey some important ideas and terminology that are used in the paper. For other, more common preliminaries, such that the definitions of cocycles and partial characteristic factors, one can turn to [FW].

8.2.1 Isometric extensions

Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system, and let $(Y, \mathcal{D}, \nu, T)$ be a factor. Observe the subspace of $L^2(X)$ spanned by finite rank modules over $L^2(Y)$. This subset naturally corresponds to another factor of X, which is called the **Maximal isometric extension** of Y in X. We will denote it by $\hat{Y}$. It is known that each such factor is isomorphic to a system of the form $(Y \times M, \tilde{B}, \nu \times m_M, \tilde{T})$, where $M := G/H$ is a homogeneous space of a compact group G, m_M is the invariant measure on M, and $\tilde{B}$ is the appropriate σ - algebra. The action is defined by $\tilde{T}(y, u) = (T(y), \rho(g)u)$, Where $\rho : Y \to G$ is some measurable map. In the case where $H = 1$, this structure is called a **group extension**. (see [FW, Chapter 5] for some more information). Let (Z, α) denote the Kronecker factor of X. Then X will be called **normal** if $\hat{Z}$ is a group extension of Z.

8.2.2 CL-cocycles AND CL-function

As noted above, we will survey a series of reductions, that will bring us to work with characteristic factors of the form $Z \times_\rho S^1$. Going further, we will reduce to the case where the cocycle defining the action $S : Z \to S^1$, satisfies the following functional equation:

Definition 2. *S will be called a* ***CL - cocycle*** *if there exist $m, \ell \in \mathbb{N}$ such that for $v, z \in Z$:*

$$\frac{S_\ell(z + mv)}{S_\ell(z)} = \Lambda_v(z + \ell Z)\frac{K_v(z + \ell\alpha)}{K_v(z)} \tag{1}$$

where Λ, K are measurable functions with values in S^1.
Also, a function $\eta \in L^2(X)$ will be called a ***CL - function*** *if there exists some CL - cocycle such that $T\eta = S\eta$.*

8.2.3 Mackey group

Let Y be an ergodic system, and let $Y \times_\rho G$ a group extension defined by a cocyle $\rho : Y \to G$. Note that if we have $Im(\rho) \subseteq H$ for some H, a proper subgroup of G, then $\nu \times m_G$ is not ergodic. This means that the group extension does not have to be ergodic by itself, but also hints us as to what its ergodic components might be:

Theorem 3. *For any cocycle $\rho : Y \to G$ there is a closed subgroup $H \subseteq G$ (called the* ***Mackey group*** *of the extension), uniquely determined up to conjugacy, such that:*
• There is a cocycle $p'(y) = \phi(Ty)\rho(y)\phi(y)^{-1}$ equivalent to ρ taking values in H (for some $\phi : Y \to G$).
• The transformation $T^{'} : Y \times G \to Y \times G$, $T'(y, g) = (Ty, \rho^{'}(y)g)$ has ergodic invariant measures $\nu \times m_{H_\beta}$ where m_{H_β} is the Haar measure m_H translated by β to the right.
• Any ergodic T'- ergodic invariant measure on $Y \times G$ has the above form, and the T - invariant measures are obtained by re-parameterization. That is, applying Ψ^{-1} to the ergodic T'- invariant measures, where $\Psi(y, g) = (y, \phi(y)g)$ for some measurable ϕ.

To prove this, for each $\gamma \in G$ define $S_\gamma(y, g) = (y, g\gamma)$. Then the appropriate group is given by $H \coloneqq \{\gamma | S_\gamma \mu = \mu\}$. The proof belongs to the theory by George Mackey. We will sketch it briefly during the talk.

8.3 Overview of the proof

We start our description with step 2. We apply the Van der Corput lemma, and want to show that $\hat{Z}$ is a characteristic factor for all schemes $\{rn,\ sn,\ tn\}$. By the mean ergodic theorem, it can then be seen that it's enough to understand $T^r \times T^s \times T^t$ - invariant functions with respect to some $T^r \times T^s \times T^t$ - invariant measure $\tilde{\mu}$, which is also a X^3 conditional product joining relative to Z^3. Such measures are closely related to maximal isometric extensions (see [FW, Theorem 5.1]). So, it follows that each such function comes from the maximal isometric extension of Z. That enables us to conclude what we wanted.
For step 3, the authors investigate the way inverse limits interact with the Kronecker factor, and with group extension. From there, for any ergodic system, they construct a normal extension, using inverse limits.
The above arguments reduce matters to the case where X is normal. Step 4 will be a reduction to the case where G (the group by which Z is extended), is an abelian group. Similarly to what was done in step 2, we define a measure. Let $W_{r,s,t} = \{(z + rz', z + sz', z + tz') | z, z' \in Z\} \subseteq Z^3$. Define $\tilde{\mu}$ to be the Haar measure on $W_{r,s,t} \times G^3$. It is $T^r \times T^s \times T^t$ - invariant. We look closely at the ergodic components of $\tilde{\mu}$ on $W_{r,s,t}$. Those are shifts of the set $Z_{r,s,t} = \{(rz', sz', tz')\}$ by an element of the form (z, z, z). For each such z,

denote by $[L_z]$ its Mackey group, defined up to conjugacy class. The function $z \to [L_z]$ will be measurable. By ergodicity, it is actually constant. Our next goal is to show that L has some structure. That is, we want to show that there exists some abelian group J and three homomorphisms $\{\psi_i\}_{i=1}^3 : G \to J$ such that $J = \{(g_1, g_2, g_3) \mid \psi_1(g_1)\psi_2(g_2)\psi_2(g_2)\}$. For this, by virtue of a group - theoretic lemma (see [FW, Lemma 9.1]), it's enough to show the for each $1 \leq i, j \leq 3$, $\pi_{i,j}(L) = G_i \times G_j$. This is done by showing the Mackey group for The action of $\pi_{i,j}(W_{r,s,t})$ on G^2, is G^2. From here, one can conclude that since $(G^{'})^3 \subseteq L$, then any $T^r \times T^s \times T^t$ - invariant function comes from an invariant function of $(G/G')^3$, and so- we can reduce to the abelian case.
Next, we look at ψ_i defined above. Since $(H_0)^3 := \bigcap \psi_i \subseteq L$, it is shown that H can be replaced by H/H_0. Then, we will have that characters of the form $\chi \circ \psi_i$ separate points, and hence, by Pontryagin duality, we will be able to reduce to the case where $H = S^1$, and the cocycle is of the form $\chi \circ \psi_i \circ \rho$. We now want to study the action by the cocycle $\tilde{\rho} := (\rho_s, \rho_r, \rho_t) : W_z \to H^3$, where W_z are the projections to $W_{r,s,t}$ of the ergodic components of $\tilde{\mu}$. Recall that the Mackey group L characterises the ergodic components. Namely, there exists a measurable $\phi : W_z \to H^3$ such that

$$\phi(z_1, z_2, z_3)\tilde{\rho}(z_1 + r\alpha, z_2 + s\alpha, z_3 + t\alpha)\phi(z_1, z_2, z_3) \in L$$

From this, we get that for every character $\chi \in \hat{J}$, we have

$$\big(\chi \circ \psi_1 \circ \rho_r(z_1)\big)\big(\chi \circ \psi_2 \circ \rho_r(z_2)\big)\big(\chi \circ \psi_3 \circ \rho_r(z_3)\big) = \frac{F_\chi((z_1, z_2, z_3) + (r\alpha, s\alpha, t\alpha))}{F_\chi((z_1, z_2, z_3))} \tag{2}$$

For some $F_\chi : ((z, z, z) + Z_{r,s,t}) \to S^1$. The next step will be to study $\chi \circ \psi_1 \circ \rho_r$. We look at Equation 2, and want to get ride of q. A main idea here is the definition of a class of functions $u_\delta : W_{r,s} \to Z$, for which the following lemma is proved:

Lemma 4. *Let $\Delta = Z_{1,1} \cap Z_{r,s} \subset Z^2$, and let $\Delta^{'} = \{\delta \in Z \mid (\delta, \delta) \in \Delta\}$. The map of $W_{r,s} \times \Delta^{'} \to W_{r,s,t}$ given by $(z_1, z_2, \delta) \to (z_1, z_2, u_\delta(z_1, z_2))$ is onto, and measure preserving with respect to the Haar measures on the groups.*

This yields the equation (depending measurably on v):

$$\frac{\sigma_r(z_1 + (r-t)v)}{\sigma_r(z_1)} \frac{\sigma_s(z_2 + (s-t)v)}{\sigma_s(z_2)} = \frac{G_v(z_1 + r\alpha, z_2 + s\alpha)}{G_v(z_1, z_2)} \tag{3}$$

Next, the following lemma is proved:

Lemma 5. *Let* $(X, \mathcal{B}, \mu, T)$, $(Y, \mathcal{D}, \nu, S)$ *be ergodic systems, and let* $f(x), g(y)$ *measurable maps taking values in* S^1. *Let* $H : X \times Y \to X \times Y$ *not* 0 *a.e. Suppose it holds a.e that* $f(x)g(y)H(x, y) = H(T(x), S(y))$. *Then there exist a constant* c *and a measurable* $K : X \to S^1$ *such that* $f(x) = c\frac{K(T(x))}{K(x)}$ *(and similarly for* g*).*

Note that the conditions of Lemma 5 are almost fulfilled in Equation 3, except we that we don't have ergodicity. Still, by treating ergodic components separately, we get the functional equation:

$$\frac{\sigma_r(z + (r - t)v)}{\sigma_r(z)} = \Lambda_v(z + rZ)\frac{K_v(z + r\alpha)}{K_v(z)} \tag{4}$$

Where the choice of Λ_v, K_v is shown to be measurable in v.
The above discussion motivates the definitions of CL - cocycles, and CL - functions (see subsection 8.2.2). It can be shown directly that CL - functions are a group, so we can use them to generate an algebra, which then corresponds to a factor of X, to be denoted by X_{CL}. We get that X_{CL} is an appropriate factor for all schemes $\{rn, sn, (r + s)n\}$. This concludes step 5.
The next step will be to investigate some conditions (for Λ_v), under which CL - functions must be degenerate. That is, defined over the Kronecker factor of the system. Those will be:

Lemma 6. *If* η *is a CL- function for a cocycle where* $\Lambda_v \equiv 1$, *then it is degenerate. If* η *is a CL- function for a cocycle where* $\Lambda_v(z + \ell Z)$ *takes the same value on a subset of positive measure of* $Z \times Z/\ell Z$, *then it is degenerate.*

Before we will be ready to prove the convergence result we were aiming for, we need another lemma, which is basically a corollary of Wiener's lemma, and is proven by mean of the Van der Corput Lemma:

Lemma 7. *Let* X *be an ergodic measure preserving system, let* $\phi(x)$ *be defined over the Kronecker factor of* X *with values in* S^1, *and assume that the distribution of* ϕ *on* S^1 *has no atoms. Let* $f \in L^2(X)$ *and let* $\eta(n)$ *be a bounded sequence. Then for any* a, b, c *with* $a \neq 0$*:*

$$\frac{1}{N}\sum_{n=1}^{N}\eta(n)\phi(x)^{an^2+bn}T^{cn}f \to 0$$

in $L^2(X)$.

By the previous discussion, and by a density argument, it's essentially enough to show the convergence of $\sum_{n=1}^{N} f(T^n)g(T^{n^2})$ for the case $g = \psi(z)$ (defined on the Kronecker), and $g = \psi(z)$, where $g = \psi(z)\eta$, where η is a CL - function for a non - degenerate CL - cocycle. Here, the result will be proved based on the form of Equation 8.2.2, along with Lemmas $6, 7$ above.

References

[HK] Host, B. and Kra, B., *Convergence of polynomial ergodic averages.* Isr. J. Math. 149, 1–19 (2005).

[FW] Furstenberg, H. and Weiss, B. 1996. A Mean Ergodic Theorem for $1/N\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$. In: Bergelson, V., March, P. and Rosenblatt, J. ed. Convergence in Ergodic Theory and Probability. Berlin, New York: De Gruyter, pp. 193-228.

Yoav Cohn, Huji
email: `Yoav.Cohn@mail.huji.ac.il`

9 A new proof of an inequality of Bourgain

After P. Durcik and J. Roos [DR]

A summary written by Leonidas Daskalakis

Abstract

We discuss an alternative proof by Durcik and Roos of a trilinear smoothing inequality originally due to Bourgain. This new approach relies on techniques from additive combinatorics developed by Peluse and Prendiville.

9.1 Introduction

Fix a compactly supported smooth function $\chi\colon \mathbb{R} \to [0,1]$ and consider the following trilinear form

$$\mathcal{I}(f_0, f_1, f_2) = \left| \int \int f_0(x) f_1(x+t) f_2(x+t^2) \chi(t) dt dx \right|.$$

In 1988 Bourgain [B] established the following trilinear smoothing inequality.

Theorem 1. *Assume $K \subseteq \mathbb{R}$ is compact. Then there exists $C_{K,\chi} > 0$ and an absolute constant $\sigma > 0$ such that*

$$\mathcal{I}(f_0, f_1, f_2) \leq C_{K,\chi} \|f_0\|_\infty \|f_1\|_2 \|f_2\|_{H^{-\sigma}} \tag{1}$$

for any $f_0 \in L^\infty$ supported on K, and any $f_1, f_2 \in L^2$.

Bourgain [B] used this theorem to prove a quantitative nonlinear Roth theorem in the real numbers. Inspired by the breakthrough work of Peluse–Prendiville [PP] and Peluse [P] on quantitative bounds for arithmetic sets lacking polynomial progressions, Durcik and Roos [DR] gave an alternative proof of Theorem 1 in order to illustrate how these new techniques from additive combinatorics can be employed to establish smoothing inequalities in harmonic analysis. Indeed, using the Peluse and Peluse–Prendiville theory, far-reaching generalizations of such smoothing inequalities for multilinear polynomial averaging operators have been established independently at the same time [KMPW].

The trilinear form considered in Theorem 1 is the simplest non-trivial example where one may apply the Peluse–Prendiville and Peluse theory, so its proof should give the key ideas with the least amount of technical complications. We discuss Durcik and Roos' proof in the following summary.

9.2 Preliminaries

We collect some useful notation and make some preliminary remarks. Any measurable function f will be called 1-bounded if $\|f\|_\infty \leq 1$. For any $f \in L^1(\mathbb{R})$, we define the Fourier transform as

$$\widehat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i \xi x} dx.$$

For any $\xi \in \mathbb{R}$, let $e_\xi(x) = e^{2\pi i \xi x}$. For any $x, h \in \mathbb{R}$ we define $\Delta_h f(x) = f(x)\overline{f(x+h)}$, and for $h \in \mathbb{R}^s$ we define $\Delta_h f(x) = \Delta_{h_1}\Delta_{h_2}\cdots\Delta_{h_s} f(x)$. For any $s \in \mathbb{N}$ we define

$$\|f\|_{u^{s+2}}^{2^s} = \int_{\mathbb{R}^s} \|\widehat{\Delta_h f}\|_\infty dh \quad \text{and} \quad \|f\|_{u^2} = \|\widehat{f}\|_\infty.$$

We remark that $\|\cdot\|_{u^s}$ should be understood as continuous variants of the Gowers uniformity norms. We remind the reader that the Sobolev norm is defined by

$$\|f\|_{H^{-\sigma}} = \Big(\int_{\mathbb{R}} |\widehat{f}(\xi)|^2 \big(1 + |\xi|^2\big)^{-\sigma} \Big)^{1/2}.$$

If A, B are two non-negative quantities, we write $A \lesssim B$ to denote that there exists a positive constant C such that $A \leq CB$, we use subscripts when C depends on parameters[3].
To establish Theorem 1, it suffices to prove the following proposition.

Proposition 2. *Assume $K \subseteq \mathbb{R}$ is compact. Then there exist two absolute constants $c, \sigma > 0$ such that for any 1-bounded functions f_0, f_1, f_2 with f_0 supported on K, we have*

$$\mathcal{I}(f_0, f_1, f_2) \lesssim_{K,\chi} \|f_2\|_{H^{-\sigma}}^c.$$

[3]For example, we could write (1) more compactly using "$\lesssim_{K,\chi}$".

One may prove that Proposition 2 implies Theorem 1 by utilizing the homogeneity of $\mathcal{I}$, Littlewood-Paley theory and interpolation using the following bound for $\mathcal{I}(f_0, f_1, f_2)$

$$\|f_0\|_\infty \int |f_1(x)| \left| \int f_2(x + t^2 - t)\chi(t)dt \right| dx \lesssim_{K,\chi} \|f_0\|_\infty \|f_1\|_{3/2} \|f_2\|_{3/2}$$

where we have used Hölder and Young's convolution inequalities.
The task is now reduced to proving Proposition 2. This will be achieved through a degree lowering argument sketched in the next section.

9.3 Degree Lowering

To establish Proposition 2, Durcik and Roos employ a degree lowering argument, adapting the ideas of Peluse and Prendiville [P, PP]. This argument relies on the following four key lemmata.

Lemma 3 (u^3-control)**.** *Assume $K \subseteq \mathbb{R}$ is compact and f_0, f_1, f_2 are 1-bounded with f_0 supported on K. Then*

$$\mathcal{I}(f_0, f_1, f_2) \lesssim_{K,\chi} \|f_0\|_{u^3}^{1/5}.$$

Lemma 4 (Dual difference interchange)**.** *Let $(F_t)_{t\in\mathbb{R}}$ be a family of jointly measurable 1-bounded functions $F_t \colon \mathbb{R} \to \mathbb{C}$ supported on a compact set K. Let*

$$F(x) = \int F_t(x)\chi(t)dt.$$

Then for any $s \in \mathbb{N}$ there exists a measurable function $\Phi \colon \mathbb{R}^s \to \mathbb{R}$ such that

$$\|F\|_{u^{s+2}} \lesssim_{K,\chi} \left(\int \left| \int \int \Delta_h F_t(x) e^{2\pi i x \Phi(h)} \chi(t) dt dx \right| dh \right)^{2^{-2s}}.$$

Lemma 5 (Bilinear case)**.** *Assume that $f, g \in L^2$, and $\xi \in \mathbb{R}$. Then*

$$\mathcal{I}(e_\xi, f, g) \lesssim_\chi \|f\|_{H^{-1/2}} \|g\|_{L^2} \quad \textit{and} \quad \mathcal{I}(f, e_\xi, g) \lesssim_\chi \|f\|_{L^2} \|g\|_{H^{-1/2}}.$$

Lemma 6. *For every $s \in \mathbb{N}$ and $\sigma > 0$, there exists $c > 0$ such that for any 1-bounded function f supported on a compact set K we have*

$$\int_{\mathbb{R}^s} \|\Delta_h f\|_{H^{-\sigma}}^2 dh \lesssim_K \|f\|_{u^{s+1}}^c$$

One can choose $c = 2^s \sigma (1 + 2\sigma)^{-1}$.

Let us see how these four intermediate results imply Proposition 2. Without loss of generality we may assume that f_1 is also supported on compact set K_1 depending only on K and χ. Consider the "dual function"

$$F_0(x) = \int \overline{f_1(x+t) f_2(x+t^2)} \chi(t) dt$$

and note that by Cauchy-Schwarz and Lemma 3 we get

$$\mathcal{I}(f_0, f_1, f_2) \lesssim_K (\mathcal{I}(F_0, f_1, f_2))^{1/2} \lesssim_{K,\chi} \|F_0\|_{u^3}^{2^{-1}/5}.$$

Lemma 4 applied for $s = 1$ guarantees the existence of a function Φ making $\mathcal{I}(f_0, f_1, f_2)$ bounded by

$$C_{K,\chi} \bigg(\int \bigg| \int \int \Delta_h f_1(x+t) \Delta_h f_2(x+t^2) e^{2\pi i x \Phi(h)} \chi(t) dt dx \bigg| dh \bigg)^{2^{-3}/5}.$$

The expression in the absolute value equals $\mathcal{I}(e_{\Phi(h)}, \Delta_h f_1, \Delta_h f_2)$, which by Lemma 5 together with Cauchy-Schwarz yields

$$\mathcal{I}(f_0, f_1, f_2) \lesssim_{K,\chi} \left(\int \|\Delta_h f_1\|_{H^{-1/2}}^2 dh \right)^{2^{-4}/5}. \tag{2}$$

Apply Lemma 6 with $\sigma = 1/2$ to (2) to obtain $\mathcal{I}(f_0, f_1, f_2) \lesssim_{K,\chi} \|f_1\|_{u^2}^{2^{-5}/5}$. Now consider the "dual function"

$$F_1(x) = \int \overline{f_0(x-t) f_2(x+t^2-t)} \chi(t) dt$$

and note that by Cauchy-Schwarz and the previous inequality we get

$$\mathcal{I}(f_0, f_1, f_2) \lesssim_{K,\chi} \mathcal{I}(f_0, F_1, f_2))^{1/2} \lesssim_{K,\chi} \|F_1\|_{u^2}^{2^{-6}/5}. \tag{3}$$

Since $\|F_1\|_{u^2} = \|\widehat{F}_1\|_\infty$, and

$$|\widehat{F}_1(\xi)| = \left| \int \int f_0(x-t) f_2(x+t^2-t) e^{2\pi i \xi x} \chi(t) dt dx \right| = \mathcal{I}(f_0, e_\xi, f_2),$$

we get that $\|F_1\|_{u^2} \lesssim \|f_2\|_{H^{-1/2}}$ by Lemma 5. Combining this with (3) gives

$$\mathcal{I}(f_0, f_1, f_2) \lesssim_{K,\chi} \|f_2\|_{H^{-1/2}}^{1/320},$$

which proves Proposition 2 with $\sigma = 1/2$ and $c = 1/320$.

9.4 Remarks on the four key lemmata

The final section is devoted to a very brief discussion on the four key lemmata.

The proof of Lemma 3 begins with the PET induction scheme of Bergelson and Leibman [BL], which reduces our task to bounding a quadrilinear form with linear patterns. PET is achieved here by repeated applications of Cauchy-Schwarz, Fubini theorem, and change of variables. Finally, appropriately bounding the resulting quadrilinear forms follows from a standard procedure relying on Cauchy-Schwarz, called Gowers differencing. We note that such a procedure can be applied in more general situations for multilinear forms containing polynomials of higher degrees.

The proof of Lemma 4 follows the ideas from Lemma 6.3 from [PP]. For simplicity, let us briefly sketch the proof only for the case $s = 1$, since this case is sufficient for our degree lowering argument. By linearizing the supremum appearing in $\|\cdot\|_{u^3}$, it suffices to show that for any measurable $\phi\colon \mathbb{R} \to \mathbb{R}$, there exists a measurable function $\Phi\colon \mathbb{R} \to \mathbb{R}$ such that

$$\int \left| \int \Delta_h F(x) e^{2\pi i x \phi(h)} dx \right| dh \lesssim_{K,\chi} \tag{4}$$

$$\left(\int \left| \int \int \Delta_h F_t(x) e^{2\pi i x \Phi(h)} \chi(t) dt dx \right| dh \right)^{1/2}.$$

One may write

$$\left| \int \Delta_h F(x) e^{2\pi i x \phi(h)} dx \right| = e^{2\pi i \Psi(h)} \int \Delta_h F(x) e^{2\pi i x \phi(h)} dx$$

for a real valued Ψ, and expand

$$\Delta_h F(x) = \int \int F_{t'}(x) \overline{F_t(x+h)} \chi(t') \chi(t) dt dt'.$$

After taking all these into consideration and applying Cauchy-Schwarz and Fubini, one can show that the expression (4) is bounded by

$$C_{K,\chi} \left(\int_{I-I} \int \left| \int \int \overline{F_t(x+h)} F_t(x+h') e^{2\pi i x (\phi(h) - \phi(h'))} \chi(t) dt dx \right| dh dh' \right)^{1/2}$$

where I is an interval where χ is supported. After a change of variables and by fixing an h' where the integrant in h' is close to its supremum, one may notice that $\Phi(h) = \phi(h + h') - \phi(h')$ has the desired properties.

Both assertions of Lemma 5 can be established using the Fourier inversion formula together with the following standard variant of van der Corput's lemma (see for example [SW, Proposition 2.1]).

Lemma 7. *For all $\alpha, \beta \in \mathbb{R}$, we have*

$$\left| \int e^{2\pi i(\alpha t + \beta t^2)} \chi(t) dt \right| \lesssim_\chi \max\{|\alpha|, |\beta|\}^{-1/2}.$$

Finally, Lemma 6 is a straightforward adaptation of Lemma 3.1 from [CDR]. Similarly to Lemma 4, we apply the result only for $s = 1$ in the degree lowering argument. The proof relies on basic properties of the Fourier transform and their interactions with the operator Δ_h, and once the situation for $s = 1$ is understood, the general case follows easily.

References

[DR] P. Durcik, J. Roos. *A new proof of an inequality of Bourgain.* To appear in Math. Res. Lett. 2210.01326

[B] J. Bourgain. *A nonlinear version of Roth's theorem for sets of positive density in the real line.* J. Anal. Math. 50 (1988), 169-181.

[KMPW] B. Krause, M. Mirek, S. Peluse, J. Wright. *Polynomial progressions in topological fields.* preprint arXiv:2210.00670.

[PP] S. Peluse, S. Prendiville. *Quantitative bounds in the non-linear Roth Theorem.* preprint arXiv:1903.02592.

[P] S. Peluse. *Bounds for sets with no polynomial progressions.* Forum Math. Pi 8 (2020), e16.

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

[SW] E. M. Stein, S. Wainger. *Oscillatory integrals related to Carleson's theorem.* Math. Res. Lett. 8 (2001), no. 5–6, 789-800.

[CDR] M. Christ, P. Durcik, J. Roos. *Trilinear smoothing inequalities and a variant of the triangular Hilbert transform.* Adv. Math. 390 (2021), Paper No. 107863, 60 pp.

Leonidas Daskalakis, Rutgers University
email: ld606@math.rutgers.edu

10 A new proof of Szemerédi's theorem for arithmetic progressions of length four; Generalized arithmetical progressions and sumsets.

After W.T. Gowers [2] and I. Z. Ruzsa [4]

A summary written by Dimas de Albuquerque and Gautam Neelakantan Memana

Abstract

In this section we will be presenting the tools required for Gowers' proof of Szemerédi's theorem for arithmetic progressions of length four as presented in [2]. One key tool among these is the improved version of Freiman's theorem on sumsets due to Ruzsa [4].

10.1 Introduction

The famous theorem of Szemerédi asserts that, for any positive integer k and any real number $\delta > 0$, there exists an $N > 0$ such that every subset of $\{1, .., N\}$ of cardinality at least δN contains an arithmetic progression of length k. In [2], Gowers extends the proof technique of Roth in [3], where the author proves Szemerédi's theorem for length three using exponential sums. This proof technique also improves the known bounds for the theorem using combinatorial proofs.

10.2 Reformulation of Szemerédi's theorem

Notation: Let $\mathbb{Z}_N$ be the group of integers $\mod N$. For any function $f : \mathbb{Z}_N \to \mathbb{C}$, $\tilde{f}(r)$ denotes the r^{th} Fourier coefficient given by $\sum_{s \in \mathbb{Z}_N} f(s)\omega^{-rs}$, where $\omega = \exp(2\pi i/N)$.

Definition 1. *1. A function f from $\mathbb{Z}_N$ to the closed unit disc (D) in $\mathbb{C}$ is called* α-uniform *if*

$$\sum_k \left| \sum_s f(s)\overline{f(s-k)} \right|^2 \leq \alpha N^3.$$

2. *A function* $f : \mathbb{Z}_N \to D$ *is called* quadratically α-uniform *if*

$$\sum_u \sum_v \left| \sum_s f(s)\overline{f(s-u)f(s-v)}f(s-u-v) \right|^2 \leq \alpha N^4.$$

3. *A set* $A \subseteq \mathbb{Z}_N$ *of size* δN *is called* α-uniform *if the following function*

$$f_A(s) = \begin{cases} 1-\delta & s \in A \\ -\delta & s \notin A \end{cases}$$

is α-uniform. Similarly, A is called quadratically α-uniform *if f_A is quadratically α-uniform. f_A is also called the* balanced *function of A.*

Remark 2. *A set satisfying (3) in Definition 1 should be seen as a "pseudorandom" set as a consequence of uncertainity principle. Analysing sets which are "pseudorandom" and not, separately, is one of the crucial ideas in the proof of the main theorem.*

Theorem 3. *([2] Corollary 8) Let* $A \subset \mathbb{Z}_N$ *be a quadratically η-uniform set of size δN, where* $\eta \leq 2^{-208}\delta^{112}$ *and* $N > 200\delta^{-3}$*. Then A contains an arithmetic progression of length four.*

The analysis of sets which are not quadratically uniform is even harder. The next section contains the main tools that show that the sets which fail to be quadratically uniform can be restricted to a large arithmetic progression where its density increases noticeably.

10.3 Application of Freiman's theorem

Definition 4. *Let* $\phi : A \to \mathbb{Z}_N$*. A quadruple* $(a, b, c, d) \in A^4$ *is called* additive *for ϕ if* $a + b = c + d$ *and* $\phi(a) + \phi(b) = \phi(c) + \phi(d)$*.*

Proposition 5. *([2] Proposition 9) Let* $f : \mathbb{Z}_N \to D$ *and* $\phi : B \to \mathbb{Z}_N$ *be a function such that*

$$\sum_{k \in B} \left| \widetilde{\Delta(f;k)}\,(\phi(k)) \right|^2 \geq \alpha N^3, \tag{1}$$

where $\Delta(f;k)(s) = f(s)\overline{f(s-k)}$*. Then, there are at least* $\alpha^4 N^3$ *additive quadruples for ϕ.*

Remark 6. *A set A with its balanced function f_A satisfying* (1) *should be seen as a "non-pseudorandom" set, which is in the same vein as Definition 1. This tells us that there are many values of k for which the function $\Delta(f_A; k)$ has large Fourier coefficient r. So, the above theorem says that the set of pairs (k, r) for which $\widehat{\Delta(f; k)}(r)$ is large is far from arbitrary (i.e we can find many additive quadruples). Also, the function ϕ should be seen as linear embedding of $A \subset \mathbb{Z}_N$ into $\mathbb{Z}_N$.*

It turns out that functions with many additive quadruples have very nice structure as a consequence of Freiman's theorem [1].

Definition 7. *A d-dimensional (generalized) arithmetic progression on a commutative group G is a set of the form $P_1 + ... + P_d$, where each P_i is an ordinary arithmetic progression on G.*

Theorem 8. *(Freiman's theorem, [1], [4] Theorem 1.1) Let A, B be finite sets in a torsion-free commutative group satisfying $|A| = |B| = n$, $|A + B| \leq \alpha n$. Then, there are numbers d, C depending on α only such that A is contained in a generalized arithmetical progression of dimension at most d and size at most Cn.*

Remark 9. *The commutative group from the above theorem will be $\mathbb{Z}^D$ for the proof in [2].*

Proposition 10. *([2] Proposition 12) Let A be a subset of $\mathbb{Z}^D$ of cardinality m such that the number of quadruples (x, y, z, w) in A^4 with $x - y = z - w$ is bigger than $c_0 m^3$ for some constant c_0. Then, there are constants c and C depending only on c_0 such that there is a subset $A'' \subset A$ of cardinality at least cm with $|A'' - A''| \leq Cm$.*

Remark 11. *The above proposition says that A has a reasonable large subset B such that $|B + B|$ is small. Now, we can apply Freiman's theorem to this B.*

Corollary 12. *([2] Corollary 14) Let $B \subset \mathbb{Z}_N$ be a set of cardinality βN, and let $\phi : B \to \mathbb{Z}_N$ be a function with at least $c_0 N^3$ additive quadruples. Then there are constants γ and η depending on β and c_0 only, a $\mod$-N arithmetic progressions $P \subset \mathbb{Z}_N$ of cardinality at least N^γ and a linear function $\psi : P \to \mathbb{Z}_N$ such that $\phi(s)$ is defined and equal to $\psi(s)$ for ar least $\eta|P|$ values of $s \in P$.*

10.4 Ideas of Ruzsa's proof of Theorem 8

Ruzsa's proof of Freiman's theorem is based on the fact that one can reduce the search for the generalized arithmetical progression to a set of residues, and this is done by the concept of Freiman isomorphy:

Definition 13. *Let G, G' be commutative groups and consider $A \subset G, A' \subset G'$. A function $\phi : A \to A'$ is called an F_r - homomorphism if*

$$a_1 + \dots a_r = b_1 + \dots b_r \implies \phi(a_1) + \dots + \phi(a_r) = \phi(b_1) + \dots + \phi(b_r) \quad (2)$$

If ϕ is bijective and its inverse is also an F_r homomorphism, we call it an F_r isomorphism.

With the above definition, we can indicate the steps in the proof of Theorem 8. Here we use the notation $kA = A + \dots + A\, k$ times.

(1) One constructs an F_8 isomorphism between A and a subset of integers A_2.

(2) One obtains a subset $A' \subset A_2 \subset \mathbb{Z}$ which is F_8 isomorphic to a set T of residues modulo m, where $m \in (16|2A - 2A|, 32|2A - 2A|)$ is a prime number. This can be achieved using the Lemma below, which is a result of Ruzsa [5].

