Higher order Fourier analysis and related topics

Summer School[*], Kopp

October 2026

Organizers:

Jop Briët, Centrum Wiskunde and Informatica, Amsterdam

Asgar Jamneshan, Universität Bonn

Luka Milićević, Serbian Academy of Sciences and Arts, Belgrade

Christoph Thiele, Universität Bonn

[*]supported by ERC Synergy grant 101224275

Contents

1 A bilinear Bogolyubov-Ruzsa lemma with polylogarithmic bounds **8**
after K. Hosseini and S. Lovett [HL] 8
Hayk Aprikyan, University of Bonn 8
1.1 Introduction 8
1.2 Preliminaries 9
1.2.1 Freiman-Ruzsa theorem 9
1.2.2 Spans of affine maps 10
1.3 Main idea of the proof 10

2 Small subalgebras of polynomial rings and Stillman's conjecture **14**
after T. Ananyan and M. Hochster [AH] 14
Noa Bihlmaier, University of Bonn 14
2.1 Introduction 14
2.2 From small subalgebras to Stillman's conjecture 15
2.3 Obtaining small subalgebras 16
2.4 Large strength implies regularity 17

3 Strong bounds for 3-progressions **19**
after Z. Kelley and R. Meka [KM] 19
Aleksandar Bulj, University of Zagreb 19
3.1 Introduction and main results 19
3.1.1 Notation 20
3.2 Proof of Theorem 1 using Theorem 2 20
3.3 Key lemmas 21
3.4 Proof of Theorem 2 22

4 The Quadratic Goldreich–Levin theorem **24**
after J. Briët and D. Castro-Silva 24
Davi Castro-Silva, University of Cambridge 24
4.1 Introduction 24
4.2 Outline of the algorithm 26

5 A sum–product estimate in finite fields, and applications **30**
after Jean Bourgain, Nets Katz and Terence Tao [BKT] 30

Jihyo Chae, Yonsei University 30
5.1 Introduction 30
5.2 Outline of the proof 31
5.3 Key Lemmas 31
5.4 Proof of Theorem 1 32

6 Stabilizer rank and higher-order Fourier analysis 35
after Farrokh Labib 35
Philippe van Dordrecht, CWI, Leiden University 35
6.1 Notions of rank 37
6.2 Sketch of the proof 38

7 The inverse theorem for the U^3 Gowers uniformity norm on arbitrary finite abelian groups: Fourier-analytic approach 40
after A. Jamneshan and T. Tao 40
Ferran Espuña, UAM 40
7.1 Introduction 40
7.2 Extracting a locally linear frequency map 41
7.3 Symmetrising and globalising the bilinear obstruction 42
7.4 From bilinear structure to quadratic correlation 44

8 A new proof of Szemerédi's theorem for arithmetic progressions of length four 46
after W. T. Gowers [1] 46
Benjamin Gillott and Fredy Yip, University of Cambridge 46
8.1 Introduction 46
8.2 Quadratic uniformity and Gowers uniformity norms 47
8.3 4-AP count 49
8.4 Density increment 51

9 Poorly-distributed polynomials over finite fields are low-rank 53
after B. Green and T. Tao [GT09] 53
V. Gladkova, Manchester 53
9.1 Introduction 53

10 Quadratic Fourier analysis and stabilizer testing 58
after S. Arunachalam and A. Dutt [AD25], Z. Bao and P. van Dordrecht and J. Helsen [BvDH25] 58

Isaac Holt, University of Cambridge . . . 58
10.1 Introduction . . . 58
10.2 Preliminaries . . . 58
10.3 The U^3 inverse theorem for quantum states . . . 59
10.4 Tolerant testing of stabiliser states . . . 61

11 Higher-order Fourier analysis and algebraic property testing 63
after H. Hatami, P. Hatami, and S. Lovett [HHL19] . . . 63
Zimrat Kaniel, The Hebrew University of Jerusalem . . . 63
11.1 Introduction and basic definitions . . . 63
11.2 The BLR test and affine linearity . . . 64
11.3 Higher-order derivatives and Gowers norms . . . 65
11.4 Low-degree testing in the 99% regime . . . 66
11.4.1 Proof architecture: reduction to flat-testing . . . 66
11.5 Analytical consequences . . . 67

12 New proofs of Plünnecke-type estimates for product sets in groups, and the theorem of Balog, Szemerédi and Gowers 69
after G. Petridis [Pe] and Y. Zhao [Zh] . . . 69
Deepanshu Kush, University of Cambridge . . . 69
12.1 Introduction . . . 69
12.2 Growth of triple products . . . 70
12.3 The Plünnecke–Ruzsa inequalities . . . 71
12.4 Additive energy and the Balog–Szemerédi–Gowers theorem . . . 72
12.4.1 Reduction to two sets, and then to a graph . . . 72
12.4.2 Paths of length two and three . . . 73

13 Roth's theorem 76
after K. Roth, following Y. Zhao [Z] . . . 76
Stefanos Lappas, Bonn . . . 76
13.1 An introduction to Roth's theorem . . . 76
13.2 Fourier analysis on finite abelian groups . . . 77
13.3 Roth's theorem in $\mathbb{F}_3^n$ . . . 77
13.4 Proof sketch of Theorem 9 . . . 78
13.5 Fourier analysis in the integers . . . 79
13.6 Roth's theorem in $\mathbb{Z}$ . . . 80
13.7 Proof sketch of Theorem 10 . . . 80
13.8 Concluding remarks . . . 81

14 The Möbius function is strongly orthogonal to nilsequences 82
after B. Green and T. Tao [GT1] 82
Tiago Moreira, SISSA 82
14.1 Introduction 82
14.2 Nilmanifolds and nilsequences 83
14.3 Factorization of polynomial orbits 83
14.4 Proof of Theorem 1 84
14.5 The Heisenberg example 85

15 The nonabelian U^2 inverse theorem 88
after W.T. Gowers and O. Hatami [1] 88
Oissin O'Donnell, Cambridge 88
15.1 Introduction 88
15.2 Preliminaries 89
15.2.1 Matrix norms 89
15.2.2 The Fourier transform 90
15.3 Remarks on the proof of the inverse theorem 91

16 Effective bounds for restricted 3-APs in $\mathbb{F}_3^n$ 93
after A. Bhangale, S. Khot, Y. P. Liu, D. Minzer [BKLMa] 93
Anthony Ostuni, UC San Diego 93
16.1 Introduction 93
16.2 Proof overview 94
16.3 Sketch of inverse theorem 96

17 The Green–Tao theorem: an exposition 98
after D. Conlon, J. Fox, and Y. Zhao [CFZ] 98
Felix Pernegger, University of Bonn 98
17.1 Overview 98
17.2 The relative Szemerédi theorem 99
17.3 A pseudorandom majorant for the primes 100
17.4 Outline of the linear-forms calculation 101
17.5 Extensions and conjectural generalizations 102

18 On the Bogolyubov–Ruzsa lemma 103
after Tom Sanders [S12] 103
António Rocha-Neves, University of Bonn 103
18.1 Main Theorem 103

18.2 Preliminaries and Fourier analysis in $\mathbb{F}_2^n$ . . . 103
18.3 The Croot–Sisask lemma . . . 104
18.4 Chang's theorem . . . 105
18.5 Extracting the subspace (main proof) . . . 106

19 An equivalence between inverse sumset theorems and inverse conjectures for the U^3 norm 108
after B. Green and T. Tao [GT] . . . 108
Nihan Tanısalı, Ecole Polytechnique and Inria . . . 108
19.1 Introduction . . . 108
19.2 PFR implies PGI(3) . . . 109
19.3 PGI(3) implies PFR . . . 111
19.4 The higher-order result . . . 112

20 An inverse theorem for the Gowers U^4-norm 114
after B. Green, T. Tao and T. Ziegler [GTZ1] . . . 114
Kit Thomas, Oxford . . . 114
20.1 Introduction . . . 114
20.1.1 Preliminaries . . . 115
20.2 Proof outline . . . 116
20.2.1 Correlation among the χ_h . . . 116
20.2.2 Reducing h-dependence . . . 117
20.2.3 Approximate linearity of h-dependent frequencies . . . 117
20.2.4 Symmetry argument . . . 117

21 Freiman's theorem in groups with bounded exponent 119
after Yufei Zhao [YZ] . . . 119
Shota Uka, TU Vienna . . . 119
21.1 Introduction . . . 119
21.2 Plünnecke's inequality and Ruzsa's coverling lemma . . . 122
21.3 Proof of Freiman's theorem in groups with bounded exponent 123

22 The sum-product problem over the reals 125
after G. Elekes [Ele] and L. A. Székely [Sz] . . . 125
Pauwel Van Den Eeckhaut, University of Bonn . . . 125
22.1 The sum-product problem . . . 125
22.2 Crossing numbers and the Szemerédi–Trotter theorem . . . 126
22.3 Elekes's sum-product bound . . . 129

23 Szemerédi's regularity lemma and the triangle removal 131
after Y. Zhao [Z] . . . 131
Hanaé Vandanjon, Univesité de Nantes . . . 131
23.1 Introduction . . . 131
23.2 Szemerédi's regularity lemma . . . 131
23.2.1 Preliminary definitions . . . 131
23.2.2 Regularity lemma . . . 132
23.3 Triangle removal lemma . . . 134

24 On a Conjecture of Marton 136
after Gowers, Green, Manners, and Tao [GGMT1] . . . 136
Jianghao Zhang, University of Edinburgh . . . 136
24.1 Introduction . . . 136
24.2 The key inductive strategy . . . 136
24.3 Entropy notion . . . 138
24.4 Endgame . . . 139

25 Quasipolynomial inverse theorem for the $U^4(\mathbb{F}_p^n)$ norm 141
after L. Milićević [Mil24+] . . . 141
Daniel Zhu, Princeton . . . 141
25.1 Background . . . 141
25.2 Philosophical remarks . . . 142
25.3 Proof overview . . . 142
25.3.1 Part I: Inductive hypothesis and Cauchy-Schwarz . . . 143
25.3.2 Part II: Some quadruples to all quadruples in a subspace 143
25.3.3 Part III: Low rank to sometimes zero on a bilinear variety 144
25.3.4 Part IV: Some quadruples to all quadruples again . . . 144
25.3.5 Part V: The finish . . . 145

1 A bilinear Bogolyubov-Ruzsa lemma with poly-logarithmic bounds

After K. Hosseini and S. Lovett [HL]

A summary written by Hayk Aprikyan

Abstract

We show that if a set $A \subset \mathbb{F} \times \mathbb{F}$ has density α, then after a constant number of horizontal and vertical sums, it contains a bilinear structure of codimension $\log^{O(1)} \alpha^{-1}$.

1.1 Introduction

Assume $\mathbb{F} = \mathbb{F}_p$ is a prime field for a fixed prime p. Given a subset $A \subset \mathbb{F}^n$, define its density $\alpha = |A|/|\mathbb{F}|^n$, and the sumset and difference set as:

$$2A = A + A = \{a + a' : a, a' \in A\}$$
$$A - A = \{a - a' : a, a' \in A\}$$

A key theorem in additive combinatorics is the (improved) Bogolyubov-Ruzsa lemma due to Sanders [San12], which states that:

Theorem 1 ([San12], [SS16])**.** *If $A \subset \mathbb{F}^n$ has density α, then $2A - 2A$ contains a subspace $V < \mathbb{F}^n$ of codimension $O(\log^4 \alpha^{-1})$. Moreover, every $y \in V$ can be expressed as $y = a_1 + a_2 - a_3 - a_4$ with $a_i \in A$ in at least $\alpha^{O(1)}|\mathbb{F}|^{3n}$ many ways.*

Let now $A \subset \mathbb{F}^n \times \mathbb{F}^n$. We define subtraction on horizontal and vertical fibers as:

$$\phi_{\mathrm{h}}(A) = \{(x_1 - x_2, y) : (x_1, y), (x_2, y) \in A\}$$
$$\phi_{\mathrm{v}}(A) = \{(x, y_1 - y_2) : (x, y_1), (x, y_2) \in A\}$$

and given a word $w \in \{\mathrm{h}, \mathrm{v}\}^k$, we define $\phi_w = \phi_{w_1} \circ \cdots \circ \phi_{w_k}$ as their composition. Finally, we say that a set $B \subset \mathbb{F}^n \times \mathbb{F}^n$ is a *bilinear variety* of codimension $r_1 + r_2 + r_3 = r$ if

$$B = \{(x, y) \in V \times W : b_1(x, y) = \cdots = b_{r_3}(x, y) = 0\}$$

where $V, W < \mathbb{F}^n$ are subspaces of codimension r_1 and r_2, respectively, and $b_i : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ are bilinear forms.

Question: For a set $A \subset \mathbb{F}^n \times \mathbb{F}^n$ of density α, does $\phi_w(A)$ contain a bilinear variety for some fixed $w \in \{\mathrm{h}, \mathrm{v}\}^k$?

In [BL17], Bienvenu and Le proved the answer to be affirmative, with a bilinear variety of codimension $r = \exp(\exp(\exp(\log^{O(1)} \alpha^{-1})))$; and in [GM17], Gowers and Milićević independently obtained a stronger bound of $r = \exp(\exp(\log^{O(1)} \alpha^{-1}))$. In this talk, we present a new proof of the result with a polynomial improvement:

Theorem 2 (**Main theorem**). *Let $A \subset \mathbb{F}^n \times \mathbb{F}^n$ be of density α and let $w = \mathrm{hvvhvvvhh}$. Then there exists a bilinear variety $B \subset \phi_w(A)$ of codimension $r = O(\log^{80} \alpha^{-1})$.*

1.2 Preliminaries

1.2.1 Freiman-Ruzsa theorem

We will use the following version of the Freiman-Ruzsa theorem in the proof of Theorem 2, to linearize maps that are approximately additive:

Theorem 3 (Freiman-Ruzsa, functional version). *Let $f : \mathbb{F}^n \to \mathbb{F}^m$ be a function such that*

$$\Pr_{y,z,z' \in \mathbb{F}^n} [f(y+z) - f(z) = f(y+z') - f(z')] \geq \alpha$$

Then there exists an affine map $L : \mathbb{F}^n \to \mathbb{F}^m$ such that

$$\Pr_{z \in \mathbb{F}^n} [L(z) = f(z)] \geq \exp(-O(\log^4 \alpha^{-1}))$$

With a slight probabilistic argument, we can prove that the result holds for any dense subset $Z \subset \mathbb{F}^n$ as well:

Lemma 4. *Let $f : \mathbb{F}^n \to \mathbb{F}^m$ be a function and $Z \subset \mathbb{F}^n$ with $|Z| \geq \alpha |\mathbb{F}|^n$ such that*

$$\Pr_{y \in \mathbb{F}^n, z,z' \in Z} [f(y+z) - f(z) = f(y+z') - f(z')] \geq \alpha$$

Then there exists an affine map $L : \mathbb{F}^n \to \mathbb{F}^m$ such that

$$\Pr_{z \in Z} [L(z) = f(z)] \geq \exp(-O(\log^4 \alpha^{-1}))$$

1.2.2 Spans of affine maps

Recall that an affine map $L : \mathbb{F}^n \to \mathbb{F}^m$ is a map of the form $L(x) = Mx + b$, where $M \in \mathbb{F}^{m\times n}$, $b \in \mathbb{F}^m$. We introduce the following notation: for a set of affine maps $\mathcal{L} = \{L_1, \dots, L_k\}$ and $y \in \mathbb{F}^n$, let

$$\mathcal{L}(y) = \{L_1(y), \dots, L_k(y)\} \subset \mathbb{F}^m$$

and define the span of $\mathcal{L}$ as the span of affine maps

$$\overline{\mathcal{L}} = \mathrm{span}(\mathcal{L})$$

or equivalently,

$$\overline{\mathcal{L}}(y) = \mathrm{span}(\mathcal{L}(y)) \subset \mathbb{F}^m$$

For notational convenience, we denote $\overline{\mathcal{L}}(y_1, y_2, y_3) = \overline{\mathcal{L}}(y_1) + \overline{\mathcal{L}}(y_2) + \overline{\mathcal{L}}(y_3)$.

As a vector space over $\mathbb{F}$, the dimension of the linear span of $\mathcal{L}$ is naturally denoted by $\dim(\overline{\mathcal{L}})$. We will use the following lemma in the proof of Theorem 2:

Lemma 5. *Let $\delta > 0$ be any number. Let $\mathcal{L}$ be a set of linear maps from $\mathbb{F}^n$ to $\mathbb{F}^m$ with $\dim(\overline{\mathcal{L}}) = k$. Then there is a subspace $Z < \mathbb{F}^n$ with*

$$\dim(Z) \le k(2k + \log \delta^{-1} + 3)$$

such that for every $y \in Z$ and for every $S \subset \mathbb{F}^n$ with $|S| \ge \delta|\mathbb{F}|^n$, at least half of the points $s, s' \in S$ satisfy

$$\big(\overline{\mathcal{L}}(s) + \overline{\mathcal{L}}(y)\big) \cap \big(\overline{\mathcal{L}}(s') + \overline{\mathcal{L}}(y)\big) \subset Z + \overline{\mathcal{L}}(y)$$

1.3 Main idea of the proof

For better readibility, we will always assume x and related sets are in $\mathbb{F}^m$, while y and related sets are in $\mathbb{F}^n$; with $n = m$ initially.

To prove Theorem 2, we perform six steps, where in each step we will pass to a certain, more suitable dense subset of A. In steps 1-2, we repeat horizontal and vertical differencing to obtain subspaces V_y inside many of the fibers. However, these subspaces may vary arbitrarily with y, so in steps 3-4 we use Freiman-Ruzsa to show that this variation contains a low-dimensional linear component $y \mapsto \mathcal{L}(y)$. Finally, in steps 5-6 we get rid of the auxiliary

representations and obtain a single subspace V such that $V \cap \mathcal{L}(y)^{\perp}$ lies in every relevant fiber.

Step 1: Linearizing x-direction (for each fixed y)

We decompose $A = \bigcup_{y \in \mathbb{F}^n} A_y \times \{y\}$, and apply ϕ_{h} twice to get the difference set in the x-direction:

$$A^1 = \phi_{\mathrm{hh}}(A) = \bigcup_{y \in \mathbb{F}^n} (2A_y - 2A_y) \times \{y\}$$

after which we apply Theorem 1 to each fiber A_y. By passing to a dense subset $S \subset \mathbb{F}^n$ such that for every $y \in S$, the fiber A_y^1 contains a subspace $V_y' < \mathbb{F}^m$ of codimension $O(\log^4 \alpha^{-1})$, we obtain a set of the form:

$$B^1 = \bigcup_{y \in S} V_y' \times \{y\} \subset A^1$$

Step 2: Linearizing y-direction

Next we use a similar strategy to obtain linear structure in the y-direction. We apply ϕ_{v} twice to get the difference set in the y-direction and obtain a subspace $W' < 2S - 2S$ of codimension $O(\log^4 \alpha^{-1})$:

$$B^2 = \bigcup_{y \in W'} V_y \times \{y\} \cong \bigcup_{y \in \mathbb{F}^n} V_y \times \{y\}$$

where $n = \dim(W')$, $\operatorname{codim}(V_y) = d = O(\log^4 \alpha^{-1})$ uniformly.

Steps 3-4: Linearizing the correspondence $y \mapsto V_y$

In order to use the Freiman-Ruzsa theorem, we need to linearize the map $y \mapsto V_y$. Denote $A^3 = \phi_{\mathrm{v}}(B^2)$, followed by $A^4 = \phi_{\mathrm{h}}(A^3)$, which results in:

$$\begin{aligned} A^4 &= \bigcup_{y,z,w \in \mathbb{F}^n} ((V_z \cap V_{y+z}) + (V_w \cap V_{y+w})) \times \{y\} \\ &= \bigcup_{y,z,w \in \mathbb{F}^n} ((U_z + U_{y+z}) \cap (U_w + U_{y+w}))^{\perp} \times \{y\} \end{aligned}$$

where $U_y = V_y^{\perp}$ are of dimension d.

To reach the desired linearization, we use Lemma 4 to prove the following key ingredient. Let $\mathcal{L}_y$ be a collection of affine maps from $\mathbb{F}^n$ to $\mathbb{F}^m$ for every $y \in \mathbb{F}^n$.

Lemma 6. *As long as*

$$\Pr_{y,z,w\in\mathbb{F}^n}\left[(U_z+U_{y+z})\cap(U_w+U_{y+w})\subset\overline{\mathcal{L}_z}(z)+\overline{\mathcal{L}_{y+z}}(y+z)+\overline{\mathcal{L}_w}(w)+\overline{\mathcal{L}_{y+w}}(y+w)\right]\leq\frac{1}{2}$$

one can find an affine map L such that

$$\Pr_{y\in\mathbb{F}^n}\left[L(y)\in U_y\setminus\overline{\mathcal{L}_y}(y)\right]\geq\exp(-O(d^4))$$

By the above result, for each $y\in\mathbb{F}^n$ we can iteratively find affine maps $\mathcal{L}'_y=\{L_1,\ldots,L_t\}$ such that the event

$$(U_z+U_{y+z})\cap(U_w+U_{y+w})\subset\overline{\mathcal{L}'}_z(z)+\overline{\mathcal{L}'}_{y+z}(y+z)+\overline{\mathcal{L}'}_w(w)+\overline{\mathcal{L}'}_{y+w}(y+w)$$

holds with probability at least $1/2$. Note that the number of such L_i is bounded by $t=\exp(d^4)$.

After a pigeonhole argument, we can collect the most popular $4d$-many affine maps in one global collection $\mathcal{L}$, to obtain a dense subset $B^4\subset A^4$ of the form:

$$B^4=\bigcup_{y\in T}\overline{\mathcal{L}}(y)^\perp\times\{y\}\subset A^4$$

Here we also restrict our attention to only those x which are orthogonal to the constant terms of the affine maps in $\mathcal{L}$, so that we can assume all maps in $\mathcal{L}$ are linear. Of course, this restriction will cost us $r_1=O(d)$ in codimension.

Step 5: Linearizing y-direction (again)

We once again apply ϕ_{v} twice on B^4 to obtain a subspace $W<2T-2T$ of codimension $r_2=O(d^{20})$, and by Theorem 1,

$$B^5=\bigcup_{y\in W}\left(\bigcup_{(y_1,y_2,y_3)\in S_y}\overline{\mathcal{L}}(y_1,y_2,y_3)^\perp\cap\overline{\mathcal{L}}(y)^\perp\right)\times\{y\}$$

where $S_y=\{(y_1,y_2,y_3):y=y_1+y_2-y_3\}$ is dense in $(\mathbb{F}^n)^3$.

Step 6: Unique subspace V for all $y\in W$

Finally, in order to apply Lemma 5, we apply ϕ_{h} once again:

$$A^6=\bigcup_{y\in W}\left(\bigcup_{\substack{(y_1,y_2,y_3)\in S_y\\(y'_1,y'_2,y'_3)\in S_y}}\left(\overline{\mathcal{L}}(y_1,y_2,y_3)^\perp\cap\overline{\mathcal{L}}(y)^\perp\right)+\left(\overline{\mathcal{L}}(y'_1,y'_2,y'_3)^\perp\cap\overline{\mathcal{L}}(y)^\perp\right)\right)\times\{y\}$$

Now by Lemma 5, we can find a subspace $V < \mathbb{F}^m$ of codimension $O(d^2 \log(\exp(-O(d^5))) = O(d^7)$ such that $V \cap \overline{\mathcal{L}}(y)^\perp$ is contained in each of the long sums above, for every $y \in W$. This gives us the desired bilinear variety

$$B^6 = \bigcup_{y \in W} \left(V \cap \overline{\mathcal{L}}(y)^\perp\right) \times \{y\} \subset A^6$$

where $(x, y) \in B^6$ if and only if $x \in V$, $y \in W$ and $\langle x, Ly \rangle = 0$ for every $L \in \mathcal{L}$. Since there are in total $r_3 = |\mathcal{L}| = O(d)$ many such conditions, B^6 is a bilinear variety of codimension $r_1 + r_2 + r_3 = O(d^7) + O(d^{20}) + O(d) = O(\log^{80} \alpha^{-1})$.

References

[HL] Hosseini, K. and Lovett, S., *A bilinear Bogolyubov-Ruzsa lemma with polylogarithmic bounds.* arXiv preprint arXiv:1905.02222, 2019.

[San12] Sanders, T., *On the Bogolyubov–Ruzsa lemma.* Analysis & PDE, 5 (2012), no. 3, pp. 627–655.

[SS16] Schoen, T. and Sisask, O., *Roth's theorem for four variables and additive structures in sums of sparse sets.* Forum Math. Sigma, 4 (2016), e5, 28 pp.

[GM17] Gowers, T. and Milićević, L., *A bilinear version of Bogolyubov's theorem.* arXiv preprint arXiv:1712.00248, 2017.

[BL17] Bienvenu, P.-Y. and Lê, T. H., *A bilinear Bogolyubov theorem.* arXiv preprint arXiv:1711.05349, 2017.

Hayk Aprikyan, University of Bonn
email: aprikyan@uni-bonn.de

2 Small subalgebras of polynomial rings and Stillman's conjecture

After T. Ananyan and M. Hochster [AH]

A summary written by Noa Bihlmaier

Abstract

We describe a solution to Stillman's conjecture, which uniformly bounds the projective dimension of R/I for any ideal I of a polynomial ring R, generated by n homogeneous polynomials of degree at most d, independent of the number of variables of the polynomial ring.

2.1 Introduction

The starting point of Stillman's conjecture is Hilbert's famous syzygy theorem, being a consequence of combining the Koszul complex with Nakayama's lemma, see, e.g., Chapter 24.4 of [Vak]:

Theorem 1 (Hilbert's syzygy theorem, 1890)**.** *For every field K, every module over the polynomial ring $R = K[T_1, \dots, T_N]$ has projective dimension at most N, i.e. admits a projective resolution of length at most N.*

An important class of such modules is formed by R/I for some ideal I. As examples, for $R = K[T_1, \dots, T_N]$ and the ideal $I = \langle T_1, \dots, T_n \rangle$ $(n \leq N)$, the projective dimension of R/I is n; and for the ideal I generated by the three homogeneous cubics T_1^3, T_2^3 and $T_1^2T_3 + T_1T_2T_4 + T_2^2T_5$ the projective dimension of $K[T_1, \dots, T_5]/I$ is 5 (see Example 3.7 of [MCS]).

Around 2000, Stillman conjectured that the projective dimension of R/I is bounded in terms of the number of generators used, and their degrees, independent of the number of variables of the ambient polynomial ring R.

Conjecture 2 (Stillman's conjecture)**.** *For every n and every d there exists a number $B(n, d)$ such that for every N, every field K and every polynomial ring $R = K[T_1, \dots, T_N]$ it holds that for every ideal I generated by n homogeneous elements of degrees at most d, the projective dimension of R/I is at most $B(n, d)$,*

$$\mathrm{pd}_R(R/I) \leq B(n, d).$$

Remark 3. *Note that one could also bound the projective dimension of I itself instead, since* $\mathrm{pd}_R(I)$ *and* $\mathrm{pd}_R(R/I)$ *only differ by one.*

In [AH], Ananyan and Hochster prove Stillman's conjecture, and actually obtain an even slightly stronger result showing that the projective dimension of an R-module may be bounded only in terms of the number of generators, relations and their degrees of a finite presentation.

2.2 From small subalgebras to Stillman's conjecture

Stillman's conjecture follows from the existence of *small subalgebras*, which allow to reduce the number of variables of the ambient polynomial ring, dependent only on the degrees and the number of generators of the ideal; not the number of original variables of the polynomial ring.

Theorem 4 (Existence of small subalgebras)**.** *For every n and d there exists a number $B(n,d)$ with the following property.*

Let K be a field and R a polynomial ring over K. Consider elements $f_1,\dots,f_n \in R$ of degree at most d. Then there exists a regular sequence[1] *of homogeneous elements $G_1,\dots,G_s \in R$ of degree at most d with $s \leq B(n,d)$ such that all f_i are contained in the polynomial subalgebra $K[G_1,\dots,G_s] \subseteq R$.*

Once this theorem is established, Stillman's conjecture can be deduced as follows (see [Sno] for more details):

From small subalgebras to Stillman's conjecture. Take an ideal I generated by $f_1,\dots,f_n$ of degrees at most d – since these lie in $S = K[T_1,\dots,T_m]$ for small $m \leq B(n,d)$, by Hilbert's syzygy theorem, the projective dimension of S/J for the generated ideal $J = \langle f_1,\dots,f_n\rangle$ in S is at most m. Using the Koszul complex of K over S and base-changing it to R one sees that all higher $\mathrm{Tor}_i^S(R,K)$ vanish, i.e. R is flat over S. Thus one can simply tensor a projective resolution of S/J over S with R to obtain that the projective dimension of $R/I = R \otimes_S S/J$ in R is at most m. □

[1] Recall that a sequence (G_i) in a ring is called *regular* if for all i, G_i is a non-zerodivisor in $R/\langle G_1,\dots,G_{i-1}\rangle$. Since regularity implies algebraic independence, for a polynomial ring the generated subalgebra of a regular sequence is itself the polynomial ring in formal variables G_i.

2.3 Obtaining small subalgebras

For proving the existence of such small subalgebras, [AH] introduce the notion of *strength*, which captures *structuredness* of a family of elements in R.

Definition 5. *The* strength *of a homogeneous polynomial $f \in R$ of degree d is the minimal number $0 \leq k \leq \infty$ such that f lies in an ideal generated by k homogeneous elements of degree strictly less than d, i.e. there exists a decomposition*

$$f = \sum_{i=1}^{k} g_i h_i, \qquad \deg g_i < d, \deg h_i < d.$$

Similarly, the collective strength *of a set of homogeneous elements $\{f_j\}$ is the minimal strength of a nontrivial homogeneous linear combination of the f_j.*

Example 6. *In $\mathbb{R}[x_1, \ldots, x_N]$, the strength of $\sum_{i \leq n} x_i^2$ is n, and the homogeneous elements of infinite strength are precisely the nonzero linear polynomials.*

Sets of large collective strength are supposed to behave like independent variables.

Theorem 7. *For every n and d there exists an integer $A(n, d)$ with the following property. For any sequence $(f_i)_{i=1}^n$ of n homogeneous elements of degree at most d, if the collective strength of (f_i) is at least $A(n, d)$, then (f_i) is a regular sequence.*

Remark 8. *Note that for (f_i) having infinite collective strength this theorem follows quite directly – if R_+ denotes the ideal of positive degree polynomials, then infinite strength of some f_i means that $f_i \notin R_+^2$, and infinite collective strength is equivalent to linear independence in R_+/R_+^2. But a vector space basis of R_+/R_+^2 is simply a choice of coordinates of the polynomial ring.*

Given this theorem, the proof of the small subalgebra theorem runs by an induction along the degrees.

Proving small subalgebras using Theorem 7. Order finite sequences of numbers in $\mathbb{N}$ by

$$(\delta_i)_{i=1}^n < (\eta_j)_{j=1}^m,$$

if $n < m$ or $n = m$ and there exists a j such that $\delta_k = \eta_k$ for all $j < k \leq m$ and $\delta_j < \eta_j$.

Let now $(\delta_i)_{i=1}^d$ be such a degree sequence, define $n = \sum \delta_i$ and take $(f_j)_{j=1}^n$ to be a sequence consisting of δ_i many homogeneous elements of degree i. Without loss of generality, assume the f_i to be linearly independent.

If only degree 1 appears the statement is clear, so let us induct along the well-ordering described above.

If the collective strength of (f_i) is at least $A(n, d)$, then (f_i) forms a regular sequence and thus $K[f_1, \ldots, f_n]$ itself forms a small subalgebra with n generators.

Else, there is a linear combination f of the f_i of low strength, i.e. expressible as sum

$$f = \sum g_j h_j$$

with at most $A(n, d)$ many g_j and h_j of strictly smaller degree. Extend f to a basis of the linear span of the (f_i), and then replace f by all the g_i and h_i to obtain a new sequence (f'_j) with smaller degree sequence. Now apply the induction hypothesis to embed (f'_j) into a polynomial subalgebra $K[G_1, \ldots, G_s] \subseteq R$ for $s \leq B(n', d')$.

But now note that the original f_i are contained in the generated subalgebra of the (f'_j) and thus also lie in $K[G_1, \ldots, G_s]$ with at most $B(n', d')$ variables. Note that only finitely many degree sequences may be reached by this process, since deleting an element of degree ℓ produces at most $2A(n, d)$ new elements and these in strictly smaller degree.

Thus, we may take the maximum over these and use this as $B(n, d)$. □

2.4 Large strength implies regularity

Ananyan and Hochster prove all of the above theorems by a complex simultaneous induction involving multiple other theorems, concentrating on the singular locus of homogeneous polynomials of large strength. One of their key results, translating strength to classical algebraic geometry, is the following.

Theorem 9. *For every $d, \eta > 0$ there exists a number $A(d, \eta)$ such that for every algebraically closed field K and every positive integer N for every*

homogeneous polynomial $F \in R = K[T_1, \ldots, T_N]$ of degree d; if the strength of F is at least $A(d, \eta)$, then $R/\langle F \rangle$ satisfies the Serre condition R_η, i.e. the codimension of the singular locus in $R/\langle F \rangle$ is bigger than η.

However, in [ESS] and [Sno], there is a short and elegant argument for proving the regularity theorem by passing to ultraproducts.

Proof sketch of Theorem 7 via ultraproducts. Assume the contrary. Then for every k there exists a sequence $(f_{i,k})_{i=1}^n$ of n homogeneous elements of degree $\leq d$ in some polynomial ring R_k, having collective strength at least k but not being regular in R_k.

Take the graded ultraproduct $R = \prod_{k \to p} R_k$ along some free ultrafilter $p \in \beta\mathbb{N}$. The ring R can be seen to be again a polynomial ring (since it has enough (Hasse) derivations), the ultraproduct of the $f_{i,k}$ yields a sequence f_i of infinite collective strength in R, and thus f_i are a regular sequence in R. This in turn implies that $(f_{i,k})_{i=1}^n$ are regular in R_k for a large set of indices k, and thereby yields a contradiction. □

References

[AH] Ananyan, T. and Hochster, M., *Small subalgebras of polynomial rings and Stillman's conjecture*, J. Amer. Math. Soc. **33** (2020), no. 1, 291–309;

[ESS] Erman, D., Sam, S. and Snowden, A., *Big polynomial rings and Stillman's conjecture*, Invent. Math., **218** (2019), pp. 413-439.

[MCS] McCullough, J. and Seceleanu, A., *Bounding projective dimension*, In: Peeva, I. (ed.) *Commutative Algebra*. Springer (2013).

[Sno] Snowden, A., *Infinite dimensional methods in commutative algebra*, Workshop on Algebra and Representation Theory, Held on Oregonian Grounds, 2022, lecture notes, available online (27.08.2026).

[Vak] Vakil, R., *The Rising Sea – Foundations for Algebraic Geometry* Princeton University Press, 2025.

Noa Bihlmaier, University of Bonn
email: nobi@math.uni-bonn.de

3 Strong bounds for 3-progressions

After Z. Kelley and R. Meka [KM]

A summary written by Aleksandar Bulj

Abstract

We present the Kelley–Meka bound for sets free of three-term progressions in the finite field model and the main structural lemma used in that proof, following the exposition given by T. F. Bloom and O. Sisask [BS].

3.1 Introduction and main results

We say that a set $A \subseteq \mathbb{F}_q^n$ has *density* α if $|A| = \alpha|\mathbb{F}_q^n|$.

The main theorem in the finite field setting is the following.

Theorem 1 ([KM], Theorem 1.3; [BS], Theorem 2)**.** *Let q be an odd prime and let $A \subseteq \mathbb{F}_q^n$ have density α. If A contains no non-trivial three-term progression then $\alpha \leq 2\exp(-c\,n^{1/9})$ for an absolute $c > 0$.*

Kelley and Meka deduced this from the structural result given by the following theorem; see Subsection 3.1.1 for the notation.

Theorem 2 ([KM], Lemma 1.14; [BS], Theorem 4)**.** *Let $\varepsilon \in (0,1)$. There is $C = C(\varepsilon) > 0$ such that the following holds. Let $G = \mathbb{F}_q^n$ for a prime q and $n \geq 1$, and let $p \geq 1$ be an integer. For a nonempty $A \subseteq G$ of density α there are a subspace $V \leq G$ with $\operatorname{codim}(V) \leq Cp^4 \log(2/\alpha)^5$ and a point $x \in G$ such that $A' = (A - x) \cap V$ satisfies*

$$\textit{(i)}\ |A'|/|V| \geq \alpha, \quad \textit{and} \quad \textit{(ii)}\ \left\|\mu_{A'}^{(V)} *_V \mu_{A'}^{(V)} - 1\right\|_{L^p(V)} \leq \varepsilon.$$

Remark. The density loss to $(1-\varepsilon)\alpha$ in the original statement (i) of [BS], Theorem 4 is an artifact of using Bourgain's narrowing trick ([BS], Lemma 16). This trick is needed for Bohr sets but not for subspaces, so no such loss appears in the finite field setting. Regardless, this difference is immaterial in the intended applications.

3.1.1 Notation

Let $G = \mathbb{F}_q^n$. For any $S \subseteq G$, we define $\mathbb{E}_{x\in S} = \frac{1}{|S|}\sum_{x\in S}$. For a linear subspace $V \leq G$ and $f, g : V \to \mathbb{C}$, we define $\|f\|_{L^p(V)} := (\mathbb{E}_{x\in V}|f(x)|^p)^{1/p}$, $\langle f, g\rangle_V = \mathbb{E}_{x\in V} f(x)\overline{g(x)}$, $f *_V g(x) = \mathbb{E}_{y\in V} f(y)g(x-y)$ and $f \circ_V g(x) = \mathbb{E}_{y\in V} f(y)\overline{g(x+y)}$. When $V = G$, we write $\|\cdot\|_p$ and omit writing the subscript G.

For $V \leq G$ and $A \subseteq V$ of density $\alpha = |A|/|V|$, we put $\mu_A^{(V)} = \alpha^{-1}1_A$; when $V = G$, we omit the superscript G. We write $\mathcal{L}(\alpha) = \log(2/\alpha)$. Finally, we use the following normalization for the Fourier transform

$$\widehat{f}(\xi) = \mathbb{E}_{x\in G} f(x)e(-x\cdot\xi), \quad \text{where} \quad e(t) = e^{2\pi it/q}.$$

3.2 Proof of Theorem 1 using Theorem 2

For random sets $A, C \subseteq G$ one expects $\langle \mu_A * \mu_A, \mu_C\rangle \approx 1$. Conclusion (ii) of Theorem 2 allows us to control such quantities. Indeed, if C has density $\gamma \geq 2^{-p}$, then $\|\mu_C\|_{p'} = \gamma^{-1/p} \leq 2$, so Hölder's inequality gives

$$|\langle \mu_A * \mu_A, \mu_C\rangle - 1| = |\langle \mu_A * \mu_A - 1, \mu_C\rangle| \leq 2\|\mu_A * \mu_A - 1\|_p.$$

We can prove Theorem 1 by counting only the progressions inside the structured set A'.

Proof of Theorem 1. Apply Theorem 2 with $\varepsilon = 1/4$ and $p = 2\lceil\mathcal{L}(\alpha)\rceil$: we obtain $V \leq \mathbb{F}_q^n$ with $s := \operatorname{codim}(V) \ll p^4\mathcal{L}(\alpha)^5 \ll \mathcal{L}(\alpha)^9$ and $A' = (A-x)\cap V$ of relative density $\alpha' \geq \alpha$ with $\|\mu_{A'}^{(V)} *_V \mu_{A'}^{(V)} - 1\|_{L^p(V)} < 1/4$. As q is odd, $C := 2\cdot A'$ is again a subset of V of relative density α', so the calculation at the beginning of this section applies inside $V \cong \mathbb{F}_q^{n-s}$ with $\gamma = \alpha'$ and gives $\langle \mu_{A'}^{(V)} *_V \mu_{A'}^{(V)}, \mu_C\rangle_V \geq 1/2$. Since A has no non-trivial 3-progression, neither does A', so only the trivial ones are counted by

$$\langle \mu_{A'}^{(V)} *_V \mu_{A'}^{(V)}, \mu_C^{(V)}\rangle_V = \alpha'^{-3}\mathbb{E}_{x,y\in V} 1_{A'}(y)1_{A'}(x-y)1_{2\cdot A'}(x) = (\alpha'^2|V|)^{-1}.$$

Hence $\alpha^2|V| \leq \alpha'^2|V| \leq 2$, and since $|V| = q^{n-s}$, a little algebra gives $n \ll \mathcal{L}(\alpha)^9$, i.e., $\alpha \leq 2\exp(-cn^{1/9})$. □

3.3 Key lemmas

Even though we state the following lemmas in terms of $\varepsilon > 0$, one does not have to worry about dependence of constants when $\varepsilon \to 0$ because we will choose some $\varepsilon \gg 1$.

The first main lemma is the following.