    Lemma 14. *Let A be a set of integers, $|A| = n, r \geq 2$ an integer and $D = rA - rA$. Write $|D| = N$. For every $m > 2r(N-1)$ there exists a set $A' \subset A, |A'| \geq n/r$ which is F_r - isomorphic to a set T of residues mod m.*

(3) There exists a generalized arithmetical progression $P \subset 2T - 2T$. This is obtained from the fact such difference sets contain Bohr sets, and Bohr sets on groups of residues contain generalized arithmetical progressions. A Bohr set on a commutative group G is a set of the form

$$B(\gamma_1, \dots, \gamma_k, \epsilon_1, \dots, \epsilon_k) = \{g \in G : |\arg\gamma_j(g)| \leq 2\pi\epsilon_j \ j = 1, \dots, k\}$$

where the γ_j are characters of the group G.

(4) Through the composition of F_r isomorphisms, one obtains a generalized arithmetical progression $P^* \subset 2A - 2A$.

(5) At last, we are able to obtain a maximal collection $\{a_1, \ldots, a_s\} \subset A$ such that $(P^* + a_i) \cap (P^* + a_j) = \emptyset$ if $i \neq j$, which in turn will give us that $A \subset \{a_1, \ldots, a_s\} + P^* - P^*$, and this last set can be covered by a generalized arithmetical progression, which concludes the Theorem.

Remark 15. *In the above route, the dimensions and sizes of the arithmetical progressions obtained are always dependent only on* α.

References

[1] G.A. Freiman, *Foundations of a Structural Theory of Set Addition(in Russian)*, Kazan Gos. Ped. Inst, Kazan, 1966.

[2] W.T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geometric and Functional Analysis **8** (1998) 529-551

[3] K.F Roth, *On certain sets of integers*, J.London Math. Soc **28** (1953), 245-252

[4] I.Z. Ruzsa, *Generalized Arithmetical progressions and sumsets*, Acta Math. Hungarica. **65** (1994), no. 4, 379-388.

[5] I. Z. Ruzsa, *Arithmetic progressions in sumsets*, Acta Arithmetica. **60** (1991), 191- 202

Dimas de Albuquerque, University of Wisconsin Madison
email: `ddealbuquerq@wisc.edu`
Gautam Neelakantan Memana, University of Wisconsin Madison
email: `neelakantanm@wisc.edu`

11 Power-type cancellation for the simplex Hilbert transform

After Polona Durcik, Vjekoslav Kovač, Christoph Thiele [1]

A summary written by Jaume de Dios Pont

Abstract

We prove L^p bounds of the truncation simplex Hilbert transform with a log-power less than one in the truncation range.

11.1 Introduction

Let $f_0, \dots, f_n$ be $n+1$ functions of n variables. Define the simplex Hilbert transform of these functions as

$$\Lambda_n(f_0, \dots, f_n) := \mathrm{p.\,v.} \int_{\mathbb{R}^{n+1}} \prod_{k=0}^{n} f_k(\mathbf{x}_0, \dots, \mathbf{x}_{k-1}, \mathbf{x}_{k+1}, \dots \mathbf{x}_n) \frac{1}{\mathbf{x}_0 + \cdots + \mathbf{x}_n} d\mathbf{x}.$$

This multilienar form is a generalization (from the case $n = 1$) of the bilinear form associated to the Hilbert transform. The Hilbert transform is bounded from L^p to itself whenever $p \in (0, \infty)$, and its bilinear form is therefore bounded on $L^{p_0} \times L^{p_1}$ whenever $p_i \in (0,1)$ and $p_0^{-1} + p_1^{-1} = 1$. This work makes partial progress toward a generalization of this boundedness result: Whether an inequality of the form

$$|\Lambda_n(f_0, \dots, f_n)| \lesssim \prod_{i=0}^{n} \|f_i\|_{p_i} \tag{1}$$

holds for any $p_0, \dots, p_n$, or more generally, for all $p_i \in (1, \infty)$ satisfying $\sum_{i=0}^{n} p_i^{-1} = 1$.

Notation: We will write $\mathbf{x}_{\hat{i}}$ as the vector $\mathbf{x}$ with the $i-$th component removed, and $\mathbf{x}_{\geq k}$ as the vector $(\mathbf{x}_k, \mathbf{x}_{k+1}, \dots \mathbf{x}_n)$. We will define $\mathbf{x}_{\leq k}$, $\mathbf{x}_{>k}$.. analogously, and combine these symbols, so that $\mathbf{z}_{>k,\hat{j}}$ denotes the vector $(\mathbf{z}_{k+1}, \dots, \mathbf{z}_{j-1}, \mathbf{z}_{j+1}, \dots \mathbf{z}_n)$. We will denote by $\bar{\mathbf{x}}$ the sum[4] of the components of $\mathbf{x}$. With this notation,

[4]The *overline* $\bar{\mathbf{x}}$ usually denotes average. Since the focus of this work is not on the explicit constants, either definition would work.

$$\Lambda_n(f_0,\ldots,f_n) := \mathrm{p.\,v.}\int_{\mathbb{R}^{n+1}} \prod_{k=0}^{n} f_k(\mathbf{x}_{\hat{k}})\frac{1}{\bar{\mathbf{x}}}d\mathbf{x} = \mathrm{p.\,v.}\int_{\mathbb{R}}\int_{\mathbb{R}^{n+1}} \prod_{k=0}^{n} f_j(\mathbf{x}_{\hat{k}})\delta_{\bar{\mathbf{x}}=t}d\mathbf{x}\frac{dt}{t}$$

A boundedness result of the form of (1) essentially states that Holder inequalities of the form

$$\int_{\mathbb{R}^{n+1}} \prod_{k=0}^{n} f_k(\mathbf{x}_{\hat{k}})(\delta_{\bar{\mathbf{x}}=t} - \delta_{\bar{\mathbf{x}}=-t})d\mathbf{x} \leq 2\prod_{i=0}^{n}\|f_i\|_{p_i}, \tag{2}$$

which hold whenever $p_i \in [1,\infty]$ satisfy $\sum_{i=0}^{n} p_i^{-1} = 1$, and which are essentially sharp for each individual value of t, cannot be simultaneously sharp for most values of t. If they were simultaneously sharp for all t, the best *truncated* inequality one would hope for is

$$\Lambda_{n,r,R}(f_0,\ldots,f_n) := \int_{r\leq|t|\leq R}\int_{\mathbb{R}^{n+1}} \prod_{k=0}^{n} f_k(\mathbf{x}_{\hat{k}})\delta_{\bar{\mathbf{x}}=t}d\mathbf{x}\frac{dt}{t} \leq 2\prod_{i=0}^{n}\|f_i\|_{p_i}\left|\log\frac{R}{r}\right|. \tag{3}$$

The main result of this work is giving a quantitative improvement on this log factor, of the form

Theorem 1 ([1, Theorem 1, Corollary 2]). *Let $p_0,\ldots,p_n \in (0,\infty)$, with $\sum_{i=0}^{n} p_i^{-1} = 1$. Then there is an $\epsilon = \epsilon(p_1,\ldots,p_n) > 0$ such that*

$$\|\Lambda_n(f_0,\ldots,f_n)\| \leq C_{p_0,\ldots,p_n}\prod_{i=0}^{n}\|f_i\|_{p_i}\left|\log\frac{R}{r}\right|^{1-\epsilon}. \tag{4}$$

When $p = (2^{-n}, 2^{-n}, 2^{-n+1}, 2^{-n+2},\ldots,2^{-1})$, ϵ can be chosen to be 2^{-n+1}.

The proof of the Theorem is entirely by studying this particular case, all the other cases are seen by interpolating with (3).

11.2 Sketch of the proof

In very broad terms, the proof studies a smoothed version of $\Lambda_{n,r,R}$, which is written as an integral over scales (parametrized by the t parameter). This smoothed version is bounded by a delicate induction procedure, in which variables are incorporated one by one, going from *free* variables to satisfying the constraints of the Hilbert transform.
The key step of this induction procedure is an equality (equation (18)) that allows one to do the integral in t explicitly.

11.3 Smoothing the cutoff

Let $g(x) := \exp(-x^2)$. Bounding $\Lambda_{n,r,R}$ is equivalent to bounding the smooth-cutoff version

$$\int_{\mathbb{R}^{n+1}} \prod_{k=0}^{n} f_k(\mathbf{x}_{\hat{k}}) \frac{g(R^{-1}\bar{\mathbf{x}}) - g(r^{-1}\bar{\mathbf{x}})}{\bar{\mathbf{x}}} d\mathbf{x} = \int_r^R \int_{\mathbb{R}^{n+1}} \prod_{k=0}^{n} f_k(\mathbf{x}_{\hat{k}}) h_t(\bar{\mathbf{x}}) d\mathbf{x} \frac{dt}{t} \quad (5)$$

where $h_t(x) = t^{-1}g'(t^{-1}x)$. This is because $1_{[r,R]} - g(R^{-1}t) - g(r^{-1}t)$ is in $L^1(dt/t)$, and the crude estimate (3) suffices to bound the difference between the smooth and non-smooth version.

11.4 Setting up the induction

Let $\mathbf{z}^{\bullet}_{>k} = (\mathbf{z}^0_{>k}, \mathbf{z}^1_{>k})$ be two lists of $n-k-1$ vectors. For $\mathbf{r}_{>k} \in \{0,1\}^{n-k+1}$ let $\mathbf{z}^{\mathbf{r}}_{>k} := (\mathbf{z}^{\mathbf{r}_k}_k, \ldots, \mathbf{z}^{\mathbf{r}_n}_n)$ be an *assignment* of the variables $\mathbf{z}^{\bullet}_{>k}$. Define

$$\mathcal{F}_k(\mathbf{x}_{<k}, y, \mathbf{z}^{\bullet}_{>k}) = \prod_{j=1}^{k} \prod_{\mathbf{r}\in\{0,1\}^{n-k-1}} F_j((\mathbf{x}_{\leq k}, y, \mathbf{z}^{\mathbf{r}}_{>k})_{\hat{j}}), \quad (6)$$

that is, the product over the first k functions and over all possible choices of the *free* variables of index larger than k, removing the appropriate variable. These *free* variables will be chosen to be close from their *true* value: Let $g_t := t^{-1}\exp(-x^2/t^2)$, the $t-$dilation of g. Given $\mathbf{x}_{>k}$ and a dilation vector $\alpha_{>k}$ of positive numbers let

$$d\gamma_{\mathbf{x}_{>k},\alpha_{>k}}(\mathbf{z}^{\bullet}_{>k}) := \prod_{s\in\{0,1\}} \prod_{j=k+1}^{n} g_{\alpha_j}(\mathbf{x}_j - \mathbf{z}^s_j) d\mathbf{z}^s_j \quad (7)$$

be the Gaussian measure (up to normalization constants) with diagonal covariance given by the vector of $\alpha^2_{>k}$.
We can now define the induction variables, which are:

$$\Lambda^k_{\alpha;\alpha_{\geq k}} := \underbrace{\int_r^R \frac{dt}{t}}_{\text{Scales}} \underbrace{\int_{\mathbb{R}^{n+1}} g_{t\alpha}(\bar{\mathbf{x}}) d\mathbf{x}}_{\text{Original variables}} \underbrace{\int_{\mathbb{R}^{2(n-k)}} d\gamma_{\mathbf{x}_{>k}, t\cdot\mathbf{a}_{>k}}(\mathbf{z}^{\bullet}_{>k})}_{\text{Free variables}} \left\{ \left| \int \underbrace{\mathcal{F}_k(\mathbf{x}_{<k}, y, \mathbf{z}^{\bullet}_{>k})}_{\text{Product of assignments}} h_{t\alpha_k}(y - \mathbf{x}_k) dy \right| \right\} \quad (8)$$

$$\tilde{\Lambda}^k_{\alpha;\alpha_{\geq k}} := \int_r^R \frac{dt}{t} \int_{\mathbb{R}^{n+1-k}} d\mathbf{x}_{\geq k} \int_{\mathbb{R}^{2(n+1-k)}} d\gamma_{\mathbf{x}_{>k}, t\cdot \mathbf{a}_{>k}}(\mathbf{z}^\bullet_{>k}) \Bigg\{ \left| \int_{\mathbb{R}^k} d\mathbf{x}_{<k} \int \mathcal{F}_k(\mathbf{x}_{<k}, y, \mathbf{z}^\bullet_{>k}) h_{t\alpha_k}(y - \mathbf{x}_k) h_{t\alpha}(\bar{\mathbf{x}}) dy \right| \Bigg\} \tag{9}$$

When $k = n$ there are no free variables in $\mathcal{F}_n$, and there is only one possible assignment, which corresponds to $\prod_{j=1}^n f_j(\mathbf{x}_{\hat{j}})$. In this case,

$$\Lambda^n_{\alpha;(\alpha_n)} := \int_r^R \frac{dt}{t} \int_{\mathbb{R}^{n+1}} g_{t\alpha}(\bar{\mathbf{x}}) d\mathbf{x} \left| \int \prod_{j=1}^n f_j(\mathbf{x}_{\hat{j}}) h_{t\alpha_n}(y - \mathbf{x}_n) dy \right|. \tag{10}$$

Setting $\alpha = \alpha_n = \frac{1}{\sqrt{2}}$ and removing the absolute values one obtains exactly the smoothed version of $\Lambda_n(f_0, \ldots, f_n)$, namely the right-hand side of (5). In particular, Theorem 1 follows from the case $k = n$ of the following lemma:

Lemma 2 ([1, Lemma 3]). *For any $2 \leq k \leq n$ and any $\alpha, \alpha_k, \ldots, \alpha_n \in [2^{-(n-k+1)/2}, \infty)$ and any $R > 2r$ we have*

$$\Lambda^n_{\alpha,(\alpha_k,\ldots,\alpha_n)}, \tilde{\Lambda}^n_{\alpha,(\alpha_k,\ldots,\alpha_n)} \lesssim (\alpha \cdot \alpha_k \ldots \alpha_n)^2 \left(\log \frac{R}{r}\right)^{1-2^{-k+1}} \tag{11}$$

If $k = 1$, we have $\tilde{\Lambda}^n_{\alpha,(\alpha_k,\ldots,\alpha_n)} \lesssim 1$

11.5 Showing $\tilde{\Lambda}^k \lesssim \Lambda^k$

By taking the absolute value inside in (9),

$$\tilde{\Lambda}^k_{\alpha;\alpha_{\geq k}} \leq \int_r^R \frac{dt}{t} \int_{\mathbb{R}^{n+1}} |h_{t\alpha}(\bar{\mathbf{x}})| d\mathbf{x} \int_{\mathbb{R}^{2(n-k)}} d\gamma_{\mathbf{x}_{>k}, t\cdot \mathbf{a}_{>k}}(\mathbf{z}^\bullet_{>k}) \Bigg\{ \left| \int \mathcal{F}_k(\mathbf{x}_{<k}, y, \mathbf{z}^\bullet_{>k}) h_{t\alpha_k}(y - \mathbf{x}_k) dy \right| \Bigg\} \tag{12}$$

where the only difference with $\Lambda^k_{\alpha;\alpha_{\geq k}}$ is the $|h_{t\alpha}(\bar{\mathbf{x}})|$ term, as opposed to $g_{t\alpha}(\bar{\mathbf{x}})$. The inequality $\tilde{\Lambda}^k \lesssim \Lambda^k$ now follows from

$$|h_{t\alpha}(\bar{\mathbf{x}})| \lesssim g_{2t\alpha}(\bar{\mathbf{x}}) \tag{13}$$

11.6 Going up in k

We apply Cauchy-Schwarz to the definiton (8) of $\Lambda^n_{\alpha;\alpha_{\geq k}}$ twice. First in t, to get

$$|\Lambda^k_{\alpha;\alpha_{\geq k}}|^2 \leq \log\frac{R}{r}\cdot\int_r^R\frac{dt}{t}\left(\int_{\mathbb{R}^{n+1}}g_{t\alpha}(\bar{\mathbf{x}})d\mathbf{x}\int_{\mathbb{R}^{2(n-k)}}d\gamma(\mathbf{z}^\bullet_{>k})\right.$$
$$\left.\left|\int\mathcal{F}_k(\dots)h_{t\alpha_k}(y-\mathbf{x}_k)dy\right|\right)^2 \tag{14}$$

(*This step is the reason the gain in the exponent is* 2^{-k+1}*: At each induction step we only keep half the gains of the previous step.*)
We then expand the product of $\mathcal{F}_k$ into the terms involving f_j for $j<k$ and the terms involving f_k only. These last terms do not depend on y, so we apply Cauchy-Schwarz in all variables but y. We obtain an expression of the form

$$|\Lambda^k_{\alpha;\alpha_{\geq k}}|^2 \leq \int_r^R \mathcal{M}_{\leq k}(t)\mathcal{N}_k(t)\frac{dt}{t} \tag{15}$$

The term $\mathcal{M}_k(T)$ only depends on the last f_k and can be dealt with (uniformly in t) by direct methods. The term $\int_r^R\mathcal{N}_{\leq k}(t)\frac{dt}{t}$ has the form

$$\int_r^R\frac{dt}{t}\int_{\mathbb{R}^{n+1}}g_{t\alpha}(\bar{\mathbf{x}})d\mathbf{x}\int_{\mathbb{R}^{2(n-k)}}d\gamma(\mathbf{z}^\bullet_{>k})\left|\int\prod_{j<k}\prod_{r}F_j((...)_{\hat{j}}),h_{t\alpha_k}(y-\mathbf{x}_k)dy\right|^2.$$

We can write a square of an integral $|\int_{\mathbb{R}}\phi(y)dy|^2$ as $\int_{\mathbb{R}^2}\phi(z^0_k)\phi(z^1_k)dz^0_kdz^1_k$. Using this, we get an equivalent expression for $N_{\leq k}$, namely $N_{\leq k}=\Theta^k_{\leq k}$, where for $j\geq k$

$$\Theta^j_{\leq k}(t) := \int_r^R\frac{dt}{t}\int_{\mathbb{R}^{n+1}}g_{t\alpha}(\bar{\mathbf{x}})d\mathbf{x}\int_{\mathbb{R}^{2(n-k)}}d\gamma(\mathbf{z}^\bullet_{\geq k,\hat{j}})\Bigg\{$$
$$\int_{\mathbb{R}^2}\mathcal{F}_{k-1}(\mathbf{x}_{<k-1},\mathbf{x}_{k-1},\mathbf{z}^\bullet_{>k-1})h_{t\alpha_j}(z^0_j-x_j)h_{t\alpha_j}(z^1_j-x_j)d\mathbf{z}^0_jd\mathbf{z}^1_j\Bigg\}. \tag{16}$$

Since $\mathcal{F}_{k-1}$ does not depend on $\mathbf{x}_{\geq k}$ or t, one can perform those integrals first, and define $d\theta^j_{\leq k}(\mathbf{x}_{\leq k},\mathbf{z}^\bullet_{>k})$ as the measure witnessing the integral over $\mathcal{F}_{k-1}$ in (16), so that one has

$$\Theta^j_{\leq k}(t) = \int_{\mathbb{R}^k\times\mathbb{R}^{2\cdot(n+1-k)}}\mathcal{F}_{k-1}(x_{<k-1},x_{k-1},z^\bullet_{>k-1})d\theta^j_{\leq k}(x_{\leq k},z^\bullet_{>k}). \tag{17}$$

Define, analogously, $d\tilde{\lambda}_{k-1}$ as the measure witnessing $\tilde{\Lambda}^k_{\alpha/\sqrt{2};\alpha_k/\sqrt{2},\alpha_{k+1}\ldots\alpha_n}$ in (9) (without absolute values). The key "integration by parts" estimate, which can be shown explicitly, where one gains some cancellation is

$$\sum_{j=k}^{n} d\theta^j_{\leq k} - \left(1 - \alpha^{-2}\sum_{j=k}^{n}\alpha_j^2\right) d\tilde{\lambda}_{k-1} = dG_R - dG_r \tag{18}$$

where dG_t is the measure induced by

$$\int \mathcal{F}_{k-1}(x_{<k-1}, x_{k-1}, z^\bullet_{>k-1}) dG_t := \int \mathcal{F}_{k-1}(x_{<k-1}, x_{k-1}, z^\bullet_{>k-1}) d\gamma(\mathbf{z}^\bullet_{>k-1}) g(\bar{\mathbf{x}}) d\mathbf{x}$$

Now the induction step follows by integrating $\mathcal{F}_k$ against both sides: The integrals against dG_R and dG_r are *single scale*, and can be controlled ($\lesssim 1$) by Cauchy-Schwartz. The integrals against $d\theta^j_{\leq k}$ give $\Theta^j_{\leq k}(t)$, which are all nonnegative because they arose from a Cauchy-Schwarz inequality. The integral over $d\tilde{\lambda}$ gives a term of the form $\tilde{\Lambda}^k_{\alpha;\alpha_{\geq k}}$. That gives

$$\Theta^k_{\leq k} \lesssim 1 + \left(1 - \alpha^{-2}\sum_{j=k}^{n}\alpha_j^2\right) \tilde{\Lambda}^k_{\alpha/\sqrt{2};\alpha_k/\sqrt{2},\alpha_{k+1}\ldots\alpha_n} \tag{19}$$

closing the induction.
The base step, corresponding to $\tilde{\Lambda}^1_{\alpha;\alpha_{\geq 1}}$, is (a simpler) variation of the induction step and can be found at the end of [1, Section 2].

References

[1] P. Durcik, V. Kovač, and C. Thiele, *Power-type cancellation for the simplex Hilbert transform*, Journal d'Analyse Mathématique, 139(1), pp.67–82, 2019. arXiv:1608.00156.

Jaume de Dios Pont, ETHZ
email: `jaumededios@gmail.com`

12 Detangling a Twisted Form in L^4

After P. Durcik [DU]

A summary written by Jacob Denson and Jacob Fiedler

Abstract

We discuss a 'twisted' singular quadrilinear form introduced by Demeter and Thiele, which relates to the almost everywhere convergence of statistics associated with commuting ergodic operators, and discuss the proof of the L^4 boundedness of this form.

Take four functions F_1, F_2, F_3, and F_4 on $\mathbb{R}^2$, and 'entangle them', forming the function

$$\mathbf{F}(x, x', y, y') \coloneqq F_1(x, y) F_2(x', y) F_3(x, y') F_4(x', y') \tag{1}$$

We will be interested in the following quadrilinear form:

$$\Lambda(F_1, F_2, F_3, F_4) \coloneqq \int_{\mathbb{R}^2} \widehat{\mathbf{F}}(\xi, -\xi, \eta, -\eta) m(\xi, \eta) d\xi d\eta,$$

where $m : \mathbb{R}^2 \to \mathbb{C}$ obeys the symbol estimates $|\partial^\alpha m(\xi, \eta)| \lesssim (|\xi| + |\eta|)^{-|\alpha|}$ for sufficiently large α. The main result of [DU] is the following L^4 bound:

Theorem 1. *The quadrilinear form Λ satisfies*

$$|\Lambda(F_1, F_2, F_3, F_4)| \lesssim \|F_1\|_{L^4(\mathbb{R}^2)} \|F_2\|_{L^4(\mathbb{R}^2)} \|F_3\|_{L^4(\mathbb{R}^2)} \|F_4\|_{L^4(\mathbb{R}^2)} \tag{2}$$

A special case of this quadrilinear form is the so-called 'twisted paraproduct' introduced by Demeter and Thiele and defined as follows:

$$T(F_1, F_2, F_3) \coloneqq \Lambda(F_1, F_2, F_3, 1). \tag{3}$$

However, using Λ brings to light certain extra symmetries in the problem not immediate obvious in the definition of T. The results of Kovač [K] and Bernicot [BE] show that for $1/p_1 + 1/p_2 + 1/p_3 = 1$ and $p_2 > 2$,

$$|T(F_1, F_2, F_3)| \lesssim_{p_1, p_2, p_3} \|F_1\|_{L^{p_1}(\mathbb{R}^2)} \|F_2\|_{L^{p_2}(\mathbb{R}^2)} \|F_3\|_{L^{p_3}(\mathbb{R}^2)}. \tag{4}$$

Bounding T has ramifications in ergodic theory, detailed in the next section.

12.1 Why do paraproducts relate to ergodic theory?

Let X be a probability space and let $T, S : X \to X$ be commuting measure-preserving transformations on X. A natural question in ergodic theory is to study, for $f, g \in L^\infty(X)$, the almost everywhere convergence of the averages

$$\frac{1}{N}\sum_{n=1}^{N} f(T^n x)g(S^{-n}x) \quad \text{as } N \to \infty. \tag{5}$$

Using a paraproduct estimate, Demeter and Thiele [DT] showed convergence of a related family of averages, including

$$\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} f(T^n S^m x)g(T^{-n}S^m x) \tag{6}$$

The basic idea is that if one can bound the oscillation of a weighted version of the ergodic averages by $C_J\|f\|_{L^{p_1}}\|g\|_{L^{p_2}}$ (where C_J is a term related to the oscillation) this is sufficient to conclude pointwise convergence on a full measure subset of X. In [DE], Demeter details this argument in the course of reproving and extending a result of Bourgain on the convergence of (5) when S is a power of T. In this case, the desired inequality is

$$\begin{aligned}\Big\|\Big(\sum_{j=1}^{J-1}\sup_{k\in[u_j,u_{j+1})}|W_k(f,g)(x)-W_{u_{j+1}}(f,g)(x)|^2\Big)^{\frac{1}{2}}\Big\|_{L^{1,\infty}(X)}& \\ \lesssim J^{\frac{1}{4}}\|f\|_{L^2(X)}\|g\|_{L^2(X)},&\end{aligned} \tag{7}$$

where the bound is uniform in J and all finite sequences $U_1, ..., U_J$, and where

$$W_k(f,g)(x) := \sum_{n\in\mathbb{Z}} w_{n,k}f(T^n x)g(T^{-n}x).$$

Connecting bounds of this type to the types of estimates in this paper requires invoking a transfer principle. Equipped with the right inequality, one can consider functions on $\mathbb{R}^2$ which are constant on all the integer lattice squares $(n, n+1)\times(m, m+1)$, essentially functions on $\mathbb{Z}^2$. To complete the transfer to X, use the functions F on $\mathbb{Z}^2$ which are of the form $F(n,m) = f(T^n S^m x)$ for some $x \in X$. For further reference, [DLTT] details the transfer of a bound on a maximal average to a bound on an ergodic average.

The salient point is that after the transfer, we have essentially the same upper bound. So, when Demeter and Thiele obtain an oscillation bound for a sum of integrals of the form

$$\int_{\mathbb{R}^2} F_1(x+t,y+s)F_2(x-t,y+s)\Psi_k(t)\Phi_k(t)dtds$$

it implies the same bound for the oscillation of (6) (note the relationship between the exponents in the ergodic average and the arguments in the above integral), which imply the required pointwise a.e. convergence. In an analogous manner, a better understanding of bounds for

$$\int_{\mathbb{R}^2} F_1(x+t,y)F_2(x,y+t)dtds$$

would improve the understanding of the more challenging average (5), and bounding this bilinear Hilbert transform is directly related to bounding the 'triangular' Hilbert transform defined in (3).

12.2 The L^4 estimate

Recall the statement of Theorem 1. Spending rescaling symmetries, we may assume that $\|F_1\|_{L^4},\dots,\|F_4\|_{L^4}=1$, and our goal is to prove that $|\Lambda|\lesssim 1$. The proof has a nice flavor, because the main tools are all very general, but used in some novel clever ways:

(A) Time-Frequency Analysis, i.e. simultaneous decompositions of functions to localize behaviour in space and frequency.

(B) Exploiting cancellation using a 'telescoping identity', which for intuition's sake behaves like a multilinear variant of an integration by parts.

(C) Using monotonicity to replacing arbitrary functions with concrete functions (e.g. Gaussians).

Let's begin with Technique (A). Without loss of generality, assume $\mathrm{supp}(m)$ is contained in a cone $\Gamma=\{(\xi,\eta):|\xi|\leq 1.001|\eta|\}$, since symmetry and the triangle inequality then give the general result. Next, perform a time-frequency decomposition of the multiplier m, writing

$$m(\xi,\eta)=\int_0^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty \mu(u,v)\widehat{\varphi}_{t,u}(t\xi)^2\widehat{\psi}_{t,v}(t\eta)^2\;dt/t\;du\;dv,$$

where $\widehat{\varphi}_{t,u}$, $\varphi_{t,u}$, $\widehat{\psi}_{t,u}$, and $\psi_{t,u}$ are concentrated on $\{\xi : |\xi| \lesssim 1/t\}$, $\{x : |x - tu| \lesssim t\}$, $\{\eta : |\eta| \sim 1/t\}$, and $\{y : |y - v| \lesssim t\}$ respectively, and $\varphi_{t,u}$ and $\psi_{t,u}$ are L^1 normalized. The squares in the exponent here are irrelevant to the existence of the decomposition, but will be necessary to get a nice convolution representation of the operator later on in equation (8). The symbol properties of m imply that the magnitude of $\mu(u, v)$ decays rapidly as $|u|, |v| \to \infty$. The result will therefore follow if we can obtain bounds on

$$\int \widehat{\varphi}_{t,u}(\xi)^2 \widehat{\psi}_{t,v}(\xi)^2 \, \widehat{\mathbf{F}}(\xi, -\xi, \eta, -\eta) \, dt/t \, d\xi \, d\eta \quad \text{uniformly in } u \text{ and } v.$$

The values of u and v are not too important to the main ideas of the problem, so we will suppress them in later notation, i.e. writing φ_t for $\varphi_{t,u}$, and ψ_t for $\psi_{t,v}$. The function ϕ_t behaves like a Gaussian supported in a neighborhood of u, and ψ_t like a modulated Gaussian supported in a neighborhood of v. In fact, in our calculations we will eventually use Technique (C) to replace these functions with Gaussians.
Writing $f^-(t) = f(-t)$ for the reflection of a function f, we can write Λ as a 'twisted convolution operator', i.e.

$$\Lambda = \int \Lambda_t \, dt/t, . \tag{8}$$

where

$$\Lambda_t := \int \mathbf{F}(x, y, x', y') \varphi_t(\tilde{x} - x) \varphi_t^-(\tilde{x} - x') \psi_t(\tilde{y} - y) \psi_t^-(\tilde{y} - y')$$

Let us write Λ_t as $\Lambda_{\varphi_t, \varphi_t^-, \psi_t, \psi_t^-}(F_1, F_2, F_3, F_4)$, where

$$\Lambda_{a,b,c,d} := \int \mathbf{F}(x, y, x', y') a(\tilde{x} - x) b(\tilde{x} - x') c(\tilde{y} - y) d(\tilde{y} - y').$$

The terms involving ψ are where the significant cancellation occurs in the integral, with ϕ providing little cancellation, and so we start by applying the triangle inequality, writing

$$\begin{aligned} |\Lambda_t| \leq \int & \left| \int F_1(x, y) F_2(x', y) [\psi_t]_t(\tilde{y} - y) \, dy \right| \\ & \left| \int F_3(x', y') F_4(x, y') [\psi_t]_t(y' - \tilde{y})] \, dy' \right| \\ & |\varphi_t(x - \tilde{x})| \, |\varphi_t(\tilde{x} - x')| \, dx \, dx' \, d\tilde{x} \, d\tilde{y}. \end{aligned} \tag{9}$$

In the worst case, the two integrals in the absolute values of (9) in could be equal to one another (e.g. if $F_1 = F_3$, $F_2 = F_4$ and $|v| \ll 1$), which means Cauchy-Schwartz in $\tilde{y}$ is likely to be efficient, and expanding out the squares that are obtained by Cauchy-Schwartz, we obtain

$$|\Lambda_t| \leq \Lambda_{\psi_t,\psi_t,|\varphi_t|,|\varphi_t^-|}(F_1, F_2, F_2, F_1)^{1/2}\Lambda_{\psi_t^-,\psi_t^-,|\varphi_t|,|\varphi_t^-|}(F_3, F_4, F_4, F_3)^{1/2}. \tag{10}$$

Notice that expanding out the square allows us to remove the absolute values we introduced via the triangle inequality. In fact, reversing this calculation shows

$$\Lambda_{a,a,b_1,b_2}(F_1, F_2, F_2, F_1) \geq 0 \qquad \text{for all } a, b_1, b_2 \text{ if } b_1, b_2 \geq 0. \tag{11}$$

By symmetry, we focus on bounding $\Lambda_{\psi_t,\psi_t,|\varphi_t|,|\varphi_t^-|}(F_1, F_2, F_2, F_1)$, which now calls for applying Technique (B) to exploit the oscillation of ψ_t.

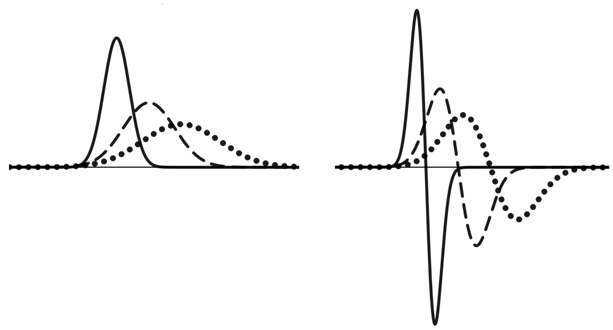
Lemma 3 of [DU]. If $-t\partial_t|\widehat{\rho}_i|^2 = |\widehat{\sigma}_i(t\tau)|^2$ for $i \in \{1, 2\}$, then

$$\int \Lambda_{\sigma_1,\sigma_1,\rho_2,\rho_2}\, dt/t = |\widehat{\rho}_1(0)|^2|\widehat{\rho}_2(0)|^2 \int_{\mathbb{R}^2} F_1F_2F_3F_4 - \int \Lambda_{\rho_1,\rho_1,\sigma_2,\sigma_2}\, dt/t. \tag{12}$$

The proof given in [DU] is very accessible, so for purposes of brevity we refer to reading that Lemma directly from the paper. This Lemma works like integration by parts, in the sense that we 'antidifferentiate' σ, at the cost of 'differentiating' ρ, negating the integral (except that these derivatives 'preserve L^1 normalization), and introducing the 'boundary term' $\int F_1F_2F_3F_4$. Abusing notation, when applying Lemma 3 we will refer to pairs ρ_i and σ_i in the theorem as 'derivatives' and 'antiderivatives' of one another respectively. To given intuition, one example of a pair ρ and ψ which satisfy

$$\rho(t, x) = t^{-1}e^{-(x/t)^2} \quad \text{and} \quad \sigma(t, x) = -(4\sqrt{\pi}x/t^2)e^{-2\pi(x/t)^2},$$

Graphs of ρ and σ for various values of t are given below, on the left and right respectively.



We wish to apply the Lemma to $\Lambda_{\psi_t,\psi_t,|\varphi_t|,|\varphi_t|^-}(F_1, F_2, F_1, F_2)$, except that $|\varphi_t|$ does not equal $|\varphi_t^-|$. But we can fix this by emplying (11), which also implies the monotonicity of $\Lambda_{a,a,b_1,b_2}(F_1, F_2, F_2, F_1)$ with respect to b_1 and b_2. This means that

$$\Lambda_{\psi_t,\psi_t,\varphi_t,\varphi_t^-}(F_1, F_2, F_2, F_1) \lesssim \Lambda_{\psi_t,\psi_t,\Phi_t,\Phi_t}(F_1, F_2, F_2, F_1), \tag{13}$$

where, roughly speaking, Φ_t is an L^1 normalized even function, a sum of two Gaussians centered at u and $-u$ and supported on a t neighborhood, chosen to dominate φ_t and φ_t^-. Thus we have applied Technique (C). If we let $D\Phi_t$ denote the 'derivative' of Φ_t, and $I\psi_t$ the 'antiderivative' of ψ_t, then we obtain that

$$\int \Lambda_{\psi_t,\psi_t,\Phi_t,\Phi_t}\, dt/t = c\int_{\mathbb{R}^2} F_1^2 F_2^2 - \int \Lambda_{I\psi_t,I\psi_t,D\Phi_t,D\Phi_t}(F_1, F_2, F_1, F_2)\, dt/t.$$

We can choose ψ_t such that $I\psi_t$ has support on a length $O(t)$ interval. By Cauchy-Schwartz, $\int F_1^2 F_2^2 \leq 1$, and so it suffices to show that

$$\left|\int |\Lambda_{I\psi_t,I\psi_t,D\Phi_t,D\Phi_t}(F_1, F_2, F_2, F_1)\, dt/t\right| \lesssim 1$$

But by differentiating Φ_t, we have juggled the oscillation from the first two functions in Λ to the latter functions in Λ, i.e. $I\psi_t$ no longer necessary oscillation, but $D\Phi_t$ is now oscillating, and so we should mirror our calculations in (10), applying Cauchy-Schwartz in $\tilde{x}$ instead of $\tilde{y}$, which yields

$$\Lambda_{I\psi_t,I\psi_t,D\Phi_t,D\Phi_t}(F_1, F_2, F_1, F_2) \leq \Lambda_{|I\Psi_t|,|I\Psi_t|,D\Phi_t,D\Phi_t}(F_1, F_1, F_1, F_1)^{1/2}$$
$$\Lambda_{|I\Psi_t|,|I\Psi_t|,D\Phi_t,D\Phi_t}(F_2, F_2, F_2, F_2)^{1/2}.$$

By symmetry, we again focus on $\Lambda_{|I\Psi_t|,|I\Psi_t|,D\Phi_t,D\Phi_t}(F_1,F_1,F_1,F_1)$. We have now succeeded at disentangling the four functions, while maintaining cancellation in the integrals studied. We now want to do a final application of the telescoping identity, which, by exploiting monotonicity, allows us to replace $|I\Psi_t|$ with Φ_t, and then the telescoping identity yields that

$$\begin{aligned} &\int \Lambda_{\Phi_t,\Phi_t,D\Phi_t,D\Phi_t}(F_1,F_1,F_1,F_1)dt/t \\ &\qquad = \int F_1^4 - \int \Lambda_{D\Phi_t,D\Phi_t,\Phi_t,\Phi_t}(F_1,F_1,F_1,F_1)dt/t. \end{aligned} \tag{14}$$

Symmetry gives $\Lambda_{D\Phi_t,D\Phi_t,\Phi_t,\Phi_t}(F_1,F_1,F_1,F_1) = \Lambda_{\Phi_t,\Phi_t,D\Phi_t,D\Phi_t}(F_1,F_1,F_1,F_1)$, and so now, rearranging the equation (14), and using the bound $\int F_1^4 = 1$, we conclude $|\int \Lambda_{D\Phi_t,D\Phi_t,\Phi_t,\Phi_t}(F_1,F_1,F_1,F_1)\ dt/t| \lesssim 1$, completing the proof.

References

[BE] Frédéric Bernicot, *Fiber-wise Calderón-Zygmund decomposition and application to a bi-dimensional para-product* llinois J. Math. 56 (2), (2012), 415-422.

[DE] Ciprian Demeter, *Pointwise convergence of the ergodic bilinear Hilbert transform.* Illinois J. Math. 51 (4) (2007), 1123 - 1158.

[DLTT] Ciprian Demeter, Michael Lacey, Terence Tao, Christoph Thiele, *Breaking the duality in the return times theorem* Duke Math. J. 143 (2), (2008), 281-355.

[DT] Ciprian Demeter, Christoph Thiele, , *On the two-dimensional bilinear Hilbert transform*, Amer. J. Math., 132 (1), (2010), 201-256.

[DU] Polona Durcik, *An L^4 estimate for a singular entangled quadrilinear form.* Mathematical Research Letters, 22 (5), (2015), 1317-1332.

[K] V. Kovač, *Boundedness of the twisted paraproduct.* Rev. Mat. Iberoam., 28 (4), (2012), 1143-1164.

Jacob Denson and Jacob Fiedler, University of Wisconsin-Madison
emails: `jcdenson@wisc.edu` and `jbfiedler2@wisc.edu`.

13 Pointwise characteristic factors for the Wiener-Wintner double recurrence theorem

After I. Assani, D. Duncan and R. Moore [ADM]

A summary written by Leon Duensing

Abstract

In [ADM] it is shown that for a standard ergodic system $(X, \Sigma, \mu, \varphi)$ with $f_1, f_2 \in L^\infty(X)$ the averages

$$\frac{1}{N}\sum_{n=1}^{N} f_1(\varphi^{an}(x))f_2(\varphi^{bn}(x))e^{2\pi int}, \quad a, b \in \mathbb{Z}$$

converge as $N \to \infty$ for almost every $x \in X$, independently of t. In the talk we prove that the Conze-Lesigne factor $\mathcal{Z}_2$ is characteristic for these averages.

13.1 Introducing the statements.

Fix a standard measure-preserving system $(X, \Sigma, \mu, \varphi)$, meaning that X is a compact, metrizable topological space, Σ is the borelean σ-algebra of X and $\varphi\colon X \to X$ is a homeomorhpism preserving the measure μ. Assume further, that this system is ergodic and take two functions $f_1, f_2 \in L^\infty(X)$ and a pair of integers $a \leq b$. Let $T\colon L^1(X) \to L^1(X)\colon f \mapsto f \circ \varphi$ be the Koopman-Operator (or pullback) of the transformation.
In this talk, we study weighted double averages of the form

$$W_{N,t}(f_1, f_2) \coloneqq \frac{1}{N}\sum_{n=1}^{N} T^{an}f_1 \cdot T^{bn}f_2 \cdot e^{2\pi int}, \quad N \in \mathbb{N}, t \in \mathbb{R}. \tag{1}$$

Assani, Duncan and Moore proved the following.

Theorem 1 (Weighted double convergence)**.** *There exists a full measure subset $X' \subseteq X$ such that the sequence $W_{N,t}(f_1, f_2)(x, x)$ converges for each $x \in X'$ and $t \in \mathbb{R}$.*

The strategy of the proof goes by first finding a factor $\mathcal{F} \subseteq L^2(X)$ of the system, which is *uniformly pointwise characteristic* (or just characteristic) for the averages $W_{N,t}$, meaning that there is a full measure subset $X' \subseteq X$ such that

$$\lim_{N\to\infty} \big|W_{N,t}(f_1, f_2)(x,x) - W_{N,t}\big(\mathbb{E}(f_1|\mathcal{F}), \mathbb{E}(f_2|\mathcal{F})\big)(x,x)\big| = 0$$

for all $x \in X'$ and $t \in \mathbb{R}$. Note, that this property is satisfied if

$$\limsup_{N\to\infty} \sup_{t\in\mathbb{R}} |W_{N,t}(f_1, f_2)(x,x)| = 0 \quad \text{for a.e. } x \in X$$

whenever either f_1 or f_2 lie in $\mathcal{F}^\perp$.
A promising candidate for this is the *Conze-Lesigne Factor* $\mathcal{Z}_2$.

Definition 2 (Host-Kra seminorms and factors)**.** *For $f \in L^2(X)$ define the Host-Kra seminorms recursively by*

$$|||f|||_1^2 := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_X |T^n f \cdot \overline{f}| \, d\mu$$

and for each $k \geq 2$

$$|||f|||_k^{2^k} := \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |||T^n f \cdot \overline{f}|||_{k-1}^{2^{k-1}}.$$

In particular, for every $k \geq 0$ we define the k-th Host-Kra factor as

$$\mathcal{Z}_k := \{f \in L^2(X) \colon |||f|||_{k+1} = 0\}^\perp.$$

In the talk we focus on proving, that $\mathcal{Z}_2$ is characteristic.

Theorem 3 (Double Uniform Wiener-Wintner Theorem)**.** *If f_1 lies in $\mathcal{Z}_2^\perp$, then*

$$W(x) := \limsup_{N\to\infty} \sup_{t\in\mathbb{R}} \left| \frac{1}{N} \sum_{n=1}^{N} T^{an} f_1(x) T^{bn} f_2(x) e^{2\pi i n t} \right| = 0 \tag{2}$$

for almost every $x \in X$.

Hence, in order to establish Theorem 1, it is sufficient to only consider functions $f_1, f_2 \in \mathcal{Z}_2$.

13.2 $\mathcal{Z}_2$ is characteristic.

We sketch the proof of Theorem 3.
In order to show that $\|W\|_2 = 0$ first, one uses the Van-der-Corput lemma on the sequence

$$\left(|T^{an} f_1(x) T^{bn} f_2(x)|\right)_{n\in\mathbb{N}}$$

and Cauchy-Schwarz to deduce

$$W(x)^2 \leq \frac{C}{H} + \frac{C}{H} \left(\sum_{h=1}^{H} \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left| T^{an}(f_1 \cdot T^{ah} f_1)(x) \cdot T^{bn}(f_2 \cdot T^{bh} f_2)(x) \right|^2 \right)^{1/2} \tag{3}$$

with $C > 0$ being a constant and $H \in \mathbb{N}$ a free parameter. Applying Van-der-Corput again on the resulting sequence of correlations

$$\left(|T^{an}(f_1 \cdot T^{ah} f_1)(x) \cdot T^{bn}(f_2 \cdot T^{bh} f_2)(x)|\right)_{n\in\mathbb{N}}$$

for each $h \in \mathbb{N}$ and setting

$$G_{h,k}^{(1)} = f_1 \cdot T^{ah} f_1 \cdot T^{ak} f_1 \cdot T^{a(k+h)} f_1, \quad G_{h,k}^{(2)} = f_2 \cdot T^{bh} f_2 \cdot T^{bk} f_2 \cdot T^{b(k+h)} f_2$$

yields the estimate

$$\begin{aligned} &\limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left| T^{an}(f_1 \cdot T^{ah} f_1)(x) \cdot T^{bn}(f_2 \cdot T^{bh} f_2)(x) \right|^2 \leq \frac{C}{K} + \\ &\qquad \frac{C}{(K+1)^2} \sum_{k=1}^{K} (K+1-k) \limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^{an} G_{h,k}^{(1)}(x) \cdot T^{bn} G_{h,k}^{(2)}(x) \end{aligned} \tag{4}$$

for any $K \in \mathbb{N}$. By a theorem of Bourgain (see [Ru] Theorem 1), the limes superior in (4) is actually a limit and we can use the mean ergodic theorem w.r.t T^{b-a} to deduce

$$\begin{aligned} &\int_X \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} T^{an} G_{h,k}^{(1)}(x) \cdot T^{bn} G_{h,k}^{(2)}(x)\, dx \\ &= \int_X \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} G_{h,k}^{(1)}(x) \cdot T^{(b-a)n} G_{h,k}^{(2)}(x)\, dx \\ &= \int_X G_{h,k}^{(1)}(x) \cdot \mathbb{E}\big(G_{h,k}^{(2)} | \mathcal{I}_{b-a}\big)(x)\, dx. \end{aligned} \tag{5}$$

Now a key step of the proof lies in representing the conditional expectation $\mathbb{E}(\cdot|\mathcal{I}_{b-a})$ as an integral operator.

Lemma 4. *Let T be the Koopman-operator of an ergodic system $(X, \Sigma, \mu, \varphi)$, $m \in \mathbb{N}$ and denote by $\mathcal{I}_m \subseteq L^2(X)$ the subspace of T^m-invariant functions. Then for every $f \in L^2(X)$*

$$\mathbb{E}(f|\mathcal{I}_m)(x) = \int_X f(y) \cdot K_m(x, y)\, dy$$

with some kernel $K_m \in L^2(X \times X)$.

Let $S\colon L^1(X \times X) \to L^1(X \times X)$ be the Koopman-Operator of the transformation $\varphi^a \times \varphi^b\colon X \times X \to X \times X$. Resuming at (5), we conclude

$$\begin{aligned}
&\int_X G^{(1)}_{h,k}(x) \cdot \mathbb{E}\big(G^{(2)}_{h,k}|\mathcal{I}_{b-a}\big)(x)\, dx \\
&= \int_X \int_X G^{(1)}_{h,k}(x) \cdot G^{(2)}_{h,k}(y) \cdot K(x, y)\, dxdy \qquad (6)\\
&= \int_{X^2} K_{b-a} \cdot f_1 \otimes f_2 \cdot \big(S^h f_1 \otimes f_2 \cdot S^k f_1 \otimes f_2 \cdot S^{k+h} f_1 \otimes f_2\big) d\mu \otimes \mu.
\end{aligned}$$

Now set $H = K$ for the parameters in (3) and (4), so that when H tends to infinity

$$\begin{aligned}
&\int_X |W(x)|^2\, d\mu \leq \int_{X^2} \underbrace{(K_{b-a} \cdot f_1 \otimes f_2)(x, y) \cdot \lim_{H\to\infty} \frac{1}{H(H+1)^2} \sum_{h,k=0}^{H-1} (H + 1 - k)\big(S^h f_1 \otimes f_2 \cdot S^k f_1 \otimes f_2 \cdot S^{k+h} f_1 \otimes f_2\big)(x, y)}_{=:F(x,y)}\, d\mu \otimes \mu
\end{aligned} \qquad (7)$$

It is left to estimate the L^2-Norm of F from (7) against the third Host-Kra seminorm of f_1.

Lemma 5. *Under the standing assumptions, there is a constant $c > 0$ such that the inequality*

$$\int_{X^2} |F|^2\, d\mu \otimes \mu \leq c|a|^{1/2} \| f_1 \|_3^2$$

yields.

Since $\| f_1 \|_3 = 0$, the claim follows by applying Hölder's inequality.

References

[ADM] Assani, I. and Duncan, D. and Moore, R. *Pointwise characteristic factors for Wiener-Wintner double recurrence theorem.* Ergodic Theory Dyn. Syst. 36 (2016), No. 4, 1037-1066.

[Ru] Rudolph, D. *A joinings proof of Bourgain's return time theorem.* Ergodic Theory Dyn. Syst. 14, No. 1, 197-203 (1994).

Leon Duensing, University of Tuebingen
email: `ledu@fa.uni-tuebingen.de`

14 A new proof of Szemerédi's theorem for arithmetic progressions of length four

After W. T. Gowers [G]

A summary written by Kornélia Héra

Abstract

We give a new proof of Szemerédi's theorem for arithmetic progressions of length four, using exponential sums.

14.1 Introduction

The famous theorem of Szemerédi states that, for any positive integer k and any real number $\delta > 0$, there exists N such that every subset of $\{1, 2, \ldots, N\}$ of cardinality at least δN contains an arithmetic progression of length k. The first progress toward the result was obtained by Roth [R], who proved it for the special case $k = 3$ using exponential sums. Szemerédi [SZ] found a more combinatorial proof for the $k = 3$ case, which he then generalized for all k. In a very influential paper, Furstenberg [F] used techniques from ergodic theory to prove Szemerédi's theorem and certain extensions.
Despite the presence of a fruitful history of the problem, a natural question remains to be asked: can Roth's method of proof for the $k = 3$ case be generalized? In this paper, this is carried our for the $k = 4$ case. The motivation to generalize Roth's method does not only stem from the fact that the argument is natural and nice, but also from the fact that bounds arising from the known proofs of Szemerédi's theorem are very weak, and in general for similar problems the use of exponential sums tend to give strong bounds. The bound appearing in our theorem below is a significant improvement over the previously known bounds.
Our theorem is the following.

Theorem 1. *There is an absolute constant C with the following property. If A is a subset of $\{1, 2, \ldots, N\}$ with cardinality δN and $N \geq \exp\exp\exp((1/\delta)^C)$, then A contains an arithmetic progression of length* 4.

As an immediate corollary, we also obtain the following.

Corollary 2. *There is an absolute constant c with the following property. If the set $\{1, 2, \ldots, N\}$ is colored with at most $(\log\log\log N)^c$ colors, then there is a monochromatic arithmetic progression of length 4.*

The rough idea of the proof of Theorem 1 is the following. We use a notion of pseudo randomness, called quadratic uniformity, and use the fact (which has been proved earlier) that quadratically uniform sets with the appropriately chosen parameters contain an arithmetic progression of length 4. Then we show that if a set fails to be quadratically uniform then it can be restricted to a large arithmetic progression where its density increases noticeably. Using the latter in an iterative fashion, the result will follow.

14.2 Preliminaries

14.2.1 Notation and definitions

Given a positive integer N, let $\mathbb{Z}_N$ denote the group of integers mod N. The cardinality of a finite set A is denoted by $|A|$.
We write $\omega = \exp(2\pi i/N)$. Given a function $f : \mathbb{Z}_N \to \mathbb{C}$, its Fourier coefficients are defined as

$$\tilde{f}(r) = \sum_{s\in\mathbb{Z}_N} f(s)\omega^{-rs} = \sum_{s\in\mathbb{Z}_N} f(s)\exp(2\pi i(-rs)/N),\ r \in \mathbb{Z}_N.$$

Moreover, for f as above and $k \in \mathbb{Z}_N$, we define

$$\Delta(f;k)(s) = f(s)\overline{f(s-k)},\ s \in \mathbb{Z}_N.$$

Let D denote the closed unit disk in $\mathbb{C}$. Let $f : \mathbb{Z}_N \to D$, and $\alpha > 0$. We say that f is quadratically α-uniform, if

$$\sum_{u\in\mathbb{Z}_N}\sum_{v\in\mathbb{Z}_N}|\sum_{s\in\mathbb{Z}_N} f(s)\overline{f(s-u)f(s-v)}f(s-u-v)|^2 \le \alpha N^4.$$

A special type of functions that we will use are balanced functions. Let $A \subset \mathbb{Z}_N$ with size δN. The balanced function of A is defined as

$$f_A(s) = \begin{cases} 1-\delta & s \in A \\ -\delta & s \notin A \end{cases}$$

A set A is quadratically α-uniform, if its balanced function f_A is.
Our last definition is the following. Let $B \subset \mathbb{Z}_N$ and let $\phi : B \to \mathbb{Z}_N$ be an arbitrary function. We say that $(a, b, c, d) \in B^4$ is an additive quadruple of ϕ, if $a + b = c + d$ and $\phi(a) + \phi(b) = \phi(c) + \phi(d)$.

14.2.2 Theorems that have been proved previously

We take the following results for granted.

Theorem 3. *[G, Corollary 8] If $A \subset \mathbb{Z}_N$ is a quadratically η-uniform set with $|A| = \delta N$ where $\eta \leq 2^{-208}\delta^{112}$ and $N > 200\delta^{-3}$, then A contains an arithmetic progression of length four.*

Theorem 4. *[G, Proposition 9] Let $\alpha > 0$, let $f : \mathbb{Z}_N \to D$, let $B \subset \mathbb{Z}_N$, and let $\phi : B \to \mathbb{Z}_N$ be a function such that*

$$\sum_{k \in B} |\Delta(f;k)\tilde{\ }(\phi(k))|^2 \geq \alpha N^3.$$

Then ϕ has at least $\alpha^4 N^3$ additive quadruples.

Theorem 5. *[G, Corollary 14] Let $B \subset \mathbb{Z}_N$ with $|B| = \beta N$, and let $\phi : B \to \mathbb{Z}_N$ be a function with at least $c_0 N^3$ additive quadruples. Then there are constants γ and η depending only on β and c_0, a mod-N arithmetic progression $P \subset \mathbb{Z}_N$ with $|P| \geq N^\gamma$ and a linear function $\psi : P \to \mathbb{Z}_N$ such that $\phi(s)$ is defined and equal to $\psi(s)$ for at least $\eta|P|$ values of $s \in P$.*
Moreover, there is an absolute constant K such that we can take $\gamma = c_0^K$ and $\eta = \exp(-(1/c_0)^K)$.

14.3 Sketch of the proof of Theorem 1

Let $N \geq \exp\exp\exp((1/\delta)^C)$, $A \subset \mathbb{Z}_N$ with $|A| = \delta N$, and suppose that A does not contain an arithmetic progression of length 4.
By Theorem 3, A is not quadratically $2^{-208}\delta^{112}$-uniform. Let $\alpha = 2^{-208}\delta^{112}$, and let f denote the balanced function of A. An equivalent formulation of the notion of quadratic uniformity (see [G, Lemma 2]) can be used to check the following. Since A is not quadratically α-uniform, there is a set $B \subset \mathbb{Z}_N$

with $|B| \geq \alpha N/2$ and a function $\phi : B \to \mathbb{Z}_N$ such that $|\Delta(f;k)\tilde{}(\phi(k))| \geq (\alpha/2)^{1/2}N$ for every $k \in B$. In particular,

$$\sum_{k\in B} |\Delta(f;k)\tilde{}(\phi(k))|^2 \geq (\alpha/2)^2 N^3.$$

By Theorem 4, the above implies that ϕ has at least $(\alpha/2)^8 N^3$ additive quadruples. Therefore, the conditions of Theorem 5 are satisfied. We derive that there is an arithmetic progression $P \subset \mathbb{Z}_N$ and a linear function $\psi : P \to \mathbb{Z}_N$ such that

$$\sum_{k\in P} |\Delta(f;k)\tilde{}(\psi(k))|^2 \geq \alpha/2N^2|B \cap P| \geq \alpha\eta/2N^2|P|$$

We now use the following theorem.

Theorem 6. *[G, Proposition 15] Let f be the balanced function of a set $A \subset \mathbb{Z}_N$. Let $P \subset \mathbb{Z}_N$ be an arithmetic progression with $|P| = T$. Suppose that there exist λ and μ such that*

$$\sum_{k\in P} |\Delta(f;k)\tilde{}(\lambda k + \mu)|^2 \geq \beta N^2 T.$$

Then there exist quadratic polynomials $\psi_0, \ldots, \psi_{N-1}$ such that

$$\sum_s \left| \sum_{z\in P+s} f(z)\omega^{-\psi_s(z)} \right| \geq \beta NT/\sqrt{2}.$$

In fact, the following slightly different inequality is also proved. There exist quadratic polynomials $\psi_0, \ldots, \psi_{N-1}$ such that for each s,

$$\left| \sum_{z\in P+s} f(z)\omega^{-\psi_s(z)} \right| \geq \gamma(s)T/\sqrt{2}$$

for some $\gamma(s)$ with $\sum_s \gamma(s) \geq \beta N$.
Lastly, we will use the following statement.

Theorem 7. *[G, Corollary 19] Let $\psi : \mathbb{Z}_N \to \mathbb{Z}_N$ be a quadratic polynomial and let $r \leq N$. There exists $m \leq Cr^{1-1/128}$ (where C is an absolute constant) and a partition $P_1, \ldots, P_m$ of $\{0, 1, \ldots, r-1\}$ such that each P_j is* *an arithmetic progression, the sizes of the* P_j *differ by at most* 1*, and if* $f : \mathbb{Z}_N \to D$ *is any function such that*

$$\left|\sum_{x=0}^{r-1} f(x)\omega^{-\psi(x)}\right| \geq \alpha r,$$

then

$$\sum_{j=1}^{m}\left|\sum_{x\in P_j} f(x)\right| \geq \alpha r/2.$$

Using Theorem 7 for each of the above $P + s$ and ψ_s as well as summing over s, we obtain the following. We can partition each $P + s$ into further progressions $P_{s1}, \ldots, P_{sm}$ of cardinalities differing by at most 1 and all at least $cT^{1/128}$, such that

$$\sum_s \sum_{j=1}^{m}\left|\sum_{x\in P_j} f(x)\right| \geq \beta NT/(2\sqrt{2}),$$

where $\beta = \exp(-(1/\delta)^K)$ for some absolute constant K.
Using the definition of the balanced function f, one can then derive that there exist s and j such that $|P_{sj}| \geq c\beta T^{1/256}$ and $|A \cap P_{sj}| \geq (\delta + c_2\beta)|P_{sj}|$. We now repeat the argument, replacing A and $\{0, 1, \ldots, N-1\}$ by $A \cap P_{sj}$ and P_{sj}, and iterate this process. Since the density of the restriction of A goes up by a multiplicative factor of at least $(1 + c_2\beta)$, one can compute that the process can be repeated at most $r = \exp((1/\delta)^K)$ times. In the replacement process, N is replaced by N^θ, where $\theta = \delta^K$. Therefore, we get that the theorem is proved if N^{θ^r} is sufficiently large. By Theorem 3, we need $N^{\theta^r} \geq 200\delta^{-3}$. A small calculation shows that the theorem follows since $N \geq \exp\exp\exp((1/\delta)^C)$.

References

[F] Furstenberg, H., *Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.

[G] Gowers, W. T., *A new proof of Szemerédi's theorem for arithmetic progressions of length four.* GAFA **8** (1998) no. 3, 529–551.

[R] Roth, K.F., *On certain sets of integers*, J. London Math. Soc. **28** (1953), 245–252.

[SZ] Szemerédi, E., *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 299–345.

Kornélia Héra, Bonn
email: herakornelia@gmail.com

15 Quantitative bounds in the nonlinear Roth theorem, part I

After S. Peluse and S. Prendiville [PP22]

A summary written by Guo-Dong Hong

Abstract

We present the quantitative bound for the nonlinear Roth configuration, which is after Bourgain and Chang [BC17], over $\mathbb{Z}$ in [PP22]. In part I, we will highlight an important technique: the degree lowering method, to implement the density increment argument.

15.1 Introduction

Bergelson and Leibman [BL96] studied the polynomial progression over $\mathbb{Z}$ for sets with positive upper density, while it leaves the question of obtaining the quantitative bound for such polynomial progression. Bourgain and Chang [BC17] studied one specific polynomial progression, or nonlinear Roth configuration, over the finite field. However, it seems their method cannot be easily generalized to the integer setting.
The goal in this paper [PP22] is to obtain the first quantitative bound for this specific configuration over $\mathbb{Z}$:

Theorem 1. *If $A \subset [N]$ does not contain the following nonlinear Roth configuration*

$$x,\ x+y,\ x+y^2 \quad (y \neq 0),$$

then $|A| \ll N(log\, log N)^{-c}$ for some constant $c > 0$.

The approach in [PP22] is the density increment argument. However, there are certain difficulties to be overcome when one considers the nonlinear configuration over $\mathbb{Z}$, and we will discuss more in the subsequent sections.

15.2 Density increment argument

The main idea in the density argument is that if a subset $A \subset [N]$ is large, then one can find a sub-progression $P \subset [N]$ in which A has an increased density. After the proper affine transformation, we can run the argument again if P is still large. However, this process can only proceed within a finite time since the density cannot exceed 1. Therefore, this means up to a certain step, the subset we are considering is no longer large, and this gives us the information on the size of the original set A.

Lemma 2. *If $A \subset [N]$ has the density at least δ and does not contain the following general nonlinear Roth configuration*

$$x,\ x+y,\ x+qy^2 \quad (y \neq 0),$$

then at least one of the following situations happens:

- $N \ll q^3\delta^{-O(1)}$
- *there exists $q' \ll \delta^{-O(1)}$ and $N' \gg \delta^{O(1)}q^{-3/2}N^{1/2}$ such that A has the increased density $\delta + \Omega(\delta^{O(1)})$ when restricted in the progression $\{a + qq' \cdot [N']\}$ for some $a \in [N]$.*

Once we have this density increment lemma, the main theorem then follows immediately.

15.3 Inverse theorem for nonlinear Roth

In order to obtain the needed density increment lemma, we need the following strong inverse theorem for the nonlinear Roth configuration:

Theorem 3. *Let $\{f_i : \mathbb{Z} \to \mathbb{C}\}_{i=0}^2$ be 1-bounded functions with supp$(f_i) \subset [N]$. If*

$$\left|\sum_{x\in\mathbb{Z}}\sum_{y\in\mathbb{N}} f_0(x)f_1(x+y)f_2(x+qy^2)\right| \geq \delta \sum_{x\in\mathbb{Z}}\sum_{y\in\mathbb{N}} 1_{[N]}(x)1_{[N]}(x+y)1_{[N]}(x+qy^2),$$

then at least one of the following situations happens:

- $N \ll q^3\delta^{-O(1)}$

- *there exists* $q' \ll \delta^{-O(1)}$ *and* $N' \gg \delta^{O(1)} q^{-3/2} N^{1/2}$ *such that for any* $i = 0, 1, 2$ *we have*

$$\sum_{x \in \mathbb{Z}} \left| \sum_{y \in \mathbb{N}'} f_i(x + q'qy) \right| \gg \delta^{O(1)} N N'$$

Motivated by Gowers's work in [G98] and [G01], Prendiville in [P17] showed that the following two ingredients are enough to deduce the density increment lemma:

- Local von Neumann theorem
- Modified Gowers's local inverse theorem

The first ingredient is how to use Gowers norm to control the counting operator, and Prendiville was able to use ideas from [BL96] to control the homogeneous polynomial progressions, even for the nonlinear Roth configuration. However, the inverse theorem for Gowers norm is only able to deal with the homogeneous polynomial progressions case. Therefore, the degree lowering method is developed to substitute the use of the inverse theorem for higher-order Gowers norm.
When we use the degree lowering method, some difficulties occur in adapting Prendiville's local von Neumann theorem. Hence, we need another variant of this local von Neumann theorem, which will be discussed more in part II.
For the sake of completeness, we include the result below:

Theorem 4. *Let* $\{f_i : \mathbb{Z} \to \mathbb{C}\}_{i=0}^2$ *be 1-bounded functions with* $supp(f_i) \subset [N]$*. If*

$$\left| \sum_{x \in \mathbb{Z}} \sum_{y \in \mathbb{N}} f_0(x) f_1(x+y) f_2(x+qy^2) \right| \geq \delta \sum_{x \in \mathbb{Z}} \sum_{y \in \mathbb{N}} 1_{[N]}(x) 1_{[N]}(x+y) 1_{[N]}(x+qy^2)$$

then at least one of the following situations happens:

- $N \ll q$
- $\sum_{u \in [q]} \|F\|_{U^5(u+q\cdot\mathbb{Z})}^{2^5} \gg \delta^{O(1)} \sum_{u \in [q]} \|1_{[N]}\|_{U^5(u+q\cdot\mathbb{Z})}^{2^5}$

15.4 Degree lowering method

Degree lowering method was originated from [P19] in order to study the polynomial progression in the finite field. As we mentioned in the previous section, degree lowering method can be regarded as a substitution for the inverse theorem of higher-order Gowers's norm.
To be more precise, one can control the U^s-Gowers norm by the U^{s-1}-Gowers norm with the unerstanding of the two-term progression and the only inverse theorem used in this approach is the U^2-inverse theorem, which is comparatively easier to understand.
However, unlike the finite field case in [P17], in order to generalize the degree lowering method to the integer setting, we had to pass to the "dual" formulation with the help of Cauchy-Schwarz inequality.