Lemma 3 (Unbalancing, [BS], Lemma 6)**.** *Let $\varepsilon \in (0,1)$. Let $f : G \to \mathbb{R}$ be such that $\widehat{f} \geq 0$. If $\|f\|_p \geq \varepsilon$ for some $p \geq 1$, then $\|f+1\|_{p'} \geq 1+\varepsilon/2$ for some $p' \ll_\varepsilon p$. In particular, if $\|\mu_A \circ \mu_A - 1\|_p \geq 1/4$, then $\|\mu_A \circ \mu_A\|_{p'} \geq 1 + 1/8$ for some $p' \ll_\varepsilon p$.*

Sketch of proof. First, observe that for any $k \in \mathbb{N}$ one has $\mathbb{E}_x f^k(x) = \widehat{f} * \cdots * \widehat{f}(0) \geq 0$, so for $p' \in 2\mathbb{N}$, using nonnegativity for odd powers and monotonicity of L^p for even powers, for $p' \geq p$ one has

$$\|1+f\|_{p'}^{p'} = \mathbb{E}_x \sum_{0 \leq k \leq p'} \binom{p'}{k} f^k(x) \geq \sum_{\substack{p \leq k \leq p' \\ k \text{ even}}} \binom{p'}{k} \varepsilon^k.$$

The last expression can be interpreted as $(1+\varepsilon)^{p'}$ times the probability that a random variable $X \sim B(p', \varepsilon/(1+\varepsilon))$ (i.e. the binomial distribution with p' trials and success probability $\varepsilon/(1+\varepsilon)$) satisfies $X \geq p$ and X is even. Using the central limit theorem, we know that the median of X is $\approx p'\varepsilon/(1+\varepsilon)$ and X is even approximately $1/2$ of the time, so if $p \leq \text{median}(X)/2$, then the probability should be at least $1/2 \cdot 1/2 = 1/4$. Thus, the last expression is bounded from below by $4^{-1}(1+\varepsilon)^{p'}$. A little algebra gives the conclusion. $\square$

Lemma 4 (Dependent random choice, [BS], Lemma 10)**.** *Let $p \geq 1$ be an integer, let $A \subseteq G$ have density $\alpha > 0$ and let $f : G \to \mathbb{R}_{\geq 0}$. Then there is $A' \subseteq G$ with $|A'|/|G| \geq \frac{1}{2}\alpha^p \|\mu_A \circ \mu_A\|_p^p$ such that*

$$\langle \mu_{A'} \circ \mu_{A'}, f \rangle \leq 2\langle (\mu_A \circ \mu_A)^p, f \rangle / \|\mu_A \circ \mu_A\|_p^p.$$

Sketch of proof. The key identity is the following. For any integer $p \geq 1$ and any $f : G \to \mathbb{R}_{\geq 0}$,

$$\langle (\mu_A \circ \mu_A)^p, f \rangle = \alpha^{-2p} \mathbb{E}_{s \in G^p} \langle 1_{(A+s_1) \cap \cdots \cap (A+s_p)} \circ 1_{(A+s_1) \cap \cdots \cap (A+s_p)}, f \rangle,$$

which follows by expansion of the p-th power of the sum. After applying this identity to the numerator and, with $f \equiv 1$, to the denominator of the expression, it remains to use the following pigeonhole principle with appropriately chosen H.

Let $(g_j)_{j=1}^n$ and $(h_j)_{j=1}^n$ be real numbers and $\eta := (\sum_{j=1}^n g_j)/(\sum_{j=1}^n h_j)$. If $H > 0$ satisfies $\sum_{j:h_j \geq H} h_j \geq (\sum_{j=1}^n h_j)/2$, then there exists $j \in [n]$ such that $\frac{g_j}{h_j} \leq 2\eta$ and $h_j \geq H$. □

Corollary 5 (Sifting, [BS], Corollary 9)**.** *Let $p \geq 1$ be an integer and $\varepsilon \in (0,1)$. If $A \subseteq G$ has density α and $\|\mu_A \circ \mu_A\|_p \geq 1+\varepsilon$, then with $S = \{x : \mu_A \circ \mu_A(x) > 1+\varepsilon/2\}$ there is $A' \subseteq G$ of density $\gg \alpha^{p+O_\varepsilon(1)}$ such that $\langle \mu_{A'} \circ \mu_{A'}, 1_S\rangle \geq 1-\varepsilon/8$.*

Proof. Apply Lemma 4 with $f = 1_{G\setminus S}$. Since $\mu_A \circ \mu_A \leq 1+\varepsilon/2$ off S, satisfies $\langle \mu_{A'} \circ \mu_{A'}, 1_{G\setminus S}\rangle \leq 2(\frac{1+\varepsilon/2}{1+\varepsilon})^{p''} \leq \varepsilon/8$ once $p'' \gg \varepsilon^{-1}\log(2/\varepsilon)$. Now use $\langle \mu_{A'} \circ \mu_{A'}, 1_S\rangle = 1 - \langle \mu_{A'} \circ \mu_{A'}, 1_{G\setminus S}\rangle$. □

The last ingredient is [SS], Theorem 3.2, in the form of [BS], Theorem 11; it combines Croot–Sisask sampling with Chang's theorem and makes no hypothesis on S. For a self-contained proof in the finite field setting, see [KM], Appendix C.

Theorem 6 (Almost periodicity)**.** *Let $\varepsilon \in (0,1)$. If $A_1, A_2, S \subseteq \mathbb{F}_q^n$ are such that A_1, A_2 have density at least α, then there is a subspace V with $\mathrm{codim}(V) \ll_\varepsilon \mathcal{L}(\alpha)^4$ such that $|\langle \mu_V * \mu_{A_1} \circ \mu_{A_2}, 1_S\rangle - \langle \mu_{A_1} \circ \mu_{A_2}, 1_S\rangle| \leq \varepsilon$.*

3.4 Proof of Theorem 2

Proposition 7 (Density increment; [BS], Prop. 15)**.** *Let $\varepsilon \in (0,1]$, let $p \geq 1$ be an integer and let $A \subseteq \mathbb{F}_q^n$ have density α. If $\|\mu_A \circ \mu_A\|_p \geq 1+\varepsilon$, then there is a subspace $V \leq \mathbb{F}_q^n$ of codimension $\ll_\varepsilon \mathcal{L}(\alpha)^4 p^4$ with $\|\mu_V * \mu_A\|_\infty \geq 1+\varepsilon/8$.*

Proof. Using Corollary 5 we obtain $A' \subseteq \mathbb{F}_q^n$ of density $\alpha' \gg \alpha^{p+O_\varepsilon(1)}$ such that $\langle \mu_{A'} \circ \mu_{A'}, 1_S\rangle \geq 1-\varepsilon/8$ for $S = \{\mu_A \circ \mu_A > 1+\varepsilon/2\}$. Theorem 6 applied to $A_1 = A_2 = A'$ with error $\varepsilon/8$ then gives $V \leq \mathbb{F}_q^n$ of codimension $\ll_\varepsilon \mathcal{L}(\alpha')^4 \ll_\varepsilon p^4\mathcal{L}(\alpha)^4$ such that

$$\langle \mu_V * \mu_{A'} \circ \mu_{A'}, 1_S\rangle \;\geq\; 1 - \frac{\varepsilon}{4}.$$

Using Hölder's inequality and then the definition of S,

$$\begin{aligned}\|\mu_V * \mu_A\|_\infty &= \|\mu_V * \mu_A\|_\infty \|\mu_A * \mu_{A'} \circ \mu_{A'}\|_1 \\ &\geq \langle \mu_V * \mu_{A'} \circ \mu_{A'},\ \mu_A \circ \mu_A\rangle \;\geq\; \Big(1+\frac{\varepsilon}{2}\Big)\langle \mu_V * \mu_{A'} \circ \mu_{A'}, 1_S\rangle \\ &\geq \Big(1+\frac{\varepsilon}{2}\Big)\Big(1-\frac{\varepsilon}{4}\Big) \;\geq\; 1+\frac{\varepsilon}{8}.\end{aligned}$$

□

Proof of Theorem 2. We may assume that p is even. Let $p' \ll_\varepsilon p$ be the exponent supplied by Lemma 3 for the pair (ε, p), and set $V_0 := \mathbb{F}_q^n$ and $A_0 := A$. Given $A_k \subseteq V_k$ of relative density α_k, identify $V_k \cong \mathbb{F}_q^{\dim V_k}$ and stop if $\|\mu_{A_k}^{(V_k)} \circ_{V_K} \mu_{A_k}^{(V_k)}\|_{L^{p'}(V_k)} < 1 + \varepsilon/2$; otherwise Proposition 7, applied inside V_k with the parameter $\varepsilon/2$, gives $V_{k+1} \leq V_k$ of relative codimension $\ll_\varepsilon \mathcal{L}(\alpha_k)^4 (p')^4 \ll_\varepsilon \mathcal{L}(\alpha)^4 p^4$, and a point $x_{k+1} \in V_k$ on whose coset A_k has relative density at least $(1 + \varepsilon/8)\alpha_k$; put $A_{k+1} := (A_k - x_{k+1}) \cap V_{k+1}$. Since $\alpha_k \geq (1 + \varepsilon/8)^k \alpha$ and $\alpha_k \leq 1$, we stop after $K \ll_\varepsilon \mathcal{L}(\alpha)$ steps, and relative codimensions add, so $\operatorname{codim}(V_K) \ll_\varepsilon \mathcal{L}(\alpha)^5 p^4$. As $x_{k+1} \in V_k$ we have $A_K = (A - x) \cap V_K$ with $x = x_1 + \cdots + x_K$, and the relative density never decreased, which is (i). Finally $\|\mu_{A_K}^{(V_K)} \circ_{V_K} \mu_{A_K}^{(V_K)}\|_{L^{p'}(V_K)} < 1 + \varepsilon/2$, so Lemma 3 gives $\|\mu_{A_K}^{(V_K)} \circ \mu_{A_K}^{(V_K)} - 1\|_{L^p(V_K)} < \varepsilon$, finally implying that $\|\mu_{A_K}^{(V_K)} * \mu_{A_K}^{(V_K)} - 1\|_{L^p(V_K)} < \varepsilon$, which follows from using the fact that p is even, passing to the Fourier side, and using the triangle inequality. □

References

[BS] Bloom, T. F. and Sisask, O., *The Kelley–Meka bounds for sets free of three-term arithmetic progressions.* `arXiv:2302.07211`

[KM] Kelley, Z. and Meka, R., *Strong bounds for 3-progressions.* In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 1–21. IEEE, 2023. `arXiv:2302.05537`

[SS] Schoen, T. and Sisask, O., *Roth's theorem for four variables and additive structures in sums of sparse sets.* Forum Math. Sigma **4** (2016), e5.

Aleksandar Bulj, University of Zagreb
email: `aleksandar.bulj@math.hr`

4 The Quadratic Goldreich–Levin theorem

After J. Briët and D. Castro-Silva

A summary written by Davi Castro-Silva

Abstract

We give a short high-level overview of the Quadratic Goldreich–Levin theorem proven by Briët and Castro-Silva.

4.1 Introduction

Let $f : \mathbb{F}_2^n \to [-1, 1]$ be a 1-bounded function. The celebrated *Goldreich–Levin theorem* [5] – in one of its guises – can be viewed as a remarkably efficient algorithm that learns the "linear structure" of f: it outputs a small list containing all non-negligible Fourier coefficients of f, in $\mathrm{poly}(n)$ time. Many important properties of bounded functions are effectively captured by this list of coefficients, which makes the algorithm widely applicable.

With the view of extending such algorithmic techniques to higher-order Fourier analysis, Tulsiani and Wolf [7] proposed a quadratic generalization of this classical result. In that setting, the analogue of a large Fourier coefficient is a quadratic polynomial $q : \mathbb{F}_2^n \to \mathbb{F}_2$ that correlates well with f:

$$|\langle f, (-1)^q \rangle| := |\mathbb{E}_{x \in \mathbb{F}_2^n} f(x)(-1)^{q(x)}| \geq \tau,$$

for some constant $\tau > 0$. While Parseval's identity implies that there are at most $1/\tau^2$ Fourier coefficients $y \in \mathbb{F}_2^n$ such that $|\widehat{f}(y)| \geq \tau$, the same does not hold in the quadratic setting: the number of large "quadratic Fourier coefficients" can grow exponentially with the dimension n. For this reason, a quadratic generalization of the Goldreich–Levin algorithm is supposed to output a single quadratic that correlates well with f, rather than a complete list (which would be impossible to do efficiently).

For convenience, let us denote by $\|f\|_{u^3}$ the maximum possible correlation of f with a quadratic phase function:

$$\|f\|_{u^3} := \max_{q \text{ quadratic}} |\mathbb{E}_{x \in \mathbb{F}_2^n} f(x)(-1)^{q(x)}|.$$

The original Quadratic Goldreich–Levin (QGL) theorem of Tulsiani and Wolf [7] gave an algorithm that, when given query access to a 1-bounded

function f satisfying $\|f\|_{u^3} \geq \tau$, outputs in $\mathrm{poly}(n)$ time a quadratic function $q : \mathbb{F}_2^n \to \mathbb{F}_2$ such that $|\mathbb{E}_{x\in\mathbb{F}_2^n} f(x)(-1)^{q(x)}| \geq 2^{-\mathrm{poly}(1/\tau)}$. This work was later built upon by Ben-Sasson, Ron-Zewi, Tulsiani and Wolf [1], who obtained the stronger guarantee $|\mathbb{E}_{x\in\mathbb{F}_2^n} f(x)(-1)^{q(x)}| \geq 2^{-\mathrm{poly}(\log(1/\tau))}$.

The main result of [2] is an essentially lossless version of the QGL theorem:

Theorem 1 (Quadratic Goldreich-Levin)**.** *Let $f : \mathbb{F}_2^n \to [-1, 1]$ be a 1-bounded function and let $\varepsilon > 0$. There is a randomized algorithm $\mathcal{A}$ that makes $n^2 \log n\,(1/\varepsilon)^{O(\log(1/\varepsilon))}$ queries to f and, with probability at least $2/3$, outputs a quadratic polynomial $p : \mathbb{F}_2^n \to \mathbb{F}_2$ such that*

$$\left|\mathbb{E}_{x\in\mathbb{F}_2^n} f(x)(-1)^{p(x)}\right| > \|f\|_{u^3} - \varepsilon.$$

In addition to the $\tilde{O}(n^2)$ queries, the algorithm $\mathcal{A}$ runs in time $O(n^3)$.

Combining this result with the polynomial U^3-inverse theorem by Gowers, Green, Manners and Tao [6], one immediately obtains the following corollary:

Corollary 2 (Algorithmic polynomial Gowers inverse theorem)**.** *Let $\gamma > 0$ and let $f : \mathbb{F}_2^n \to [-1, 1]$ be a function with $\|f\|_{U^3} \geq \gamma$. There is a randomized algorithm $\mathcal{A}$ that makes $n^2 \log n/\gamma^{O(\log(1/\gamma))}$ queries to f, runs in time $O(n^3)$ and that, with probability at least $2/3$, outputs a quadratic polynomial $p : \mathbb{F}_2^n \to \mathbb{F}_2$ satisfying $\left|\mathbb{E}_{x\in\mathbb{F}_2^n} f(x)(-1)^{p(x)}\right| \geq (\gamma/2)^C$ where $C > 0$ is some absolute constant.*

The techniques used in the proof of these results differ significantly from the earlier approaches in [7, 1]. The earlier QGL algorithms operate by making constructive each step in the proofs of the U^3-inverse theorem that were known at the time, and the bounds obtained reflected the then-current state of the art. A similar strategy, however, seems harder to implement for the recent proof in [6], which is information-theoretic in nature.

Instead, the algorithm in Theorem 1 borrows techniques from *quantum learning theory*, in particular those for learning stabilizer states. In fact, this algorithm may be seen as a "dequantization" of the recent agnostic stabilizer tomography algorithm of Chen, Gong, Ye and Zhang [4]. The reader is referred to [3, Section 2] for a mathematical overview of stabilizer states and the connection between symplectic concepts and quadratic Fourier analysis.

4.2 Outline of the algorithm

Recall that the *characteristic distribution* of a function $f : \mathbb{F}_2^n \to \mathbb{C}$ is the probability distribution P_f over $\mathbb{F}_2^n \times \mathbb{F}_2^n$ given by

$$P_f(a,b) = \frac{|\widehat{\Delta_a f}(b)|^2}{2^n \|f\|_2^4}.$$

Recall also that *stabilizer states* are the extremizers of the U^3 norm under L^2 normalization:

$$\mathrm{Stab}(\mathbb{F}_2^n) = \big\{\phi : \mathbb{F}_2^n \to \mathbb{C} \mid \|\phi\|_{U^3} = \|f\|_2 = 1\big\}.$$

The algorithm of Theorem 1 operates on the standard symplectic vector space $V = (\mathbb{F}_2^{2n}, [\cdot,\cdot])$. The key mathematical idea underlying this algorithm is the following fact: the *quadratic structure* of a function $f : \mathbb{F}_2^n \to \mathbb{C}$ is reflected on a Lagrangian subspace $L \leq V$ that has high characteristic weight $P_f(L)$. Indeed, this quantity provides a natural bridge between the U^3 norm of f and its maximal correlation with a stabilizer state: we have

$$\max_{\phi \in \mathrm{Stab}(\mathbb{F}_2^n)} \left(\frac{|\langle f,\, \phi\rangle|}{\|f\|_2}\right)^4 \leq \max_{L \text{ Lagrangian}} P_f(L) \leq \left(\frac{\|f\|_{U^3}}{\|f\|_2}\right)^4. \tag{1}$$

(This is proven in [3, Section 2.7].) By the polynomial U^3-inverse theorem on L^2, it follows that the maximum P_f-weight of a Lagrangian subspace is polynomially related to $\|f\|_{U^3}/\|f\|_2$.

Crucially, knowing a Lagrangian of high P_f-weight provides significant information about a stabilizer state that correlates well with f. Indeed, stabilizer states are naturally associated with Lagrangian subspaces: a function $\phi : \mathbb{F}_2^n \to \mathbb{C}$ is a stabilizer state iff

$$|\widehat{\Delta_a \phi}(b)| = \mathbf{1}_L(a,b) \quad \text{for all } (a,b) \in V$$

holds for some Lagrangian $L = \mathcal{L}(\phi) \leq V$. This equation determines, for each Lagrangian L, (scalar multiples of) an orthonormal basis of stabilizer states; the elements of this basis share the same "quadratic component", but differ by a linear phase and a translate. The maximum correlation of f with an element of this basis is tightly connected to the value of $P_f(L)$:

$$\max_{\phi:\, \mathcal{L}(\phi)=L} \left(\frac{|\langle f,\, \phi\rangle|}{\|f\|_2}\right)^4 \leq P_f(L) \leq \max_{\phi:\, \mathcal{L}(\phi)=L} \left(\frac{|\langle f,\, \phi\rangle|}{\|f\|_2}\right)^2. \tag{2}$$

Suppose then that we can find a Lagrangian subspace $L \leq V$ with high P_f-weight. By equation (2) above, it follows that some stabilizer state ϕ in the basis determined by L has high correlation with f. Finally, if the original function f is L^∞-bounded, one can show that any stabilizer state that correlates well with f can be "rounded" to a quadratic phase function $(-1)^q$ without loosing much in terms of correlation.

These ideas inform the algorithmic strategy in [2, 3], starting from a function $f : \mathbb{F}_2^n \to [-1, 1]$ with $\|f\|_{U^3} \geq \tau$:

Step 1. Find a Lagrangian subspace $L \leq V$ with high characteristic weight: $P_f(L) \geq \text{poly}(\tau)$. Such a Lagrangian exists due to the polynomial U^3-inverse theorem and equation (1).

This is the central step of the algorithm. The intuition here is that, since $P_f(L)$ is large, one might find a basis for L by sampling many times from P_f and keeping only vectors whose pairwise symplectic inner product is zero. This intuition can be transformed into an *efficient* algorithm using three key ideas:

(*i*) A "boosting" argument, which makes it easier to sample elements from L after randomly projecting f onto suitable subspaces of the form

$$V_{a,b}^{\zeta} = \left\{ g : \mathbb{F}_2^n \to \mathbb{C} \mid g(x+a) = \zeta \cdot (-1)^{b \cdot x} g(x) \text{ for all } x \in \mathbb{F}_2^n \right\},$$

for some $\zeta \in \{1, i, -1, -i\}$ and $(a, b) \in L$. Note that every stabilizer state in the basis determined by L belong to one such subspace $V_{a,b}^{\zeta}$ for each $(a, b) \in L$, which explains why they are considered. The parameters determining these good subspaces can be sampled with non-negligible probability; by repeating the process several times, we will be able to get them right at least once with high probability.

(*ii*) The following "uncertainty principle" that relates P_f-weight with isotropy: for any function $f : \mathbb{F}_2^n \to \mathbb{C}$, the set $\{v \in V \mid P_f(v) > 0.5/2^n\}$ is isotropic. As $2^n P_f(a, b) = |\widehat{\Delta_a f}(b)|^2 / \|f\|_2^4$ can be easily estimated for any $(a, b) \in V$, this provides an efficient way to decide which of the sampled vectors to keep for the desired basis of the Lagrangian.

(*iii*) An efficient sampling procedure for a suitable probability distribution μ_f closely related to P_f. In [2, 3], this distribution satisfies $\mu_f(v) \approx \|f\|_2^8 P_f * P_f(v)$, and sampling from it relies on the (linear) Goldreich–Levin algorithm.