Lemma 5. *Let $\{f_i : \mathbb{Z} \to \mathbb{C}\}_{i=0}^2$ be 1-bounded functions with $supp(f_i) \subset [N]$. Define the dual function*

$$F(x) \coloneqq \mathbb{E}_{y\in[M]} f_0(x - qy^2) f_1(x + y - qy^2),$$

where $M = \sqrt{N/q}$.
If for $s \geq 3$, we have

$$\sum_{u\in[q]} \|F\|_{U^s(u+q\cdot\mathbb{Z})}^{2^s} \geq \delta \sum_{u\in[q]} \|1_{[N]}\|_{U^s(u+q\cdot\mathbb{Z})}^{2^s},$$

then at least one of the following situations happens:

- $N \ll_s q^3 \delta^{-O_s(1)}$
- $\sum_{u\in[q]} \|F\|_{U^{s-1}(u+q\cdot\mathbb{Z})}^{2^{s-1}} \gg_s \delta^{O_s(1)} \sum_{u\in[q]} \|1_{[N]}\|_{U^{s-1}(u+q\cdot\mathbb{Z})}^{2^{s-1}}$,

Repeating this process, with the control for the U^5-Gowers norm from the variant of the local von Neumann theorem, will give us the control for the U^1-Gowers norm in the end, and this is the desired inverse theorem for the nonlinear Roth configuration.

References

[BC17] J. Bourgain and M.-C. Chang. *Nonlinear Roth type theorems in finite fields.* Israel J. Math., 221(2):853–867, 2017.

[BL96] V. Bergelson and A. Leibman. *Polynomial extensions of van der Waerden's and Szemerédi's theorems.* J. Amer. Math. Soc., 9(3):725–753, 1996.

[G98] W. T. Gowers. *A new proof of Szemerédi's theorem for arithmetic progressions of length four.* Geom. Funct. Anal., 8(3):529–551, 1998.

[G01] W. T. Gowers. *A new proof of Szemerédi's theorem.* Geom. Funct. Anal., 11(3):465–588, 2001.

[P17] S. Prendiville. *Quantitative bounds in the polynomial Szemerédi theorem: the homogeneous case.* Discrete Anal., pages 34, Paper No. 5, 2017.

[P19] S. Peluse. *On the polynomial Szemerédi theorem in finite fields.* Duke Math. J., 168(5):749–774, 04 2019.

[PP22] S. Peluse and S. Prendiville. *Quantitative bounds in the nonlinear Roth theorem.* arXiv:1903.02592.

Guo-Dong Hong, Caltech
email: ghong@caltech.edu

16 Norm-variation of ergodic averages with respect to two commuting transformations

After P. Durcik, V. Kovač, K.A. Škreb, and C. Thiele[1]

A summary written by Martin Hsu, Fred Lin

Abstract

We present a quantitative result on the norm convergence of double ergodic averages with respect to two commuting transformations. In [1], the authors first reduce the estimate of a discrete model to the estimate of a continuous model. Then the estimate of the continuous model can be done with the aid of the twisted technology, a method which first developed to estimate some multilinear singular integrals with entangled structure.

16.1 Introduction

Let $(X, \mathcal{F}, \mu)$ be a σ-finite measure space and let $S, T : X \to X$, be two commuting measure-preserving transformations. For two measurable functions f, g on X and a positive integer n we define the double ergodic average:

$$M_n(f,g)(x) := \frac{1}{n} \sum_{i=0}^{n-1} f(S^i x) g(T^i x). \tag{1}$$

In [2], Conze and Lesigne show the L^2 convergence of the sequence of double ergodic averages $\{M_n(f,g)\}_{n\in\mathbb{N}}$ on a probability space $(X, \mathcal{F}, \mu)$ for functions $f, g \in L^\infty(X)$. The main result in this paper [1] is to quantify such convergence through a norm-variation estimate on $\{M_n(f,g)\}_{n\in\mathbb{N}}$.

Theorem 1. *For every choice of increasing sequence $n_0 < n_1 < \cdots < n_j < \cdots$, we have the following bound*

$$\sum_{j=1}^{m} \left\| M_{n_j}(f,g) - M_{n_{j-1}}(f,g) \right\|_{L^2(X)}^2 \lesssim \|f\|_{L^4(X)}^2 \|g\|_{L^4(X)}^2 \tag{2}$$

with the implicit constant independent of m and the choice of the sequence.

The proof of **Theorem 1.** can be summarized as follow:

- **Three main reduction steps** to a continuous model
- **Two key estimates** on the continuous model via twisted technology

16.2 Reduction to Continuous Model

The series of reduction steps go through several models, including the following: for $\widetilde{F}, \widetilde{G} \in \ell^4\left(\mathbb{Z}^2\right)$, $F, G \in L^4\left(\mathbb{R}^2\right)$, and $\varphi \in L^1(\mathbb{R})$, we define

- **Discrete Bilinear Average:**

$$\widetilde{A}_n\left(\widetilde{F}, \widetilde{G}\right)(k,l) := \frac{1}{n}\sum_{i=0}^{n-1} \widetilde{F}(k+i,l)\,\widetilde{G}(k,l+i) \tag{3}$$

- **Continuous Bilinear φ-Average:**

$$A_t^{\varphi}(F,G)(x,y) := \int_{\mathbb{R}} F(x+s,y)\,G(x,y+s)\,t^{-1}\varphi\left(t^{-1}s\right)ds \tag{4}$$

We provide a sketch of the reduction process among the four different models:

$$M_n(f,g) \overset{\mathbf{I}}{\rightsquigarrow} \widetilde{A}_n\left(\widetilde{F}, \widetilde{G}\right) \overset{\mathbf{II}}{\rightsquigarrow} A_t^{\mathbb{1}_{[0,1)}}(F,G) \overset{\mathbf{III}}{\rightsquigarrow} A_t^{\varphi}(F,G) \text{ with } \varphi \in S(\mathbb{R}). \tag{5}$$

- **Step I:** We reinterpret the action of T, S on a fixed reference point $x \in X$ as two independent shifts on the integer grid $\mathbb{Z}^2$ by considering the following two double sequences:

$$\widetilde{F}_x(k,l) \approx f\left(T^k S^l x\right) \text{ and } \widetilde{G}_x(k,l) \approx g\left(T^k S^l x\right). \tag{6}$$

  This allows us to pass the norm-variation estimate on $\{M_n(f,g)\}_{n\in\mathbb{N}}$ to the corresponding estimate on $\left\{\widetilde{A}_n\left(\widetilde{F}_x, \widetilde{G}_x\right)\right\}_{n\in\mathbb{N}}$.

- **Step II:** We perform two parallel changes of variables:

$$\begin{cases} \widetilde{A}_n\left(\widetilde{F}, \widetilde{G}\right)(k,l) = \dfrac{1}{n}\displaystyle\sum_{i=k+l}^{k+l+n-1} \widetilde{F}(i-l,l)\,\widetilde{G}(k,i-k) \\ A_t^{\mathbb{1}_{[0,1)}}(F,G)(x,y) = \dfrac{1}{t}\displaystyle\int_{x+y}^{x+y+t} F(s-y,y)\,G(x,s-x)\,ds \end{cases} \tag{7}$$

  The two similar formulations suggest that we set:

$$\begin{cases} F(s-y,y) := \displaystyle\sum_{i,l\in\mathbb{Z}} \widetilde{F}(i-l,l)\,\mathbb{1}_{[i,i+1)\times[l,l+1)}(s,y) \\ G(x,s-x) := \displaystyle\sum_{i,k\in\mathbb{Z}} \widetilde{G}(k,i-k)\,\mathbb{1}_{[i,i+1)\times[k,k+1)}(s,x) \end{cases} \tag{8}$$

to derive the following approximation.

$$A_n^{\mathbb{1}_{[0,1)}}(F,G)(x,y) \approx \sum_{k,l\in\mathbb{Z}} \widetilde{A}_n\left(\widetilde{F},\widetilde{G}\right)(k,l)\cdot \mathbb{1}_{[k,k+1)\times[l,l+1)}(x,y) \qquad (9)$$

This allows us to pass the norm-variation estimate on $\left\{\widetilde{A}_n\left(\widetilde{F},\widetilde{G}\right)\right\}_{n\in\mathbb{N}}$ to the corresponding estimate on $\left\{A_t^{\mathbb{1}_{[0,1)}}(F,G)\right\}_{t\in\mathbb{R}_+}$.

- **Step III:** This step causes all the technicality regarding the decay control on φ. In short, we perform a **Littlewood-Paley** decomposition on $\mathbb{1}_{[0,1)}$ to separate information with different regularity control:

$$\mathbb{1}_{[0,1)} = \mathbb{1}_{[0,1)} * \chi + \sum_{k<0} \mathbb{1}_{[0,1)} * \theta_{2^k}. \qquad (10)$$

With careful book-keeping on the dependency of the norm-variation of $\{A_t^{\varphi}(F,G)\}_{t\in\mathbb{R}_+}$ on $\varphi = \mathbb{1}_{[0,1)} * \chi$ and $\varphi = \mathbb{1}_{[0,1)} * \theta_{2^k}$, we derive bounds that are summable over k. Via triangle inequality, we have the desired estimate on the norm-variation of $\left\{A_t^{\mathbb{1}_{[0,1)}}(F,G)\right\}_{t\in\mathbb{R}_+}$ and thus, complete the reduction step.

16.3 Key Estimates: Long and Short Variation

We aim to derive the analogous statement of **Theorem 1.** for $\{A_t^{\varphi}(F,G)\}_{t\in\mathbb{R}_+}$:

Theorem 2. *For a fixed Schwartz function $\varphi \in S(\mathbb{R})$ and an arbitrary chosen increasing sequence $t_0 < t_1 < \cdots < t_j < \cdots$ in $\mathbb{R}_+$, we have*

$$\sum_{j=1}^{m} \left\| A_{t_j}^{\varphi}(F,G) - A_{t_{j-1}}^{\varphi}(F,G)\right\|_{L^2(\mathbb{R}^2)}^2 \lesssim_{\varphi} \|F\|_{L^4(\mathbb{R}^2)}^2 \|G\|_{L^4(\mathbb{R}^2)}^2 \qquad (11)$$

with the implicit constant independent of m and the choice of the sequence ***but dependent on φ in a controlled manner.***

A standard procedure for norm-variation estimates is to divide the analysis into two parts: **Long variation** and **Short variation**. Roughly speaking, one can decompose the sequence into **dyadic segments**:

$$2^{k_i-1} < \cdots < t_{j-1} < t_j < \cdots \leq 2^{k_i}. \tag{12}$$

- **Long variation** controls the long jumps across dyadic segments by measuring the jumps in the lacunary sequence $\left\{2^{k_i}\right\}_{i=0}^{\infty}$:

  Lemma 3. *For a fixed Schwartz function $\varphi \in S(\mathbb{R})$ and an arbitrary chosen increasing sequence $k_0 < k_1 < \cdots < k_i < \cdots$ in $\mathbb{Z}$, we have*

$$\sum_{i=1}^{m} \left\| A^{\varphi}_{2^{k_i}}(F,G) - A^{\varphi}_{2^{k_{i-1}}}(F,G) \right\|^2_{L^2(\mathbb{R}^2)} \lesssim_{\varphi} \|F\|^2_{L^4(\mathbb{R}^2)} \|G\|^2_{L^4(\mathbb{R}^2)} \tag{13}$$

  with the implicit constant independent of m and the choice of the sequence ***but dependent on φ in a controlled manner.***

- **Short variation** controls the overall effect of the short jumps within the dyadic segments by summing over all norm-variation estimates within the dyadic segments:

  Lemma 4. *For a fixed Schwartz function $\varphi \in S(\mathbb{R})$ and a collection of increasing sequence*

$$2^{k-1} < t_0^{(k)} < \cdots < t_{j-1}^{(k)} < t_j^{(k)} < \cdots < t_{m_k}^{(k)} \leq 2^k \text{ for } k \in \mathbb{Z}, \tag{14}$$

  we have the following estimate:

$$\left\| \left\| A^{\varphi}_{t_j^{(k)}}(F,G) - A^{\varphi}_{t_{j-1}^{(k)}}(F,G) \right\|_{\ell^2(k,j)} \right\|_{L^2(\mathbb{R}^2)} \lesssim_{\varphi} \|F\|^2_{L^4(\mathbb{R}^2)} \|G\|^2_{L^4(\mathbb{R}^2)} \tag{15}$$

  with the implicit constant independent of m and the choice of the sequence ***but dependent on φ in a controlled manner.***

The proof of the above-mentioned estimates mainly relies on the **Twisted Technology**. The rest of the technicality arises from all the bookkeeping on the dependency of the estimates on the Schwartz function φ.

References

[1] DURCIK, P., KOVAČ , V., ŠKREB, K. & THIELE, C. Norm variation of ergodic averages with respect to two commuting transformations. *Ergodic Theory And Dynamical Systems*. **39**, 658-688 (2017,8), https://doi.org/10.1017

[2] Conze, J. & Lesigne, E. Théoremes ergodiques pour des mesures diagonales. *Bulletin De La Société Mathématique De France*. **112** pp. 143-175 (1984)

Martin Hsu, Purdue University
email: `hsu263@purdue.edu`
Fred Lin, University of Bonn
email: `fredlin@math.uni-bonn.de`

17 Nonconventional ergodic averages and nilmanifolds

After B. Host and B. Kra [HK]

A summary written by Henrik Kreidler

Abstract

We discuss a proof for the convergence of nonconventional ergodic averages via a structure theorem for certain factors of a measure-preserving system.

Motivated by Furstenberg's seminal work [F] on an ergodic theoretic approach to Szemerédi's theorem on arithmetic progressions, the authors study, for a measure-preserving automorphism $\varphi_X \colon X \to X$ of a probability space X and $k \in \mathbb{N}$, the asymptotic behavior of the nonconventional ergodic averages

$$\frac{1}{N}\sum_{n=0}^{N-1} T_{\varphi_X}^{n} f_1 \cdots T_{\varphi_X}^{kn} f_k$$

for $f_1, \ldots, f_k \in \mathrm{L}^\infty(X)$ as $N \to \infty$, where $T_{\varphi_X} f = f \circ \varphi_X$ for $f \in \mathrm{L}^\infty(X)$. To prove convergence of these means for every measure-preserving system $\mathbf{X} = (X, \varphi_X)$, they use the classical idea to orthogonally decompose $\mathrm{L}^2(X)$ into a structured part, which can be identified with the L^2-space of a well-understood factor $\mathbf{Z}_k$ of $\mathbf{X}$, and a stable part, for which an application of the so-called van der Corput inequality (see [HK, Appendix D]) shows convergence to zero. In this case, $\mathbf{Z}_k$ is called a **characteristic factor**. Earlier results (see, e.g., [L]) show convergence for a nilsystem $\mathbf{X}$, i.e., $\mathbf{X}$ is given as the homogeneuous space $X = G/H$ of a nilpotent Lie group G modulo a discrete cocompact subgroup H with the (normalized) Haar measure, and $\varphi_X \colon G/H \to G/H$, $xH \mapsto axH$ is the rotation by a fixed element $a \in G$. Using ergodic decomposition and approximation, it is therefore enough to show that every ergodic measure-preserving system $\mathbf{X}$ admits a characteristic factor $\mathbf{Z}_k$ which is an inverse limit of nilsystems. In their article [HK] (and their book [HK2]), Host and Kra construct such factors (these even govern the asymptotic behavior of other ergodic averages).

To explain their construction, start from a dynamical system $\mathbf{X}$ and consider its invariant factor $\mathbf{X}_{\mathrm{inv}}$, i.e., the largest factor[5] of $\mathbf{X}$ on which the dynamics are trivial. We then form the relatively independent joining[6] $\mathbf{X}^{[1]} \coloneqq \mathbf{X} \times_{\mathbf{X}_{\mathrm{inv}}} \mathbf{X}$ with respect to the invariant factor. One can then apply the same construction to the system $\mathbf{X}^{[1]}$. By iterating, we arrive at the **Host-Kra cubes**

$$\mathbf{X}^{[k]} \coloneqq \mathbf{X}^{[k-1]} \times_{\mathbf{X}^{[k-1]}_{\mathrm{inv}}} \mathbf{X}^{[k-1]} \text{ for } k \in \mathbb{N} \quad \text{and} \quad \mathbf{X}^{[0]} \coloneqq \mathbf{X}.$$

The coordinates of its elements are indexed over the set $\{0,1\}^k$. Note that the construction is functorial: Every factor map $\pi\colon \mathbf{X} \to \mathbf{Y}$ of measure-preserving systems gives rise to a factor map $\pi^{[k]}\colon \mathbf{X}^{[k]} \to \mathbf{Y}^{[k]}$.
A short definition of the kth **Host-Kra factor** $\mathbf{Z}_k$ of an ergodic system $\mathbf{X}$ for $k \in \mathbb{N}_0$ is now the following: It is the smallest factor $\mathbf{Y}$ of $\mathbf{X}$ for which the factor map $\pi^{[k]}$ is **relatively ergodic** meaning that on the level of invariant factors $\pi^{[k]}$ defines an isomorphism from $(\mathbf{X}^{[k]})_{\mathrm{inv}}$ to $(\mathbf{Y}^{[k]})_{\mathrm{inv}}$.[7]
Note that, by ergodicity of $\mathbf{X}$, the factor $\mathbf{Z}_0$ is trivial.
A further short, but equivalent definition uses the **Host-Kra-seminorms**, inspired by seminorms introduced earlier by Gowers in [G]: For $f \in \mathrm{L}^\infty(X)$ we form the product $f^{[k]} \coloneqq \bigotimes_{\epsilon\in\{0,1\}^k} C^{|\epsilon|}(f) \in \mathrm{L}^\infty(X^{[k]})$ where $C(f) \coloneqq \overline{f}$ and $|\epsilon| = \sum_{j=1}^k \epsilon_j$ for $\epsilon \in \{0,1\}^k$. We then set

$$\|f\|_k \coloneqq \left(\int_{X^{[k]}} f^{[k]} \,\mathrm{d}\mu^{[k]}\right)^{\frac{1}{2^k}}$$

where $\mu^{[k]}$ denotes the measure of the kth cube $\mathbf{X}^{[k]}$. One can check that $\|f\|_k > 0$ holds precisely when the conditional expectation $\mathbb{E}_{\mathbf{Z}_{k-1}} f$ to the factor $\mathbf{Z}_{k-1}$ is non-zero (see [HK, Lemma 4.3]).
The Host-Kra factors form an increasing sequence

$$\mathbf{Z}_0 \leftarrow \mathbf{Z}_1 \leftarrow \mathbf{Z}_2 \leftarrow \mathbf{Z}_3 \leftarrow \cdots \leftarrow \mathbf{X},$$

see [HK, Corollary 4.4], and the system $\mathbf{X}$ is said to have **order** k if it agrees with its kth Host-Kra factor $\mathbf{Z}_k$. The first major structural result

[5]Strictly speaking this is only determined up to isomorphism, but there is a canonical choice.

[6]One can think of this as the measure-theoretic replacement for a fiber product.

[7]It is a priori not clear that such a factor exists. In fact, Host and Kra use a different definition for their factors which is equivalent by [HK, Proposition 4.7].

now provides a representation of a factor from this sequence with respect to its predecessor (see [HK, Subsection 2.1 and remarks after Definition 4.1] for $k = 1$ and [HK, Proposition 6.3] for $k \geq 2$).

Proposition 1. *For every $k \in \mathbb{N}$ the factor map $\mathbf{Z}_k \to \mathbf{Z}_{k-1}$ is an abelian group extension: There is a compact abelian group U and a measurable map $\varrho\colon Z_{k-1} \to U$ such that $\mathbf{Z}_k$ is a skew-product*
$\mathbf{Z}_{k-1} \rtimes_\varrho \mathbf{U} = (Z_{k-1} \times U, \varphi_{Z_{k-1}} \rtimes \varrho)$ where
$(\varphi_{Z_{k-1}} \rtimes \varrho)(z,u) = (\varphi_{Z_{k-1}}(z), \varrho(z)u)$ for $(z,u) \in Z_{k-1} \times U$.

In particular, $\mathbf{Z}_1$ is given by a rotation of a compact abelian group. Since every such group can be writen as a projective limit of factors which are compact abelian Lie groups, $\mathbf{Z}_1$ is an inductive limit of nilsystems of 1-step nilpotent Lie groups.
The idea is now to prove the structure theorem inductively where the order of the involved nilpotent Lie groups is allowed to grow by one in each step. To do so, we have to make further progress on the structure of the factor maps $\mathbf{Z}_k \to \mathbf{Z}_{k-1}$ between consecutive factors. For this we need some cohomological considerations.
For a measure-preserving system $\mathbf{X}$ and a compact abelian group U we write $\mathrm{C}(\mathbf{X}, U)$ for the equivalence classes of measurable maps $\varrho\colon X \to U$ (with two such maps being equivalent if they agree almost everywhere). The elements $\varrho \in \mathrm{C}(\mathbf{X}, U)$ are called **cocycles**. Given $\varrho \in \mathrm{C}(\mathbf{X}, U)$ we can form the associated **coboundary** $\partial\varrho \coloneqq (\varrho \circ \varphi_X) \cdot \varrho^{-1} \in \mathrm{C}(\mathbf{X}, U)$ and we write $\partial\mathrm{C}(\mathbf{X}, U)$ for the subgroup of $\mathrm{C}(\mathbf{X}, U)$ of all coboundaries arising in this way. The quotient group

$$\mathrm{H}^1(\mathbf{X}, U) \coloneqq \mathrm{C}(\mathbf{X}, U)/\partial\mathrm{C}(\mathbf{X}, U)$$

is the **first cohomology group** of $\mathbf{X}$ with respect to the group U.[8] Two cocycles $\varrho_1, \varrho_2\colon X \to U$ with $[\varrho_1] = [\varrho_2]$ in $\mathrm{H}^1(\mathbf{X}, U)$ define isomorphic skew-products $\mathbf{X} \rtimes_{\varrho_1} \mathbf{U}$ and $\mathbf{X} \rtimes_{\varrho_2} \mathbf{U}$.
From a cocycle $\varrho \in \mathrm{C}(\mathbf{X}, U)$ we can construct a cocycle $\Delta\varrho \in \mathrm{C}(\mathbf{X}^{[1]}, U)$ via $\Delta\varrho(x_0, x_1) = \varrho(x_0)\varrho(x_1)^{-1}$ for $(x_0, x_1) \in X^{[1]}$. We can apply the same procedure to cocycles of $\mathbf{X}^{[k]}$ for $k \in \mathbb{N}_0$ and then obtain a chain of group

[8]We remark that Host and Kra in [HK] do not explicitly introduce the cohomology group, but it will be convenient to do so in this summary. Also, in the article, abelian groups are usally written additively, whereas here multiplicative notation is used.

homomorphisms

$$\mathrm{H}^1(\mathbf{X},U) \to \mathrm{H}^1(\mathbf{X}^{[1]},U) \to \mathrm{H}^1(\mathbf{X}^{[2]},U) \to \mathrm{H}^1(\mathbf{X}^{[3]},U) \to \cdots$$

For $k \in \mathbb{N}$ the composition of the first k of these maps is explicitly given by

$$\Delta^k \colon \mathrm{H}^1(\mathbf{X},U) \to \mathrm{H}^1(\mathbf{X}^{[k]},U), \quad [\varrho] \mapsto \left[\prod_{\epsilon \in \{0,1\}^k} (\varrho \circ \mathrm{pr}_\epsilon)^{(-1)^{|\epsilon|}} \right]$$

where $\mathrm{pr}_\epsilon \colon X^{[k]} \to X$ is the projection onto the component of $\epsilon \in \{0,1\}^k$. A cocycle $\varrho \in \mathrm{C}(\mathbf{X},U)$ which is trivialized by this homomorphism, i.e., $\Delta^k[\varrho] = [1]$, is **of type k** (see [HK, Definition 7.1]). The following is a consequence of [HK, Proposition 6.4].

Proposition 2. *For every $k \in \mathbb{N}$ the factor map $\mathbf{Z}_k \to \mathbf{Z}_{k-1}$ is an abelian group extension by a cocycle of type k.*

Proposition 2 is a crucial observation for the structure theorem representing the Host-Kra factors $\mathbf{Z}_k$. We will sketch the proof of this representation theorem in the case $k = 2$ which is also discussed in [HK4, Section 4.5], see also [JST] (the result for a general $k \in \mathbb{N}$ requires much more work, but its proof follows similar steps). The idea is to first approximate a system of order 2 by the following particularly nice ones (see [HK, Definition 8.5]).

Definition 3. *A system $\mathbf{X}$ of type 2 is **toral** if $\mathbf{Z}_1$ is a compact abelian Lie group and $\mathbf{Z}$ is a group extension of $\mathbf{Z}_1$ by a torus and a cocycle of type 2.*

To obtain the desired approximation, one first proves the following additional result on the structure group (see [HK4, Corollary 8.4]).

Proposition 4. *The factor map $\mathbf{Z}_2 \to \mathbf{Z}_1$ is an extension by a* connected *compact abelian group and a cocycle of type 2.*

As a consequence the extending group U can be represented as a projective limit of tori. However, in order to obtain toral systems, we also have to find an approximation for the Kronecker factor $\mathbf{Z}_1$. This is done via [HK, Lemma 8.3]:

Lemma 5. *Let $\mathbf{Z}$ be an ergodic rotation on a compact abelian group, U a torus and $\varrho \in \mathrm{C}(\mathbf{Z},U)$ a cocycle of type 2. Then there is a closed subgroup $Z_0 \subseteq Z$ such that Z/Z_0 is a Lie group, and a cocycle $\varrho' \in \mathrm{C}(\mathbf{Z}/\mathbf{Z_0},U)$ such that $[\varrho' \circ \mathrm{pr}_{Z/Z_0}] = [\varrho]$ in $\mathrm{H}^1(\mathbf{Z},U)$.*

With Proposition 4 and Lemma 5 one readily obtains the following approximation result (see [HK, Proposition 8.6]).

Proposition 6. *Every system of order* 2 *is an inverse limit of toral systems of order* 2*.*

The proof of Lemma 5 rests on a number of cohomological considerations. We highlight one of particular interest (see [HK, Lemma 8.1]). Observe here that for an ergodic rotation $\mathbf{Z}$ on a compact abelian group every $s \in Z$ induces a map

$$L_s\colon \mathrm{H}^1(\mathbf{Z}, U) \to \mathrm{H}^1(\mathbf{Z}, U), \quad [\varrho] \mapsto [L_s\varrho]$$

where $L_s\varrho(x) = \varrho(sx)$ for $x \in Z$ and $\varrho \in \mathrm{C}(\mathbf{Z}, U)$.

Proposition 7 (Conze-Lesigne equation)**.** *Consider an ergodic rotation* $\mathbf{Z}$ *on a compact abelian group, a torus* U *and a cocycle* $\varrho\colon Z \to U$ *of type* 2*. Then for every* $s \in Z$ *there is* $c \in U$ *with* $L_s[\varrho] = [c\varrho]$*.*

The equation had already been studied earlier in work of Conze and Lesigne (see [CL]) as well as Furstenberg and Weiss (see [FW]). It motivates the following definition of a group associated with a toral system.

Definition 8. *For a toral system* $\mathbf{Z} \rtimes_\varrho \mathbf{U}$ *we let* G *be the group of all skew-rotations* $s \rtimes \vartheta\colon Z \times U \to Z \times U$ *with* $L_s\varrho = c\varrho \cdot \partial\vartheta$ *for some* $c \in U$*.*[9]

Thus, the group consists of those skew-rotations which yield coboundaries "implementing the Conze-Lesigne equation". We now represent toral systems (see [HK, Lemma 8.8] and [HK3], as well as [JST, Section 4.2]).

Proposition 9. *For a toral system* $\mathbf{X}$ *the group* G*, equipped with the topology of convergence in probability, is a* 2*-step nilpotent Lie group. Moreover,* $\mathbf{X}$ *is isomorphic to a nilsystem induced by* G*.*

The combination of Propositions 6 and 9 finally yields the structure theorem for $k = 2$.

Theorem 10. *Every ergodic system of order* 2 *is an inverse limit of* 2*-step nilsystems.*

[9]For general $k \in \mathbb{N}$ the group is constructed more abstractly by using transformations of X which give rise to measure-preserving transformations of the cube $\mathbf{X}^{[k]}$ respecting the invariant factor $(\mathbf{X}^{[k]})_{\mathrm{inv}}$, see [HK, Section 5].

References

[CL] Conze, J.-P. and Lesigne, E., *Sur un théorème ergodique pour des mesures diagonales.* Publications de l'Institut de recherche mathématiques de Rennes, no. 1 (1987), 1–31.

[F] Furstenberg, H., *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions.* J. Anal. Math. **31** (1977), 204–256.

[FW] Furstenberg, H. and Weiss, B., *A mean ergodic theorem for* $\frac{1}{N}\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$. In Convergence in Ergodic Theory and Probability, De Gruyter (1996).

[G] Gowers, T., *A new proof of Szemerédi's theorem.* Geom. Funct. Anal. **11** (2001), 465–588.

[HK] Host, B. and Kra, B., *Nonconventional ergodic averages and nilmanifolds.* Ann. of Math. (2) **161** (2005), 387–488.

[HK2] Host, B. and Kra, B., *Nilpotent Structures in Ergodic Theory.* American Mathematical Society (2018).

[HK3] Host, B. and Kra, B., *Correction and addition to chapter 19.* `https://sites.math.northwestern.edu/~kra/papers/Chapter19.pdf` (accessed 08/08/2023).

[HK4] Host, B. and Kra, B., *Averaging along cubes.* In Modern Dynamical Systems and Applications, Cambridge University Press (2004).

[JST] Jamneshan, A. and Shalom, O. and Tao, T., *The structure of arbitrary Conze-Lesigne systems.* arXiv:2112.02056 (2021).

[L] Leibman, A., *Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold.* Ergodic Theory Dynam. Systems **25** (2005), 201–213.

Henrik Kreidler, University of Wuppertal
email: `kreidler@uni-wuppertal.de`

18 Quantitative bounds in the polynomial Szemerédi theorem: the homogeneous case

After S. Prendiville [5]

A summary written by Borys Kuca

Abstract

We give an exposition of the result of Prendiville that all subsets of $\{1, \ldots, N\}$ lacking ℓ-term arithmetic progressions with differences of the form n^k have at most $O(N/(\log\log N)^c)$ elements.

18.1 Introduction

The celebrated theorem of Szemerédi on arithmetic progressions has inspired numerous far-reaching generalisations. Among its most famous extensions is the following result of Bergelson and Leibman.

Theorem 1 (Polynomial Szemerédi theorem [1])**.** *Let $p_1, \ldots, p_\ell \in \mathbb{Z}[n]$ be polynomials with zero constant terms. Then each subset of $\mathbb{N}$ of positive upper density contains a progression*

$$x,\ x + p_1(n),\ \ldots,\ x + p_\ell(n) \quad \textit{for some} \quad n \neq 0.$$

After Gowers gave his quantitative proof of the Szemerédi theorem [3], the question of quantifying Theorem 1 has come onto the agenda. For a long time, no bounds were known for even the simplest polynomial configurations, as the only existing proofs of Theorem 1 relied on infinitary methods of ergodic theory. The first such bounds were obtained by Prendiville about 15 years after Gowers, and the exposition of his result is the primary goal of this chapter.

Theorem 2 (Bounds for arithmetic progressions with higher power differences [5])**.** *Let $k, \ell \in \mathbb{N}$. There exist $C, c > 0$ such that for each sufficiently large $N \in \mathbb{N}$, every subset of $[N] := \{1, \ldots, N\}$ with at least $CN/(\log\log N)^c$ elements contains*

$$x,\ x + n^k,\ \ldots,\ x + \ell n^k \quad \textit{with} \quad n \neq 0. \tag{1}$$

Prendiville's argument broadly follows Gowers' strategy with several adjustments necessary for the polynomial case. As such, it comprises three steps that consist in proving the following three statements:

1. *(Gowers norm control)* For all 1-bounded functions $f_0, \ldots, f_\ell : \mathbb{Z} \to \mathbb{C}$ supported on $[N]$, we have

$$\left| \sum_{x \in \mathbb{Z}} \sum_{n \in [N]} f_0(x) f_1(x+n) \cdots f_\ell(x+\ell n) \right| \ll \min_j \|f_j\|_{U^\ell[N]}$$

   (here, $\|\cdot\|_{U^\ell}$ is the unnormalised degree ℓ Gowers norm, $\|f\|_{U^\ell(A)} = \|f \cdot 1_A\|_{U^\ell}$ is its localisation to a set A, and $\ll$ is the Vinogradov notation[10]);

2. *(Local inverse theorem for Gowers norms)* If $f : \mathbb{Z} \to \mathbb{C}$ is 1-bounded and supported on $[N]$, and $\|f\|_{U^\ell[N]} \geq \delta N^{(d+1)/2^d}$, then one can partition $[N]$ into arithmetic progressions P_i of average length at least $c\delta^C N^{c\delta^C}$ such that

$$\sum_i \left| \sum_{x \in P_i} f(x) \right| \geq c\delta^C N.$$

3. *(Density increment)* Let $\delta > 0$ and $N \geq \exp\exp(C\delta^{-C})$. If $A \subset [N]$ of size $|A| \geq \delta N$ contains no $(\ell+1)$-term arithmetic progression with $n \neq 0$, then there exists an arithmetic progression P of length $|P| \geq N^{\exp(-C\delta^{-C})}$ such that

$$\frac{|A \cap P|}{|P|} \geq \delta + c\delta^C,$$

   i.e. A has an increased density on P.

Letting $M_1 = |P| \geq N^{\exp(-1/c\delta^C)}$, $P = \{q_1 n + r_1 :\ n \in [M_1]\}$ and $B_1 = \{n \in [M_1] :\ q_1 n + r_1 \in A\}$, the density increment allows Gowers to pass to a set $B_1 \subset [M_1]$ of increased density $|B_1|/M_1 \geq \delta + c\delta^C$. Iterating this step d times and noting that density cannot exceed 1, Gowers

[10] Meaning that $f \ll g$ if there exists $C > 0$ such that $|f(x)| \leq C|g(x)|$ for sufficiently large x.

eventually arrives at a set $B_d \subset [M_d]$ whose density on $[M_d]$ is so close to 1 that it has to contain an $(\ell+1)$-term arithmetic progression, as the density increment step can no longer be performed. Since the set B_d obtained this way takes the form $B_d = \{n \in [M_d] : \; q_d n + r_d \in A\}$ for some q_d, r_d, and the family of $(\ell+1)$-term arithmetic progressions is invariant under the affine maps $n \mapsto q_d n + r_d$, it follows that A contains an $(\ell+1)$-term arithmetic progression as well. The bound on the size of A follows from carefully estimating the size of $|P_d|$ from below.
Prendiville's modifications of steps (ii) and (iii) are straightforward, so we briefly discuss them first before moving to his more involved adaptation of step (i). By a Diophantine approximation argument, Prendiville ensures that the common differences of the progressions appearing in the local inverse theorem for Gowers norm can be taken to be k-th powers, and so can be the common difference of the progression obtained in the density increment step. Thus, the set B_d obtained in the last iteration of density increment takes the form $B_d = \{n \in [M_d] : \; q_d^k n + r_d \in A\}$ for some q_d, r_d, and if (1) is the arithmetic progression with k-th power difference lying inside B_d, then

$$q_d^k x + r_d, \; q_d^k x + r_d + (q_d n)^k, \; \ldots, \; q_d^k x + r_d + \ell (q_d n)^k$$

is the arithmetic progression with k-th power difference inside A.
Prendiville's adaptation of step (i) is the following Gowers norm estimate.

Theorem 3. *Let $k, \ell \in \mathbb{N}$. There exist $s, d \in \mathbb{N}$ and $C > 0$ such that for every $\delta > 0$, integer $N \gg \delta^{-C}$ and 1-bounded functions $f_0, \ldots, f_\ell : \mathbb{Z} \to \mathbb{C}$ supported on $[N]$, the lower bound*

$$\left| \sum_{x \in \mathbb{Z}} \sum_{n \in [N^{1/k}]} f_0(x) f_1(x + n^k) \cdots f_\ell(x + \ell n^k) \right| \gg \delta N^{1+1/k} \tag{2}$$

implies

$$\sum_{x \in \mathbb{Z}} \|f_j\|_{U^s(x+[M])} \gg \delta^C N M^{(s+1)/2^s}$$

for every $j \in \{0, \ldots, \ell\}$ and some $\delta^C N^{1/k} \ll M \ll \delta^{-C} N^{1/k}$.

The crucial difference between Theorem 3 and the analogous result of Gowers is that Prendiville has only managed to control the counting operator for (1) by an average of Gowers norms localised to intervals of size $\sim N^{1/k}$.

The proof of Theorem 3 follows from a variant of the classical PET induction argument which is the main tool used to control the counting operators of polynomial progressions by Gowers norms and relies on the following standard lemma of van der Corput.

Lemma 4 (van der Corput lemma [4, Lemma 4.1])**.** *Let $f : \mathbb{Z} \to \mathbb{C}$ be 1-bounded. Then for any $H, M \in \mathbb{N}$, we have*

$$\left| \frac{1}{M} \sum_{n \in [M]} f(n) \right|^2 \ll \frac{M+H}{M^2} \sum_{h \in \mathbb{Z}} \mu_H(h) \sum_{n \in [M] \cap ([M]-h)} f(n) \overline{f(n+h)},$$

where $\mu_H(h) = \left(1 - \frac{|h|}{H}\right)_+$ and $[M] - h := \{m - h : \ m \in [M]\}$.

The PET argument consists of a number of steps in which one replaces the counting operator for the original polynomial progression by the progression which is somehow "less complex". Each of these steps involves an application of the Cauchy-Schwarz inequality and Lemma 4 followed by a change of variables. Repeating this procedure finitely many times[11], one eventually arrives at a counting operator for a linear configuration that can be directly controlled using Gowers' estimates. We illustrate the technically involved proof of Theorem 3 in the simple case $k = \ell = 2$. Starting with (2) and letting $M = \lfloor N^{1/2} \rfloor$, we apply the Cauchy-Schwarz inequality and Lemma 4 to conclude that

$$\sum_{x \in \mathbb{Z}} |f_0(x)|^2 \cdot \sum_{x \in \mathbb{Z}} (M+H) \sum_{h_1 \in \mathbb{Z}} \mu_H(h_1) \sum_{n \in [M] \cap ([M]-h)} f_1(x+n^2) \overline{f_1(x + (n+h_1)^2)}$$
$$f_2(x + 2n^2) \overline{f_2(x + 2(n+h_1)^2)} \geq \delta^2 N^3$$

for some $H \in \mathbb{N}$ to be chosen later. The 1-boundedness of f_0 and the fact that it is supported on $[N]$ imply $\sum_{x \in \mathbb{Z}} |f_0(x)|^2 \leq N$. For the second term, we shift $x \mapsto x - n^2$, and assume $H \leq \delta^2 N^{1/2}/8$ so that the condition

[11] The number of steps, and hence also the degree of the Gowers norm obtained this way, can be bounded purely in terms of the length and degree of the original polynomial progression.

$n \in [M] \cap ([M] - h)$ can be replaced by $n \in [M]$, obtaining

$$\sum_{x\in\mathbb{Z}} \sum_{h_1\in\mathbb{Z}} \mu_H(h_1) \sum_{n\in[M]} f_1(x)\overline{f_1(x + 2h_1 n + h_1^2)}$$
$$f_2(x + n^2)\overline{f_2(x + n^2 + 4h_1 n + 2h_1^2)} \geq \delta^2 N^{3/2}/4.$$

We repeat the same procedure (the Cauchy-Schwarz inequality, Lemma 4 and the change of variables $x \mapsto x - 2h_1 n$) to remove the first f_1, getting

$$\sum_{x\in\mathbb{Z}} \sum_{h_1,h_2\in\mathbb{Z}} \mu_H(h_1, h_2) \sum_{n\in[M]} f_1(x + h_1^2)\overline{f_1(x + 2h_1 h_2 + h_1^2)} f_2(x + n^2 - 2h_1 n)$$
$$\overline{f_2(x + (n + h_2)^2 - 2h_1 n) f_2(x + n^2 + 2h_1 n + 2h_1^2)}$$
$$f_2(x + (n + h_2)^2 + 2h_1 n + 4h_1 h_2 + 2h_1^2) \geq \delta^4 N^{3/2}/64$$

as long as $H \leq \delta^4 N^{1/2}/128$ (here, $\mu_H(h_1, \dots h_s) = \mu_H(h_1) \cdots \mu_H(h_s)$). Since f_1's do not depend on n, we remove them through one more iteration of the argument, obtaining

$$\sum_{h_1,h_2,h_3\in\mathbb{Z}} \mu_H(h_1, h_2, h_3) \sum_{x\in\mathbb{Z}} \sum_{n\in[M]} \prod_{\epsilon\in\{0,1\}^3} \tilde{f}_{\epsilon,h}(x + 2(\epsilon \cdot h)n) \geq \delta^8 N^{3/2}/2^{14}$$

for various functions $\tilde{f}_{\epsilon,h}(x) = \mathcal{C}^{|\epsilon|} f_2(x + p_\epsilon(h))$ with $p_\epsilon \in \mathbb{Z}[h]$, where $\mathcal{C}z = \overline{z}$ and $|\epsilon| = \epsilon_1 + \cdots + \epsilon_s$. The polynomials appearing in this new expression are linear in n, and an argument similar to Gowers' from [3] gives the claimed estimate

$$\sum_{x\in\mathbb{Z}} \|f_2\|_{U^7(x+[M'])} \gg \delta^C N M'^{8/2^7} \quad \text{for some} \quad M' \gg \delta^{-C} N^{1/2}.$$

References

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

[2] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Anal. Math.* **71** (1977), 204–256.

[3] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.* **11** (2011), 465–588.

[4] S. Peluse. Bounds for sets with no polynomial progressions. *Forum of Mathematics, Pi* **8** (2020), e16.

[5] S. Prendiville. Quantitative bounds in the polynomial Szemerédi theorem: the homogeneous case. *Discrete Anal.* **5** (2017).

Borys Kuca, Jagiellonian University
email: boryskuca@gmail.com

19 Norm convergence of commutative ergodic averages

After T. Tao [T1]

A summary written by James Leng

Abstract

We give a brief summary of Tao's proof of norm convergence of commutative ergodic averages.

19.1 Introduction

Theorem 1 (Tao)**.** *Let $(X, \mu, \mathcal{B})$ be a probability space and let $T_1, \dots, T_\ell : X \to X$ be measure preserving. Then for $f_1, \dots, f_\ell \in L^\infty(X)$,*

$$\lim_{N \to \infty} \mathbb{E}_{n \in [N]} T_1^n f_1 \cdots T_\ell^n f_\ell$$

converges.

For the purposes of this note, we shall work on the first non-classical case of case of $\ell = 2$ in [T1], as the argument for that nearly captures the entire argument. There is a striking similarity between the argument given in [A], which will also be presented in the summer school, and the argument in [T1], and this is no coincidence. According to Tao (private communication), [A] is very much inspired by [T1], which is inspired by more combinatorial contexts such as cut norms and hypergraph regularity. Tao's proof is a "finitary" proof, with many quantifiers whereas Austin's proof is infinitary, which has the advantage of removing many of the quantifiers. One striking thing about both proofs [A, T1] is that they avoid the heavy machinary of nilsystems as previous works of [HK, Z] did. This illustrates that merely proving norm convergence is much weaker than having a good understanding of the structure of measure preserving systems as the deep works of [HK, Z] try to do.
On the way to proving 1, Tao proves a finitary version. In order to state the theorem, we need some terminology.

Definition 2. *Let $\ell, P \geq 1$, and let $e_1, \ldots, e_\ell$ be the standard generators for $\mathbb{Z}_P^\ell$. For any functions $f_1, \ldots, f_\ell : \mathbb{Z}_P^\ell \to \mathbb{R}$, we define*

$$A_N(f_1, \ldots, f_\ell)(x) := \mathbb{E}_{n \in [N]} \prod_{i=1}^{\ell} f_i(x + e_i n).$$

For the case we will be considering, we have

$$A_N(f_1, f_2)(x) = \mathbb{E}_{n \in [N]} f_1(x + e_1 n) f_2(x + e_2 n).$$

Theorem 3. *Let $\ell \geq 1$ and $F : \mathbb{N} \to \mathbb{N}$ any function, and $\epsilon > 0$. Then there exists an integer $M^* > 0$ with the following property: if $P \geq 1$ and $f_1, \ldots, f_\ell : \mathbb{Z}_P^\ell \to [-1, 1]$ functions, then there exists an integer $1 \leq M \leq M^*$ such that we have the "L^2-metastability"*

$$\|A_N(f_1, \ldots, f_\ell) - A_{N'}(f_1, \ldots, f_\ell)\|_{L^2(\mathbb{Z}_P^\ell)} \leq \epsilon$$

for all $M \leq N, N' \leq F(M)$.

Via a simple argument, one can deduce Theorem 1 from Theorem 3.

19.2 Measurability

In the proof, it is convenient to work with a measure space $\mathbf{X} = (X, \mathcal{X}, \mu)$. Given a finite index set I and for each $i \in I$, measure spaces $\mathbf{Y}_i = (Y_i, \mathcal{Y}_i, \nu_i)$, we define the measure space $\mathbf{X} \times \mathbf{Y}_I$ as the space

$$X \times \prod_{i \in I} Y_i$$

equipped with the obvious (I-fold) tensor product sigma algebra and (I-fold) tensor product measure. An integral notion of the proof is the following:

Definition 4. *Given a measurable function $g : Y_I \times X \to [-1, 1]$, we say g is a primitive function of complexity at most d if g is $\mathcal{Y}_e \otimes \mathcal{X}$-measurable, or simply e-measurable for some $e \subseteq I$ is of size d. We say that g is of complexity at most (J, d) if it can be expressed as a sum of at most J many primitive functions of complexity at most d.*

The proof of Theorem 3 will involve an induction on ℓ and on the complexity of a function.
Given an additive group $(G,+)$, let $G^I = \prod_{i\in I} G$ and given $v \in G^I$, we define $\Sigma(v) := \sum_{i\in I} v_i$. Recall our setup

$$A_N(f_1,f_2)(v_1,v_2) = \mathbb{E}_{n\in[N]} f_1(v_1+n,v_2) f_2(v_1,v_2+n)$$

$$= \mathbb{E}_{n\in[N]} f_1(-v_2-(n-v_1-v_2),v_2) f_2(v_1,-v_1-(n-v_1-v_2)).$$

We see here that f_1 depends only on v_2 and $-v_1-v_2-n$ and f_2 depends only on v_1 and $-v_1-v_2-n$. We define the $\{2,3\}$ and $\{1,3\}$ measurable functions

$$g_{\{2,3\}}(v_1,v_2,v_3) := f_1(-v_2-v_3,v_2)$$
$$g_{\{1,3\}}(v_1,v_2,v_3) := f_2(v_1,-v_1-v_3).$$

Thus

$$A_N(f_1,f_2) = \mathbb{E}_{n\in[N]} g_{\{1,3\}} g_{\{2,3\}}(v_1,v_2,-v_1-v_2-n).$$

Thus, A_N is an average of a product of a $\{2,3\}$ and a $\{1,3\}$-measurable function along $\{(v_1,v_2,v_3) : v_3 \in -v_1-v_2-[N]\}$. For the sake of this exposition, the definition above is sufficient, but for the general case, we give the following definition:

Definition 5. *We define the diagonally averaged projection*

$$\Delta_N f(v,x) := \mathbb{E}_{n\in[N]} f((v,-\Sigma(v)-n),x).$$

We observe that if $g_{\{1,\dots,\ell\}}$ is $\{1,\dots,\ell\}$-measurable, it follows that

$$\Delta_N(g_{\{1,\dots,\ell\}} h) = g_{\{1,\dots,\ell\}} \Delta_N(h).$$

Under this notation, we have

$$A_N(f_1,\dots,f_\ell) = \Delta_N(\prod_{i=1}^{\ell} g_{\{1,\dots,\ell+1\}\setminus\{i\}})$$

where

$$g_{\{1,\dots,\ell\}\setminus\{i\}}(v_1,\dots,v_{\ell+1}) = f_i(v_1,\dots,v_{i-1},-\sum_{j\neq i} v_j, v_{i+1},\dots,v_{\ell+1}).$$

Remark 6. *According to [T1], these operations are morally equivalent to the hypergraph approaches of Szemerédi's theorem.*

Under this notation, [T1] proves the following theorem:

Theorem 7. *Let $1 \leq d \leq \ell$, $M_* \geq 1$, $J \geq 1$ be integers, $F : \mathbb{N} \to \mathbb{N}$ a function, and $\epsilon > 0$ real. Then there exists an integer $M^* \geq M_*$ with the following property: if $P \geq 1$ and $(X, \mathcal{X}, \mu)$ is a probability space, and $g : \mathbf{Z}^{\ell+1} \times X \to \mathbb{R}$ is an elementary function of complexity at most (d, J), then there exists an integer $M_* \leq M \leq M^*$ such that*

$$\|\Delta_N(g) - \Delta_{N'}(g')\|_{L^2(\mathbb{Z}_P^{\ell+1} \times X)} \leq \epsilon$$

whenever $M \leq N, N' \leq F(N)$.

19.3 A Sketch of the proof of Theorem 7

We shall prove this via induction on d, J, and ℓ. Since we are only treating the case of $\ell = 2$, we will make some notational simplifications from [T1].

19.3.1 Base case: $d = 1$

In the base case of $d = 1$, we make reductions to $\ell = 1$, $M_* = 1$, and $J = 1$. Since g has complexity $(1, J)$, we may write $g = g_1 + \cdots + g_J$ where g_i are basic. We now define $\tilde{X} = X \times \{1, \ldots, J\}$ and $\tilde{g} : \mathbb{Z}_P^3 \times \tilde{X} \to [-1, 1]$ by $\tilde{g}(v, (x, k)) = g_k(v, x)$. Thus,

$$\|\Delta_N(\tilde{g}) - \Delta_{N'}(\tilde{g})||_{L^2(\tilde{X} \times \mathbb{Z}_{P^3})} = J^{1/2} \|\Delta_N(g) - \Delta_{N'}(g)\|_{L^2(X \times \mathbb{Z}_P^3)}.$$

Hence, we can reduce to the $J = 1$ case since $\tilde{g}$ is primitive of complexity d. Thus, we may write $g = g_{\{1\}} g_{\{2\}} g_{\{3\}}$ where $g_{\{i\}}$ is $\{i\} \times X$-measurable, it follows that we may discard $g_{\{i\}}$ for $i \neq 3$, since they don't change under any terms of the averaging operator Δ_N. Thus, we may just focus on a single function $g_{\{3\}}$. Since

$$\Delta_N(g_{\{3\}})(v_1, v_2, v_3, x) = \mathbb{E}_{n \in [N]} g_{\{3\}}(-v_1 - v_2 - v_3 - n, x)$$

which only depends on $v_1 + v_2 + v_3$, we may quotient by $v_1 + v_2 + v_3 = 0$ so that $g_{\{3\}}$ only depends on $\mathbb{Z}_P$. We have thus reduced to the case of $\ell = 1$. Thus it remains to prove the following:

Theorem 8. *Let $F : \mathbb{N} \to \mathbb{N}$ a function, and $\epsilon > 0$ real. Then there exists an integer $M^* \geq 1$ with the following property: if $P \geq 1$ and $(X, \mathcal{X}, \mu)$ is a probability space, and $g : \mathbf{Z} \times X \to \mathbb{R}$, then there exists an integer $1 \leq M \leq M^*$ such that*

$$\|S_N(g) - S_{N'}(g')\|_{L^2(\mathbb{Z}_P^{\ell+1} \times X)} \leq \epsilon$$

whenever $M \leq N, N' \leq F(N)$ where $S_N(f)(x,v) = \mathbb{E}_{n \in [N]} f(v+n, x)$

This can be deduced via some quantitative Lebesgue dominated convergence theorem [T1, Theorem A.2] from the following (using the function $f_{N,N'}(x) = \|S_N g(\cdot, x) - S_{N'}(\cdot, x)\|_{L^2(\mathbb{Z}_P)}$):

Theorem 9. *Quantitative convergence of a single ergodic average Let $F : \mathbb{N} \to \mathbb{N}$ a function, and $\epsilon > 0$ real. Then there exists an integer $M^* \geq 1$ with the following property: if $P \geq 1$ and $(X, \mathcal{X}, \mu)$ is a probability space, and $g : \mathbb{Z}_P \to \mathbb{R}$, then there exists an integer $1 \leq M \leq M^*$ such that*

$$\|S_N(g) - S_{N'}(g')\|_{L^2(\mathbb{Z}_P^{\ell+1} \times X)} \leq \epsilon$$

whenever $M \leq N, N' \leq F(N)$ where $S_N(f)(x,v) = \mathbb{E}_{n \in [N]} f(v+n)$.

The proof of this theorem proceeds via an *energy increment argument*. To see a similar argument, see [T2, Chapter 1.2]. To save space, we only provide a sketch. We encourage the reader to compare this proof with the proof of the von Neumann ergodic theorem. First, a definition:

Definition 10. *Basic $\{1\}$-anti-uniform function Let $M \geq 1$. A basic $\{1\}$-anti-uniform function on scale M is any function $\varphi : \mathbb{Z}_P \to \mathbb{R}$ of the form*

$$\varphi(v) = \mathbb{E}_{n \in [M]} b(v-n)$$

for some function $b : \mathbb{Z}_P \to [-1, 1]$.

These anti-uniform functions satisfy a Lipschitz bound of $|\varphi(v+n) - \varphi(v)| \leq \frac{|n|}{M}$. The point is that if g is $\{1\}$-anti-uniform on scale M_1, then expressing

$$g(v) = \mathbb{E}_{n \in [M_1]} b(v-n)$$

$$S_N g - S_{N'} g = \mathbb{E}_{n \in [N]} \mathbb{E}_{m \in [M_1]} g(n-m) - \mathbb{E}_{n \in [N']} \mathbb{E}_{m \in [M_1]} g(n-m) = O\left(\frac{F(M)}{M_1}\right)$$

so choosing M_1 sufficiently large, we obtain the desired inequality. One can then use the energy increment argument to prove that each g can be decomposed as a sum of $\{1\}$-anti-uniform-measurable function $g^{U^\perp}$ and a "uniform part" g^U, for which

$$\|S_N g^U - S_N g^U\|_{L^2}$$

is small anyways.

19.3.2 Induction Step (i.e., $\ell = 2$ and $d = 2$)

For the $\ell = 2$ case, we can once again reduce to $M_* = 1$ and $J = 1$, so we may assume that f takes the form

$$f(v_1, v_2, v_3, x) = g_{\{1,2\}}(v_1, v_2, x) g_{\{2,3\}}(v_2, v_3, x) g_{\{1,3\}}(v_1, v_3, x).$$

Under this identification, we see that

$$\Delta_N(f) = \mathbb{E}_{n\in[N]} g_{\{1,2\}}(v_1, v_2) g_{\{2,3\}}(v_2, -v_1 - v_2 - n) g_{\{1,3\}}(v_1, -v_1 - v_2 - n).$$

As before, we can pull out the $g_{\{1,2\}}$ term so we may assume it is constant. If $g_{\{2,3\}}$ and $g_{\{1,3\}}$ can be written as a tensor products

$$g_{\{2,3\}}(v_2, v_3) = h_{\{2\}}(v_2) h_{\{3\}}(v_3), \quad g_{\{1,3\}}(v_1, v_3) = k_{\{1\}}(v_1) k_{\{3\}}(v_3)$$

then the average simplifies to

$$g_{\{1,2\}} h_{\{2\}}(v_2) k_{\{1\}}(v_1) \mathbb{E}_{n\in[N]} h_{\{3\}} k_{\{3\}}(-v_1 - v_2 - n)$$

which can be taken care of by the base case.
If, however, $g_{\{2,3\}}$ is "orthogonal" to these tensor products, i.e., for "most" w_2, w_3, we have

$$\mathbb{E}_{v_2\in w_2+[N']} \mathbb{E}_{v_3\in w_3+[N']} g_{\{2,3\}}(v_2, v_3) h_2(v_2) h_3(v_3)$$

are small, then we are analogous to the "weakly mixing case" in the ergodic theoretic proofs, and so

$$\|\Delta_N(f)\|_{L^2}^2 = \mathbb{E}_{v_1,v_2} \Delta_N(f)(v_1, v_2) \mathbb{E}_n g_{\{2,3\}}(v_2, -v_1 - v_2 - n) g_{\{1,3\}}(v_1, -v_1 - v_2 - n)$$

is small since we may rewrite the above as

$$\mathbb{E}_{v_1,v_2,v_3 : -\Sigma(v)\in[N]} g_{\{2,3\}}(v_2, v_3) \Delta_N(f)(v_1, v_2) g_{\{3,1\}}(v_3, v_1).$$

Similarly to the base case, one may also decompose $g_{\{2,3\}}$ as of tensor products of $\{2\}$-measurable and $\{3\}$-measurable functions (i.e., a "structured" piece), and an "orthogonal" piece (i.e., a "random" piece). By treating the tensor piece similarly as the anti-uniform case in the above base case, and the "orthogonal" piece similarly with the "uniform part," we may conclude the case for $\ell = 2$ and $d = 2$. The general case follows a similar procedure.

References

[A] T. Austin. *On the norm convergence of non-conventional ergodic averages. Ergodic Theory and Dynamical Systems*. 30(2), 321-338. doi:10.1017/S014338570900011X

[HK] B. Host and B. Kra. *Nonconventional Ergodic Averages and Nilmanifolds*. Annals of Mathematics, vol. 161, no. 1, 2005, pp. 397–488.

[T1] T. Tao. *Norm convergence of multiple ergodic averages for commuting transformations. Ergodic Theory and Dynamical Systems*. 28(2), 657-688. doi:10.1017/S0143385708000011.

[T2] T. Tao, *Higher Order Fourier Analysis*. Graduate Studies in Mathematics. Vol 42. AMS, Providence, RI, 2012.

[Z] T. Ziegler. *Universal Characteristic Factors and Furstenberg Averages*. J. Amer. Math. Soc. 20 (2007), 53-97.

James Leng, UCLA
email: `jamesleng@math.ucla.edu`

20 Bourgain's return time theorem

After D. Rudolph [R]

A summary written by Zi Li Lim

Abstract

Bourgain's return time theorem investigates the correlation of the time average of two dynamical systems. Rudolph had given a simplified proof of the return time theorem based on the machinery of joinings. This is a sketch of Rudolph's proof.

20.1 Introduction

Let $(X, \mathcal{F}, \mu, T)$ be a dynamical system, i.e., $(X, \mathcal{F}, \mu)$ is a Lebesgue probability space and T is a measure preserving transformation on X. Bourgain's return time thoerem states that the correlation of time average of $(X, \mathcal{F}, \mu, T)$ and any other dynamical systems is well-defined. More precisely, we have the following theorem.

Theorem 1. *Let $(X, \mathcal{F}, \mu, T)$ be a dynamical system and $f \in L^p(\mu)$ for some $1 \leq p \leq \infty$. There exists a subset $X(f) \subset X$ of full measure such that for any dynamical system $(Y, \mathcal{G}, \nu, S)$ and $g \in L^q(\nu)$ with $1/p + 1/q = 1$, for any $x \in X(f)$ and for ν-a.e. y,*

$$\frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) \overline{g}(S^i(y))$$

converges as $n \to \infty$.

Recall the Birkhoff pointwise ergodic theorem says that the time average along the orbit of a typical point is well-defined. Intuitively, Bourgain's return time theorem tells us that the correlation of the time average of two dynamical systems is also well-defined to a great extent. In fact, it is well-defined in a universal sense, the subset $X(f)$ does not depend on the other dynamical system $(Y, \mathcal{G}, \nu, S)$.
Bourgain's return time theorem was first proved by Bourgain in [B1]. Furstenberg, Katznelson and Ornstein gave a different proof in the appendix to [B2]. Later, Rudolph found a proof based on the machinery of

joinings in [R]. We were hoping to present a sketch of Rudolph's proof in these expository notes.
The proof consists of three ingredients: the construction of the measures to keep track of the orbits, the reduction to the good enough dynamical system, and building the measures inductively that lead to a contradiction.

20.2 Construction of the measures

Instead of working with the orbits of the typical points, working with measures is much more flexible. In this section, we will explain the dictionary that allows us to pass from the orbits to the measures on certain spaces.
Let $D = \{z \in \mathbb{C} : |z| \leq 1\}$ be the closed unit disk in the complex plane. Define $Z = D^{\mathbb{Z}_{\geq 0}}$ be the countable product of the unit disk and $Z^{(k)} = Z_0 \times Z_1 \times \cdots \times Z_{k-1}$, where each Z_i is a copy of Z, that is, $Z^{(k)}$ is the k-fold product of Z. The space of all Borel probability measures on $Z^{(k)}$ could be regarded as a weak* compact subset of the dual space of the space of all continuous functions on $Z^{(k)}$.
Let $(X, \mathcal{F}, \mu, T)$ be a dynamical system and $f_0, f_1, ..., f_{k-1}$ be complex functions on X with $|f_i| \leq 1$, the dynamic of X could be kept track by considering the map $F : X \longrightarrow Z^{(k)}$

$$F(x) = ((f_0(T^i(x))), (f_1(T^i(x))), ..., (f_{k-1}(T^i(x))))$$

and we would denote the push-forward of the measure μ with respect to F by $m((X, \mathcal{F}, \mu, T), f_0, f_1, ..., f_{k-1})$.
Given an element $(d_0, d_1, ..., d_{n-1})$ in $(D^{(k)})^{(n)}$, let $m(d_0, d_1, ..., d_{n-1})$ be the atomic measure on $Z^{(k)}$, uniformly supported on the subset

$$\{z(t) \in Z^{(k)} : z(t)_i = d_{t+i \ mod \ n}\}$$

Informally speaking, these measures could keep track of the orbits. For example, assume the dynamical system $(X, \mathcal{F}, \mu, T)$ is ergodic, by Birkhoff pointwise ergodic theorem, for a.e. $x \in X$, the measures $m(f(x), f(T(x)), ..., f(T^{n-1}(x)))$ converges to $m((X, \mathcal{F}, \mu, T), f_0, f_1, ..., f_{k-1})$ in weak* topology as $n \to \infty$.
To summarise, if we are interested in the asymptotic behaviour of $f(T^i(x))$, we shall investigate the weak* limit of the associated measures

$m(f(x), f(T(x)), ..., f(T^{n-1}(x)))$. These measures can 'detect' the orbits, and we could recover the information about orbits by integrations:

$$\int_{Z^{(k)}} z_{i,0} dm \quad , \int_{Z^{(k)}} z_{i,0} z_{j,0} dm$$

where $z_{i,0}$ is the zero-th coordinate of the i-component function.

20.3 Reduction to the best case

Fix a dynamical system $(X, \mathcal{F}, \mu, T)$ and $f \in L^p(\mu)$, given any other arbitrary dynamical system $(Y, \mathcal{G}, \nu, S)$ and $g \in L^q(\nu)$, we were hoping to reduce the proof to the 'best' case. What is the best scenario that the dynamical systems are sufficiently good enough?
First, we can assume the dynamical systems $(X, \mathcal{F}, \mu, T)$ and $(Y, \mathcal{G}, \nu, S)$ are ergodic just like many other proofs in ergodic theory. This is reasonable since ergodic dynamical systems are the building blocks of the general systems, thanks to ergodic decomposition.
Next, we shall assume the test functions f and g are good enough, that is, they decay sufficiently fast. Let's assume that f and g are in $L^\infty(\mu)$ and $L^\infty(\nu)$ respectively. In fact, we could even normalize the functions such that $|f|, |g| \leq 1$.
What are some other reasonable assumptions that could simplify the situation? Through normalization, we might assume the space average $\int_X f d\mu = 0$ as well. In order to consider the correlation of time average with arbitrary systems, we should pretend that we understand the self correlation of the function f well enough, say

$$\frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x_1)) \overline{f}(T^i(x_2)) \to 0$$

as $n \to \infty$ for $\mu \times \mu$-a.e. $(x_1, x_2) \in X \times X$. In this ideal case, we have a suitable canditate for $X(f)$, let $G(f)$ consists of the points $x_1 \in X$ such that

$$\frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x_1)) \to 0$$

as $n \to 0$ and

$$\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x_1))\overline{f}(T^i(x_2)) \to 0$$

as $n \to 0$ for μ-a.e. x_2. Now, we have reduced the proof to the following proposition.

Proposition 2. *Suppose the Bourgain's return time theorem is false, then there exist ergodic dynamical systems $(X, \mathcal{F}, \mu, T)$ and $(Y, \mathcal{G}, \nu, S)$, the functions f and g as above, a point x' in $G(f)$, a positive measure subset $B \subset Y$ and a positive real number a such that*

$$\limsup_{n\to\infty} |\frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x'))\overline{g}(S^i(y))| > a$$

for all $y \in B$.

20.4 The key step

Suppose the Bourgain's return time theorem is false, we could construct a sequence of measures inductively based on the machinery introduced in previous sections. For any integer $k \geq 2$, there exists a measure $m^{(k)}$ on $Z^{(k)}$ such that

1. The projection of $m^{(k)}$ to Z_0 is some fixed measure, say m_0.
2. The projection of $m^{(k)}$ to other coordinate Z_i is another fixed measure, say m_1, for all $1 \leq i \leq k-1$.
3. $\int_{Z^{(k)}} z_{i,0}z_{j,0}dm^{(k)}$=0 for all $1 \leq i, j \leq k-1, i \neq j$.
4. $|\int_{Z^{(k)}} z_{0,0}z_{i,0}dm^{(k)}| > a$ for all $1 \leq i \leq k-1$.

Heuristically, the condition $\int_{Z^{(k)}} z_{i,0}z_{j,0}dm^{(k)}$=0 holds since the self correlation of f is zero and the condition $|\int_{Z^{(k)}} z_{0,0}z_{i,0}dm^{(k)}| > a$ holds due to the reduction to the best case in last section. Choose constants c_i with $|c_i| = 1$ such that $\int_{Z^{(k)}} c_i z_{0,0}z_{i,0}dm^{(k)} > a$. Note that

$$||\sum_{i=1}^{k} c_i z_{i,0}||_{L^2(m^{(k+1)})} = \sqrt{k}||z_0||_{L^2(m_1)}$$

and

$$\langle \sum_{i=1}^{k} c_i z_{i,0}, z_{0,0} \rangle_{L^2(m^{(k+1)})} > ka$$

However, by Cauchy-Schwartz inequality, we have

$$\langle \sum_{i=1}^{k} c_i z_{i,0}, z_{0,0} \rangle_{L^2(m^{(k+1)})} \leq \sqrt{k} ||z_0||_{L^2(m_1)} ||z_0||_{L^2(m_0)}$$

this leads to a contradiction when k is sufficiently large.

References

[R] Rudolph, D., *A joinings proof of Bourgain's return time theorem.* Ergodic Theory Dynam. Systems, 14 (1994), no.1, 197–203.

[B1] Bourgain, J., *Return time sequences of dynamical systems.* IHES Preprint (3/1988).