Step 2. Among all stabilizer states in the orthonormal basis determined by L, find one that maximizes the correlation with f. For such a stabilizer state ϕ, equation (2) establishes that $|\langle f, \phi\rangle| \geq \mathrm{poly}(\tau)$.

This step essentially amounts to Fourier analysis to determine the optimal "linear part" of ϕ from among all relevant stabilizer states. This can be done with help from the original Goldreich–Levin algorithm.

Step 3. Find a quadratic function $q : \mathbb{F}_2^n \to \mathbb{F}_2$ that "matches" the phase of ϕ, and which satisfies

$$|\mathbb{E}_{x\in\mathbb{F}_2^n} f(x)(-1)^{q(x)}| = |\langle f, (-1)^q\rangle| \geq |\langle f, \phi\rangle|^2/\sqrt{2}.$$

Recall that stabilizer states essentially amount to a quadratic phase function $(-1)^{p(x)}$ supported on an affine subspace – up to some normalizing factor $\sqrt{2^d}$ and a linear-like modulation $i^{\ell(x)}$ on the fourth-roots of unity. If one randomly extends the domain of the "quadratic part" p to the whole space $\mathbb{F}_2^n$, on average the correlation with f does not change; we thus obtain a randomized quadratic function p' such that $|\langle f, (-1)^{p'} i^{\ell}\rangle| \geq 2^{-d/2}|\langle f, \phi\rangle|$ in expectation. It is not hard to show that the normalizing factor $2^{d/2}$ is at most $\|f\|_\infty/|\langle f, \phi\rangle|$, from which we conclude that $|\langle f, (-1)^{p'} i^{\ell}\rangle| \geq |\langle f, \phi\rangle|^2$. Finally, it is possible to replace the fourth-root modulation $i^{\ell(x)}$ by a quadratic phase $(-1)^{r(x)}$ while losing at most a $\sqrt{2}$-factor in correlation, from which we obtain the claimed inequality for $q = p' + r$. By step 2, it then follows that $|\mathbb{E}_{x\in\mathbb{F}_2^n} f(x)(-1)^{q(x)}| \geq \mathrm{poly}(\tau)$.

We note that a more detailed analysis of the algorithm outlined above guarantees that the output is essentially optimal, rather than merely polynomially strong, thus obtaining Theorem 1.

References

[1] E. Ben-Sasson, N. Ron-Zewi, M. Tulsiani, and J. Wolf, Sampling-based proofs of almost-periodicity results and algorithmic applications, in *Automata, Languages, and Programming (ICALP 2014)*, Lecture Notes in Computer Science 8572, Springer, 2014, pp. 955–966.

[2] J. Briët and D. Castro-Silva, A near-optimal quadratic Goldreich–Levin algorithm, in *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '26)*, 2026, pp. 6233–6239.

[3] D. Castro-Silva, J. Briët, S. Arunachalam, A. Dutt, and T. Gur, An algorithmic Polynomial Freiman–Ruzsa theorem, *arXiv preprint arXiv:2604.04547*, 2026.

[4] S. Chen, W. Gong, Q. Ye, and Z. Zhang, Stabilizer bootstrapping: A recipe for efficient agnostic tomography and magic estimation, in *Proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC '25)*, 2025, pp. 429–438.

[5] O. Goldreich and L. A. Levin, A hard-core predicate for all one-way functions, in *Proceedings of the 21st Annual ACM Symposium on Theory of Computing (STOC '89)*, 1989, pp. 25–32.

[6] W. T. Gowers, B. Green, F. Manners, and T. Tao, On a conjecture of Marton, *Annals of Mathematics* 201(2) (2025), 515–549.

[7] M. Tulsiani and J. Wolf, Quadratic Goldreich–Levin theorems, *SIAM Journal on Computing* 43(2) (2014), 730–766.

Davi Castro-Silva, University of Warwick
email: davisilva15@gmail.com

5 A sum–product estimate in finite fields, and applications

After Jean Bourgain, Nets Katz and Terence Tao [BKT]

A summary written by Jihyo Chae

Abstract

In this summary, we present the proof of the sum–product estimate of Bourgain–Katz–Tao, stating that any subset A of a prime field F such that $|F|^\delta \leq |A| \leq |F|^{1-\delta}$ satisfies the bound

$$\max(|A+A|, |A \cdot A|) \geq c(\delta)|A|^{1+\varepsilon}.$$

5.1 Introduction

One of the most studied phenomena in arithmetic combinatorics is the sum–product phenomenon, which says that the sumset and the product set of a set in a ring cannot be both small unless the set is close to being a subring.

Bourgain, Katz and Tao [BKT] first discovered this phenomenon in prime fields, and found its connections to some geometric problems via a Szemerédi-Trotter type bound. Formally, they proved the following bound on sumsets and product sets in finite fields. In the following discussion, F denotes the prime field of order q.

Theorem 1. *Let $A \subseteq F$ be a set such that*

$$|F|^\delta \leq |A| \leq |F|^{1-\delta}$$

holds for some $\delta > 0$. Then there exists $\varepsilon = \varepsilon(\delta) > 0$ such that the bound

$$\max(|A+A|, |A \cdot A|) \geq c(\delta)|A|^{1+\varepsilon}$$

holds, with $c(\delta)$ being a constant depending solely on δ.

It is worth noting that the same result holds in the regime $1 < |A| < |F|^\delta$; see [BGK]. Also, the same bound does not hold for non-prime finite fields due to the existence of proper subfields. This theorem is interesting in its own sake as it reveals a fundamental connection between combinatorial number theory and projection theory. Below, we introduce some ideas and key machinery of the proof.

5.2 Outline of the proof

First, we prove that both $A + A$ and $A \cdot A$ being small implies that for any multivariate polynomial $P(x_1, \ldots, x_n)$, the size of the set

$$P(A, \ldots, A) \coloneqq \{P(a_1, \ldots, a_n) : a_1, \ldots, a_n \in A\}$$

is strictly smaller than the size of the whole field. This is proved by the Balog–Szemerédi–Gowers theorem and double counting.

Next, we prove that for any subset $A \subseteq F$ satisfying the size condition $|F|^\delta \leq |A| \leq |F|^{1-\delta}$, there exists some positive integer d and a polynomial P_A on d variables such that the image of A^d under P_A equals the whole field. To prove this, we construct a linear surjection from A^k to F and compress this surjection to be a function from $\tilde{A}^{k-1}$ to F, while $\tilde{A}$ is defined as $p(A, \ldots, A)$ with p being a simple, given polynomial. By reducing the rank of this surjection k times, we obtain a polynomial P_A whose image of A^d under the map equals F. This technique originates from [EM]. Now we get a contradiction to the size bound on the polynomial expression of A with assumed properties, thereby proving Theorem 1.

5.3 Key Lemmas

A key ingredient in the proof of this theorem is the Balog–Szemerédi–Gowers Theorem, which tells the existence of a large subset with small doubling in a set with large additive energy.

Lemma 2. *Let A, B be finite subsets of an abelian group Γ with cardinality $|A| = |B| = n$ and G be a subset of $A \times B$ such that both $|G| \geq n^2/K$ and $|\{a + b : (a, b) \in G\}| \leq Kn$ hold. Then there exists $A' \subseteq A$ and $B' \subseteq B$ such that both*

$$|A' - B'| \leq CK^C n,\ |A'|, |B'| \geq cK^{-C} n$$

holds.

In some cases, we need the following stronger version of the same theorem.

Lemma 3. *Under the same assumptions, for any $a' \in A$ and $b' \in B$, there exist at least $cK^{-C}|A|^5$ solutions $(a_1, a_2, a_3, b_1, b_2, b_3) \in A \times A \times A \times B \times B \times B$ to the equation*

$$a' - b' = (a_1 - b_1) - (a_2 - b_2) + (a_3 - b_3).$$

If the underlying group is taken to be $(F^*, \cdot)$, we say that we are applying the multiplicative version of the theorem. These theorems are used to guarantee the existence of a large subset $A' \subseteq A$ that satisfies $|P(A', \dots, A')| \leq c|A'|^{1+C'\epsilon}$ for any multivariate polynomial P with $c = c(\delta, \epsilon)$.

Our next ingredient is the Cauchy–Davenport inequality and a variant of it, which will be used to show the existence of a linear function on k variables whose image of A^k under the function equals the whole field, F.

Lemma 4. *For any nonempty finite subsets A, B of F,*

$$|A + B| \geq \min(|A| + |B| - 1, |F|).$$

Lemma 5. *Let A, B be nonempty finite subsets of F. Then there exists $\xi \in F^*$ such that*

$$|A + B\xi| \geq \min(\frac{1}{2}|A||B|, \frac{1}{10}|F|).$$

Next, we introduce some definitions before bringing our last ingredient. Assume that $1_F \in A$ and $|A \cdot A - A \cdot A| \leq K|A|$.

Definition 6. *Let A, $B \subseteq F$. If there exists $X \subseteq F$ such that $|X| \leq CK^C$ and $A \subseteq X + B$, then A is* essentially contained *in B, and we write $A \Subset B$.*

Definition 7. *If $x \in F$ satisfies $x \cdot A \Subset A - A$, then x is* good.

Lemma 8. *Let x, y be good elements. Then $x + y$, $x - y$, xy are all good.*

5.4 Proof of Theorem 1

In the following discussion, we say $x \lessapprox y$ if there exists a constant C such that $x \leq CK^C y$ and $x \approx y$ if both $x \lessapprox y$ and $y \lessapprox x$ hold. Define the dilate of a set $A \subseteq F$ by $x \in F$ as $xA \coloneqq \{xa : a \in A\}$.

Lemma 9. *Let A be a nonempty subset of F such that*

$$|A + A|,\ |A \cdot A| \leq K|A|.$$

Then there exists a subset A' of A with $|A'| \geq cK^{-C}|A|$ such that

$$|A' \cdot A' - A' \cdot A'| \leq CK^C|A'|.$$

Proof. From Lemma 3, there exists C, $D \subseteq A$ of size comparable to that of A such that for any $c, d \in C, D$, there exists more than $cK^{-C}|A|^5$ solutions of the equation $c - d = a_1 - a_2 + a_3 - a_4 + a_5 - a_6$ in A^6. This implies that any element x in $(C-D) \cdot A \cdot A \cdot A/A \cdot A$ has at least $c'K^{-C'}|A|^5$ solutions of the equations $x = b_1 - b_2 + b_3 - b_4 + b_5 - b_6$ with each b_i being an element of $A \cdot A \cdot A \cdot A/A \cdot A$. Plünnecke–Ruzsa type bound on iterated products in $(F^*, \cdot)$ gives the estimate $|A \cdot A \cdot A \cdot A/A \cdot A| \approx |A|$, and then we have

$$|(C-D) \cdot (A \cdot A \cdot A)/(A \cdot A)| \lessapprox |A|,$$

from double counting. As C and D are subsets of A whose size is comparable of that of A, we observe that size of CD is also comparable to that of A. This tells that C, D satisfies the assumptions of the multiplicative version of Lemma 2. Let the subsets found in this step be denoted as C' and D'. Then we obtain a size bound on $C'D' - C'D'$ by verifying an algebraic identity and applying combinatorial tools such as double counting and pigeonholing. Again, the Plünnecke–Ruzsa type bound and the injectivity of the map $x \mapsto ax$ in F gives us the existence of an element $\xi \in F^*$ such that $|C' \cap D'\xi| \approx |A|$. Taking $C' \cap D'\xi = A'$ proves the lemma.

□

By bootstrapping this bound on $|A' \cdot A' - A' \cdot A'|$, we bound the size of any polynomial expressions of A'. The following lemma is proved by considering the collection of monomials A^k and performing induction on the degree k. This induction relies heavily on the fact that good elements are closed under summation, subtraction, and multiplication.

Lemma 10. *Let A be a nonempty subset of F such that $1_F \in A$ and $|A \cdot A - A \cdot A| \leq K|A|$ holds for some K. Then for any multivariate polynomial P, there exists a constant C that satisfies*

$$|P(A, \ldots, A)| \leq CK^C|A|.$$

We now prove the existence of a polynomial such that the image of A^d under the map equals F. Let $F^* \coloneqq (F \backslash \{0\}, \cdot)$.

Lemma 11. *There exists a positive integer $k \sim 1/\delta$ and $\xi_1, \ldots, \xi_k \in F^*$ such that $F = A\xi_1 + \cdots A\xi_k$.*

Proof. We may assume that $|A| \geq 3$. Iteratively apply Lemma 5 until we find $\xi_1, \ldots, \xi_k \in F^*$ such that $|A\xi_1 + \cdots + A\xi_k| \geq |F|/10$. This is possible as $|A| \geq |F|^\delta$. Applying Lemma 4 $O(1)$ times finishes the proof. □

Lemma 12. *Let $B \subseteq F$ be a nonempty set and $\phi : B^k \to F$ be a linear surjection. Define $\tilde{B} \coloneqq B \cdot (B - B) + B \cdot (B - B)$. Then there is a linear surjection from $\tilde{B}^{k-1}$ to F.*

Proof. As $|F|$ is prime and $|B|^k$ isn't, this map ϕ is not an injection and there exist distinct elements $(b_1, \ldots, b_k)$, $(b'_1, \ldots, b'_k) \in B^k$ such that

$$(b_1 - b'_1)\xi_1 + \cdots + (b_k - b'_k)\xi_k = 0$$

holds. Without loss of generality, assume that $b_k \neq b'_k$. As $x \mapsto ax$ is a permutation on F for any $a \in F^*$, $F = B\xi_1(b_k - b'_k) + \cdots + B\xi_k(b_k - b'_k)$ holds. Substuting $(b_k - b'_k)\xi_k$ then gives

$$\begin{aligned} F &= B\xi_1(b_k + b_1 - b'_k - b_1) + \cdots + B\xi_{k-1}(b_k + b_{k-1} - b'_k - b_{k-1}) \\ &\subset \tilde{B}\xi_1 + \cdots \tilde{B}\xi_k. \end{aligned}$$

□

Now we prove the main theorem. Let k be a positive integer given from the application of Lemma 11 on the set A, and apply Lemma 12 k times. Then the image of the polynomial $P(A, \ldots, A)$, which is found as a linear function on some polynomial expression of A, equals the entire field F. This contradicts the conclusion of Lemma 10.

References

[BKT] Bourgain, J., Katz, N. and Tao, T., *A sum-product estimate in finite fields, and applications.* Geometric & Functional Analysis GAFA **14** (2004), no. 1 27-57;

[BGK] Bourgain, J. and Glibichuk, A. A. and Konyagin, S. V. *Estimates for the number of sums and products and for exponential sums in fields of prime order* Journal of the London Mathematical Society. Second Series **73** (2), 380-398;

[EM] Edgar, G. A. and Miller, C., *Borel subrings of the reals.* Proc. Amer. Math. Soc. **131** (2003) no. 4 1121–1129.

Jihyo Chae, Yonsei University
email: jihyochae@yonsei.ac.kr

6 Stabilizer rank and higher-order Fourier analysis

After Farrokh Labib

A summary written by Philippe van Dordrecht

Abstract

The paper proves a linear lower bound on the stabilizer rank of the magic state over finite fields of prime order. This result was known before only for qubits, in the characterstic $p = 2$. The proof uses methods from higher order Fourier analysis

In quantum computing, we are interested in quantum states. These states can be thought of as vectors in the space $\mathbb{C}^{p^n}$, where p is a prime. some of these states exhibit a polynomial structure. Roughly, stabilizer states are the states that exhibit a quadratic polynomial structure, and a state known as the "magic state" exhibits a cubic polynomial structure. The stabilizers can be efficiently simlated on a classical computer, whereas computations involving the magic state are widely believed to be intractable for classical computers. So with the introduction of this magic state, we can hopefully go beyond the power of classical computing, hence the name.

The article [Lab22] investigates something called the stabilizer rank of the magic state, which is defined as the number of stabilizer states you need in a linear combination to produce a magic state. The stabilizer rank of the magic state is proved to be at least linear in n. The best known upper bounds are of the form 2^{cn} where c is some constant with $0 < c < 1$, see for example [Qas21]. Mehraban and Tahmasbi have since proved a quadratic lower bound on the approximate stabilizer rank [MT24]. It is widely believed that the stabilizer rank of the magic state is superpolynomial.

Let p be a prime. Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. Let $P : \mathbb{F}_p^n \to \mathbb{T}$ be a function. We define the discrete derivative of a function P in direction h as

$$\Delta_h P(x) = P(x + h) - P(x).$$

The nonclassical polynomials of degree at most d are functions $P : \mathbb{F}_p^n \to \mathbb{T}$ that satisfy, for all $h_1, \ldots, h_{d+1}$

$$\Delta_{h_1} \ldots \Delta_{h_{d+1}} P(x) = 0.$$

Examples of these functions include the so-called classical polynomials. These classical polynomials are elements of $\mathbb{F}_p[x_1, \ldots, x_n]$ embedded in $\mathbb{T}$ by the function $\iota : x \mapsto |x|/p$ where $|\cdot|$ is the natural map from $\mathbb{F}_p$ to $\{0, \ldots, p-1\}$. The set of nonclassical polynomials is a strict superset of the set of classical polynomials. An explicit form of all nonclassical polynomials was derived by Tao and Ziegler [TZ12].

We will refer to nonclassical polynomials simply as polynomials from here on out, since we never require a polynomial to be classical.

We can think of the polynomials as vectors in a p^n dimensional vector space over $\mathbb{C}$, by considering the exponential map $e : \mathbb{T} \to \mathbb{C}$. This exponential map can be defined as

$$e(\alpha) = e^{2\pi i \alpha},$$

where the e on the right hand side is Euler's number. This allows us broaden our view to functions that take values not just on the unit circle, but on the entire complex plane. For functions $f : \mathbb{F}_p^n \to \mathbb{C}$, define the multiplicative discrete derivative in direction $h \in \mathbb{F}_p^n$ as

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

Then we see that

$$e(\Delta_h P) = \delta_h e(P).$$

Definition 1. *The function $f : \mathbb{F}_p \to \mathbb{C}$ is called a stabilizer if f is of the form*

$$f(x) = \sqrt{\frac{p^n}{|H|}} 1_H(x) e(Q(x))$$

where H is an affine subspace of $\mathbb{F}_p^n$, and Q is a polynomial of degree at most 2.

The terminology comes from quantum computing, where the stabilizer states play an important role in for example quantum error correction.

Another important state that we consider in quantum computing is the magic state. For the sake of brevity we will only define it for $p = 2$, but there are analogues for all primes p.

Definition 2. *The magic state* $\psi : \mathbb{F}_2^n \to \mathbb{T}$ *is defined as*

$$\psi(x) = e((|x_1| + \cdots + |x_n|)/8)$$

This is an example of an exponent of a polynomial of degree 3 that is not classical.

6.1 Notions of rank

Definition 3. *The d-rank of a function* $P : \mathbb{F}_p^n \to \mathbb{T}$*, denoted* $\mathrm{rank}_d(P)$*, is defined as the minimal number* r *such that there exists a function* $\Gamma : \mathbb{T}^r \to \mathbb{T}$ *and polynomials* $Q_1, \ldots, Q_r$ *of degree at most* $d-1$ *such that for all* $x \in \mathbb{F}_p^n$*,*

$$P(x) = \Gamma(Q_1(x), \ldots, Q_r(x)).$$

The rank of a polynomial of degree d will mean d-rank of this polynomial.

Definition 4. *The Fourier d-rank of a function* $f : \mathbb{F}_p^n \to \mathbb{T}$*, denoted* $\mathrm{frank}_d(f)$ *is the minimal number* r *such that there exist polynomials* $g_1, \ldots, g_r$ *of degree at most* $d-1$ *and complex numbers* $c_1, \ldots, c_r$ *such that*

$$e(f) = \sum_{i=1}^{r} c_i e(g_i)$$

The following proposition shows that if P has low d-rank, then it must correlate with some degree $d-1$ polynomial.

Proposition 5. *Let* $d \geq 2$ *and* $P : \mathbb{F}_p^n \to \mathbb{T}$ *a nonclassical polynomial with* $r := \mathrm{rank}_d(P)$*. Then there exists a polynomial* Q *of degree at most* $d-1$ *such that*

$$|\langle e(P), e(Q)\rangle| \geq p^{-(1+\lceil (d-1)/(p-1)\rceil)r}.$$

The following lemma is an immediate consequence of the definitions.

Lemma 6. *Let* $P \colon \mathbb{F}_p^n \to \mathbb{T}$ *be a polynomial and* $d \geq 2$*. Then* $\mathrm{frank}_d(e(P)) \geq \mathrm{rank}_d(P)$*.*

The stabilizer rank is very similar to this definition, only now we replace the nonclassical polynomials with stabilizers.

Definition 7. *The stabilizer rank of a function $\psi : \mathbb{F}_p^n \to \mathbb{C}$ is the minimal number r such that there exist stabilizers $\phi_1, \ldots, \phi_r$ and complex numbers $c_1, \ldots, c_r$ such that*

$$\psi = \sum_{i=1}^{r} c_i \phi_i$$

Theorem 8. *Let p be a prime and let $\psi : \mathbb{F}_p^n \to \mathbb{C}$ be the magic state over $\mathbb{F}_p$. The stabilizer rank of ψ is $\Omega(n)$.*

6.2 Sketch of the proof

Let $P : \mathbb{F}_P^n \to \mathbb{T}$ be defined as $P(x) = (|x_1| + \cdots + |x_n|)/8$. A relatively straightforward calculation, done in Proposition 4.2 in [Lab22], shows that there is a constant $c > 0$ such that for any nonclassical polynomial Q of degree at most 2,

$$|\langle e(P), e(Q) \rangle| \leq 2^{-cn}.$$

Since the magic state correlates so poorly with degree 2 polynomials, by the contrapositive of Proposition 5, it must have rank $\Omega(n)$.

Now we assume that the stabilizer rank of $e(P)$ is at most r. Then there exist for all $i \in [r]$ a complex number c_i, a quadratic polynomial Q_i and an affine subspace $H_i \subseteq \mathbb{F}_p^n$.

$$e(P) = \sum_{i=1}^{r} c_i e(Q_i) 1_{H_i}$$

Now Claim 5.1 in [Lab22] lets us find a subspace U such that there is a set $S \subset [r]$ such that for $i \in S$, U is a subset of H_i whereas for $i \notin S$, U is disjoint from H_i. Therefore, letting P' be the restriction from P to U, we will find that

$$e(P') = \sum_{i \in S} c_i e(Q_i'),$$

where for all i, Q_i' is equal to Q_i up to an affine linear transformation of the input. This implies that the Fourier 2-rank of P' is at most $|S|$, which is at most r.

By Lemma 3.11 from [Lab22], since P' is the restriction of a rank $\Omega(n)$ polynomial of degree 3 to an affine subspace that is not too small, P' is still cubic and still has rank $\Omega(n)$. Therefore, we have found that

$$r \geq \operatorname{frank}_2(P') \geq \operatorname{rank}_2(P') = \Omega(n),$$

which completes the proof.

References

[Qas21] Qassim, H., Pashayan, H. and Gosset, D., *Improved upper bounds on the stabilizer rank of magic states.* Quantum **5** (2021), 606.

[TZ12] Tao, T. and Ziegler, T., *The inverse conjecture for the Gowers norm over finite fields in low characteristic.* Ann. Comb. **16** (2012), no. 1, 121–188.

[MT24] Mehraban, S. and Tahmasbi, M., *Quadratic lower bounds on the approximate stabilizer rank: A probabilistic approach.* arXiv:2305.10277 (2024).

[Lab22] Labib, F., *Stabilizer rank and higher-order Fourier analysis.* Quantum **6** (2022), 645.

Philippe van Dordrecht, Leiden University
email: philippevdord@gmail.com

7 The inverse theorem for the U^3 Gowers uniformity norm on arbitrary finite abelian groups: Fourier-analytic approach

After A. Jamneshan and T. Tao [JT]

A summary written by Ferran Espuña

Abstract

We outline Jamneshan and Tao's Fourier-analytic proof of their quantitative U^3 inverse theorem for arbitrary finite abelian groups.

7.1 Introduction

Earlier talks linked Gowers norms to arithmetic progressions. In particular, Talk 10 bounded four-term progression counts by the U^3 norm [Go]. Essentially, if A is a subset of a finite abelian group G with density α and no nontrivial four-term arithmetic progressions, then its balanced indicator function $f = 1_A - \alpha$ has large U^3 norm, and the inverse theorem turns this into a structural description of f, which feeds back into a density increment argument in α. The theorem below is uniform in G. Earlier proofs covered odd-order groups [GT, Theorem 2.7] and $G = \mathbb{F}_2^n$ [S, Theorem 2.3], but relied on features unavailable in arbitrary torsion.

Write $\widehat{G} = \mathrm{Hom}(G, \mathbb{R}/\mathbb{Z})$, $\xi \cdot x = \xi(x)$, and $e(\theta) = e^{2\pi i\theta}$. For a finite $S \subset \widehat{G}$, put

$$\|x\|_S = \sup_{\xi \in S} \|\xi \cdot x\|_{\mathbb{R}/\mathbb{Z}}, \qquad \mathrm{Bohr}(S, \rho) = \{x \in G : \|x\|_S < \rho\}.$$

A Bohr set is *regular* if a small relative change in its radius produces only a comparably small relative change in its cardinality. For a map $\phi : U \to H$, where U lies in an additive group, define its additive derivative by $\partial_h \phi(x) = \phi(x) - \phi(x-h)$ whenever both terms are defined. The map is *locally linear* if every second derivative vanishes on its natural domain, and *locally quadratic* if every third derivative does.

Theorem 1 (Locally quadratic inverse theorem, Theorem 1.6 of [JT]). *Let G be a finite abelian group, let $0 < \eta \le 1/2$, and let $f : G \to \mathbb{C}$ satisfy $|f| \le 1$ and $\|f\|_{U^3(G)} \ge \eta$. Then there are a regular Bohr set $\mathrm{Bohr}(S, \rho)$ satisfying*

$$|S| \ll \eta^{-O(1)}, \qquad \exp\bigl(-\eta^{-O(1)}\bigr) \ll \rho \le \frac{1}{100},$$

a locally quadratic map $\phi : \mathrm{Bohr}(S, \rho) \to \mathbb{R}/\mathbb{Z}$, and a map $\xi : G \to \widehat{G}$ such that

$$\mathbb{E}_{x \in G} \bigl|\mathbb{E}_{h \in \mathrm{Bohr}(S,\rho)} f(x+h) e\bigl(-\phi(h) - \xi(x) \cdot h\bigr)\bigr| \gg \eta^{O(1)}.$$

Two algebraic black boxes enter near the end: globalisation of a locally bilinear form on a lifted group G_S, and integration of a symmetric global bilinear form to a quadratic map. They are stated when used. Suppressing averages and auxiliary 1-bounded factors, the obstruction evolves as

$$\begin{aligned} \Delta_h f(x) e(-\xi_h \cdot x) \quad &\rightsquigarrow \quad f(x+h) e(-M(h) \cdot x) \\ \rightsquigarrow \quad f(x+h+k) e(-B(h,k)) \quad &\rightsquigarrow \quad f(x+z) e(-\phi(z)) \\ &\rightsquigarrow \quad f(x+z) e(-\phi(z) - \xi(x) \cdot z). \end{aligned}$$

Sections 7.2–7.4 pass successively to M, to a symmetric local form B and its global lift, and to ϕ and the final frequency.

7.2 Extracting a locally linear frequency map

For the remainder, G, f, η are as in Theorem 1, and $b(\,\cdot\,)$ denotes a changing 1-bounded function. The identity

$$\|f\|_{U^3(G)}^8 = \mathbb{E}_{h \in G} \|\Delta_h f\|_{U^2(G)}^4, \qquad \Delta_h f(x) = f(x+h)\overline{f(x)},$$

combined with the U^2 inverse theorem and averaging shows that for many increments h there is a frequency $\xi_h \in \widehat{G}$ such that

$$|\mathbb{E}_{x \in G} \Delta_h f(x) e(-\xi_h \cdot x)| \gg \eta^{O(1)}.$$

We call such increments h *good*. At this stage ξ_h may depend arbitrarily on h; the next lemma organises the family into one locally linear map.

Lemma 2 (Locally linear frequency map, Lemma 2.5 of [JT]). *There are a regular Bohr set $\mathrm{Bohr}(S, \rho)$ with $|S| \ll \eta^{-O(1)}$ and $1 \ll \rho \le 1/8$, and a locally linear map $M : \mathrm{Bohr}(S, 2\rho) \to \widehat{G}$, such that*

$$\left|\mathbb{E}_{\substack{x \in G \\ h \in \mathrm{Bohr}(S,\rho)}} b(h) b(x) f(x+h) e\bigl(-M(h) \cdot x\bigr)\right| \gg \eta^{O(1)}.$$

For $\Gamma = \{(h, \xi_h)\} \subset G \times \widehat{G}$, additive quadruples encode the same relations among the chosen frequencies as among their good directions.

Cauchy–Schwarz supplies many such quadruples. The theorem of Balog, Szemerédi and Gowers isolates a large piece, while Plünnecke controls its iterated sums and differences. After a refinement, $4\Gamma'' - 4\Gamma''$ is still a graph [GT, Propositions 5.1, 5.4, and 9.1]. Here

$$4\Gamma'' - 4\Gamma'' = \{\gamma_1 + \cdots + \gamma_4 - \gamma_5 - \cdots - \gamma_8 : \gamma_i \in \Gamma''\}.$$

Saying that this set is a graph means that projection onto G is injective: if two such signed sums have the same first coordinate, then they also have the same frequency coordinate. Thus the resulting frequency depends only on the combined increment, not on how that increment is represented.

Bogolyubov places a bounded-rank Bohr set in the projection of $2\Gamma'' - 2\Gamma''$ [GT, Lemma 6.3]. Consequently, each h in this Bohr set has a unique frequency $M(h)$ with $(h, M(h)) \in 2\Gamma'' - 2\Gamma''$. If $h_1, h_2, h_1 + h_2$ all lie in the Bohr set, then

$$(h_1, M(h_1)) + (h_2, M(h_2)) \quad \text{and} \quad (h_1 + h_2, M(h_1 + h_2))$$

may both be viewed as elements of $4\Gamma'' - 4\Gamma''$ with the same first coordinate. The graph property therefore gives $M(h_1 + h_2) = M(h_1) + M(h_2)$, which is precisely local linearity. Regularisation and pigeonholing then recover the displayed correlation [GT, Proposition 9.3]. M replaces the unrelated ξ_h, while b absorbs the other factors: the phase is now linear in x and locally linear in h.

7.3 Symmetrising and globalising the bilinear obstruction

For U in an additive group and H another, a map $B : U \times U \to H$ is *locally bilinear* if it is additive in each variable whenever all points involved lie in U, and *symmetric* if $B(h, k) = B(k, h)$. Normalise $M(0) = 0$ by absorbing the constant frequency into $b(x)$. Then local linearity makes $M(h) \cdot k$ locally bilinear, and averaging over a second increment exposes this phase. The next lemma symmetrises it so that it can be a quadratic second derivative.

Lemma 3 (Symmetric locally bilinear form, Lemma 2.6 of [JT])**.** *There are a regular Bohr set* $\mathrm{Bohr}(S, \rho)$ *with* $|S| \ll \eta^{-O(1)}$ *and* $\eta^{O(1)} \ll \rho \leq 1/8$*, and a*

symmetric locally bilinear map

$$B : \mathrm{Bohr}(S, 2\rho) \times \mathrm{Bohr}(S, 2\rho) \longrightarrow \mathbb{R}/\mathbb{Z}$$

such that

$$\left| \mathbb{E}_{\substack{x \in G \\ h,k \in \mathrm{Bohr}(S,\rho)}} b(x,h)b(x,k)f(x+h+k)e\big(-B(h,k)\big) \right| \gg \eta^{O(1)}.$$

The symmetry argument of [GT, Lemma 9.4], with the odd-order factor 2 omitted, makes $D(y,z) = M(y)\cdot z - M(z)\cdot y$ uniformly small on a smaller Bohr set. Its small real lift $d(y,z)$ is locally bilinear, so the midpoint

$$B(y,z) = M(y)\cdot z - \tfrac{1}{2}d(y,z) = M(z)\cdot y + \tfrac{1}{2}d(y,z) \pmod 1$$

is symmetric and locally bilinear. Averaging Lemma 2 after small shifts in h and x, then expanding M, leaves the coupled phase $M(y)\cdot z$; one-variable terms enter b. Replacing it by $B(y,z)$ and pigeonholing gives the displayed correlation. The obstruction is now the fixed symmetric bilinear phase $e(-B(h,k))$, but it is defined only on a Bohr set. For $S \subset \widehat{G}$, introduce

$$G_S = \left\{(x,\theta) \in G \times \mathbb{R}^S : (\xi\cdot x)_{\xi\in S} = \theta \pmod{\mathbb{Z}^S}\right\}, \qquad \|\theta\|_{\mathbb{R}^S} = \sup_{\xi\in S}|\theta_\xi|.$$

Lemma 4 (Globalising a locally bilinear form, Lemma 2.2 of [JT])**.** *Let G be a finite abelian group, $S \subset \widehat{G}$, and $0 < \rho < 1/2$. If $B : \mathrm{Bohr}(S,\rho)^2 \to H$ is locally bilinear, then there is a globally bilinear map $\widetilde{B} : G_S^2 \to H$ such that $\widetilde{B}((x,\theta),(y,\sigma)) = B(x,y)$ whenever $\|\theta\|_{\mathbb{R}^S}, \|\sigma\|_{\mathbb{R}^S} \le \exp(-C|S|^C)\rho$ for some absolute $C > 0$. If B is symmetric, $\widetilde{B}$ may also be chosen symmetric.*

The lift is essential: even for a cyclic group, a locally bilinear form need not extend to G [JT, Example 2.3]. Apply Lemma 4 to the form above, with $H = \mathbb{R}/\mathbb{Z}$ and 2ρ as the input radius, and set $\rho_0 = \exp(-C|S|^C)\rho$. For $h \in \mathrm{Bohr}(S,\rho_0)$, let $\widetilde{h} = (h,\theta_h) \in G_S$ be its canonical small lift, so that $\|\theta_h\|_{\mathbb{R}^S} < \rho_0$. The conclusion is a symmetric globally bilinear map $\widetilde{B} : G_S^2 \to \mathbb{R}/\mathbb{Z}$ satisfying

$$\widetilde{B}(\widetilde{h},\widetilde{k}) = B(h,k) \qquad (h,k \in \mathrm{Bohr}(S,\rho_0)).$$

Thus, on the smaller Bohr set, the phase $e(-B(h,k))$ is literally the restriction of the global bilinear phase $e(-\widetilde{B})$ on G_S. Global bilinearity is what allows the next lemma to integrate the obstruction.

7.4 From bilinear structure to quadratic correlation

Lemma 5 (Integrating a symmetric bilinear form, Lemma 2.4 of [JT])**.** *Let G be a finitely generated abelian group.*

(i) If $B : G \times G \to \mathbb{R}/\mathbb{Z}$ is symmetric and globally bilinear, there is a globally quadratic map $\phi : G \to \mathbb{R}/\mathbb{Z}$ satisfying $\phi(x+y) = \phi(x) + \phi(y) + B(x,y)$ for every $x, y \in G$.

(ii) Suppose in addition that G is finite, $S \subset \widehat{G}$, and $\rho > 0$. If

$$B : \mathrm{Bohr}(S,\rho)^2 \longrightarrow \mathbb{R}/\mathbb{Z}$$

is symmetric and locally bilinear, then there is a radius ρ' satisfying $\exp(-|S|^{O(1)})\rho \ll \rho' \leq \rho$, and a locally quadratic map $\phi : \mathrm{Bohr}(S,\rho') \to \mathbb{R}/\mathbb{Z}$ for which $\phi(x+y) = \phi(x) + \phi(y) + B(x,y)$ whenever $x, y \in \mathrm{Bohr}(S,\rho'/2)$. Moreover, ϕ has a globally quadratic lift $\widetilde{\phi} : G_S \to \mathbb{R}/\mathbb{Z}$ such that $\widetilde{\phi}(x,\theta) = \phi(x)$ whenever $\|\theta\|_{\mathbb{R}^S} < \rho'$.

Part (ii) records exactly this lift–integrate–restrict argument. Applied to Lemma 3, it rewrites the obstruction as the second derivative of a locally quadratic map ϕ:

$$B(y,z) = \phi(y+z) - \phi(y) - \phi(z).$$

Average Lemma 3 after $(h,k) \mapsto (h+y+z, k-y)$; regularity controls the error. Bilinearity leaves as the only coupled y,z-phase $e(B(y,z)) = e(\phi(y+z) - \phi(y) - \phi(z))$. The first two factors enter b, leaving $e(-\phi(z))$. Setting $x' = x+h+k$ and pigeonholing h, k gives, on regular Bohr sets $\mathcal{B}_2 \subset \mathcal{B}_1$ chosen so that $|\mathcal{B}_1 + \mathcal{B}_2| \ll |\mathcal{B}_1|$,

$$\left| \mathbb{E}_{\substack{x' \in G \\ y \in \mathcal{B}_1,\ z \in \mathcal{B}_2}} b(x', y+z) b(x', y) f(x'+z) e(-\phi(z)) \right| \gg \eta^{O(1)}.$$

For many x', the corresponding (y,z)-average is large. The local Fourier lemma [GT, Lemma 4.4] converts this final two-point obstruction into a Fourier coefficient: it supplies a frequency $\xi(x')$ for which $f(x'+z)e(-\phi(z) - \xi(x')\cdot z)$ has large mean on $\mathcal{B}_2$. Averaging over x' gives Theorem 1.

References

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

[GT] B. Green and T. Tao, *An inverse theorem for the Gowers $U^3(G)$ norm*, Proc. Edinb. Math. Soc. (2) **51** (2008), no. 1, 73–153.

[JT] A. Jamneshan and T. Tao, *The inverse theorem for the U^3 Gowers uniformity norm on arbitrary finite abelian groups: Fourier-analytic and ergodic approaches*, Discrete Analysis 2023:11 (2023), 48 pp.

[S] A. Samorodnitsky, *Low-degree tests at large distances*, in *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, ACM, New York, 2007, pp. 506–515.

Ferran Espuña, UAM
email: ferranespuna@gmail.com

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

After W. T. Gowers [1]

A summary written by Benjamin Gillott and Fredy Yip

Abstract

We give a summary exposition of Gowers's improvement of the bound for Szemerédi's theorem for arithmetic progressions of length four.

8.1 Introduction

Szemerédi's theorem asserts that for any real number $\delta > 0$ and positive integer k, if a positive integer N is sufficiently large with respect to δ, then any $A \subseteq [N]$ of at least δN elements contains a non-trivial k-term arithmetic progression ("k-AP"). In other words, Szemerédi's theorem is the density version of Van der Waerden's celebrated theorem on monochromatic arithmetic progressions.

A quantitative question remains, namely how large does N need to be for a given choice of δ. Prior to the work of Gowers, the best known dependency of N on δ^{-1} is wowzer-type even in the special case where $k = 4$. On the other hand, the Fourier-analytic method of Roth for the first non-trivial case $k = 3$, which predates Szemerédi's work, yields a much sharper double exponential upper bound.

Theorem 1 (Roth's theorem [3])**.** *Every subset $A \subseteq [N]$ which avoids non-trivial 3-APs contains at most $O(N/\log\log N)$ elements.*

Roth's proof of Theorem 1 introduced the now classical framework of density increment. His argument leveraged a dichotomy of randomness and structure. For a "uniform" subset A of $[N]$ with δN elements, Roth showed that the number of 3-APs in A is, in fact, approximately $\delta^3 N^2$, as one would expect if elements of A were randomly chosen. On the other hand, when A is not "uniform", Roth showed that one may instead restrict to an arithmetic progression $P \subseteq [N]$ on which A is denser, and iterate the argument.

One major difficulty of adapting Roth's argument for 3-APs to the setting of avoiding non-trivial 4-APs was identifying the correct notion of "uniformity". Gowers' introduction of quadratic uniformity allows this scheme of density increment to be executed in the case of 4-APs, albeit requiring significantly more intricate arguments.

Theorem 2 (Theorem 20 of [1]). *Every subset $A \subseteq [N]$ which avoids non-trivial 4-APs contains at most $N/(\log\log N)^{\Omega(1)}$ elements.*

8.2 Quadratic uniformity and Gowers uniformity norms

We shall work over $\mathbb{Z}/N\mathbb{Z}$ instead of $[N]$, which offers a slightly more convenient setting for some of our tools. However, we must be careful to ensure that the arithmetic progressions we identify in $\mathbb{Z}/N\mathbb{Z}$ are genuine, that is, they are in fact arithmetic progressions in $[N]$. This applies to both the 4-APs we identify within A, and the eventual densification step. In either case, the requisite argument is standard and will not be the focus of this summary.

Without loss of generality, we assume that N is a sufficiently large prime number. We fix $\omega = \omega_N = \exp(2\pi i/N)$ to be a primitive Nth root of unity.

We identify a set A with its indicator function $g = 1_A$. The number of k-APs in A may then be expressed as

$$\sum_{a,d} g(a)g(a+d)\cdots g(a+(k-1)d). \tag{1}$$

We may model a random set of δN elements by the constant function δ, in which case the above count gives $\delta^k N^2$. This is the expected number of k-APs, should A be sufficiently "uniform". In order to study the deviation of 1_A from the constant function modelling a random set, we define the balanced function of A to be $f_A := 1_A - \delta$, where $\delta = |A|/N$ is the density of A. A notion of uniformity then corresponds to a sense in which f_A is small.

For a polynomial P, we call ω^P a polynomial phase of degree $\deg P$. An example of an obstruction to uniformity comes from the correlation with polynomial phases of degree at most $k-2$. More precisely, if f_A is sufficiently well correlated with a polynomial phase ω^P of degree at most $k-2$, then the k-AP count within A may deviate significantly from that of the random model.