[B2] Bourgain, J., *Pointwise ergodic theorems for arithmetic sets.* Inst. Hautes Études Sci. Publ. Math.(1989), no.69, 5–45.

ZI LI LIM, UCLA
email: zililim@math.ucla.edu

21 On the convergence of multiple Ergodic averages

After G. Karagulyan, M. Lacey and V. Martirosyan [KLM]

A summary written by Gevorg Mnatsakanyan

Abstract

Let $\mathcal{U} := \{U_j :\ j = 1, \ldots, n\}$ be a sequence of invertible, commuting measure preserving transformation on a measureable space $(X, d\mu)$. We prove the almost everywhere convergence of averages

$$\frac{1}{s_1 \ldots s_n} \sum_{j_1=0}^{s_1-1} \cdots \sum_{j_n=0}^{s_n-1} f(U_1^{j_1} \cdots U_n^{j_n} x),$$

as $\min s_j \to \infty$, for f in $L \log^{d-1} L$ where $d \leq n$ is the rank of $\mathcal{U}$.

21.1 Introduction

Let $(X, \mathcal{B}, d\mu)$ be a probability space and T be a measure-preserving transformation. The famous ergodic theorem of Birkoff states that for $f \in L^1(X)$ the averages

$$\frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) \tag{1}$$

converge almost everywhere to a T invariant function. A generalization of this result for multiple transformations goes back to Dunford [D] and Zygmund [Z]. Let us first introduce the spaces $L \log^n L$. For a non-decreasing function $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ we define the class $L_\Phi(X)$ of $\mathcal{B}$ measurable functions f on X so that $\Phi(|f|) \in L^1(X)$. The class corresponding to the function $Log_n t := t(1 + \max(0, \log^n t))$ will be denoted by $L \log^n L$, for $n \geq 1$.
Henceforth, $\mathcal{U} := \{U_j :\ j = 1, \ldots, n\}$ will be a sequence of invertible measure-preserving transformations on X. Then, Dunford and Zymgund independently proved that for $f \in L \log L^{n-1}$ the averages

$$\frac{1}{s_1 \ldots s_n} \sum_{j_1=0}^{s_1-1} \cdots \sum_{j_n=0}^{s_n-1} f(U_1^{j_1} \cdots U_n^{j_n} x) \tag{2}$$

converge almost everywhere as $\min_j s_j \to +\infty$. Hagelstein and Stokolos [HS] proved that the class of functions above is sharp in the following sense. If $\mathcal{U}$ is additionally commuting and non-periodic, i.e. for any non-trivial collection of integers p_k, $k = 1, \dots, n$,

$$\mu\{U_1^{p_1} \circ U_2^{p_2} \circ \dots \circ U_n^{p_n} x = x\} = 0, \tag{3}$$

then, for any non-decreasing function $\Phi : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\Phi = o(t \log^{n-1} t)$ as $t \to +\infty$, there exists a function $f \in L_\Phi(X)$ so that the averages (2) unboundedly diverge at almost every point.
It turns out that if the transformations $\mathcal{U}$ are not independent, then the class $L \log^{n-1} L$ can be improved.

Definition 1. *A set of commuting transformations $\mathcal{U}$ is called independent, if for any non-trivial collection of integers p_k, $k = 1, \dots, n$ we have*

$$\mu\{U_1^{p_1} \circ U_2^{p_2} \circ \dots \circ U_n^{p_n} x = x\} < 1. \tag{4}$$

The rank of $\mathcal{U}$, denoted $\mathrm{rank}(\mathcal{U})$, is the largest integer $r \le n$ so that there is an independent subset of $\mathcal{U}$ of cardinality r.

Note, that non-periodicity implies independence. Let us also introduce the following maximal function

$$\mathcal{M}_{\mathcal{U}} f(x) := \sup_{s_j \in \mathbb{N}} \frac{1}{s_1 \dots s_n} \sum_{j_1=0}^{s_1-1} \cdots \sum_{j_n=0}^{s_n-1} |f(U_1^{j_1} \cdots U_n^{j_n} x)|. \tag{5}$$

The following theorem is the main result of [KLM].

Theorem 2. *Let $\mathcal{U}$ be also commuting and of rank d. Then, for any function $f \in L \log^{d-1} L(X)$ and $\lambda > 0$, we have*

$$\mu\{x \in X : \mathcal{M}_{\mathcal{U}} f(x) > \lambda\} \lesssim_{\mathcal{U}} \int_X Log_{d-1}\Big(\frac{|f|}{\lambda}\Big). \tag{6}$$

The convergence (2) follows from the above maximal estimate by a standard density argument.
We will reduce Theorem 2 to a bound for a strong maximal function in the Euclidean setting. Let $A : \mathbb{R}^n \to \mathbb{R}^d$ be a linear operator. We consider

$$M_A f(x) := \sup_R \frac{1}{|R|} \int_R f(x + At) dt, x \in \mathbb{R}^d, \tag{7}$$

where the sup is taken over all symmetric intervals

$$R = \{t = (t_1, \dots, t_n) : \ t_j \in [-r_j, r_j] \text{ for } j = 1, \dots, n\} \subset \mathbb{R}^n. \tag{8}$$

Theorem 3. *If $rank(A) = r$, then for any $f \in L\log^{r-1} L(\mathbb{R}^d)$ and $\lambda > 0$*

$$|\{x \in \mathbb{R}^d : M_A f(x) > \lambda\}| \lesssim_A \int_{\mathbb{R}^d} Log_{r-1}\Big(\frac{|f|}{\lambda}\Big). \tag{9}$$

.

When $n = d = r$ and $A = I_n$, the identity matrix, this is a well-known theorem of Guzmán [G]. The general case will be reduced to the latter.

21.2 Sketch of proof of Theorem 2 assuming Theorem 3

The maximal inequality (9) has a discrete analog, that is easily deduced from it. Let $\phi : \mathbb{Z}^d \to \mathbb{R}$ and let $A = \{a_{kj}, 1 \le j \le n, 1 \le k \le d\}$ be an integer matrix. We consider the maximal operator

$$\mathcal{D}_A\phi(n) = \sup_{s_j \in \mathbb{N}} \frac{1}{s_1 \dots s_n} \sum_{k_1=0}^{s_1-1} \cdots \sum_{k_n=0}^{s_n-1} \phi(n + Ak), \ n \in \mathbb{Z}^d, \tag{10}$$

then

$$\#\{n \in \mathbb{Z}^d : \ \mathcal{D}_A\phi(n) > \lambda\} \lesssim_A \sum_{n \in \mathbb{Z}^d} Log_{r-1}\left(\frac{|\phi(n)|}{\lambda}\right). \tag{11}$$

As rank of $\mathcal{U}$ is d, let us assume $U_1, \dots, U_d$ are independent, and

$$U_k^{l_k} = U_1^{a_{1,k}} \circ \cdots \circ U_d^{a_{d,k}}, \ d < k \le n. \tag{12}$$

Assume $l_k = 1$. The general case can be deduced from this. We write

$$\begin{aligned} &f(U_1^{k_1} \circ \cdots \circ U_n^{k_n} x) \\ &\quad = f\left(U_1^{k_1 + a_{1,d+1}k_{d+1} + \cdots + a_{1,n}k_n} \circ \cdots \circ U_d^{k_d + a_{d,d+1}k_{d+1} + \cdots + a_{d,n}k_n}\right) = \phi(x, A \cdot k), \end{aligned}$$

where

$$\phi(x, n) = f(U_1^{n_1} \circ \cdots \circ U_d^{n_d} x), \ x \in X, \ n \in \mathbb{Z}^d, \tag{13}$$

and

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 & a_{1,d+1} & \cdots & a_{1,n} \\ 0 & 1 & \cdots & 0 & a_{2,d+1} & \cdots & a_{2,n} \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \dots & 1 & a_{d,d+1} & \dots & a_{d,n} \end{pmatrix}. \tag{14}$$

We consider the truncated and translated maximal function

$$f^*_M(x,n) := \max_{1\le s_j\le M} \sum_{k=0}^{s-1} |\phi(x, n + Ak)|. \tag{15}$$

Then, $\mathcal{M}_{\mathcal{U}} f(x) = \lim_{M\to\infty} f^*_M(x,0)$. Now let

$$\begin{aligned} E_\lambda(x) &= \{n : \ 1 \le n_j \le N \text{ and } f^*_M(x,n) > \lambda\}, \\ F_\lambda(n) &= \{x : \ f^*_M(x,n) > \lambda\}, \ n \in \mathbb{Z}^d. \end{aligned}$$

So the inequality (6) would following from

$$\mu(F_\lambda(0)) \lesssim_{\mathcal{U}} \int_X Log_{d-1}\Big(\frac{|f|}{\lambda}\Big), \tag{16}$$

where the implicit constant is independent of M.
In (15), the coordinates of Ak may vary in $[-R, R]$ where $R = R(A, M)$. So by the discrete inequality (11), we have

$$\# E_\lambda(x) \lesssim_A \sum_{1\le n_j\le N+R} Log_{r-1}\left(\frac{|\phi(x,n)|}{\lambda}\right). \tag{17}$$

Since U_j are measure preserving $F_\lambda(n)$ have the same measure for all n, hence

$$\begin{aligned} \mu(F_\lambda(0)) &= \frac{1}{N^d} \sum_{1\le n_j\le N} \mu(F_\lambda(n)) = \frac{1}{N^d} \int_X \# E_\lambda(x) \\ &\lesssim_A \frac{1}{N^d} \sum_{1\le n_j\le N+R} \int_X Log_{r-1}\left(\frac{|\phi(x,n)|}{\lambda}\right) = \frac{(N+R)^d}{N^d} \int_X Log_{r-1}\left(\frac{|f|}{\lambda}\right). \end{aligned}$$

In the last line we again used that U_j are measure preserving. Fixing M and letting $N \to +\infty$ we get the desired bound (16).

21.3 Reducing Theorem 3 to the case $n = d = r$

Let $\mathfrak{U} = \{u_j : j = 1, \dots, n\}$ be a set of vectors in $\mathbb{R}^d$ and $\mathbb{R}_{\mathfrak{U}} := span(\mathfrak{U}) \subset \mathbb{R}^d$. We call a parallelipiped in $\mathbb{R}^d$ a set of the form $R = \{x \in \mathbb{R}^d : x = t_1 u_1 + \dots + t_n u_n, t_j \in [-r_j, r_j]\}$. The family of such paralleipipeds is denoted by $\mathcal{P}_{\mathfrak{U}}$. Further, let us associate a probability measure μ_R with R. Let μ_j be the uniform probability measure on one-dimensional parallelipiped $\{tu_j : t \in [-r_j, r_j]\}$. Then, we define μ_R by a convolution, namely,

$$\mu_R(E) = \int_{\mathbb{R}_{\mathcal{U}}} \dots \int_{\mathbb{R}_{\mathcal{U}}} \mathbf{1}_E(v_1 + \dots + v_n) d\mu_1(v_1) \dots d\mu_n(v_n). \tag{18}$$

Let f_R be the density of μ_R. Observe, that if $\mathfrak{U}$ is indepedent, then

$$f_R(x) = \begin{cases} 1/|R|, & \text{if } x \in R \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

The following lemma can be proved by some not very difficult geometric considerations and is somewhat intuitive.

Lemma 4. *Let $\mathfrak{U}$ be arbitrary and $R \in \mathcal{P}_{\mathfrak{U}}$. Then, there exists an independent subset $\mathfrak{V} \subset \mathfrak{U}$ of maximal rank and a parallelipiped $R' \in \mathcal{P}_{\mathfrak{V}}$ such that*

$$\mu_R \lesssim_{\mathfrak{U}} \mu_{R'}. \tag{20}$$

Let u_j be the jth column of the matrix A and $\mathfrak{U} = \{u_j : 1 \le j \le n\}$, so $rank(A) = rank(\mathfrak{U})$. We abuse the notation and write $M_{\mathfrak{U}}$ for M_A. The integral in (7) can be rewritten through the measure μ_R, namely, we have

$$\frac{1}{|R|} \int_R |f(x{+}At)| dt = \int_{\mathbb{R}^d} |f(x{+}v)| d\mu_R(v) \lesssim_{\mathfrak{U}} \int_{\mathbb{R}^d} |f(x{+}v)| d\mu_{R'} \lesssim_{\mathfrak{U}} M_{\mathfrak{V}} f(x), \tag{21}$$

wherethe first inequality follows from the lemma above. So we conclude

$$M_{\mathfrak{U}} f(x) \lesssim \sup_{\mathfrak{V}} M_{\mathfrak{V}} f(x). \tag{22}$$

If we assume the theorem in the case $n = d = r$, then $M_{\mathfrak{V}}$ satisfies the bound (9) in $\mathbb{R}_{\mathfrak{V}}$ and there are finitely many independent subset $\mathfrak{V}$ of $\mathfrak{U}$, so we are done.

21.4 The Theorem of Guzman

We call $\mathcal{R} \subset 2^{\mathbb{R}^n}$ a differentiation basis if for each $x \in \mathbb{R}^n$ there is an arbitrarily small $R \in \mathcal{R}$ that contains x. The maximal function with respect to $\mathcal{R}$ is defined by

$$M_{\mathcal{R}} f(x) := \sup_{R \ni x} \frac{1}{m_n(R)} \int_R |f(x)| dm_n(x). \tag{23}$$

For a strictly increasing continuous function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ with $\phi(0) = 0$, we say that $M_{\mathcal{R}}$ has type ϕ if it satisfies the following weak type bound

$$m_n(\{M_{\mathcal{R}} f > \lambda\}) \lesssim \int \phi \left(\frac{|f(x)|}{\lambda} \right) dm_n(x). \tag{24}$$

Theorem 5. *Let $\mathcal{R}_i$, $i = 1, 2$, be two differentiation bases in $\mathbb{R}^{n_i}$ with maximal operators M_i. Let M_i have type ϕ_i. Consider in $\mathbb{R}^n = \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ the differentiation bases $\mathcal{R} = \mathcal{R}_1 \times \mathcal{R}_2$. Then, the corresponding maximal operator M satisfies the following inequality*

$$\begin{aligned} m_n(\{x :\ Mf(x) > \lambda\} &\le \phi_2(1) \int\limits_{\mathbb{R}^n} \phi_1 \left(\frac{2|f(x)|}{\lambda} \right) dm_n(x) \\ &\qquad + \int\limits_{\mathbb{R}^n} \left[\int\limits_1^{\frac{4|f(x)|}{\lambda}} \phi_1 \left(\frac{4|f(x)|}{\lambda\sigma} d\phi_2(\sigma) \right) \right]. \end{aligned}$$

Applying the above inequality with $\mathcal{R}_i$, $i = 1, \ldots, n$, the family of bounded intervals in $\mathbb{R}$, and using the weak-L^1 bounded of the one-dimensional Hardy-Littlewood maximal function we arrive at Theorem 3 with $n = d = r$ and $A = I_n$. Then, the equivalence of the paralleipiped and rectangular bases imply the result for general invertible A.

References

[KLM] Grigori Karagulyan, Michael Lacey and Vahan Martirosyan, *On the convergence of multiple Ergodic averages,* New York J. Math. **28** (2022), 1448-1462.

[G] Miguel de Guzmán, *An inequality for Hardy-Littlewood maximal operator with respect to a product of differentiation bases*, Studia Math. **49** (1973/74), 185-194.

[D] Nelson Dunford, *An individual ergodic theorem for non-commutative transformations*, Acta Sci. Math. (Szeged) **14** (1951), 1-4.

[Z] Antoni Zygmund, *An individual ergodic theorem for non-commutative transformations*, Acta Sci. Math. (Szeged) **14**, 103-110.

[HS] Paul Hagelstein and Alexander Stokolos, *Weak type inequalities for ergodic strong maximal operators*, Acta Sci. Math. (Szeged) **76** (2010), no. 3-4, 427-441.

Gevorg Mnatsakanyan, University of Bonn
email: gevorg@math.uni-bonn.de

22 Joint ergodicity of sequences

After N. Frantzikinakis [Fr1]

A summary written by Andreas Mountakis

Abstract

A collection of integer sequences is jointly ergodic if for every ergodic measure preserving system the multiple ergodic averages, with iterates given by this collection of sequences, converge in the mean to the product of the integrals. We give necessary and sufficient conditions for joint ergodicity that are flexible enough to recover the known examples of jointly ergodic sequences and also allow us to answer some related open problems. An interesting feature of our arguments is that they avoid deep tools from ergodic theory that were previously used to establish similar results. Our approach is primarily based on an ergodic variant of a technique pioneered by Peluse and Prendiville in order to give quantitative variants for the finitary version of the polynomial Szemerédi theorem.

22.1 Introduction

The study of multiple ergodic averages was initiated in the seminal work of Furstenberg [Fu1], where an ergodic theoretic proof of Szemerédi's theorem on arithmetic progressions was given. Since then, the study of different types of multiple ergodic averages has been a central object in ergodic theory, resulting in many more combinatorial consequences. A rather general family of problems is as follows: We are given a collection of integer sequences $a_1, \ldots, a_\ell : \mathbb{N} \to \mathbb{Z}$ and an invertible measure preserving system (X, μ, T). We would like to understand the behaviour, as $N \to \infty$, of the ergodic averages

$$\mathbb{E}_{n\in[N]}\, T^{a_1(n)} f_1 \cdot \ldots \cdot T^{a_\ell(n)} f_\ell \tag{1}$$

for all functions $f_1, \ldots, f_\ell \in L^\infty(\mu)$.

The question the author tries to answer in [Fr1] is under what conditions, the iterates $T^{a_1(n)}, \ldots, T^{a_\ell(n)}$ behave independently enough, so that the averages in (1) converge in $L^2(\mu)$ to the product of the integrals of the f_i's.

22.2 Definitions and results

Notation. Whenever we say that $(X, \mathcal{X}, \mu, T)$ is a system, we mean that $(X, \mathcal{X}, \mu)$ is a probability space and $T : X \to X$ is an invertible, measurable and measure preserving transformation. From now on, we will usually omit writing the σ-algebra $\mathcal{X}$. For $n \in \mathbb{Z}$ and $f \in L^\infty(\mu)$, we denote $f \circ T^n$ by $T^n f$. In addition, for $N \in \mathbb{N}$, $[N]$ denotes $\{1, \dots, N\}$ and $\mathbb{E}_{n\in[N]}$ denotes the average $\frac{1}{N}\sum_{n=1}^{N}$. Lastly, for $t \in \mathbb{R}$, $e(t)$ denotes $e^{2\pi i t}$ and $\lfloor t \rfloor$ denotes the integer part of t.
Let us start by defining joint ergodicity, which is one of the central notions in the paper.

Definition 1. *A collection of sequences $a_1, \dots, a_\ell : \mathbb{N} \to \mathbb{Z}$ is called*

- *jointly ergodic for the system (X, μ, T) if for all $f_1, \dots, f_\ell \in L^\infty(\mu)$ we have*

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]} T^{a_1(n)} f_1 \cdot \ldots \cdot T^{a_\ell(n)} f_\ell = \int_X f_1\, d\mu \cdot \ldots \cdot \int_X f_\ell\, d\mu \tag{2}$$

  where convergence takes place in $L^2(\mu)$.

- *jointly ergodic, if it is jointly ergodic for every ergodic system.*

Examples. For $\ell \in \mathbb{N}$, and $c_1, \dots c_\ell \in (0, +\infty) \setminus \mathbb{N}$, it was proved in [Fr2] that the collection of sequences $\lfloor n^{c_1} \rfloor, \dots, \lfloor n^{c_\ell} \rfloor$ is jointly ergodic. In addition, a collection of polynomial sequences $p_1, \dots, p_\ell \in \mathbb{Z}[t]$ is jointly ergodic for all totally ergodic systems if and only if the polynomials are rationally independent (see [FK]).

Definition 2. *If (X, μ, T) is a system, then*

- *$Spec(T) := \{t \in [0, 1) : Tf = e(t)f$ for some non-zero $f \in L^2(\mu)\} =$ the set of eigenvalues of T.*

- *$\mathcal{E}(T) := \{f \in L^\infty(\mu) : Tf = e(t)f$ for some $t \in [0, 1)$ and $|f| = 1\}$.*

Now, let us define what it means for a collection of sequences to be *good for equidistribution* and *good for seminorm estimates.*

Definition 3. *A collection of sequences $a_1, \dots, a_\ell : \mathbb{N} \to \mathbb{Z}$ is called*

- *good for seminorm estimates for the system (X, μ, T), if there is $s \in \mathbb{N}$ such that whenever $f_1, \ldots, f_\ell \in L^\infty(\mu)$ satisfy $|||f_m|||_s = 0$ for some $m \in \{1, \ldots, \ell\}$ and $f_{m+1}, \ldots, f_\ell \in \mathcal{E}(T)$, then*

$$\lim_{N \to \infty} \mathbb{E}_{n \in [N]} T^{a_1(n)} f_1 \cdot \ldots \cdot T^{a_\ell(n)} f_\ell = 0 \ \text{ in } L^2(\mu).$$

  It is called good for seminorm estimates, if it is good for seminorm estimates for every ergodic system. In the previous, $|||\cdot|||_s$ denotes the s-th Gowers-Host-Kra seminorm.

- *good for equidistribution for the system (X, μ, T), if for all $t_1, \ldots, t_\ell \in Spec(T)$, not all of them 0, we have*

$$\lim_{N \to \infty} \mathbb{E}_{n \in [N]} e(a_1(n)t_1 + \ldots + a_\ell(n)t_\ell) = 0. \tag{3}$$

  It is called good for equidistribution, if it is good for equidistribution for every ergodic system, or equivalently if (3) holds for all $t_1, \ldots, t_\ell \in [0, 1)$, not all of them 0.

It turns out that the notions of joint ergodicity, good for seminorm estimates and good for equidistribution are closely connected, as shown by the main theorem of the paper:

Theorem 4. *Let (X, μ, T) be an ergodic system and $a_1, \ldots, a_\ell : \mathbb{N} \to \mathbb{Z}$ be a collection of sequences. Then the following conditions are equivalent:*

(i) $a_1, \ldots, a_\ell$ are jointly ergodic for (X, μ, T)

(ii) $a_1, \ldots, a_\ell$ are good for seminorm estimates and equidistribution for (X, μ, T).

If $a_1, \ldots, a_\ell : \mathbb{N} \to \mathbb{Z}$ are jointly ergodic for the ergodic system (X, μ, T), then by applying (2) for eigenfunctions of they system, we get that (3) holds for all $t_1, \ldots, t_\ell \in Spec(T)$, not all 0. Hence $a_1, \ldots, a_\ell$ are good for equidistribution for (X, μ, T). In addition, since $|||f|||_1 = 0$ implies $\int_X f d\mu = 0$, again using (2) we get that $a_1, \ldots, a_\ell$ is good for seminorm estimates for (X, μ, T). This proves the implication $(i) \implies (ii)$.
It is the converse implication that is a rather surprising fact, and this is really the context of Theorem 4.
As an application of Theorem 4, one gets the following strong multiple recurrence property, which is not shared for example by linear sequences:

Corollary 5. *Let $a_1, \ldots, a_\ell : \mathbb{N} \to \mathbb{Z}$ be sequences that are good for equidistribution and good for seminorm estimates for the system (X, μ, T). Then for every set $A \in \mathcal{X}$ we have*

$$\lim_{N\to\infty} \mathbb{E}_{n\in[N]}\mu\left(A \cap T^{-a_1(n)}A \cap \ldots \cap T^{-a_\ell(n)}A\right) \geq (\mu(A))^{\ell+1}.$$

Invoking Furstenberg's correspondence principle, (see [Fu2], Lemma 3.17), and using Corollary 5, one gets that for every $\Lambda \subset \mathbb{N}$,

$$\liminf_{N\to\infty} \mathbb{E}_{n\in[N]}\overline{d}\left(\Lambda \cap (\Lambda - a_1(n)) \cap \ldots \cap (\Lambda - a_\ell(n))\right) \geq \left(\overline{d}(\Lambda)\right)^{\ell+1},$$

where for a set $E \subset \mathbb{N}$, $\overline{d}(E)$ denotes its upper density and is defined by $\overline{d}(E) := \limsup_{N\to\infty}|\Lambda \cap [1, N]|/N$.
We will now state another application of Theorem 4, which is about nilsystems. Before stating it, let us first remind the reader of what a nilsystem is: A k-step nilsystem is a system of the form (X, m_X, T_a), where $X = G/\Gamma$ is a k-step nilmanifold (i.e. G is a k-step nilpotent Lie group and Γ is a discrete cocompact subgroup of G), $a \in G$, $T_a : X \to X$ is defined by $T_a(g\Gamma) = (ag)\Gamma$, for every $g \in G$, and m_X is the normalised Haar measure on X.
If (X, μ, T) is an ergodic k-step nilsystem, then $|||\cdot|||_{k+1}$ is a norm on $L^\infty(\mu)$. Therefore, if $f \in L^\infty(\mu)$ has $|||f|||_{k+1} = 0$, then $f = 0$ μ-almost everywhere on X. Hence, any collection of sequences is good for seminorm estimates for the system (X, μ, T). Combining the previous with Theorem 4, we obtain the following:

Corollary 6. *Let (X, μ, T) be an ergodic nilsystem, and let $a_1, \ldots, a_\ell : \mathbb{N} \to \mathbb{Z}$ be a collection of sequences. Then the following are equivalent:*

(i) $a_1, \ldots, a_\ell$ *are jointly ergodic for* (X, μ, T).

(ii) $a_1, \ldots, a_\ell$ *are good for equidistribution for* (X, μ, T).

References

[Fr1] N. Frantzikinakis, *Joint ergodicity of sequences.* Advances in Mathematics **417** (2023), no. 108918, 63 pp.;

[Fr2] N. Frantzikinakis, *Multiple recurrence and convergence for Hardy sequences of polynomial growth.* J. Analyse Math. **112** (2010), pp. 79-135;

[FK] N. Frantzikinakis, B. Kra, *Polynomial averages converge to the product of integrals.* Isr. J. Math. **148** (2005), pp. 267-276; fkanm

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

[Fu2] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory.* Princeton University Press, Princeton, 1981

Andreas Mountakis, University of Warwick
email: andreas.mountakis.1@warwick.ac.uk

23 Norm convergence of nilpotent ergodic averages

After M. N. Walsh [W]

A summary written by Lars Niedorf and Chiara Paulsen

Abstract

We prove L^2-convergence for multiple polynomial ergodic averages associated with a sequence of measure preserving transformations possessing a nilpotent group structure.

The main result of [W] is the following:

Theorem 1. *Let G be a nilpotent group of measure preserving transformations on a probability space $(X, \mathcal{X}, \mu)$. Then, for every $T_1, \dots, T_l \in G$, the averages*

$$\frac{1}{N} \sum_{n=1}^{N} \prod_{j=1}^{d} \left(T_1^{p_{1,j}(n)} \cdots T_l^{p_{l,j}(n)} \right) f_j$$

always converge in $L^2(X, \mathcal{X}, \mu)$ for every $f_1, \dots, f_d \in L^\infty(X, \mathcal{X}, \mu)$ and every set of integer valued polynomials $p_{i,j}$.

We write the action of $T_1^{p_{1,j}(n)} \cdots T_l^{p_{l,j}(n)}$ on f_j as the (componentwise) action of a system $\mathbf{g} = (g_1, \dots, g_j)$ of G-sequences $g_i : \mathbb{Z} \to G$ on the vector of functions $(f_1, \dots, f_j)$. We set $[N] = \{1, \dots, N\}$ and use the notations

$$\mathcal{A}_N^{\mathbf{g}}[f_1, \dots, f_j] = \mathbb{E}_{n \in [N]} \prod_{i=1}^{j} g_i(n) f_i = \frac{1}{N} \sum_{n=1}^{N} \prod_{i=1}^{j} g_i(n) f_i \quad \text{and}$$

$$\mathcal{A}_{N,N'}^{\mathbf{g}}[f_1, \dots, f_j] = \mathcal{A}_{N'}^{\mathbf{g}}[f_1, \dots, f_j] - \mathcal{A}_N^{\mathbf{g}}[f_1, \dots, f_j].$$

To explain the idea of the proof, we start with some manipulations. Given some functions $f_1, \dots, f_{j-1}, u \in L^\infty(X)$, we observe that

$$\|\mathcal{A}_N^{\mathbf{g}}[f_1, \dots, f_{j-1}, u]\|_2^2 = \langle u, \sigma \rangle,$$

where

$$\sigma = \mathbb{E}_{n\in[N]} g_j(n)^{-1} \mathcal{A}_N^{\mathbf{g}}[f_1,\ldots,f_{j-1},u] \prod_{i=1}^{j-1} g_j(n)^{-1} g_i(n) f_i.$$

Replacing $[N]$ by $[N]+l$ in the summation, we see that the right-hand side changes only by a small magnitude if $l/N \ll 1$, meaning that

$$\left\| \sigma - \mathbb{E}_{n\in[N]} g_j(l+n)^{-1} \mathcal{A}_N^{\mathbf{g}}[f_1,\ldots,f_{j-1},u] \prod_{i=1}^{j-1} g_j(l+n)^{-1} g_i(l+n) f_i \right\|_\infty$$

is small. Applying $g_j(l)$, this equals

$$\left\| g_j(l)\sigma - \mathbb{E}_{m\in[M]} \left(\langle g_j | 1_G\rangle_m(l)\right) b_0 \prod_{i=1}^{j-1} \left(\langle g_j | g_i\rangle_m(l)\right) b_i \right\|_\infty \tag{1}$$

where $b_0 := \mathcal{A}_N^{\mathbf{g}}[f_1,\ldots,f_{j-1},u]$ and $b_i := f_i$, and where we use for a given pair $g,h:\mathbb{Z}\to G$ of G-sequences the notation

$$\langle g|h\rangle_m(n) := g(n)g(n+m)^{-1}h(n+m), \quad m\in\mathbb{Z}.$$

In (1), we see that, up to some small error, we may recover the action of g_j on σ in terms of the system

$$\mathbf{g}_m^* = \left(g_1,\ldots,g_{j-1},\langle g_j|1_G\rangle_m,\langle g_j|g_1\rangle_m,\ldots,\langle g_j|g_{j-1}\rangle_m\right),$$

which is called the *m-reduction* of $\mathbf{g}$.

23.1 Convergence holds for systems of finite complexity

In the following, $C^* > 0$ is a constant depending only on some $\varepsilon > 0$. Given an integer $L > 0$, we say that $\sigma \in L^\infty(X)$ with $\|\sigma\|_\infty \le 1$ is *L-reducible* (with respect to $\mathbf{g}$) if there exists some integer $M > 0$ and $b_0, b_1, \ldots, b_{j-1} \in L^\infty(X)$ with $\|b_i\|_\infty \le 1$ such that for every positive integer $l \le L$ the quantity (1) is smaller than $\varepsilon/(16C^*)$.
By similar arguments as above, one can show the following:

Theorem 2 (Weak inverse result for ergodic averages)**.** *Let $u \in L^\infty(X)$, $\varepsilon > 0$, $1 \le C \le C^*$ and $f_1,\ldots,f_{j-1}$ with $\|f_i\|_\infty < 1$ such that $\|u\|_\infty < 3C$ and $\|\mathcal{A}_N^{\mathbf{g}}[f_1,\ldots,f_{j-1},u]\|_2 > \varepsilon/6$. Then there exists $0 < c_1 < 1$ such that for every $L < c_1 N$ there exists an L-reducible function σ with $\langle u,\sigma\rangle > \varepsilon^2/(2^2 3^3 C)$.*

We say a system $\mathbf{g} = (g_1, \ldots, g_j)$ has *complexity* 0 if $\{g_1, \ldots, g_j\} = \{1_G\}$. Moreover, we say that $\mathbf{g}$ has *complexity* $d \in \mathbb{N}$ if it is not of complexity d' for any $d' \in \{0, \ldots, d-1\}$ and if it is equivalent to a system $\mathbf{h} = (h_1, \ldots, h_k)$ (i.e. $\{g_1, \ldots, g_j\} = \{h_1, \ldots, h_k\}$) for which the m-reduction $\mathbf{h}_m^*$ is of complexity $\leq d-1$ for every $m \geq 0$.

Theorem 3. *Let $F : \mathbb{N} \to \mathbb{N}$ non-decreasing with $F(N) \geq N$ for all $N \in \mathbb{N}$, $\varepsilon > 0$ and $d, M \in \mathbb{N}$. Then there exist $M^{\varepsilon,F,d}, K_{\varepsilon,d} \in \mathbb{N}$ and a sequence of integers*

$$M \leq M_1^{\varepsilon,F,d} \leq \cdots \leq M_{K_{\varepsilon,d}}^{\varepsilon,F,d} \leq M^{\varepsilon,F,d} \tag{2}$$

such that for any system $\mathbf{g} = (g_1, \ldots, g_j)$ of complexity $\leq d$ and $f_1, \ldots, f_j \in L^\infty(X)$ with $\|f_1\|_\infty, \ldots, \|f_j\|_\infty \leq 1$, there exists some $1 \leq i \leq K_{\varepsilon,d}$ such that

$$\|\mathcal{A}_{N,N'}^{\boldsymbol{g}}[f_1, \ldots, f_j]\|_2 \leq \varepsilon$$

for every $M_i^{\varepsilon,F,d} \leq N, N' \leq F(M_i^{\varepsilon,F,d})$.