In the case of Roth's theorem, where $k = 3$, this obstruction of correlations with linear phases translates to ℓ_∞ bounds on the Fourier transform

$$\widehat{f_A}(r) = \mathbb{E}_{x \sim \mathbb{Z}/N\mathbb{Z}} \left(\omega^{-rx} f_A(x) \right)$$

of f_A. The lack of correlation with linear phases is captured by Fourier uniformity under the norm $f_A \to \|\widehat{f_A}\|_\infty$. We call a set A linearly uniform if

$$\|\widehat{f_A}\|_\infty = o(1).$$

In other words, $\|\widehat{f_A}\|_\infty$ is small relative to the trivial upper bound of 1. In general, uniformity is defined with such qualitative conditions. This ensures that the failure of uniformity produces a significant deviation from the random baseline which may then be used to produce a density increment.

It turns out that this notion of uniformity suffices to ensure that the 3-AP count given by (1) lies within $o(N^2)$ of the random $\delta^3 N^2$.

For longer arithmetic progressions, one may consider generalisations of this argument by studying the correlation of f_A with higher-degree phases. However, these methods only produce bounds of the form $\delta^k N^2 + o(N^{k-1})$, which are wholly insufficient when $k \geq 4$.

In other words, for $k \geq 4$, controlling the correlation of f_A with phases of degree at most $k-2$, whilst necessary, is not sufficient.

Gowers's introduction of uniformity norms in [1] (and later in full generality in [2]) addresses this problem.

For a function $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, and an element $y \in \mathbb{Z}/N\mathbb{Z}$, let $\Delta_y f$ denote the function given by

$$\Delta_y f(x) = f(x)\overline{f(x-y)}.$$

Note that Δ_y for $y \in \mathbb{Z}/N\mathbb{Z}$ defines a family of commuting operators. A key property of these operators is that for any $y \neq 0$, Δ_y reduces the degree of a polynomial phase by one. Therefore, for any choice of $k-1$ non-zero elements $y_1, \ldots, y_{k-1}$ in $\mathbb{Z}/N\mathbb{Z}$, the function $\Delta_{y_1} \cdots \Delta_{y_{k-1}} f$ is the constant function 1 if and only if f is a phase of degree at most $k-2$. To capture this behaviour, Gowers [2] defined the $(k-1)$th uniformity norm U_{k-1} of a function $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ as

$$\|f\|_{U_{k-1}} = \left(\mathbb{E}_{x, y_1, \ldots, y_{k-1}} \Delta_{y_1} \cdots \Delta_{y_{k-1}} f(x) \right)^{2^{1-k}}.$$

Relative to an ℓ_∞ uniform control on the correlation of f_A with polynomial phases of degree at most $k-2$, the uniformity norm U_{k-1} captures the synergy between numerous individually small correlations with these polynomial phases.

We call a set A U_{k-1}-uniform if

$$\|f_A\|_{U_{k-1}} = o(1).$$

Here we again demand that $\|f_A\|_{U_{k-1}}$ is small with respect to the trivial upper bound 1.

To see how the Gowers uniformity norms generalise Fourier uniformity, we note that when $k = 3$, the relevant uniformity norm U_2 is given by

$$\|f\|_{U_2} = (\mathbb{E}_{x,y_1,y_2} \Delta_{y_1} \Delta_{y_2} f(x))^{1/4} = \|\hat{f}\|_4.$$

Therefore, by the monotonicity and log-convexity of ℓ_p norms, U_2-uniformity is equivalent to Fourier uniformity.

The introduction of U_{k-1}-uniformity allows Roth's density increment scheme for $k = 3$ to be extended to the case where $k \geq 4$. In particular, the k-AP count in a U_{k-1}-uniform set A of δN elements may be shown to lie within $o(N^2)$ of $\delta^k N^2$. The difficulty in completing this argument lies in extracting structural information when A fails to be U_{k-1}-uniform. Results establishing such structure are now known as inverse theorems for Gowers uniformity norms. Following Gowers [1], we shall focus on the first non-trivial case $k = 4$.

For 4-APs, the relevant uniformity norm U_3 is given by

$$\|f\|_{U_3} = (\mathbb{E}_{x,y_1,y_2,y_3} \Delta_{y_1} \Delta_{y_2} \Delta_{y_3} f(x))^{1/8}.$$

We call a set $A \subseteq [N]$ quadratically uniform if it is uniform with respect to the U_3 norm. Since

$$\|f\|_{U_3} = \left(\mathbb{E}_y \|\Delta_y f\|_{U_2}^4\right)^{1/8},$$

a set A is quadratically uniform if and only if $\Delta_y f_A$ is Fourier uniform for $N - o(N)$ choices of $y \in \mathbb{Z}/N\mathbb{Z}$.

8.3 4-AP count

We need to show that a quadratically uniform set contains the expected number of 4-APs. To do this, we prove the following lemma. Since N is a sufficiently large prime, all linear maps we consider from $(\mathbb{Z}/N\mathbb{Z})^k$ to $(\mathbb{Z}/N\mathbb{Z})^k$ will be invertible.

Lemma 3. *For any four functions, f_1, f_2, f_3, f_4, from $\mathbb{Z}/N\mathbb{Z}$ to $\mathbb{C}$ we have the following:*

$$|\mathbb{E}_{a,d} f_1(a) f_2(a+d) f_3(a+2d) f_4(a+3d)| \leq ||f_1||_2 ||f_2||_2 ||f_3||_{U_3} ||f_4||_{U_3}$$

We will prove an analogous statement for 3-APs instead since it demonstrates the way one uses Cauchy-Schwarz and is shorter.

Proof of the 3-AP version. First we apply Cauchy-Schwarz to the expectation over a in

$$|\mathbb{E}_a f_1(a) \mathbb{E}_d f_2(a+d) f_3(a+2d)|$$

to get an upper bound of

$$(\mathbb{E}_a |f_1(a)|^2)^{1/2} (\mathbb{E}_a |\mathbb{E}_d f_2(a+d) f_3(a+2d)|^2)^{1/2}.$$

The first term is $||f_1||_2$. We expand out the square in the second term to get

$$(\mathbb{E}_{d,e} \mathbb{E}_a f_2(a+d) f_3(a+2d) \overline{f_2(a+e) f_3(a+2e)})^{1/2}.$$

We change the coordinates to $(x, y, z) = (a+d, a+2d, d-e)$.

$$(\mathbb{E}_z \mathbb{E}_x f_2(x) \overline{f_2(x-z)} \mathbb{E}_y \overline{f_3(y)} f_3(y-2z))^{1/2}$$

Now we apply Cauchy-Schwarz to the expectation over z to obtain

$$(\mathbb{E}_z |\mathbb{E}_x f_2(x) \overline{f_2(x-z)}|^2)^{1/4}$$

And the same term for f_3 but with $2z$. This is the U_2 norm. So putting it all together we get that

$$|\mathbb{E}_{a,d} f_1(a) f_2(a+d) f_3(a+2d)| \leq ||f_1||_2 ||f_2||_{U_2} ||f_3||_{U_2}$$

□

One can prove the 4-AP version with a similar argument. The first step is the same. After expanding, we control the more complicated term using the 3-AP version that we have just proved.

Let A be quadratically uniform with density δ, and let f_A be its balanced function. The proportion of 4-APs in A is

$$\mathbb{E}_{a,d} (\delta + f_A(a))(\delta + f_A(a+d))(\delta + f_A(a+2d))(\delta + f_A(a+3d)).$$

This expands into 16 terms. The main term is δ^4. The five error terms containing at least three copies of f_A can be bounded above by $||f_A||_{U_3}$, since $||f_A||_2 \leq 1$. The other error terms can be seen to be zero using the fact that $\mathbb{E}_x f_A(x) = 0$. Thus, the proportion of 4-APs in A is between $\delta^4 - 5||f_A||_{U_3}$ and $\delta^4 + 5||f_A||_{U_3}$.

8.4 Density increment

This part of the proof is the most involved and contains most of the technical details. We obtain a density increment by proving an inverse theorem. We take a bounded function f with $||f||_{U_3} \geq \alpha$ and obtain some structural information about f. For the purposes of a density increment, we will need the structural information that we obtain to be applied to f_A to reveal an arithmetic progression P of length at least N^γ such that $|A \cap P|/|P| \geq \delta + \epsilon$, where ϵ and γ depend only on α.

What we show is that there is a uniform multicover of $\mathbb{Z}/N\mathbb{Z}$ by arithmetic progressions Q_i of length at least N^γ and a choice of quadratic polynomials p_i such that f correlates with $\omega^{p_i(n)}$ on Q_i, on average. Precisely, this means that

$$\mathbb{E}_i |\mathbb{E}_{x \in Q_i} f(x) \omega^{p_i(x)}| \geq \epsilon$$

for some ϵ depending only on δ. This is not an ideal inverse theorem, since it does not "go both ways", by which we mean that a function satisfying the conclusions of the inverse theorem must have a large U_3 norm. However, it is enough to deduce a density increment in a way similar to the proof of Roth's theorem.

In this overview, we suppress the dependencies of the constants in order to focus on the main ideas. Thus, we use c and C to denote constants that depend only on α.

- The first step is to use the fact that the U_2 norm of $\Delta_a(f)$ must be large somewhat often. This is equivalent to $\Delta_a(f)$ having a large Fourier coefficient somewhat often. Thus, we can define a function ϕ from $\mathbb{Z}/n\mathbb{Z}$ to $\widehat{\mathbb{Z}/n\mathbb{Z}}$ such that

  $$\mathbb{E}_a |\widehat{\Delta_a f}(\phi(a))| \geq c.$$

  This implies that ϕ must have a lot of structure. By applying Cauchy–Schwarz many times along with orthogonality of characters one can deduce that the graph $(x, \phi(x))$ has many additive quadruples.

- The second step is to use Balog–Szemerédi–Gowers [1] to show that there is a subset B of $[N]$ such that the graph of ϕ, $\Gamma = (x, \phi(x))$, restricted to B has a doubling constant of at most C and $|B| \geq cN$. We let Γ_B denote the restriction of the graph to B.

- One now uses Freiman's theorem, with Ruzsa's proof [4], on sets with small doubling in $\mathbb{Z}^d$ to obtain an arithmetic progression $\tilde{P}$ in $\mathbb{Z}^2$ of length at least N^c such that $|\Gamma_B \cap \tilde{P}| \geq c|\tilde{P}|$. This means that there is a linear function λ and an arithmetic progression in $\mathbb{Z}$, P of length at least N^c such that
$$\mathbb{E}_{x \in P} |\widehat{\Delta_x f}(\lambda(x))|^2 \geq c.$$
- The last step is to expand this expression and perform manipulations to obtain the conclusion. The arithmetic progressions in the covering are the translates of P.

Many important details have been omitted. In particular, obtaining a density increment from this weak inverse theorem requires further ideas and the manipulations in the last step deserve more attention. We hope this overview of the density increment step is enough to give one an idea of the combinatorial ideas which drive the inverse theorem.

References

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

[2] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geometric & Functional Analysis GAFA, **11** (2001), no. 3, 465–588.

[3] K. F. Roth, *On certain sets of integers*, Journal of the London Mathematical Society, **1** (1953), no. 1, 104–109.

[4] I. Z. Ruzsa, *Generalized arithmetical progressions and sumsets*, Acta Mathematica Hungarica, **65** (1994), no. 4, 379–388.

Benjamin Gillott, Fredy Yip, University of Cambridge
email: bg449@cam.ac.uk, fy276@cam.ac.uk

9 Poorly-distributed polynomials over finite fields are low-rank

After B. Green and T. Tao [GT09]

A summary written by V. Gladkova

Abstract

Higher-order Fourier analysis over a finite field $\mathbb{F}$ concerns itself with measuring bias of functions with respect to polynomial phases, typically over partitions of $\mathbb{F}^n$ defined by polynomials of bounded degree. In order to leverage information about bias for the purposes of counting solutions to linear equations, it is important to understand the structure of the underlying partition. In this talk, we will see that, for fields of sufficiently high characteristic, a collection of polynomials defines approximately an equipartition of $\mathbb{F}^n$, unless the polynomials have 'low rank', meaning that the collection can be expressed in terms of a bounded number of lower-degree polynomials.

9.1 Introduction

Given a function $P : \mathbb{F}^n \to \mathbb{F}$ and an $h \in \mathbb{F}^n$, define the *additive derivative* D_h *of* P as $D_h(P)(x) = P(x+h) - P(x)$. By analogy with real analysis, polynomials over $\mathbb{F}$ in n variables can be defined as precisely those functions P that satisfy

$$D_{h_1} \dots D_{h_{d+1}} P \equiv 0 \tag{1}$$

for all choices of $h_1, \dots, h_{d+1} \in \mathbb{F}^n$. The smallest such d, if it exists, is the *degree* of P, and we will write $\mathcal{P}_d(\mathbb{F}^n)$ for the set of all polynomials of degree at most d.

One can check that when $\mathrm{char}(\mathbb{F}) > d$, $\mathcal{P}_d(\mathbb{F}^n)$ looks precisely as one would expect, meaning that all such polynomials are linear combinations of monomials $x_1^{i_1} \dots x_n^{i_d}$ with $i_1, \dots, i_n \in \{0, \dots, \mathbb{F} - 1\}$ and $\sum i_j \leq d$.

Fun fact: These are so-called *classical* polynomials. For lower characteristics, the class of functions defined by (1) is somewhat broader and includes *non-classical* polynomials. You can find out more about these in [TZ12], but [GT09] only deals with classical polynomials as they assume $\mathrm{char}(\mathbb{F}) > d$ throughout.

Definition 1. *For $j \in \{0, \ldots, p-1\}$, write $e_{\mathbb{F}}(j) = e^{2\pi ij/|\mathbb{F}|}$. A function $f : \mathbb{F}^n \to \mathbb{C}$ is called a* polynomial phase of degree d *if there is a polynomial $P : \mathbb{F}^n \to \mathbb{F}$ of degree d such that $f(x) = e_{\mathbb{F}}(P(x))$.*

Recall that in Fourier analysis, the Fourier transform is precisely the bias of a function with respect to a polynomial phase of degree 1, i.e. a linear phase:

$$\hat{f}(r) = \mathbb{E}_{x \in \mathbb{F}^n} f(x) e_{\mathbb{F}}(r^T x).$$

A function is considered Fourier-uniform if it has low bias with respect to all non-zero linear phases. It would be natural to define higher-order uniformity as a lack of bias with respect to all polynomial phases of some degree d. Indeed, there is a norm for measuring this quantity.

Definition 2 (Weak Gowers norm)**.** *Given a function $f : \mathbb{F}^n \to \mathbb{C}$ and an integer $d \geq 0$, the* weak Gowers norm $\|f\|_{u^{d+1}}$ of f *is defined as*

$$\|f\|_{u^{d+1}} = \max_{P \in \mathcal{P}_d(\mathbb{F}^n)} \mathbb{E}_{x \in \mathbb{F}^n} f(x) e_{\mathbb{F}}(P(x)).$$

In other words, it is the maximum correlation of f with a polynomial phase of degree at most d.

However, for technical reasons, this norm is rarely used directly in applications. Instead, we typically use a proxy quantity, known as the Gowers uniformity norm – this is closely related to the weak Gowers norm via the Inverse Theorem for Gowers Uniformity Norms and will be (or will have been) expanded upon in other talks.

Another object that is central to higher-order Fourier analysis is a type of partition called *a polynomial factor*.

Definition 3 (Polynomial factor)**.** *Given non-negative integers $M_1, \ldots, M_d$ and a collection of polynomials $(P_{i,j})_{1 \leq i \leq d, 1 \leq j \leq M_i}$ where $P_{i,j} \in \mathcal{P}_i(\mathbb{F}^n)$ for every (i, j), define the evaluation map $\Phi : \mathbb{F}^n \to \mathbb{F}_1^M \times \ldots \times \mathbb{F}_d^M$ by setting $\Phi(x) = (P_{i,j}(x))_{i,j}$. A* polynomial factor $\mathcal{F}$ of degree d and dimension $\dim(\mathcal{F}) = \sum M_i$ *is a partition of $\mathbb{F}^n$ into the level sets of Φ. These level sets are referred to as* atoms *and have the form $\{x \in \mathbb{F}^n : \Phi(x) = \underline{c}\}$, for each $\underline{c} \in \mathbb{F}_1^M \times \ldots \times \mathbb{F}_d^M$.*

Given two factors $\mathcal{F}_1$ and $\mathcal{F}_2$, we will write $\mathcal{F}_1 \preceq \mathcal{F}_2$ to mean that $\mathcal{F}_1$ refines $\mathcal{F}_2$.

Note that a polynomial factor of degree 1, or linear factor, is simply a partition of $\mathbb{F}^n$ into cosets of some subspace, which is precisely the type of partition given by the (linear) arithmetic regularity lemma. In particular, a linear factor is an equipartition, as all cosets have exactly the same size.

However, in general, the atoms of a polynomial factor may vary in size considerably, including being empty, which is not ideal for counting purposes. To understand what collections of polynomials lead to more or less equally-sized atoms, we consider the notion of *rank*.

Definition 4 (Rank of a polynomial)**.** *Fix $d \geq 0$, and let $P : \mathbb{F}^n \to \mathbb{F}$ be a function. The* degree-d rank *of P, denoted by* $\mathrm{rank}_d(P)$*, is the least integer $R \geq 0$ (if it exists) for which there are polynomials $Q_1, \ldots, Q_R \in \mathcal{P}_d(\mathbb{F}^n)$ and a function $F : \mathbb{F}^R \to \mathbb{F}$ satisfying*

$$P(x) = F(Q_1(x), \ldots, Q_R(x))$$

for all $x \in \mathbb{F}^n$. If such an integer R does not exist, $\mathrm{rank}_d(P) = \infty$.

Example 5. *The degree-0 rank of P is 1 if P is constant and ∞ otherwise.*

Example 6. *When P is a quadratic polynomial, it can always be written in the form $P(x) = x^T Mx + L(x)$ where M is an $n \times n$ matrix and L is a linear function. Then the degree-1 rank of P is at most the matrix rank of M plus 1 (to account for L).*

Remark: If P is a polynomial of degree $d+1$ that has low degree-d rank, it means that P can be expressed in terms of a small number of lower-degree polynomials. In some way, such a polynomial is not 'truly' degree $d+1$.

What makes high rank a desirable property is that it ensures that polynomial 'Gauss sums', expressions of the form $|\mathbb{E}_{x \in \mathbb{F}^n} e_{\mathbb{F}}(P(x))|$, are small. Recall that when P is a linear polynomial, this quantity is 1 if an only if P is a constant, i.e. $\mathrm{rank}_0(P) = 1$, and is zero otherwise (when $\mathrm{rank}_0(P) = \infty$). It is also not hard to check that when $P(x) = x^t Mx + L(x)$ is a quadratic polynomial, $|\mathbb{E}_{x \in \mathbb{F}^n} e_{\mathbb{F}}(P(x))| \leq |\mathbb{F}|^{-\mathrm{rank}(M)/2}$, which approaches zero as the degree-1 rank of the polynomial grows (see Example 6).

In [GT09], Green and Tao show that for finite fields of high enough characteristics, this pattern continues holding for polynomials of higher degree. This is the main theorem of the paper and the talk.

Theorem 7 (Lack of equidistribution implies bounded rank [GT09])**.** *Let $0 \leq d < |\mathbb{F}|$ and $\delta \in (0, 1]$. Suppose there is a polynomial $P \in \mathcal{P}_d(\mathbb{F}^n)$ such that $|\mathbb{E}_{x \in \mathbb{F}^n} e_{\mathbb{F}}(P(x))| \geq \delta$. Then* $\mathrm{rank}_{d-1}(P) \leq O_{\mathbb{F},\delta,d}(1)$.

Before saying anything about the proof, it is worth noting how Theorem 7 relates to the earlier assertion that atoms of a high-rank polynomial factor are all approximately the same size. Indeed, first, one has to define what it means for a polynomial factor to have high rank.

Definition 8 (Rank of a polynomial factor)**.** *Let $d \geq 0$ be an integer and let $F : \mathbb{Z}_{\geq 0} \to \mathbb{R}_{\geq 0}$ be a non-decreasing function. Suppose $\mathcal{F}$ is a polynomial factor of degree d defined by polynomials $(P_{i,j})_{1 \leq i \leq d, 1 \leq j \leq M_i}$ with $P_{i,j} \in \mathcal{P}_i(\mathbb{F}^n)$. We will say that $\mathcal{F}$* is F-regular *if*

$$\operatorname{rank}_{i-1}(\sum_{j=1}^{M_i} \lambda_{i,j} P_{i,j}) \geq F(\dim(\mathcal{F}))$$

for all $1 \leq i \leq d$ and all $\lambda_{i,1}, \ldots, \lambda_{i,M_i} \in \mathbb{F}$ not simultaneously zero.