Sketch of proof. For sufficiently large L, one can show that f_j can be decomposed into a 'structured part' $\sum_{t=0}^{k-1} \lambda_t \sigma_t$ of L-reducible functions σ_t with $\sum_{t=0}^{k-1} |\lambda_t| < C^*$ and a well-behaved part, called 'random part' (which is a similar approach to that of [T]). We first assume that f_j consists only of the structured part and prove the statement by induction over the complexity d of the system $\mathbf{g}$.
For systems of complexity 0, the statement is trivially true. For the inductive step, let $M_0 \in \mathbb{N}$. For every $t \in \{0, \ldots, k-1\}$, let $b_0^{(t)}, \ldots, b_{j-1}^{(t)} \in L^\infty(X)$ and $M^{(t)} \in \mathbb{N}$ be the objects from the definition of L-reducibility corresponding to σ_t. If we replace every instance of $g_j(l)\sigma_t$ in $\mathcal{A}_N^{\mathbf{g}}[f_1, \ldots, f_{j-1}, \sigma_t]$ with

$$\mathbb{E}_{m \in [M^{(t)}]} \left(\langle g_j | 1_G \rangle_m(l)\right) b_0^{(t)} \prod_{i=1}^{j-1} \left(\langle g_j | g_i \rangle_m(l)\right) b_i^{(t)},$$

we get by the definition of L-reducibility that

$$\begin{aligned} &\|\mathcal{A}_{N,N'}^{\mathbf{g}}[f_1, \ldots, f_j]\|_2 \\ &\leq \sum_{t=0}^{k-1} |\lambda_t| \mathbb{E}_{m \in [M^{(t)}]} \|\mathcal{A}_{N,N'}^{\mathbf{g}_m^*}[f_1, \ldots, f_{j-1}, b_0^{(t)}, \ldots, b_{j-1}^{(t)}]\|_2 + \varepsilon/8. \end{aligned} \tag{3}$$

Since $\mathbf{g}_m^*$ is a system of complexity $\leq d-1$, the induction hypothesis yields for every summand in (3) a sequence of integers $i \mapsto M_i^{\gamma,F,d-1}$ as in (2) starting at M_0 such that the contribution by the averages is smaller than $\gamma = \varepsilon/(16C^*)$ for all $N, N' \in [M_i^{\gamma,F,d-1}, F(M_i^{\gamma,F,d-1})]$ and for some $1 \leq i \leq K_{\gamma,d-1} =: K$. However, the index i of the interval depends on t and therefore one has to construct a finer sequence of integers to gain control over every average simultaneously. Roughly speaking, this is done by selecting $r \in \mathbb{N}$ large enough and suitable non-decreasing functions $F_1, ..., F_r : \mathbb{N} \to \mathbb{N}$ with $F_r = F$. Via the pigeonhole principle, there is an interval $[M_{i_1}^{\gamma,F_1,d-1}, F_1(M_{i_1}^{\gamma,F_1,d-1})]$ where the averages on the right-hand side of (3) that aren't bounded by γ sum up to at most $(K-1)C^*/K$. Set $M^{(i_1)} := M_{i_1}^{\gamma,F_1,d-1}$. Repeating this process for the sequence in (2) that starts at $M^{(i_1)}$ and corresponds to the parameters $\gamma, d-1$ and F_2 yields an integer $M^{(i_1,i_2)} := (M^{(i_1)})_{(i_2)}^{\gamma,d-1,F_2}$ such that $[M^{(i_1,i_2)}, F_2(M^{(i_1,i_2)})] \subseteq [M^{(i_1)}, F_1(M^{(i_1)})]$ and the averages in (3) that aren't bounded by γ sum up to at most $C^*(K-1)^2/K^2$. After r repetitions, one obtains an integer $M^{(i_1,\ldots,1_r)}$ such that the left-hand side of (3) is bounded by $(K-1)^r/K^r + \varepsilon/8$ for all $N, N' \in [M^{(i_1,\ldots,1_r)}, F(M^{(i_1,\ldots,1_r)})]$.
Let now $f_j \in L^\infty(X)$. For $L \in \mathbb{N}$, let Σ_L^+ be the set of L-reducible functions with L^2-norm smaller than $\varepsilon/(2^5 3C^*)$. It can be shown that $(\|\cdot\|_L)_{L\in\mathbb{N}}$ with

$$\|f\|_L := \inf\left\{ \sum_{t=0}^{k-1} |\lambda_t| : f = \sum_{t=0}^{k-1} \lambda_t \sigma_t, \sigma_0, \ldots, \sigma_{k-1} \in \Sigma_L^+ \right\}$$

defines an equivalent family of norms on $L^2(X)$ and its dual norms satisfy $\|f\|_L^* = \sup_{\sigma\in\Sigma_L^+} |\langle f, \sigma\rangle|$. Furthermore, there exists $0 < c_1 < 1$ and integers $M \geq M_0$, $1 \leq C_i \leq C^*$ such that f_j decomposes into

$$f_j = \tilde{f} + u + v$$

where $\|\tilde{f}\|_B < C_i, \|u\|_A^* < \varepsilon^2/(2^3 3^3 C_i), \|v\|_2 < \varepsilon/(2^5 3)$ for $1 \leq A < c_1 M < M \leq B$. This follows from a general result about families of equivalent norms on Hilbert spaces whose dual norms are decreasing, which is proven by a version of the Hahn-Banach separation theorem [W, Proposition 2.3]. We obtain $\tilde{f} = \sum_{t=0}^{k-1} \lambda_t \sigma_t$ with $\sigma_0, \ldots, \sigma_{k-1} \in \Sigma_B^+$ and $\sum_{t=0}^{k-1} |\lambda_t| \leq C_i$, and $|\langle \sigma, u\rangle| < \varepsilon^2/(2^3 3^3 C_i)$ for every $\sigma \in \Sigma_A$. For u and v, we get

$$\|\mathcal{A}_{N,N'}^{\mathbf{g}}[f_1, \ldots, f_{j-1}, v]\|_2 \leq \varepsilon/(2^4 3) \quad \text{and} \quad \|\mathcal{A}_{N,N'}^{\mathbf{g}}[f_1, \ldots, f_{j-1}, u]\|_2 \leq 2\varepsilon/3,$$

where the second estimate follows from the weak inverse result above. The 'structured part' $\tilde{f}$ may now be treated as in the beginning of the proof. □

23.2 Every polynomial system has finite complexity

Given $m \in \mathbb{Z}$, we consider the operator D_m acting on G-sequences given by

$$(D_m g)(n) = g(n)g(n+m)^{-1}.$$

We say that a G-sequence g is *polynomial* if there is some integer $d > 0$ such that $D_{m_1} \dots D_{m_d} g = 1_G$ for every choice of integers $m_i > 0$. Accordingly, we call a system $\mathbf{g} = (g_1, \dots, g_j)$ *polynomial* if every entry is polynomial.
Now suppose that G is nilpotent of order s with lower central series $G = G_1 \supset \dots \supset G_s \supset G_{s+1} = \{1_G\}$. We say that a polynomial sequence g has *degree* $\leq d = (d_1, \dots, d_s)$ if $D_{m_1} \dots D_{m_{d_k+1}} g(n) \in G_{k+1}$ for every $n \in \mathbb{Z}$, $1 \leq k < s$ and $m_i \in \mathbb{Z}$. Accordingly, we say that system $\mathbf{g} = (g_1, \dots, g_j)$ has *degree* $\leq d$ if every entry has degree $\leq d$.

Theorem 4. *If $\mathbf{g} = (g_1, \dots, g_j)$ is a polynomial system of size $|\mathbf{g}| := j \leq C_1$ and degree $\leq \overline{d}$, then it has finite complexity of order $O_{C_1,\overline{d}}(1)$.*

Proof. The proof works by induction on the degree $\overline{d}$ by showing simultaneously the following two statements:

(i) one can go from $\mathbf{g}$ to the trivial system in $O_{C_1,\overline{d}}(1)$ steps,

(ii) one can go from $\mathbf{g}$ to a system consisting of a single sequence of degree $\leq \overline{d}$ in $O_{C_1,\overline{d}}(1)$ *complete* steps,

where ???going from one system to another??? is performed by appropriate m-reductions $\mathbf{g}_m^*$, while ???going from one system to another in complete steps??? has to performed by the so-called *complete reductions* $\mathbf{g}_m^{**}$ given by $\mathbf{g}_m^{**} = \mathbf{g}_m^* \setminus \{\langle g_j | 1_G \rangle\}$. Now, for the inductive step from $\overline{d} - 1$ to $\overline{d}$, we write

$$\mathbf{g} = \mathbf{h}_0 \cup \bigcup_{i=1}^{l} s_i \mathbf{h}_i \tag{4}$$

where s_i has degree $\overline{d}$ and each $\mathbf{h}_i$ has degree $\leq \overline{d} - 1$. (For instance, one may choose $\mathbf{h}_0 = \emptyset$, $s_i = g_i$, and $\mathbf{h}_i = (1_G)$ being the system containing only the constant sequence 1_G.) For passing from $\mathbf{g}$ to its m-reduction $\mathbf{g}_m^*$, we observe that $\langle sh_j|sh_i\rangle = s\langle h_j|h_i\rangle$, as well as

$$\begin{aligned}\langle s_jh_j|s_ih_i\rangle_m &= D_m(s_jh_j)(D_m(s_ih_i))^{-1}s_ih_i\\ &= s_iD_m(s_jh_j)(D(s_ih_i))^{-1}[D_m(s_jh_j)(D_m(s_ih_i))^{-1}, s_i]h_i = s_ih^{j,i},\end{aligned}$$

where $[\cdot,\cdot]$ denotes the commutator of two group elements. One can show that the degree of a polynomial sequence gets lowered by 1 when applying D_m, is closed under multiplication, and behaves additive under the commutator relation $[\cdot,\cdot]$, see [W, Lemma 4.1], which implies in particular that the polynomial sequence $h^{j,i}$ has degree $\overline{d}-1$. We obtain that the reduction $\mathbf{g}_m^*$ of $\mathbf{g}$ is equivalent to

$$\mathbf{h}_0^{(1)} \cup \left(\bigcup_{i=1}^{l-1} s_i\mathbf{h}_i^{(1)}\right) \cup s_l\mathbf{h}_l^{**},$$

where $\mathbf{h}_l^{**} = (h_{l,1},\ldots,h_{l,j_l-1},\langle h_{l,j_l}|h_{l,1}\rangle,\ldots,\langle h_{l,j_l}|h_{l,j_l-1}\rangle)$. As $\mathbf{h}_l^{**}$ has degree $\leq \overline{d}-1$, we may apply the induction hypothesis (ii) to pass to the system

$$\mathbf{h}_0^{(2)} \cup \left(\bigcup_{i=1}^{l-1} s_i\mathbf{h}_i^{(2)}\right) \cup s_l\mathbf{h}, \tag{5}$$

where $\mathbf{h}_i^{(2)}$ has degree $\leq \overline{d}-1$ and $\mathbf{h}$ is a system consisting of a single sequence of degree $\leq \overline{d}-1$. Because of that, the reduction of (5) will be of the form

$$\mathbf{h}_0^{(3)} \cup \left(\bigcup_{i=1}^{l-1} s_i\mathbf{h}_i^{(3)}\right). \tag{6}$$

Comparing (4) with (6), we see that we have ultimately discarded the sequence s_l in (4). Iterating this argument, we may thus also discard the sequences s_{l-1}, s_{l-2}, etc. until we end up with a system $\mathbf{h}_0^{(k)}$, where $k = O_{C_1,\overline{d}}(1)$, to which we can apply the induction hypothesis (i). The inductive step for showing (ii) for polynomial sequences of degree $\overline{d}$ follows similar. □

References

[T] Tao, T., *Norm convergence of multiple ergodic averages for commuting transformations.* Ergodic Theory Dynam. Systems **28** (2008), no. 2, 657–688.

[W] Walsh, M. N., *Norm convergence of nilpotent ergodic averages.* Ann. of Math. (2) **175** (2012), no. 3, 1667–1688.

Lars Niedorf, Kiel University
email: niedorf@math.uni-kiel.de

Chiara Paulsen, Kiel University
email: stu217380@mail.uni-kiel.de

24 Cancellation for the multilinear Hilbert transform

After T. Tao [T]

A summary written by Wojciech Słomian

Abstract

We obtain some improvement over the trivial upper bound for the k-linear truncated Hilbert transform. We obtain the bound of order $o(\log \frac{R}{r})$ as $\frac{R}{r} \to \infty$ instead of the trivial $\mathcal{O}(\log \frac{R}{r})$. This shows that we may expect some cancellation properties in the k-linear Hilbert transform H_k.

24.1 Introduction

Let $k \geq 1$ be a fixed natural number. For any sequence of the Schwartz functions $f_1, \dots, f_k : \mathbb{R} \to \mathbb{C}$ we define the *k-linear Hilbert transform* by setting

$$H_k(f_1, \dots, f_k)(x) := \text{p.v.} \int_{\mathbb{R}} f_1(x+t) \dots f_k(x+kt) \frac{\mathrm{d}t}{t}, \quad x \in \mathbb{R}.$$

In the case $k = 1$ we get that H_1 the standard Hilbert transform. It is well-known that it maps $L^p(\mathbb{R})$ to itself for any $p > 1$. For $k = 2$ we obtain the bilinear Hilbert transform given by

$$H_2(f, g)(x) := \text{p.v.} \int_{\mathbb{R}} f(x+t) g(x+2t) \frac{\mathrm{d}t}{t}, \quad x \in \mathbb{R}.$$

In their groundbreaking work Lacey and Thiele [LT] have shown that H_2 maps $L^{p_1}(\mathbb{R}) \times L^{p_2}(\mathbb{R})$ to $L^p(\mathbb{R})$ whenever $1 < p, p_1, p_2 < \infty$[12] and $\frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p}$. The question about boundedness of H_k for $k > 3$ remains open. One way to approach the problem of the boundedness of the k-linear Hilbert transform is to study its truncated form defined by

$$H_{k,r,R}(f_1, \dots, f_k)(x) := \int_{r \leq |t| \leq R} f_1(x+t) \dots f_k(x+kt) \frac{\mathrm{d}t}{t}, \quad x \in \mathbb{R}. \tag{1}$$

[12] Actually they have shown that we may take $p > 2/3$ instead of $p > 1$.

Then one has

$$H_k(f_1,\dots,f_k)(x) = \lim_{r\to 0, R\to\infty} H_{k,r,R}(f_1,\dots,f_k)(x), \quad x\in\mathbb{R},$$

which allows to study the operator $H_{k,r,R}$ instead of H_k. Namely, the $L^p(\mathbb{R})$ boundedness of H_k may be deduced from the uniform inequality

$$\|H_{k,r,R}(f_1,\dots,f_k)\|_{L^p(\mathbb{R})} \le C_{k,p,p_1,\dots,p_k}\|f_1\|_{L^{p_1}(\mathbb{R})}\dots\|f_k\|_{L^{p_k}(\mathbb{R})}, \tag{2}$$

where $1 < p_1,\dots,p_k, p < \infty$ are such that $\frac{1}{p} = \frac{1}{p_1} + \cdots \frac{1}{p_k}$ and the constant $C_{k,p,p_1,\dots,p_k}$ is independent of R and r. Clearly, by Minkowski's integral inequality and Hölder's inequality we may get the *trivial bound*, that is

$$\frac{\|H_{k,r,R}(f_1,\dots,f_k)\|_{L^p(\mathbb{R})}}{\|f_1\|_{L^{p_1}(\mathbb{R})}\dots\|f_k\|_{L^{p_k}(\mathbb{R})}} \le \int_{r\le|t|\le R}\frac{\mathrm{d}t}{|t|} = 2\log(R/r).$$

Obviously, this does not proves the inequality (2) since this bound depends on R and r. The following theorem is a non-trivial improvement of the above estimate.

Theorem 1. *Let $k\ge 1$ be fixed and let $1 < p_1,\dots,p_k, p < \infty$ be such that $\frac{1}{p} = \frac{1}{p_1} + \cdots + \frac{1}{p_k}$, and let $\varepsilon > 0$. Then, if R/r is sufficiently large depending on $\varepsilon, k, p_1,\dots,p_k, p$, one has*

$$\|H_{k,r,R}(f_1,\dots,f_k)\|_{L^p(\mathbb{R})} \le \varepsilon\log\frac{R}{r}\prod_{i=1}^k\|f_i\|_{L^{p_i}(\mathbb{R})}$$

for all $f_i\in L^{p_i}(\mathbb{R})$, $i=1,\dots,k$.

24.2 Basic reductions

By using the standard transference arguments it is enough to prove the discrete counterpart of Theorem 1. Namely, we prove the following result.

Theorem 2. *Let $k\ge 1$ be fixed and let $1 < p_1,\dots,p_k, p < \infty$ be such that $\frac{1}{p} = \frac{1}{p_1} + \cdots + \frac{1}{p_k}$, and let $\varepsilon > 0$. Then, if $R\ge r\ge 1$ and R/r is sufficiently large depending on $\varepsilon, k, p_1,\dots,p_k, p$, one has*

$$\left\|\sum_{\substack{t\in\mathbb{Z}\\ r\le|t|\le R}}\frac{f_1(x+t)\dots f_k(x+kt)}{t}\right\|_{\ell^p(\mathbb{Z},\mathrm{d}x)} \le \varepsilon\log\frac{R}{r}\prod_{i=1}^k\|f_i\|_{\ell^{p_i}(\mathbb{Z})}$$

for all $f_i\in\ell^{p_i}(\mathbb{Z})$, $i=1,\dots,k$.

The advantage of the discrete setting is the possibility of using arithmetic regularities and counting lemmas which are not available in the continuous setting. Next, by duality and multilinear interpolation we are reduced to show that for any collection of finite subsets $E_0, E_1, \dots E_k \subset \mathbb{Z}$ which satisfy[13]

$$|E_i| \eqsim_\varepsilon |E_0|, \quad \text{for each } i = 1, \dots, k \tag{3}$$

we have

$$\left| \sum_{x\in\mathbb{Z}} \sum_{\substack{t\in\mathbb{Z} \\ r\le|t|\le R}} \frac{\mathbb{1}_{E_0}(x)\mathbb{1}_{E_1}(x+t)\dots\mathbb{1}_{E_k}(x+kt)}{t} \right| \lesssim \varepsilon \log\frac{R}{r}|E_0| \tag{4}$$

with the implicit constant being independent of $\varepsilon > 0$.
In the first step we localize the quantity on the left hand side of (4) in the t variable. Let $\psi : \mathbb{R} \to \mathbb{R}$ be a fixed smooth odd function supported on $[-2, -1/2] \cup [1/2, 2]$ which satisfies

$$\sum_{n\in\mathbb{Z}} 2^{-n}\psi(2^{-n}t) = \frac{1}{t}, \quad t \neq 0.$$

Then in order to prove (4) it is enough to establish

$$\sum_{n:r\le 2^n\le R} 2^{-n}\left| \sum_{x,t\in\mathbb{Z}} \mathbb{1}_{E_0}(x)\mathbb{1}_{E_1}(x+t)\dots\mathbb{1}_{E_k}(x+kt)\psi(t/2^n) \right| \lesssim \varepsilon\log\frac{R}{r}|E_0|.$$

In the next step we localize the x variable by using dyadic intervals. Let $\varphi\colon \mathbb{R}\to\mathbb{R}$ be a smooth function supported on $[-1, 1]$ such that $\sum_{j\in\mathbb{Z}} \varphi(x-j) = 1$ for each $x \in \mathbb{R}$. For any dyadic interval $I := \{x \in \mathbb{Z} : j2^n < x \le (j+1)2^n\}$ with $n \ge 0$ we set

$$a_I := 2^{-2n}\left| \sum_{x,t\in\mathbb{Z}} \mathbb{1}_{E_0}(x)\mathbb{1}_{E_1}(x+t)\dots\mathbb{1}_{E_k}(x+kt)\psi(t/2^n)\varphi(2^{-n}x-j) \right|$$

so by triangle's inequality in order to prove (4) it suffices to show that

$$\sum_{\substack{I-\text{dyadic} \\ r\le|I|\le R}} a_I|I| \lesssim \varepsilon\log\frac{R}{r}|E_0| \tag{5}$$

where the sum taken over dyadic intervals I of length between r and R.

[13] We write $A \eqsim_\varepsilon B$ if there exist constants $c_\varepsilon, C_\varepsilon$ such that $c_\varepsilon A \le B \le C_\varepsilon A$.

24.3 Arithmetic regularity

Now our aim is to show (5). For this purpose we make use of the following result.

Lemma 3. *Let $R \geq r \geq 1$. Then for any $\theta \in (1, \infty)$ we have*

$$\sum_{\substack{I-\text{dyadic}\\ r\leq |I|\leq R}} a_I^{\theta/2}|I| \leq C_{\varepsilon,\theta} \log \frac{R}{r}|E_0|,$$

for some constant $C_{\varepsilon,\theta} > 0$.

Proof. For any I we have $a_I \lesssim_{\varphi,\psi} \inf_{y\in I} \mathcal{M}_{\mathrm{HL}}\mathbb{1}_{E_0}(y)\mathcal{M}_{\mathrm{HL}}\mathbb{1}_{E_1}(y)$ where $\mathcal{M}_{\mathrm{HL}}f$ is the discrete Hardy–Littlewood maximal function. Consequently, we may write

$$a_I^{\theta/2}|I| \lesssim \sum_{x\in I} \mathcal{M}_{\mathrm{HL}}\mathbb{1}_{E_0}(x)^{\theta/2}\mathcal{M}_{\mathrm{HL}}\mathbb{1}_{E_1}(x)^{\theta/2}.$$

Now since every $x \in \mathbb{Z}$ is an element of $\mathcal{O}(\log \frac{R}{r})$ dyadic intervals we estimate

$$\sum_{\substack{I-\text{dyadic}\\ r\leq |I|\leq R}} a_I^{\theta/2}|I| \lesssim \log\frac{R}{r} \sum_{x\in I} \mathcal{M}_{\mathrm{HL}}\mathbb{1}_{E_0}(x)^{\theta/2}\mathcal{M}_{\mathrm{HL}}\mathbb{1}_{E_1}(x)^{\theta/2}.$$

Now the desired result follows by the Cauchy–Schwarz inequality, the Hardy–Littlewood maximal estimate, and by condition (3). □

We use the above result with $\theta = 3/2$ to get that for any $\delta > 0$ we have

$$\sum_{\substack{I-\text{dyadic}\\ r\leq |I|\leq R, a_I\leq\delta}} a_I^{\theta/2}|I| \lesssim_\varepsilon \delta^{1/4} \log\frac{R}{r}|E_0|$$

so it is enough to prove that for any $\delta > 0$ one has

$$\sum_{\substack{I-\text{dyadic}\\ r\leq |I|\leq R, a_I>\delta}} a_I^{\theta/2}|I| \lesssim \varepsilon \log\frac{R}{r}|E_0|, \tag{6}$$

whenever R/r is sufficiently large depending on ε, δ.

In order to establish (6) we make use of the arithmetic regularity lemma proven by Green and Tao [GT]. Before we state it we need to introduce the notion of the *Gowers uniformity norms.* Let $f\colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ be a function on the cyclic group $\mathbb{Z}/N\mathbb{Z}$. We define the Gowers norm of order $j \geq 1$ of the function f as

$$\|f\|_{U^j(\mathbb{Z}/N\mathbb{Z})} := \Big(\frac{1}{N^{j+1}} \sum_{h_1,\dots,h_j,x\in\mathbb{Z}/N\mathbb{Z}} \Delta_{h_1}\dots\Delta_{h_j} f(x)\Big)^{2^j}$$

where $\Delta_h f(x) := f(x+h)\overline{f(x)}$. If the function $f\colon \{1,2,\dots,N\} \to \mathbb{C}$ is defined on the set $\{1,2,\dots,N\}$ we define the Gowers norm of f as

$$\|f\|_{\mathcal{U}^j(N)} := \frac{\|f\|_{U^j(\mathbb{Z}/N'\mathbb{Z})}}{\|\mathbb{1}_{\{1,2,\dots,N\}}\|_{U^j(\mathbb{Z}/N'\mathbb{Z})}}, \qquad \text{for any } N' \geq 2^j N,$$

where we have embed the set $\{1,2,\dots,N\}$ into $\mathbb{Z}/N\mathbb{Z}$ and extent f by zero outside. Let us note that the above definition does not depend on the choice of N'.

Theorem 4. *Let $d \geq 1$ and let $f\colon \{1,2,\dots,N\} \to [0,1]^d$. Let $s \geq 1$ be an integer, let $\varepsilon > 0$, and let $\mathcal{F} : \mathbb{R}^+ \to \mathbb{R}^+$ be a monotone increasing function with $\mathcal{F}(M) \geq M$ for all $M > 0$. Then there exists a quantity $M = \mathcal{O}_{s,\varepsilon,\mathcal{F},d}(1)$ and a decomposition*

$$f = f_{\text{nil}} + f_{\text{sml}} + f_{\text{unf}}$$

of function f into functions $f_{\text{nil}}, f_{\text{sml}}, f_{\text{unf}}\colon \{1,2,\dots,N\} \to [-1,1]^d$ which satisfy the following conditions:

1. *The function f_{nil} is of the following form $f_{\text{nil}}(n) := F\big(g(n), n \bmod q, \frac{n}{N}\big)$, where $F : X \times \mathbb{Z}/q/\mathbb{Z} \times \mathbb{R} \to \mathbb{C}^d$ is a function of the Lipschnitz norm[14] at most M, $q \in \{1,2,\dots,M\}$ and $g\colon \mathbb{Z} \to X$ is a "well behaved" function into some filtered manifold X;*

[14] For function $F\colon X \to \mathbb{C}$ defined on a metric space (X,d) the Lipschnitz norm of F is defined as

$$\|f\|_{\text{Lip(X)}} := \|f\|_{L^\infty(X)} + \sup_{\substack{x,y\in X\\ x\neq y}} \frac{|F(x)-F(y)|}{d(x,y)}.$$

2. For each coordinate of the function f_{sml} we have

$$\|(f_{\mathrm{sml}})_j\|_{\ell^2(\{1,2,\ldots,N\})} \leq \varepsilon N^{1/2}, \quad j = 1, \ldots, d,$$

where $(f_{\mathrm{sml}})_j$ denotes the j-th coordinate of f_{sml};

3. For each coordinate of the function f_{sml} we have

$$\|(f_{\mathrm{unf}})_j\|_{\mathcal{U}^{s+1}(N)} \leq \frac{1}{\mathcal{F}(M)}, \quad j = 1, \ldots, d,$$

where $\|\cdot\|_{\mathcal{U}^{s+1}(N)}$ is the Gowers norm of order $s+1$;

4. The functions f_{nil} and $f_{\mathrm{nil}} + f_{\mathrm{sml}}$ take values in $[0,1]^d$.

We apply the above theorem to the function

$$f(n) := (\mathbb{1}_{E_0}(n), \ldots, \mathbb{1}_{E_k}(n)), \text{ for } n \in \{1, 2, \ldots, N\}$$

for some large $N \in \mathbb{N}$. Then we obtain that for each set E_i we have the following decomposition $\mathbb{1}_{E_i} = f_{\mathrm{nil},i} + f_{\mathrm{sml},i} + f_{\mathrm{unf},i}$. We replace each $\mathbb{1}_{E_i}$ in the definition of a_I by the above decomposition. Any term which contains $f_{\mathrm{sml},i}$ or $f_{\mathrm{unf},i}$ can be estimated by using the properties (2) and (3) from Theorem 4. Consequently, we are left to estimate the left hand side of (6) with a'_I defined by

$$a'_I := 2^{-2n} \left| \sum_{x,t \in \mathbb{Z}} f_{\mathrm{nil},0}(x) f_{\mathrm{nil},1}(x+t) \ldots f_{\mathrm{nil},k}(x+kt) \psi(t/2^n) \varphi(2^{-n}x - j) \right|.$$

Since each function $f_{\mathrm{nil},i}$ has a "good structure" we may use the counting lemma of Green and Tao [GT, Theorem 1.11] to, roughly speaking, replace the sum by integrals (with some error terms). Since ψ is odd the one integral is zero while the other one turns out to be a constant. Then after appropriate selection of constants in Theorem 4 we get the desired result.

References

[T] Tao, T., *Cancellation for the multilinear Hilbert transform*, Collect. Math., **67** (2016), 191–206.

[GT] Green, B., Tao, T., An Arithmetic Regularity Lemma, An Associated Counting Lemma, and Applications. An Irregular Mind, 261–334, *Bolyai Soc. Math. Stud.*, J´anos Bolyai Math. Soc., Budapest, 2010.

[LT] Lacey, M. and Thiele, C., *On Calderón's conjecture*, Ann. of Math. (2) **149** (1999), no. 2, 475–496.

Wojciech Słomian, WUST
email: wojciech.slomian@pwr.edu.pl

25 A variational proof of the ergodic theorem

After R. L. Jones, A. Seeger, and J. Wright [JSW]

A summary written by Joe Trate

Abstract

We give a simple proof of Birkhoff's ergodic theorem which is based on Lépingle's inequality and Calderón's transference principle.

25.1 Introduction

Let $(\Omega, \mathcal{F}, \mathbb{P}, T)$ denote a dynamical system. For a function $f : \Omega \to \mathbb{R}$, let

$$A_N f(x) = \frac{1}{n} \sum_{j=1}^{N} f\left(T^j x\right). \tag{1}$$

Birkhoff's ergodic theorem states that if $f \in L^1(\Omega)$, then the sequence of ergodic averages $(A_N f)_{N \in \mathbb{N}}$ converges almost surely. The goal of this talk is to give a proof of this theorem based on variational inequalities. We begin by giving an overview of our proof strategy. First we will use Lépingle's inequality to obtain variational estimates on $L^p(\mathbb{Z})$ for the averaging operators $A_N = \frac{1}{N} \sum_{k=1}^{N} f(m+k)$. We will then use Calderón's transference principle to translate these estimates to the setting of a general dynamical system to obtain the same types of bounds for the strong $r-$variations of the corresponding ergodic averaging operators in that context. Doing so gives a proof of Birkhoff's ergodic theorem, as given any $x \in \Omega$, if $V_r(\mathcal{A}f)(x)$ is finite for some r, then the limit of the ergodic averages will exist as $N \to \infty$.

25.2 Ingredient 1: Lépingle's Inequality

Perhaps the most important tool used to prove variational inequalities is a result due to Lépingle which gives a moment estimate for the pathwise $r-$variation of a martingale. We give a generalization of this result, which we derived by simplifying the proof from, [ZK], utilizing the lack of any weights in our context. Let $\left(\Omega, (\mathcal{F}_n)_{n \in \mathbb{N}}, \mathbb{P}\right)$ be a filtered probability space and $F_\infty = \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n\right)$. For $0 < r < \infty$, a sequence of random variables $X = (X_n)$, and $\omega \in \Omega$, the $r-$variation of X at ω is defined by:

$$V^r X(\omega) = \sup_{u_1 < u_2 < \ldots} \left(\sum_j \left| X_{u_j} - X^2_{u_{j-1}} \right| \right)^{1/r} \tag{2}$$

where the supremum is taken over all increasing sequences. Given any integrable $\mathcal{F}_\infty-$measurable function $X : \Omega \to \mathbb{R}$, we define an associated martingale as follows. Let $(\mathbb{E}_n)_{n\in\mathbb{N}}$ be the sequence of conditional expectation operators with respect to the filtration of the probability space. Then, letting $X_n = \mathbb{E}_n X$ gives a martingale sequence which converges to X. For $r > 2$, we have:

$$\|V^r X\|_{L^p} \le C_p \sqrt{\frac{r}{r-2}} \|X\|_{L^p} \tag{3}$$

Like many results in the theory of martingales, an efficient proof of this result can be obtained from a judicious choice of stopping times. Then, we estimate the $r-$variation pathwise with a linear combination of square functions. these stopping times are:

Definition 1. *Let $M_t = \sup_{t'' \le t' \le t} |X_{t'} - X_{t''}|$. For each $m \in \mathbb{N}$, letting $\tau_0^{(m)}(\omega) = 0,$ and*

$$\tau_{j+1}^{(m)}(\omega) = \inf\{t \ge t_j^{(m)}(\omega) : \left| X_t(\omega) - X_{\tau_j(\omega)}(\omega) \right| \ge 2^{-m} M_t(\omega)\} \tag{4}$$

gives an increasing sequence of stopping times.

25.3 Ingredient 2: Variational Estimates of Averaging Operators:

The second ingredient in our proof are the weak and strong type inequalities for the ergodic averaging operators on $\mathbb{Z}$. To derive them, we follow the structure of [JKR], whose work is based on [Bou89] We deal first with inequalities in L^2. To do so we connect differentiation to martingales by proving the L^2-boundedness of several ergodic square functions. Then we prove that these square functions are of weak type $(1, 1)$, and finally, that several of them map L^∞ to BMO, and are type (p, p) for $1 < p < \infty$.

Letting our dynamical system be as above, and given an increasing sequence $(n_k)_{k\in\mathbb{N}}$ we can define the operator:

$$Of(x) = \mathcal{O}(A_n f(x)) = \left(\sum_{k=1}^{\infty} \sup_{n_{k-1}\le n\le m<n_k} |A_n f(x) - A_m f(x)|^2\right)^{1/2} \tag{5}$$

Post transference, the main results of this paper can be stated as:

Theorem 1. *In any dynamical system, the operator O is weak type (1,1), type (p,p) for $1<p<\infty$ and maps L^∞ into BMO. The constants in the inequality do not depend on the sequence $(n_k)_{k\in\mathbb{N}}$.*

and letting V^r be a above, we have

Theorem 2. *Let $r>2$. Then in any dynamical system the operator V^r is weak type $(1,1)$, type (p,p) for $1<p<\infty$, and maps L^∞ to BMO.*

In the proofs of these theorems, Lépingle's inequality plays a critical role.