It is not hard to show that for a function F that grows sufficiently fast, F-regular polynomial factors are approximately equipartitions, provided Theorem 7 holds.

Corollary 9 (Lemma 4.1 of [GT09])**.** *Let $d \geq 1$ and $\epsilon \in (0, 1]$. Suppose Theorem 7 holds for all degrees up to d. Then there exists a non-decreasing function $F : \mathbb{Z}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that if $\mathcal{F}$ is an F-regular polynomial factor of degree d, then every atom A of $\mathcal{F}$ satisfies*

$$\left| \mathbb{E}_{x \in \mathbb{F}^n} \mathbb{1}_A(x) - \frac{1}{|\mathbb{F}|^{dim(\mathcal{F})}} \right| \leq \epsilon.$$

In other words, all atoms of $\mathcal{F}$ are approximately the same size

Proof sketch. Write $D = \dim(\mathcal{F})$. Recall that $A = \{x \in \mathbb{F}^n \colon \Phi(x) = \underline{c}\}$ for some $\underline{c} \in \mathbb{F}^D$, and observe that $\mathbb{1}_A(x) = \mathbb{E}_{\underline{\lambda} \in \mathbb{F}^D} e_{\mathbb{F}}((\Phi(x) - \underline{c})^T \underline{\lambda})$ since $L(\underline{\lambda}) = (\Phi(x) - \underline{c})^T \underline{\lambda}$ is a linear function for a fixed x. Then

$$\begin{aligned} \mathbb{E}_{x \in \mathbb{F}^n} \mathbb{1}_A(x) &= \mathbb{E}_{\underline{\lambda} \in \mathbb{F}^D} \mathbb{E}_{x \in \mathbb{F}^n} e_{\mathbb{F}}(\Phi(x)^T \underline{\lambda}) \\ &= \frac{1}{|\mathbb{F}|^D} + \frac{1}{|\mathbb{F}|^D} \sum_{\underline{\lambda} \neq 0} \mathbb{E}_{x \in \mathbb{F}^n} e_{\mathbb{F}}(\Phi(x)^T \underline{\lambda}). \end{aligned}$$

The first term comes from $\underline{\lambda} = 0$ and the second term can be made as small as necessary by choosing F to grow sufficiently fast, by Theorem 7. $\square$

The proof of Theorem 7 itself is by induction on the degree, and will utilise the following lemma due to Bogdanov and Viola [BV07].

Lemma 10 (Bogdanov-Viola, Lemma 25 of [BV07])**.** *Let $d \geq 0$ be an integer and $\delta, \sigma \in (0, 1]$. Given a polynomial $P \in \mathcal{P}_d(\mathbb{F}^n)$, if $|\mathbb{E}_{x \in \mathbb{F}^n} e_{\mathbb{F}}(P(x))| \geq \delta$, then there is a function $\tilde{P} : \mathbb{F}^n \to \mathbb{F}$ with $\mathrm{rank}_{d-1}(\tilde{P}) \leq |\mathbb{F}|/\delta^2\sigma$ such that P and $\tilde{P}$ disagree on at most a σ-proportion of $\mathbb{F}^n$.*

Note that $\mathrm{rank}_{d-1}(\tilde{P}) \leq |\mathbb{F}|/\delta^2\sigma$ implies that $\tilde{P}$ is constant on the atoms of some polynomial factor $\mathcal{F}$ of degree $d-1$ and dimension at most $|\mathbb{F}|/\delta^2\sigma$, i.e. $\tilde{P}$ is $\mathcal{F}$-measurable. The missing step is a lemma that shows that if a polynomial P agrees with such a $\tilde{P}$ almost everywhere, then P itself must be low-rank.

Lemma 11 (Proposition 5.1 of [GT09])**.** *Let $d \geq 1$ be an integer and $\sigma_d > 0$ a small quantity to be specified later. Suppose that $\mathcal{F}$ is an F-regular polynomial factor of degree $d-1$ for some non-decreasing F that grows sufficiently fast in terms of d. Suppose further that $P \in \mathcal{P}_d(\mathbb{F}^n)$ and there exists an $\mathcal{F}$-measurable $\tilde{P} : \mathbb{F}^n \to \mathbb{F}$ such that P and $\tilde{P}$ disagree on at most a σ_d-proportion of $\mathbb{F}^n$. If Theorem 7 holds for all degrees up to $d-1$, then P itself is $\mathcal{F}$-measurable.*

References

[BV07] Bogdanov, A. and Viola, E., *Pseudorandom bits for polynomials*, Proc. of FOCS (2007), pp. 41–51;

[GT09] Green, B. J. and Tao, T. C., *The distribution of polynomials over finite fields, with applications to the Gowers norms*, Contrib. Discrete Math. **4** (2009), no. 2, pp. 1–36;

[TZ12] Tao, T. C. and Ziegler, T. D., *The inverse conjecture for the Gowers norm over finite fields in low characteristic*, Ann. Comb. **16** (2012), no. 1, pp. 121–188.

V. Gladkova, University of Manchester
email: valeriia.gladkova@manchester.ac.uk

10 Quadratic Fourier analysis and stabilizer testing

After S. Arunachalam and A. Dutt [AD25], Z. Bao and P. van Dordrecht and J. Helsen [BvDH25]

A summary written by Isaac Holt

Abstract

We give an outline of the proof of the U^3 inverse theorem for quantum states based on the work of [BvDH25] and [AD25], and explore an application of this result to tolerant testing of stabiliser states.

10.1 Introduction

Stabiliser states are important class of quantum states, as the outputs of Clifford circuits. A wealth of literature exists around *property testing* of stabiliser states. Property testing is central framework in theoretical computer science, in which we are given some form of access to an object X belonging to a space $\mathcal{X}$, and we want to determine whether $X \in \mathcal{P}$ or X is ϵ-far from $\mathcal{P}$ according to some distance metric, where $\mathcal{P}$ is a given subset of $\mathcal{X}$. Given that processes on real-world quantum computers are subject to noise, a more physically realistic task is *tolerant* stabiliser property testing, where we are promised that either a quantum state $|\psi\rangle$ is ϵ_1-close to a stabiliser state or is ϵ_2-far from all stabiliser states. Previous tolerant stabiliser testing algorithms required ϵ_1 to be at least some absolute threshold; however, the recent breakthrough of the polynomial Freiman-Ruzsa theorem means tolerant stabiliser testing can be performed efficiently for any $\epsilon_2 \leq \text{poly}(\epsilon_1)$.

10.2 Preliminaries

An n-qubit *Clifford circuit/unitary* is a quantum circuit can be written as a combination of Hadamard, CNOT and S gates. An n-qubit *stabiliser state* $|\mathcal{S}\rangle$ is any quantum state which is the output of an n-qubit Clifford unitary applied to the all-zeros state $|0^n\rangle$.

Stabiliser states can be characterised as non-classical quadratic phase functions supported on an affine subspace of $\mathbb{F}_2^n$:

$$|\mathcal{S}\rangle = \frac{1}{\sqrt{|V|}} \sum_{x \in a+V} (-1)^{q(x)} i^{\ell(x)} |x\rangle \qquad \text{for every stabiliser state } |\mathcal{S}\rangle,$$

where $a \in \mathbb{F}_2^n$, V is a subspace of $\mathbb{F}_2^n$, $q : \mathbb{F}_2^n \to \mathbb{F}_2$ is quadratic and $\ell : \mathbb{F}_2^n \to \mathbb{F}_2$ is linear. This suggests that a quantum notion of Gowers U^3 norm (which measures how close a function is to being quadratic) may be useful in the context of stabiliser testing.

Indeed, the definition of the *Gowers U^3 norm* of a quantum state $|\psi\rangle = \sum_{x \in \mathbb{F}_2^n} f(x) |x\rangle$, introduced in [AD25], is defined by

$$\|\psi\|_{U^3}^8 \coloneqq 2^{4n} \cdot \mathbb{E}_{x,a,b,c \in \mathbb{F}_2^n} f(x) \overline{f(x+a) f(x+b)} f(x+c)$$

i.e. $\|\psi\|_{U^3} = 2^{n/2} \|f\|_{U^3}$. It is fairly straightforward to show that $\|\psi\|_{U^3} = 1$ if and only if $|\psi\rangle$ is a stabiliser state. Thus, if we can estimate the U^3 norm of $|\psi\rangle$, then we can test (non-tolerantly) whether $|\psi\rangle$ is a stabiliser state. The U^3 *inverse theorem* is essentially a robust version of this result: if $\|\psi\|_{U^3}$ is large, then $|\psi\rangle$ is close to a stabiliser state. The notion of "closeness" is captured by the *stabiliser fidelity* $\mathcal{F}_S(|\psi\rangle)$ of $|\psi\rangle$, which is the maximum fidelity of $|\psi\rangle$ with a stabiliser state:

$$\mathcal{F}_S(|\psi\rangle) = \max_{|\mathcal{S}\rangle \text{ stabiliser state}} |\langle \psi | \mathcal{S} \rangle|^2.$$

For $x \in \mathbb{F}_2^n$, we write $p_\psi(x) = |\langle \psi | W_x | \psi \rangle|^2$, where $\{W_x : x \in \mathbb{F}_2^{2n}\}$ label the 2^{2n} phaseless n-qubit Pauli operators. A subspace $V \leq \mathbb{F}_2^{2n}$ is *isotropic* if $[x, y] = 0$ for all $x, y \in V$, where $[x, y]$ is the symplectic inner product. V is *Lagrangian* if it is isotropic and $\dim(V) = n$.

10.3 The U^3 inverse theorem for quantum states

In this section, we prove the following inverse theorem from [BvDH25]:

Theorem 1. *If $\|\psi\|_{U^3} \geq \gamma$, then $\mathcal{F}_S(|\psi\rangle) \geq \mathrm{poly}(\gamma)$.*

To prove the theorem, we consider a quantity $G(|\psi\rangle)$ which can be shown to satisfy $G(|\psi\rangle) \geq \|\psi\|_{U^3}^{16}$; hence, it suffices to prove a lower bound of the

form $\mathcal{F}_S(|\psi\rangle) \geq \text{poly}(G(|\psi\rangle))$. So for the rest of the proof, we assume instead that $G(|\psi\rangle) \geq \gamma$.

A folklore result tells us that for any Lagrangian $L \leq \mathbb{F}_2^{2n}$, $\mathcal{F}_S(|\psi\rangle) \geq \sum_{x\in L} p_\psi(x)$. Since p_ψ is non-negative and any isotropic subspace $I \leq \mathbb{F}_2^{2n}$ can be extended to a Lagrangian subspace, this also means that $\mathcal{F}_S(|\psi\rangle) \geq \sum_{x\in I} p_\psi(x)$ for any isotropic I. So our goal is to find an isotropic $V' \leq \mathbb{F}_2^{2n}$ such that

$$\sum_{x\in V'} p_\psi(x) \geq \text{poly}(\gamma).$$

The proof of the existence of V' begins with showing the existence of a "probabilistically approximate subspace" $S \subseteq \mathbb{F}_2^{2n}$ such that $|S|/2^n \geq \gamma/2$, $2^n p_\psi(x) \geq \gamma/4$ for all $x \in S$, and $\mathbb{P}_{s,s'\in S}[s + s' \in S] \geq \gamma/6$. To prove this, we consider the set X of all x such that p_ψ is large; specifically,

$$X = \{x \in \mathbb{F}_2^{2n} : 2^n p_\psi(x) \geq \gamma/4\}.$$

S is constructed as a random subset of X by independently including $x \in X$ in S with probability $2^n p_\psi(x)$. Since this random S satisfies the three desired properties simultaneously with non-zero probability, there must exist a deterministic S satisfying the three properties.

Now by the Balog-Szemerédi-Gowers theorem, there is a large subset $S' \subseteq S$ which is an approximate subspace, in the sense that it has small doubling constant; in particular,

$$|S'| \geq \frac{\gamma}{18}|S| \quad \text{and} \quad |S' + S'| \leq (36/\gamma)^6|S'|$$

The polynomial Freiman-Ruzsa theorem then implies there is a subspace $V \leq \mathbb{F}_2^{2n}$ such that $|V| \leq |S'|$ and S' is covered by (i.e. is contained in a union of) $O(1/\gamma^{54})$ cosets $y + V$ of V. It can then be deduced that

$$\sum_{x\in V} p_\psi(x) \geq \frac{|V|}{2^n}\,\text{poly}(\gamma) \quad \text{and} \quad \sum_{x\in V} p_\psi(x) \geq \text{poly}(\gamma).$$

Finally, we seek an *isotropic* subspace $V' \leq V$ such that $\sum_{x\in V'} p_\psi(x) \geq \text{poly}(\gamma)$, which will complete the proof. By applying a Clifford unitary to $|\psi\rangle$ (which does not change the stabiliser fidelity), we may assume WLOG that $V = \langle Z_1, X_1, \ldots, Z_k, X_k, Z_{k+1}, \ldots, Z_{k+m}\rangle$ for some k, m. The structure of this V means that it can be covered by $2^k + 1$ isotropic subspaces

$V_1, \ldots, V_{2^k+1} \leq V$ of V. By a simple averaging argument, there must exist $j \in [2^k + 1]$ such that $\sum_{x \in V_j} p_\psi(x) \geq \frac{1}{2^k+1} \sum_{x \in V} p_\psi(x)$. This V_j will be our desired isotropic subspace V'. So it suffices to show that $2^k \leq 1/\operatorname{poly}(\gamma)$.

We already have $\sum_{x \in V} 2^n p_\psi(x) \geq |V| \cdot \operatorname{poly}(\gamma)$, so it is enough to show that $\sum_{x \in V} 2^n p_\psi(x) \leq |V|/2^k$. For this, we use the following *uncertainty relation* for Pauli operators proven in [BvDH25]:

Lemma 2. *Let $A \subseteq \mathbb{F}_2^{2n}$ be a subset of $2n$-bit strings which label the subset of phaseless Pauli operators $\{W_x : x \in A\}$. Let Γ_A denote the anti-commutation graph of A, i.e. the graph with vertex set A and edge set $E = \{(x, y) \in A^2 : W_x W_y = -W_y W_x\}$. Then for any quantum state $|\psi\rangle$,*

$$\sum_{x \in A} |\langle \psi | W_x | \psi \rangle|^2 \leq \vartheta(\Gamma_A),$$

where $\vartheta(\Gamma_A)$ is the Lovász theta function *of Γ_A, which is defined as*

$$\vartheta(\Gamma_A) \coloneqq \max_{\rho \in \mathbb{R}^{A \times A}} \{\|\rho\|_1 : \rho \geq 0, \operatorname{Tr}(\rho) = 1, \rho_{jk} = 0 \text{ for all } (j, k) \in E\}.$$

The proof is completed by applying Lemma 2 and using some simple properties of the Lovász theta function to show that $\vartheta(\Gamma_V) \leq |V|/2^k$.

10.4 Tolerant testing of stabiliser states

We now show the connection of the inverse theorem to tolerant stabiliser testing.

Theorem 3 ([AD25, BvDH25])**.** *There is a quantum algorithm which determines with high probability whether $\mathcal{F}_S(|\psi\rangle) \geq \epsilon_1$ (accept) or $\mathcal{F}_S(|\psi\rangle) \leq \epsilon_2$ (reject), provided that $\epsilon_2 \leq \operatorname{poly}(\epsilon_1)$. It uses $\operatorname{poly}(1/\epsilon_1)$ copies of $|\psi\rangle$ and runs in $n \cdot \operatorname{poly}(1/\epsilon_1)$ time.*

Via the inverse theorem, we have that

$$\mathcal{F}_S(|\psi\rangle) \geq \operatorname{poly}(G(|\psi\rangle)).$$

We also have

$$\mathcal{F}_S(|\psi\rangle) \leq G(|\psi\rangle)^{1/6}$$

which is proven much more simply using some basic analysis.

Since $\mathcal{F}_S(|\psi\rangle)$ is upper and lower bounded by polynomials in $G(|\psi\rangle)$, we can reduce the task of determining whether $\mathcal{F}_S(|\psi\rangle) \geq \epsilon_1$ or $\mathcal{F}_S(|\psi\rangle) \leq \epsilon_2$ to determining whether $G(|\psi\rangle) \geq \text{poly}(\epsilon_1)$ or $G(|\psi\rangle) \leq \text{poly}(\epsilon_2)$ (these implicit polynomials correspond to those in the upper and lower bounds above).

$G(|\psi\rangle)$ can be efficiently estimated using *Bell difference sampling*, an algorithmic primitive which samples from the probability distribution $p_\psi * p_\psi$ using 4 copies of the state $|\psi\rangle$. The procedure to estimate $G(|\psi\rangle)$ is:

- Perform Bell difference sampling to obtain a sample $z \in \mathbb{F}_2^{2n}$.
- Estimate $|\langle\psi|W_z|\psi\rangle|^2$ using the SWAP test.
- Repeat the above steps $\text{poly}(1/\epsilon_1)$ times and take the mean of the estimates, which quickly converges to the true mean $G(|\psi\rangle)$ by a simple Chernoff bound argument.

Therefore, the algorithm simply estimates $G(|\psi\rangle)$ and accepts if the estimate is above a certain threshold depending on ϵ_1 and ϵ_2, and rejects otherwise.

Note that the Gowers norm $\|\psi\|_{U^3}$ is not directly but indirectly estimated by the algorithm, since it satisfies the quadratic relationship

$$\|\psi\|_{U^3}^{16} \leq G(|\psi\rangle) \leq \|\psi\|_{U^3}^{8}.$$

Thus, we can think of $\|\psi\|_{U^3}$ as being a good "proxy" for $G(|\psi\rangle)$; moreover, the Gowers norm is a more "natural" quantity with an intuitive definition, so it is often helpful to think of this instead of $G(|\psi\rangle)$.

References

[AD25] Arunachalam, S. and Dutt, A., *Polynomial-time tolerant testing stabilizer states.* In *STOC '25: Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 1234–1241. ACM, New York, 2025. https://doi.org/10.1145/3717823.3718277.

[BvDH25] Bao, Z. and van Dordrecht, P. and Helsen, J., *Tolerant testing of stabilizer states with a polynomial gap via a generalized uncertainty relation.* In *STOC '25: Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 1254–1262. ACM, New York, 2025. https://doi.org/10.1145/3717823.3718201.

Isaac Holt, University of Cambridge
email: isaac.holt@cst.cam.ac.uk

11 Higher-order Fourier analysis and algebraic property testing

After H. Hatami, P. Hatami, and S. Lovett [HHL19]

A summary written by Zimrat Kaniel

Abstract

These notes survey algebraic property testing over $\mathbb{F}_2^n$. We analyze the BLR affine linearity test, link the AKKLR low-degree test to Gowers uniformity norms, and sketch the analysis of the AKKLR tester by Bhattacharyya et al. [BKSSZ10]. Finally, we show that high Gowers uniformity implies structural closeness to low-degree polynomials.

11.1 Introduction and basic definitions

Property testing studies randomized algorithms that determine whether an object satisfies a global structural property using only a few local queries. In algebraic property testing, the objective is to verify whether an unknown function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ belongs to a specific algebraic family, such as affine linear functions ($\mathcal{P}_1$) or polynomials of degree at most d ($\mathcal{P}_d$). While classical Fourier analysis perfectly captures linear properties, it fails to detect higher-degree structures. This limitation naturally motivates the development of higher-order Fourier analysis and Gowers uniformity norms.

Given $f, g : \mathbb{F}_2^n \to \mathbb{F}_2$, their *distance* is $\operatorname{dist}(f, g) := \mathbb{P}_x[f(x) \neq g(x)]$ for a uniform $x \in \mathbb{F}_2^n$. For a property $\mathcal{P} \subset \{f : \mathbb{F}_2^n \to \mathbb{F}_2\}$, the *distance from f to $\mathcal{P}$* is $\operatorname{dist}(f, \mathcal{P}) := \min_{g \in \mathcal{P}} \operatorname{dist}(f, g)$. We say f is *ε-close* to $\mathcal{P}$ if $\operatorname{dist}(f, \mathcal{P}) \leq \varepsilon$, and *$\varepsilon$-far* otherwise. We denote $\delta_d(f) := \operatorname{dist}(f, \mathcal{P}_d)$, where $\mathcal{P}_d$ is the set of polynomials of degree at most d. A property $\mathcal{P}$ is *testable with one-sided error* if there are functions $q : (0, 1) \to \mathbb{Z}_{>0}$, $\delta : (0, 1) \to (0, 1)$, and an algorithm T making $q(\varepsilon)$ queries to f, that always accepts if $f \in \mathcal{P}$, and rejects with probability $\geq \delta(\varepsilon)$ if f is ε-far from $\mathcal{P}$. If q is constant (independent of n and ε), $\mathcal{P}$ is *proximity-obliviously testable*.

11.2 The BLR test and affine linearity

A function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is *affine linear* ($f \in \mathcal{P}_1$) if $f(x) = a_0 + \sum_{i=1}^n a_i x_i$ for $a_0, a_i \in \mathbb{F}_2$. Affine linearity is locally characterized by the identity $f(x) + f(y) + f(z) = f(x + y + z)$ for all $x, y, z \in \mathbb{F}_2^n$.