25.4 Ingredient 3: Calderón's Transference Principle:

The final ingredient in our proof, is the Calderón tranference principle, which allows us to lift a strong or weak-type inequality from a family of operators on the locally integrable functions on R to a far more general dynamical system. Doing so allows us to, for example, reduce the maximal ergodic theorem to Hardy-Littlewood's maximal theorem. We follow [C], here is the setup: Let X be a $\sigma-$finite measure space and U^t be a one-parameter group of measure-preserving transformations of X. Suppose that for every measurable function $f: X\to\mathbb{R}$ the function $F:\mathbb{R}\times X\to\mathbb{R}$ defined by the equation $F(t,x) = f(U^t x)$ is measurable. We denote by T an operator defined on $L^1_{loc}(\mathbb{R})$ which satisfies the following four properties:

1. T takes values in the space of continuous functions on the real line.
2. T is sublinear.
3. T commutes with translations.
4. T is semilocal, in the sense that there exists an $\epsilon>0$, such that for every $f\in L^1_{loc}(\mathbb{R})$, the support of Tf lies within an $\epsilon-$neighborhood of the support of f.

We define an operator $T^\sharp$ on the space of functions on X as follows. Given a function f defined on X, let $F(x,t) = f\left(U^t x\right)$. When f is the sum of two bounded integrable functions, $F(x,t)$ is locally integrable in t for almost every x, thus the function $G(t,x) = T\left(F\left(t,x\right)\right)$ is a well-defined, $t-$continuous function at almost every x. Thus $g(x) = G(0,x)$ is well-defined and we define $T^\sharp f = g(x)$, then the following theorem holds:

Theorem 3. *Let T_n be a sequence of operators that satisfy the four properties above and suppose that the operator $Sf = \sup|T_n f|$ is of strong or weak type (p,p), $1 \leq p \leq \infty$. Then the same holds for the operator $S^\sharp f = \sup|T_n^\sharp f|$ and $\|^\sharp\| \leq \|S\|$.*

Notice that when

$$\left(Sf\right)(t) = \sup_s \left| \frac{1}{s} \int_0^s f(t+u)du \right|$$

$S^\sharp$ is the maximal ergodic operator, and so, by Hardy-Littlewood's maximal theorem, this operator is of weak type (p,p) for all $p \geq 1$, and strong type (p,p) for $p > 1$. To obtain almost sure convergence results we utilize the following theorem.

Theorem 4. *Let $T_n f = k_n * f$, where k_n is bounded and has bounded support. Suppose that $Sf = \sup_n |T_n f|$ is of weak type (p,p), $1 < p < \infty$, and that $\int k_n(t)\,dt$ converges and $k_n * \phi$ converges in L^1, as $n \to \infty$, for every infinitely differentiable ϕ with compact support and vanishing integral. Then $T_n^\sharp f$ converges almost everywhere in Ω for every f in $L^p(\Omega)$. If, in addition, the operator S is of weak type $(1,1)$, then $T_n^\sharp f$ converges almost everwhere in Ω, for every integrable f.*

Setting $k_n = \frac{1}{n}(-n,0)$ in the preceding theorem reduces proving the ergodic theorem to verifying that the conditions of the above theorem are satisfied, i.e.

1. Observing that that $\int k_n(t)dt$ converges and that $k_n * \phi$ converges in L^1 as $n \to \infty$ for every infinitely differentiable ϕ with compact support and vanishing integral.

2. Proving that the operator S defined by $Sf = \sup_n |\frac{1}{n}\sum_{k=1}^n f(T^k)f|$ is of weak type (p,p) for $1 < p < \infty$.

3. Proving that the operator S is of weak type $(1, 1)$.

We note that the lion's share of the work toward this goal (corresponding to the second and third items on our list) was completed in the second section of our talk.

References

[JSW] Jones, R., Seeger, A., & Wright, J., *Strong variational and jump inequalities in harmonic analysis.* Transactions of the American Mathematical Society 360(12) pp. 6711-6742.

[ZK] Zorin-Kranich, P. , *Weighted Lépingle inequality.* Bernoulli Society for Mathematical Statistics and Probability. 26 (2020), no. 3, pp. 2311-2318.

[Bou89] J. Bourgain, *Pointwise ergodic theorems for arithmetic sets* Inst. Hautes E'tudes Sci. Publ. Math. 69 (1989), pp. 5–45.

[JKR] Jones, R. L., Kaufman, J., and Wierdl M. *Oscillation in ergodic theory,* Erg. Th. & Dyn. Sys., 18 (1998), pp. 889–936.

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

Joe Trate, Bonn
email: `s72jtrat@uni-bonn.de`

26 Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold

After A. Leibman [Le]

A summary written by Kostas Tsinas

Abstract

We examine the pointwise convergence of the ergodic averages of a continuous function on a nilmanifold, evaluated along a polynomial sequence. Central to this topic is an equidistribution property of polynomial sequences on nilmanifolds.

26.1 Introduction

Let G be a nilpotent Lie group of degree s and let Γ be a discrete cocompact subgroup. Then, the space $X = G/\Gamma$ is called an *s-step nilmanifold.* The group G acts on the space X by left multiplication, so that for any $g \in G$ and point $x = b\Gamma$, we have $gx = gb\Gamma$. The natural projection of the Haar measure m_G of G on X will be denoted by m_X. For this presentation, a *polynomial sequence* on the group G is a sequence of the form

$$v(n) = g_1^{p_1(n)} \dots g_k^{p_k(n)}, \tag{1}$$

where $g_1, \dots, g_k$ are elements in G and $p_1, \dots, p_k$ are polynomials taking integer values on the integers. Our main problem is to investigate the limiting behavior of the ergodic averages $\frac{1}{2N+1} \sum_{n=-N}^{N} F(g(n)x)$, where $F : X \to \mathbb{C}$ is a continuous function and $x \in X$. This is the content of the first theorem.

Theorem 1. *Let $g(n)$ be a polynomial sequence in G. For any continuous function $F : X \to \mathbb{C}$ and any point $x \in X$, the averages $\frac{1}{2N+1} \sum_{n=-N}^{N} F(g(n)x)$ converge.*

The previous theorem is an immediate corollary of the following theorem, which asserts that (the projection of) any polynomial sequence on $X = G/\Gamma$ is equidistributed on a subspace of X, as long as we restrict n to appropriate arithmetic progressions. A closed subset Y of X of the form Hx, where H is a closed subgroup of G will be called a *sub-nilmanifold* of X. A sequence x_n in a nilmanifold X is *equidistributed* on X (with respect to m_X) if for any continuous function $F : X \to \mathbb{C}$, we have

$$\lim_{N\to+\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} F(x_n) = \int F\, dm_X.$$

Furthermore, we will say that the action of an element $g \in G$ is *ergodic* or that g is *ergodic* (with respect to m_X), if the sequence $(g^n x)_{n\in\mathbb{N}}$ is equidistributed (with respect to m_X) for all $x \in X$.

Theorem 2. *Let $g(n)$ be a polynomial sequence on a nilpotent Lie group F and and let $x \in G/\Gamma$. There exists a connected closed subgroup H of G and points $x_0, \dots, x_{Q-1} \in G/\Gamma$ (not necessarily distinct), such that the sets $Y_j = Hx_j,\ j \in \{0, \dots, Q-1\}$ are closed sub-nilmanifolds of G/Γ, $\overline{\{g(n)x\colon n \in \mathbb{N}\}} = \bigcup_{j=0}^{Q-1} Y_j$, the sequence $g(n)x$ visits the sets $Y_0, \dots, Y_{Q-1}$ cyclically and the sequence $g(Qn+j)$ is equidistributed in Y_j for every $j \in \{0, \dots Q-1\}$.*

In the case that X is connected, we can write $X = G_0/(\Gamma \cap G_0)$, where G_0 denotes the connected component of the identity on G. In this case, the equidistribution of a polynomial sequence on X is controlled by its projection on the *maximal factor torus* $Z = [G_0, G_0]/X = G_0/\Big([G_0, G_0](\Gamma \cap G_0)\Big)$. Observe that Z is a compact, connected, and abelian Lie group, thus a finite-dimensional torus.

Theorem 3. *Assume that $X = G/\Gamma$ is a connected nilmanifold and let $x \in X$ and $g(n)$ be a polynomial sequence. Let Z be the maximal factor torus of X and let $\pi : X \to Z$ denote the projection map. The following are equivalent:*
i) the sequence $\big(g(n)x\big)_{n\in\mathbb{Z}}$ is dense on X,
ii) the sequence $\big(g(n)x\big)_{n\in\mathbb{Z}}$ is equidistributed on X,
iii) the sequence $\{g(n)\pi(x)\}_{n\in\mathbb{Z}}$ is dense on Z.

The proof of Theorem 2 contains two main ingredients. Firstly, we need to investigate the linear case (when the polynomials p_i have degree 1). This is essentially the study of the ergodicity properties of the map $x \to gx$. This argument was carried out originally by Parry in [Pa] in the case where the group G is connected and simply connected. Following that, a lifting argument due to Furstenberg allows one to express a polynomial orbit as a linear orbit on an extension of the original system. In the abelian setting where the nilmanifold is a finite-dimensional torus, the argument is well-known. Suppose we are given a real number a and the polynomial orbit $\frac{n(n-1)}{2}a$ on the torus $\mathbb{T}$ equipped with the transformation $x \to x + a$. Considering the affine system on the torus $\mathbb{T}^2$ with the map $S : (x, y) \to (x + a, y + x)$, we readily notice that $n^2 a = F(S^n(0, 0))$, where F is the projection on the second coordinate. In particular, the sequence $n^2 a$ is expressed as linear orbit on the system $(\mathbb{T}^2.S)$. This lifting trick is adapted to the nilpotent setting to deduce the general case of Theorem 2 from the linear one. A generalization of the affine system $(\mathbb{T}^2, S)$ is prominent in these arguments.

26.2 Deducing the general case from the linear one

In this section, we show that the general case follows from the linear one. We will assume that the following proposition holds and then deduce the general case in the case that the Lie group G has some nice connectedness assumptions.

Proposition 4. *Let $X = G/\Gamma$ be an s-step nilmanifold. Then, for any $g \in G$, there exists a closed subgroup H of G, such that $\overline{\{g^n x \colon n \in \mathbb{Z}\}} = Hx$ for any $x \in X$. In particular, the orbit $\overline{\{g^n x \colon n \in \mathbb{Z}\}}$ is a sub-nilmanifold of X.*

We say that an automorphism ϕ of a group G is *unipotent* of degree r if the map ξ^r is the identity, where $\xi(g) = \phi(g) \cdot g^{-1}$. Using induction on the nilpotency degree of a nilpotent group G, it is easy to check that an automorphism is unipotent if and only if the induced automorphism on $G/[G, G]$ is unipotent. The first step is to establish the result for transformations on a nilmanifold induced by a unipotent automorphism using the previous proposition. The general case will then follow by showing that polynomial orbits can be realized as orbits of unipotent transformations.

Lemma 5. *Let $X = G/\Gamma$ be a nilmanifold and let ϕ be a unipotent measure-preserving automorphism on G such that $\phi(\Gamma) = \Gamma$. For any $x \in X$, there exists a connected, closed subgroup H of G and points*

$x_0, \ldots, x_{Q-1} \in X$, *such that* Hx_j *are subnilmanifolds of* X *and for each* $j \in \{0, \ldots, Q-1\}$, *the sequence* $\left(\phi^{j+Qn}x\right)_{n\in\mathbb{Z}}$ *is equidistributed on* Hx_j.

Proof of Lemma 5. Firstly, observe that since $\phi(\Gamma) = \Gamma$, ϕ induces a homeomorphism S of X onto itself, so that the notation $\phi^n x$ in the statement is always well-defined (it denotes the point $S^n x$). We consider the semi-direct product $\widetilde{G} = G \times_\phi \mathbb{Z}$, that is we have the set $G \times \mathbb{Z}$ with the operation $(g_1, n_1) * (g_2, n_2) = (g_1 \phi^{n_1}(g_2), n_1 + n_2)$. The group $\widetilde{G}$ is nilpotent[15]. Furthermore, the semi-direct product $\widetilde{\Gamma} = \Gamma \times_\phi \mathbb{Z}$ is a discrete subgroup of $\widetilde{G}$ and $\widetilde{G}$ acts on X through the map $(g, n) * x = g \cdot S^n x$. This action is transitive and the stabilizer of the base point Γ in X is the group $\widetilde{\Gamma}$. Therefore, $\widetilde{\Gamma}$ is also co-compact in $\widetilde{G}$ and X can be identified with $\widetilde{G}/\widetilde{\Gamma}$. Then, the action of S on X is represented by the action of the element $(e_G, 1) \in \widetilde{G}$ on X. Now, we apply Proposition 4 to deduce that there exists a closed, connected subgroup H of $\widetilde{G}$, such that Hx is a subnilmanifold of X and the sequence $\left(S^n x\right)_{n\in\mathbb{Z}}$ is equidistributed on Hx. Let H_0 denote the connected component of the identity in H. The group H_0 is a normal subgroup of H and is also open in H. Since G is open in $\widetilde{G}$, we get $H_0 \subset G$. In addition, since $H_0 x$ is a connected component of Hx and Hx is compact, we have that Hx is comprised out of finitely many translates of $H_0 x$ and, thus, the stabilizer of $H_0 x$ has finite index in H. We let $b_0, \ldots, b_{Q-1} \in H$ be representatives of $H/\mathrm{Stab}(H_0 x)$ and let $x_j = b_j x$ for all $j \in \{0, \ldots, Q-1\}$. Since H_0 is normal in H, we easily deduce that $b_j H_0 x = H_0 b_j x = H_0 x_j$. We set $Y_j = H_0 x_j$, which are connected sub-nilmanifolds of X, whose union is Hx. Observe that S acts transitively on the set $\{Y_0, \ldots Y_{Q-1}\}$, so that the sequence $S^n x$ visits the sub-nilmanifolds cyclically. Reordering the Y_j appropriately, we conclude that $S^{j+Qn} \in Y_j$ for all $j \in \{0, \ldots, Q-1\}$ and the sequence $\left(\phi^{j+Qn}x\right)_{n\in\mathbb{Z}} = \left(S^{j+Qn}x\right)_{n\in\mathbb{Z}}$ is equidistributed on Y_j. □

We proceed towards the proof of Theorem 2. In this exposition, we shall work under the assumption that the Lie group G is connected and simply connected. The assumption of simple connectedness is not restrictive, since any nilpotent Lie group G is a factor of a simply connected nilpotent Lie group $\widetilde{G}$ by considering the universal cover. In addition, one can show that there exists a connected simply connected Lie group G' with a discrete

[15] It can be shown through calculations with commutators and induction on s that if G is s-step nilpotent and the unipotent automorphism ϕ has degree r, then $\widetilde{G}$ is sr-step nilpotent.

co-compact subgroup Γ', such that $X = G/\Gamma$ is a sub-nilmanifold of $X' = G'/\Gamma'$ and every translation on X is represented in X'. For more details, we also refer the reader to [Le, Subsection 1.11] or [HK, Chapter 10, Corollary 26].
In any connected, simply connected nilpotent Lie group G, the map exp from G to its Lie algebra $\mathfrak{B}$ is a diffeomorphism. In particular, it is a bijection between G and $\mathfrak{B}$. For $g \in G$ and $t \in \mathbb{R}$ we can then define the element g^t as the unique element of G satisfying $g^t = \exp(tY)$, where $\exp(Y) = g$.
Any connected, simply connected nilpotent Lie group possesses a *Mal'cev* basis so that every element g can be written uniquely in the form $a_1^{t_1} \dots a_m^{t_m}$, where $t_1, \dots, t_m \in \mathbb{R}$. The numbers $(t_1, \dots, t_m)$ will be called the *coordinates* of g. Furthermore, elements Γ are precisely those whose coordinates consist of integers. It can be shown that multiplication in G is given by polynomial mappings of the coordinates. In particular, we readily deduce that any polynomial sequence of the form (1) can be written in the form $a_1^{q_1(n)} \cdot \ldots \cdot a_m^{q_m(n)}$, where q_i are polynomials with real coefficients.
Let $\mathcal{F}$ denote the free group on the continuous generators $a_1, \dots, a_m$ and let $\mathcal{F}_i$ be its commutator subgroups. Given $s \in \mathbb{N}$, the nilpotent Lie group $F_s = \mathcal{F}/\mathcal{F}_{s+1}$ will be called the free s-step nilpotent Lie group with continuous generators $a_1, \dots, a_m$. The discrete subgroup of F_s generated by $a_1, \dots, a_m$ will be denoted by $\Gamma(F_s)$ (it can be shown that it is cocompact in F_s). It is standard to deduce that F_s possesses a universal property: a connected, simply connected, s-step nilpotent Lie group with a Mal'cev basis $(a_1, \dots, a_m)$ is a factor of F_s with continuous generators $a_1, \dots, a_m$.

Proof of Theorem 2. Let s and $(a_1, \dots, a_m)$ be the nilpotency degree and the Mal'cev basis of G respectively, so that the polynomial sequence $g(n)$ can be written in the form $a_1^{q_1(n)} \cdot \ldots \cdot a_m^{q_m(n)}$, where q_i are polynomials with real coefficients. We also denote $\theta : G \to X$ the natural projection. We may assume that $x = \theta(e_G) = \Gamma$ because otherwise, we can replace the sequence $g(n)$ with the sequence $g(n)\alpha$, where $\alpha \in G$ is any element for which $\theta(\alpha) = x$ and reduce to the aforementioned case.
Let F_s be the free Lie group of degree s on the continuous generators $a_1, \dots, a_m$ and let $\pi_1 : F_s \to G$ denote the associated epimorphism. It is immediate that $\pi_1(\Gamma(F_s)) \subseteq \Gamma$ since the elements of Γ are those whose coordinates in the basis are integers. We consider the free Lie group $\widetilde{G}$ with generators $\{b_{i,0} = a_i, b_{i,1} \dots, b_{i,\deg(q_i)}\}_{i=1,\dots,m}$ and let B be the normal closure

in $\widetilde{G}$ of the group generated by $\{b_{i,1}^t, \dots, b_{i,\deg(q_i)}^t\}_{i=1,\dots,m,t\in\mathbb{R}}$. Observe that $F_s = \widetilde{G}/B$. If π_2 denotes the corresponding projection from $\widetilde{G}$ to F_s, then $\pi = \pi_1 \circ \pi_2$ is an epimorphism from $\widetilde{G}$ to G. Additionally, the group $\Gamma(\widetilde{G})$ satisfies $\pi(\Gamma(\widetilde{G})) \subseteq \Gamma$.
We define the automorphism ϕ on $\widetilde{G}$ by $\phi(a_i) = a_i$ and $\phi(b_{i,k}) = b_{i,k} \cdot b_{i,k-1}$ for $k = 1, \dots, \deg(q_i)$. Thus, ϕ is a unipotent automorphism on $\widetilde{G}/[\widetilde{G}, \widetilde{G}]$, so that it is also a unipotent automorphism on $\widetilde{G}$. If we write the polynomial q_i in the form

$$q_i(x) = c_{i,0} + c_{i,1}\binom{n}{1} + \dots + c_{i,\deg(q_i)}\binom{n}{\deg(q_i)}, c_{i,0}, c_{i,1}, \dots, c_{i,\deg(q_i)} \in \mathbb{R},$$

and set $u_i = a_i^{c_{i,0}} \cdot b_{i,1}^{c_{i,1}} \cdot \dots \cdot b_{i,\deg(q_i)}^{c_{i,\deg(q_i)}}$, then it follows that

$$\phi^n(u_i) = a_i^{c_{i,0}+c_{i,1}\binom{n}{1}+\dots+c_{i,\deg(q_i)}\binom{n}{\deg(q_i)}} w_i(n) = a_i^{q_i(n)} w_i(n)$$

for every $n \in \mathbb{Z}$, where $w_i(n) \in B$. Thus, if $u = u_1 \dots u_m$, we conclude that $g(n) = \pi(\phi^n(u))$. Thus, we have expressed the polynomial sequence as the orbit of a unipotent transformation on the Lie group $\widetilde{G}$.
Let $\widetilde{X} = \widetilde{G}/\Gamma(\widetilde{G})$. The epimorphism $\pi : \widetilde{G} \to G$ induces a factor map from $\widetilde{X}$ to $X = G/\Gamma$, which we denote by $\widetilde{\pi}$. Thus, we deduce that $\pi(\phi^n(u\Gamma(\widetilde{G}))) = g(n)\Gamma$ for all $n \in \mathbb{Z}$. Applying Lemma 5, we deduce that there exists a closed, connected subgroup $\widetilde{H}$ of $\widetilde{G}$ and points $\widetilde{x}_0, \dots, \widetilde{x}_{Q-1} \in \widetilde{X}$, such that $\phi^{j+Qn}(u\Gamma(\widetilde{G}))$ is equidistributed on $\widetilde{H}\widetilde{x}_j$ for all admissible values of j. Thus, if we naturally define $H = \pi(\widetilde{H})$ and $x_j = \widetilde{\pi}(\widetilde{x}_j)$, we have that $Y_j = Hx_j$ is a connected sub-nilmanifold of X and the H-invariant measure of Y_j is the image under $\widetilde{\pi}$ of the $\widetilde{H}$-invariant measure of $\widetilde{H}\widetilde{x}_j$. Consequently, the sequence $\big(\widetilde{\pi}(\phi^{j+Qn}\Gamma(\widetilde{G}))\big)_{n\in\mathbb{Z}} = \big(g(j+Qn)\big)_{n\in\mathbb{Z}}$ is equidistributed on Hx_j for every $j = 0, \dots, Q-1$, which is the desired result. $\square$

References

[HK] Host B. and Kra B., *Nilpotent structures in ergodic theory*, AMS (2018).

[Le] Leibman A., *Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold.* Ergodic Theory and Dynamical Systems, **25** (2005) no. 1, pp. 201-213.

[Pa] Parry W., *Ergodic properties of affine transformations and flows on nilmanifolds*, Amer. J. Math. **91** (1969), pp. 757-771.

Kostas Tsinas, University of Crete
email: kon.tsinas@gmail.com

27 Cancellation for the simplex Hilbert transform

After P. Zorin-Kranich [ZK]

A summary written by Jianghao Zhang

Abstract

The regularity lemma from additive combinatorics [G] is also useful in Euclidean harmonic analysis. We discuss how to attack the simplex Hilbert transform by using the regularity lemma inductively.

27.1 Introduction

Consider the $(d+1)$-linear operator

$$\Lambda^K(f_0, \cdots, f_d) := \int_{\mathbb{R}^{d+1}} \prod_{j=0}^{d} f_j(x_{j^c}) K(\sum_{j=0}^{d} x_j) dx,$$

where $j^c := \{0, \cdots, d\} \backslash \{j\}, x_{j^c} := (x_0, \cdots, x_{j-1}, x_{j+1}, \cdots, x_d) \in \mathbb{R}^d$. It's of a long-standing interest in harmonic analysis to study the boundedness of such operators when K is a one-dimensional Calderón-Zygmund kernel. The $d \geq 3$ cases are still widely open.
Take an even, smooth function ψ supported on $\left[\frac{1}{2}, 2\right]$ such that $\sum_{k \in \mathbb{Z}} \psi(\frac{t}{2^k}) = 1, \forall t \neq 0$. Truncate K into

$$\psi_k(t) := \psi(\frac{t}{2^k}) K(t).$$

For any interval S of $k \in \mathbb{Z}$, let $\psi_S := \sum_{k \in S} \psi_k$. We have the trivial bound

$$|\Lambda^{\psi_S}(f_0, \cdots, f_d)| \lesssim (\#S) \prod_{j=0}^{d} \|f_j\|_{p_j}$$

for any Hölder tuple of exponents $1 \leq p_j \leq \infty$. Our theorem is the first-step nontrivial improvement of this bound.

Theorem 1. *Let $d \geq 1$. For any Hölder tuple of exponents $1 < p_j < \infty$ we have*

$$|\Lambda^{\psi_S}(f_0, \cdots, f_d)| \leq o_{d,p_0,\cdots,p_d}(\#S) \prod_{j=0}^{d} \|f_j\|_{p_j}.$$

27.2 An outline of the argument

It suffices to tackle the case when $f_j = 1_{E_j}$ and $p_j > d$. The argument itself doesn't give a new method to directly attack this operator. Instead, we keep reducing it to the lower dimensional cases and finally use the known result of $d = 1$ (i.e. induction on d).

27.2.1 The regularity lemma

To go back to the $d-1$ case, f_d should be like

$$\prod_{A \subsetneq \{0,\cdots .d-1\}} f_A(x|_A) \tag{1}$$

so that $\prod_{j=0}^{d-1} f_j$ can absorb it. This is where the regularity lemma comes into play. Denotes the set of Functions like (1) by Σ. We can write $f_d = f_{str} + f_{uni}$ in the Hilbert space L^2. Here morally $f_{str} \sim \Sigma, f_{uni} \sim \Sigma^{\perp}$. ($\sim$ means the element is somehow near the set but not necessary to lie in it.) Write $x = (x^{'}, x_d)$. Then

$$\int_{\mathbb{R}^d} (\prod_{j=0}^{d-1} f_j) f_{str} \psi_S dx^{'} \tag{2}$$

can be handled by the induction hypothesis. On the other hand,

$$\int_{\mathbb{R}^{d+1}} (\prod_{j=0}^{d-1} f_j) f_{uni} \psi_S dx$$

looks like $\langle \prod_{A \subsetneq \{0,\cdots .d-1\}} f_A, f_{uni} \rangle$. It also has a good bound since $f_{uni} \sim \Sigma^{\perp}$. Turns out the bound of f_{uni} is not satisfying, and we need a more flexible version of the decomposition

$$f_d = f_{str} + f_{uni} + f_{er}, \tag{3}$$

where we allow an error term for better control on f_{uni}.

27.2.2 Trees

We've seen how the induction process goes. But there are still obstacles. We only integrate the first d coordinates in (2). The induction hypothesis

gives a bound independent with x_d, so a further integration w.r.t. x_d will become $+\infty$. Moreover, the error term f_{er} of (3) is small in the L^2 sense, but we will encounter $\|f_{er}\|_1$ when estimating $|\Lambda^{\psi_S}(f_0, \cdots, f_{er})|$.
To settle these problems, we do everything locally. Consider

$$\Lambda_Q(f_0, \cdots, f_d) := \int_Q \prod_{j=0}^{d} f_j(x_{j^c})\psi_k(\sum_{j=0}^{d} x_j)dx, \tag{4}$$

where $Q = Q^{'} \times Q_d$ is a dyadic cube in $\mathbb{R}^{d+1}$ such that $s(Q) = k$. ($s(Q)$ means the scale of Q.) To get some cancellation from the induction hypothesis, we sum k over an interval and restrict $x^{'}$ inside $Q^{'}$. We will call such patterns trees. This definition is designed to facilitate our induction argument, quite different from the trees in time-frequency analysis [LT].
Fix $Q^{'} = 2^k(m_0, \cdots, m_{d-1}) + \big[0, 2^k\big)^d$. If $x^{'} \in Q^{'}$, then x_d must lie near $2^k m_d := 2^k(-\sum_{j=0}^{d-1} m_j)$ to make the integrand of (4) nonzero. Thus we only need to consider $Q = 2^k(m_0, \cdots, m_d) + \big[0, 2^k\big)^{d+1}$ such that $\sum_{j=0}^{d} m_j = 0$. All cubes mentioned below are of this type. If redefine

$$\Lambda_Q(f_0, \cdots, f_d) := \int_{Q^{'}} (\prod_{j=0}^{d} f_j)\psi_k dx = \int_{Q^{'} \times 10Q_d} (\prod_{j=0}^{d} f_j)\psi_k dx, \tag{5}$$

we get

$$\Lambda^{\psi_S}(f_0, \cdots, f_d) = \sum_{k \in S} \sum_{s(Q)=k} \Lambda_Q(f_0, \cdots, f_d).$$

Definition 2. *Call a collection of cubes $\mathscr{T}$ a tree with the top Q_{top} if there exists an interval Υ such that $s(Q_{top}) \geq \max \Upsilon$ and $\mathscr{T} = \{Q : s(Q) \in \Upsilon, Q^{'} \subset Q^{'}_{top}\}$.*

For each tree $\mathscr{T}$ we have

$$\begin{aligned}
\sum_{Q \in \mathscr{T}} \Lambda_Q(f_0, \cdots, f_d) &= \sum_{k \in \Upsilon} \int_{Q^{'}_{top}} \prod_{j=0}^{d} f_j(x_{j^c})\psi_k(\sum_{j=0}^{d} x_j)dx \\
&= \int_{Q^{'}_{top}} \prod_{j=0}^{d} f_j(x_{j^c})\psi_\Upsilon(\sum_{j=0}^{d} x_j)dx \\
&= \int_{Q^{'}_{top} \times 10(Q_{top})_d} \prod_{j=0}^{d} f_j(x_{j^c})\psi_\Upsilon(\sum_{j=0}^{d} x_j)dx.
\end{aligned}$$

Then we can do the procedure in 27.2.1 to each tree.

Theorem 3. *Given $\delta > 0$, there exists $C_d(\delta)$ such that for any tree with $\#\Upsilon = C_d(\delta)$, the following estimate holds:*

$$|\sum_{Q\in\mathscr{T}} \Lambda_Q(f_0, \cdots, f_d)| \lesssim \delta |Q'_{top}|(\#\Upsilon).$$

This bound easily extends to trees with general length $\#\Upsilon$ by dividing Υ into pieces of the length $C_d(\delta)$.

Corollary 4. *Given $\delta > 0$, there exists $C_d(\delta)$ such that for any tree with $\#\Upsilon > \delta^{-1}C_d(\delta)$, the following estimate holds:*

$$|\sum_{Q\in\mathscr{T}} \Lambda_Q(f_0, \cdots, f_d)| \lesssim \delta |Q'_{top}|(\#\Upsilon).$$

27.2.3 The selection scheme

Our task now is to sum up these tree estimates. Note that we cannot simply put each cube in a tree, because there will then be infinite trees with tops of the largest scale, and the summation of their tree estimates is ∞. A more efficient way is to consider the density of Λ_Q over each Q. Let $a_Q = \frac{\Lambda_Q}{|Q'|}$. We have

$$|a_Q| \lesssim \prod_{j=0}^{d} \inf_{x_{j^c}\in\pi_{j^c}Q} M_d f_j(x_{j^c}). \tag{6}$$

by the Loomis-Whitney inequality.
Now only take trees with large densities on tops. For each $Q, s(Q) \in S$ satisfying $|a_Q| > \lambda$ (λ depends on δ) and Q' is maximal, we construct a tree $\mathscr{T}(Q) := \{P : s(P) \in S, P' \subset Q'\}$. (6) implies the weak type estimate for these trees:

$$\sum_{Q:\ |a_Q|>\lambda,\ Q' \text{ is maximal}} |Q'| \lesssim \frac{1}{\lambda} \prod_{j=0}^{d} |E_j|^{\frac{1}{p_j}}. \tag{7}$$

Thus summing over these trees is acceptable. We briefly explain why (7) is true. First take a maximal sub-collection $\{\widetilde{Q}\}$ from

$\{Q: \ |a_Q| > \lambda, \ Q' \text{ is maximal}\}$ such that $(C\widetilde{Q}')$'s don't include each other. We have

$$\sum_{Q:\ |a_Q|>\lambda,\ Q' \text{ is maximal}} |Q'| \lesssim \sum_{\widetilde{Q}} |\widetilde{Q}'|.$$

Then we use (6) to estimate

$$\begin{aligned}
\sum_{\widetilde{Q}} |\widetilde{Q}'| \lesssim \frac{1}{\lambda} \sum_{\widetilde{Q}} \prod_{j=0}^{d} \|M_d f_j\|_{L^{p_j}(\pi_{j^c}\widetilde{Q})} &\leq \frac{1}{\lambda} \prod_{j=0}^{d} \Big(\sum_{\widetilde{Q}} \|M_d f_j\|^{p_j}_{L^{p_j}(\pi_{j^c}\widetilde{Q})}\Big)^{\frac{1}{p_j}} \\
&\leq \frac{1}{\lambda} \prod_{j=0}^{d} \|M_d f_j\|_{L^{p_j}} \\
&\lesssim \frac{1}{\lambda} \prod_{j=0}^{d} |E_j|^{\frac{1}{p_j}},
\end{aligned}$$

where we use $(\pi_{j^c}\widetilde{Q})$'s are disjoint since $(C\widetilde{Q}')$'s don't include each other. All cubes left must have small densities, and we can use (6) as above to directly control them.

References

[G] T. Gowers, *Decompositions, approximate structure, transference, and the Hahn-Banach theorem.* Bull. Lond. Math. Soc. **42** (2010), no. 4, 573–606.

[LT] M. Lacey and C. Thiele, *A proof of boundedness of the Carleson operator.* Math. Res. Lett. **7** (2000), no. 4, 361-370.

[ZK] P. Zorin-Kranich, *Cancellation for the simplex Hilbert transform.* Math. Res. Lett. **24** (2017), no. 2, 581-592.

Jianghao Zhang, Bonn University
email: Jianghao@uni-bonn.de