Algorithm 1: BLR test for affine linearity
Given query access to $f : \mathbb{F}_2^n \to \mathbb{F}_2$:

1. Sample $x, y, z \in \mathbb{F}_2^n$ independently and uniformly at random.
2. Query the values $f(x), f(y), f(z)$, and $f(x + y + z)$.
3. **Accept** if $f(x) + f(y) + f(z) = f(x + y + z)$; **Reject** otherwise.

Theorem 1 (Theorem 2.3 in [HHL19])**.** *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ s.t. $\mathrm{dist}(f, \mathcal{P}_1) = \varepsilon$. Then the BLR test rejects f with probability between ε and 6ε.*

Proof. Let $F(x) = (-1)^{f(x)} = \sum_\alpha \widehat{F}(\alpha)(-1)^{\langle \alpha, x\rangle}$, where the Fourier coefficients are $\widehat{F}(\alpha) = \mathbb{E}_x[F(x)(-1)^{\langle \alpha, x\rangle}]$. The test accepts if and only if $F(x)F(y)F(z)F(x+y+z) = 1$. Substituting the Fourier expansion and using character orthogonality, which ensures $\mathbb{E}_{x,y,z}[(-1)^{\langle \alpha+\lambda, x\rangle + \langle \beta+\lambda, y\rangle + \langle \gamma+\lambda, z\rangle}]$ is 1 if $\alpha = \beta = \gamma = \lambda$ and 0 otherwise, yields:

$$\mathbb{P}[\text{accept}] = \frac{1}{2} + \frac{1}{2}\mathbb{E}_{x,y,z}\left[F(x)F(y)F(z)F(x+y+z)\right] = \frac{1}{2} + \frac{1}{2}\sum_{\alpha \in \mathbb{F}_2^n} \widehat{F}(\alpha)^4.$$

By Parseval's identity, $\sum_\alpha \widehat{F}(\alpha)^2 = 1$. Since $\mathrm{dist}(f, \mathcal{P}_1) = \varepsilon$, every Fourier coefficient satisfies $|\widehat{F}(\alpha)| \le 1 - 2\varepsilon$. Let α^* attain the maximum $|\widehat{F}(\alpha^*)| = 1 - 2\varepsilon$.

For the **upper bound on rejection**, note that $\widehat{F}(\alpha^*)^4 \ge 1 - 8\varepsilon$, and the remaining terms satisfy $\sum_{\alpha \ne \alpha^*} \widehat{F}(\alpha)^4 \ge 0$. Thus,

$$\mathbb{P}[\text{accept}] \ge \frac{1}{2}\big(1 + (1 - 8\varepsilon) + 0\big) = 1 - 4\varepsilon \implies \mathbb{P}[\text{reject}] \le 4\varepsilon \le 6\varepsilon.$$

For the **lower bound on rejection**, bounding the sum via the maximum gives:

$$\mathbb{P}[\text{accept}] \le \frac{1}{2}\Big(1 + \max_\alpha |\widehat{F}(\alpha)|^2 \sum_\alpha \widehat{F}(\alpha)^2\Big) \le 1 - \varepsilon \implies \mathbb{P}[\text{reject}] \ge \varepsilon. \quad \square$$

11.3 Higher-order derivatives and Gowers norms

To capture polynomial structure beyond degree 1, we utilize directional derivatives.

Definition 2 (Directional derivatives)**.** *For $f : \mathbb{F}_2^n \to \mathbb{F}_2$ and $y \in \mathbb{F}_2^n$, the* additive directional derivative *is $D_y f(x) := f(x+y) + f(x)$. The* iterated derivative *in directions $y_1, \dots, y_k$ is $D_{y_1,\dots,y_k} f(x) = \sum_{S \subseteq [k]} f\big(x + \sum_{i \in S} y_i\big)$. For complex functions $F : \mathbb{F}_2^n \to \mathbb{C}$, the* multiplicative derivative *is $\Delta_y F(x) := F(x+y)\overline{F(x)}$. Setting $F = (-1)^f$ gives $\Delta_y F(x) = (-1)^{D_y f(x)}$.*

A fundamental algebraic fact is that $f \in \mathcal{P}_d$ if and only if $D_{y_1,\dots,y_{d+1}} f(x) = 0$ for all $x, y_1, \dots, y_{d+1} \in \mathbb{F}_2^n$. This local characterization motivates both Gowers norms and higher-degree testing.

Definition 3 (Gowers uniformity norm)**.** *For $F : \mathbb{F}_2^n \to \mathbb{C}$ and integer $d \geq 1$, the* Gowers uniformity norm of order d *is*

$$\|F\|_{U_d}^{2^d} := \big|\mathbb{E}_{x,y_1,\dots,y_d}\big[(\Delta_{y_1} \cdots \Delta_{y_d} F)(x)\big]\big| = \left|\mathbb{E}\Big[\prod_{S \subseteq [d]} \mathcal{C}^{d-|S|} F\Big(x + \sum_{i \in S} y_i\Big)\Big]\right|$$

where $\mathcal{C}$ denotes complex conjugation.

The local characterization leads directly to the AKKLR(d) test of Alon et al. [AKKLR05].

Algorithm 2: AKKLR(d) test for $\mathcal{P}_d$
Given query access to $f : \mathbb{F}_2^n \to \mathbb{F}_2$:

1. Sample $x, y_1, \dots, y_{d+1} \in \mathbb{F}_2^n$ independently and uniformly at random.
2. Query the values $f\big(x + \sum_{i \in S} y_i\big)$ for all subsets $S \subseteq [d+1]$.
3. **Accept** if $\sum_{S \subseteq [d+1]} f\big(x + \sum_{i \in S} y_i\big) = 0$; **Reject** otherwise.

Proposition 4 (Analytical bridge)**.** *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ and $F = (-1)^f$. The rejection probability of AKKLR(d) satisfies:*

$$\|(-1)^f\|_{U_{d+1}}^{2^{d+1}} = 1 - 2 \cdot \mathbb{P}[\text{AKKLR}(d) \text{ rejects } f].$$

Proof. The multiplicative derivative $(\Delta_{y_1} \dots \Delta_{y_{d+1}} F)(x)$ evaluates to $(-1)^{\sum_{S \subseteq [d+1]} f(x + \sum_{i \in S} y_i)}$.
Taking expectations gives $(+1)\mathbb{P}[\text{accept}] + (-1)\mathbb{P}[\text{reject}] = 1 - 2\mathbb{P}[\text{reject}]$. $\square$

11.4 Low-degree testing in the 99% regime

In the 99% regime, the goal is to reject functions slightly far from $\mathcal{P}_d$ with optimal query complexity.

Theorem 5 (Theorem 3.6 in [HHL19]; [BKSSZ10])**.** *Fix $1 \le d \le n$. There exists an absolute constant $c > 0$ such that for all $f : \mathbb{F}_2^n \to \mathbb{F}_2$,*

$$\mathbb{P}[\text{AKKLR}(d) \text{ rejects } f] \ge c \cdot \min\{2^d \delta_d(f), 1\}.$$

11.4.1 Proof architecture: reduction to flat-testing

The analysis reduces the AKKLR(d) test to the *k-flat test*, which samples a uniform k-dimensional affine subspace $A \subset \mathbb{F}_2^n$ and rejects if $f|_A \notin \mathcal{P}_d$, or equivalently if $\sum_{x \in A} f(x) \neq 0$. Because $d+1$ uniform directions are linearly independent with high probability, they span a uniform $(d+1)$-flat. Thus, the AKKLR(d) test is asymptotically equivalent to the $(d+1)$-flat test. The analysis splits into two proximity regimes depending on $\delta_d(f) = \text{dist}(f, \mathcal{P}_d)$:

Regime 1: small proximity ($\delta_d(f) \le 2^{-(d+2)}$). Let $g \in \mathcal{P}_d$ be the closest polynomial to f. On a random $(d+1)$-flat A, the 2^{d+1} points are pairwise independent. The test rejects whenever $f|_A$ and $g|_A$ differ on an odd number of points. To establish a lower bound, it suffices to calculate the probability that they differ on *exactly one* point By inclusion-exclusion, this probability is at least:

$$\mathbb{P}[(d+1)\text{-flat rejects}] \ge 2^{d+1}\delta_d(f)\big(1 - 2^{d+1}\delta_d(f)\big).$$

For $\delta_d(f) \le 2^{-(d+2)}$, we have $1 - 2^{d+1}\delta_d(f) \ge 1/2$, yielding $\mathbb{P}[\text{reject}] \ge 2^d \delta_d(f)$.

Regime 2: large proximity ($\delta_d(f) > 2^{-(d+2)}$). When $\delta_d(f) > 2^{-(d+2)}$, the bound from Regime 1 becomes negative. To establish a constant lower bound independent of n and d, the proof uses a two-step reduction:

Step-down reduction (Lemma 3.9 in [HHL19]): For $k \ge k' \ge d+1$,

$$\mathbb{P}[k'\text{-flat test rejects}] \ge 2^{-(k-k')}\mathbb{P}[k\text{-flat test rejects}].$$

If $f|_A \notin \mathcal{P}_d$ on a k-flat A, yet a majority of its hyperplanes $H \subset A$ satisfy $f|_H \in \mathcal{P}_d$, we can find $k \ge d+2$ independent hyperplanes where the degree-$(d+1)$ component of $f|_A$ vanishes. This forces $\deg(f|_A) \le d$, a contradiction.

Thus, a random $(k-1)$-flat rejects with probability at least $1/2$, yielding a constant penalty factor $2^{-(k-d-1)}$ for $k = d + O(1)$.

Inductive bound (Lemma 3.10 in [HHL19]): Inducting on n, the hyperplanes of $\mathbb{F}_2^n$ are partitioned into K "close" hyperplanes (where $\delta_d(f|_H) < \beta 2^{-d}$) and remaining "far" hyperplanes. The far hyperplanes provide a constant rejection probability, which suffices if $K \leq \gamma 2^d$. If $K > \gamma 2^d$, the Hyperplane Gluing Lemma below shows f is actually globally close ($\delta_d(f) \leq 2^{-(d+2)}$), reducing the analysis back to Regime 1.

Lemma 6 (Hyperplane gluing)**.** *Let $A_1, \ldots, A_K$ be distinct hyperplanes in $\mathbb{F}_2^n$. If $\delta_d(f|_{A_i}) \leq \alpha < 2^{-(d+2)}$ for each $i \in [K]$ and $K > 2^{d+1}$, then $\delta_d(f) \leq \frac{3}{2}\alpha + \frac{9}{K}$.*

Sketch of gluing lemma. Let $P_i \in \mathcal{P}_d(A_i)$ be the local polynomial α-close to $f|_{A_i}$. On the intersection $A_i \cap A_j$, the distance between P_i and P_j is at most $4\alpha < 2^{-d}$. Since any two distinct degree-d polynomials over $\mathbb{F}_2$ must differ on at least a 2^{-d} fraction of points (see MacWilliams and Sloane [MS77]), the distance constraint forces $P_i \equiv P_j$ identically on $A_i \cap A_j$.

Since $K > 2^{d+1}$, we can select $\ell > d$ independent hyperplanes. The absolute agreement on their intersections allows us to uniquely *glue* these local polynomials into a single global polynomial $P \in \mathcal{P}_d(\mathbb{F}_2^n)$ such that $P|_{A_i} \equiv P_i$ for all $i \in [K]$.

Finally, let $\mathrm{BAD} \subset \mathbb{F}_2^n$ be the points contained in fewer than $K/3$ hyperplanes. By Chebyshev's inequality, $|\mathrm{BAD}|/2^n \leq 9/K$. For any point $z \notin \mathrm{BAD}$, $f(z)$ agrees with $P(z)$ on a majority of the hyperplanes containing z. Averaging the local errors over $z \notin \mathrm{BAD}$ yields:

$$\frac{\alpha}{2} \geq \frac{1}{3}\Big(\mathrm{dist}(f,P) - \frac{|\mathrm{BAD}|}{2^n}\Big) \implies \mathrm{dist}(f,P) \leq \frac{3}{2}\alpha + \frac{9}{K}. \qquad \square$$

11.5 Analytical consequences

As a direct consequence of Theorem 5 and Proposition 4, we establish that functions with Gowers uniformity norm close to 1 are structurally close to low-degree polynomials.

Theorem 7 (Theorem 3.17 in [HHL19])**.** *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$. For any $1 \leq d \leq n$, if $\|(-1)^f\|_{U_{d+1}} \geq 1 - \varepsilon$, then $\delta_d(f) \leq C\varepsilon$ for an absolute constant $C > 0$.*

Proof. By the analytical bridge (Proposition 4):

$$1 - 2 \cdot \mathbb{P}[\text{AKKLR}(d) \text{ rejects } f] = \|(-1)^f\|_{U_{d+1}}^{2^{d+1}} \geq (1-\varepsilon)^{2^{d+1}} \geq 1 - 2^{d+1}\varepsilon.$$

Thus $\mathbb{P}[\text{AKKLR}(d) \text{ rejects } f] \leq 2^d\varepsilon$. Applying Theorem 5 gives $c \cdot \min\{2^d\delta_d(f), 1\} \leq 2^d\varepsilon$. For $\varepsilon < c/2^d$, this yields $c \cdot 2^d\delta_d(f) \leq 2^d\varepsilon \implies \delta_d(f) \leq \frac{1}{c}\varepsilon$. □

References

[AKKLR05] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn, and D. Ron, *Testing Reed-Muller codes.* IEEE Trans. Inform. Theory, **51** (2005), no. 11, 4032–4039.

[BKSSZ10] A. Bhattacharyya, S. Kopparty, G. Schoenebeck, M. Sudan, and D. Zuckerman, *Optimal testing of Reed-Muller codes.* In: 2010 IEEE 51st Annual Symposium on Foundations of Computer Science FOCS 2010. IEEE Computer Soc., Los Alamitos, CA, 488–497.

[HHL19] H. Hatami, P. Hatami, and S. Lovett, *Higher-order Fourier Analysis and Applications.* Foundations and Trends in Theoretical Computer Science, **13** (2019), no. 4, 247–448.

[MS77] F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes.* Elsevier, 1977.

Zimrat Kaniel, The Hebrew University of Jerusalem
email: zimrat.kniel@mail.huji.ac.il

12 New proofs of Plünnecke-type estimates for product sets in groups, and the theorem of Balog, Szemerédi and Gowers

After G. Petridis [Pe] and Y. Zhao [Zh]

A summary written by Deepanshu Kush

Abstract

We present Petridis' short proof of the Plünnecke–Ruzsa sumset inequalities, and Gowers' proof of the Balog–Szemerédi–Gowers theorem in Zhao's exposition. The halves are more closely related than they appear: the one Ruzsa-type triangle inequality not accessible by reflecting sets is exactly what converts the two-set form of the Balog–Szemerédi–Gowers theorem into the one-set form, and its known proof runs through Petridis' lemma.

12.1 Introduction

A recurring theme in additive combinatorics is that a single crude statistic of a finite set forces a great deal of structure; we treat two instances. Throughout A, B are finite non-empty subsets of a group, written multiplicatively in general and additively when it is commutative; kB is the k-fold sumset and B^k the k-fold product set. The first statistic is the doubling of A relative to B, the least α with $|A+B| \le \alpha|A|$. Plünnecke [Pl] showed in 1970 that this alone controls every iterated sumset of B: some non-empty $X \subseteq A$ satisfies $|X+hB| \le \alpha^h|X|$ for all h. Ruzsa [Ru] extended this to difference sets.

Theorem 1 (Ruzsa)**.** *Let A, B be finite non-empty sets in a commutative group with $|A+B| \le \alpha|A|$. Then $|kB-lB| \le \alpha^{k+l}|A|$ for all integers $k, l \ge 0$.*

Both original proofs were graph-theoretic, via magnification ratios in commutative layered graphs and Menger's theorem. In 2012 Petridis [Pe] found a proof occupying barely a page, which moreover works in an arbitrary group up to the exact point where commutativity becomes indispensable. Sections 12.2–12.3 present Sections 2 and 3 of [Pe], culminating in the proof of Theorem 1.

The second statistic is the additive energy $E(A) = |\{(a,b,c,d) \in A^4 : a+b = c+d\}|$, which counts additive quadruples rather than sums. Small doubling implies large energy, but not conversely; the Balog–Szemerédi–Gowers theorem repairs the converse at the cost of passing to a subset.

Theorem 2 (Balog–Szemerédi–Gowers)**.** *Let A be a finite subset of a commutative group with $E(A) \geq |A|^3/K$. Then there is $A' \subseteq A$ with $|A'| \geq K^{-O(1)}|A|$ and $|A' + A'| \leq K^{O(1)}|A'|$.*

The original proof [BS] used the regularity method and so gave tower-type bounds; Section 12.4 reproduces Gowers' polynomial bounds [Go] following [Zh], the form in which the theorem became a basic tool and an ingredient of his proof of Szemerédi's theorem for progressions of length four. The two topics are linked by more than proximity in the syllabus; see Section 12.4.1.

12.2 Growth of triple products

Petridis' method rests on a single well-chosen subset: calling $|ZB|/|Z|$ the expansion ratio of Z under multiplication by B, choose a non-empty $X \subseteq A$ minimising it. Minimality is a hypothesis one may use, not merely a normalisation, and it is what makes the following induction close.

Proposition 3 ([Pe], Proposition 2.1)**.** *Let X and B be finite non-empty sets in a group with $K := |XB|/|X| \leq |ZB|/|Z|$ for every non-empty $Z \subseteq X$. Then $|CXB| \leq K|CX| = |CX||XB|/|X|$ for every finite set C.*

Proof. Write $C = \{c_1, \dots, c_r\}$ and, using this arbitrary ordering, disjointify $CX = \bigcup_{i \leq r} c_i X$ by setting $X_1 = X$ and, for $i > 1$, $X_i = \{x \in X : c_i x \notin \{c_1, \dots, c_{i-1}\}X\}$. Then $\{c_1, \dots, c_j\}X = \bigcup_{i \leq j} c_i X = \bigcup_{i \leq j} c_i X_i$ for every j, and the sets $c_i X_i$ are pairwise disjoint, so

$$|\{c_1, \dots, c_j\}X| \;=\; \sum_{i=1}^{j} |c_i X_i| \;=\; \sum_{i=1}^{j} |X_i|. \tag{1}$$

We induct on r. For $r = 1$, $|c_1 XB| = |XB| = K|X| = K|c_1 X|$. For $r > 1$ put $X_r^c = X \setminus X_r$. By definition of X_r we have $c_r X_r^c \subseteq \{c_1, \dots, c_{r-1}\}X$, hence $c_r X_r^c B \subseteq \{c_1, \dots, c_{r-1}\}XB$, and so $CXB \subseteq (\{c_1, \dots, c_{r-1}\}XB) \cup ((c_r XB) \setminus (c_r X_r^c B))$. As $|(c_r XB) \setminus (c_r X_r^c B)| = |XB| - |X_r^c B|$, this yields

$$|CXB| \;\leq\; |\{c_1, \dots, c_{r-1}\}XB| + \big(|XB| - |X_r^c B|\big). \tag{2}$$

The first term of (2) is at most $K\sum_{i<r}|X_i|$ by the induction hypothesis and (1). For the second, the hypothesis applied to $Z = X_r^c$ gives $|X_r^cB| \geq K|X_r^c|$ (trivially if $X_r^c = \emptyset$), whence $|XB| - |X_r^cB| \leq K(|X| - |X_r^c|) = K|X_r|$. Adding, $|CXB| \leq K\sum_{i\leq r}|X_i|$, and (1) concludes. □

Note how little is used: associativity, and the minimality of X at the one moment where a subset of X appears. Minimising over all non-empty subsets of A, so that $K \leq |AB|/|A| \leq \alpha$, gives Theorem 1.5 of [Pe]: a single X satisfies $|CXB| \leq \alpha|CX|$ for *all* finite C at once, and it is this uniformity in C that permits the iteration below.

12.3 The Plünnecke–Ruzsa inequalities

Theorem 4 ([Pe], Theorem 3.1)**.** *Let A, B be finite non-empty sets in a commutative group with $|A+B| \leq \alpha|A|$. Then there is a non-empty $X \subseteq A$ with $|X+hB| \leq \alpha^h|X|$ for every positive integer h.*

Proof. Choose a non-empty $X \subseteq A$ minimising $|Z+B|/|Z|$ over non-empty $Z \subseteq A$, and induct on h. For $h = 1$, $|X+B| \leq |X||A+B|/|A| \leq \alpha|X|$. For $h > 1$ apply Proposition 3 with $C = (h-1)B$, whose hypothesis holds by the choice of X:

$$|X+hB| = |C+X+B| \leq \alpha|C+X| = \alpha|X+(h-1)B| \leq \alpha^h|X|. \quad \square$$

Commutativity is used once, and decisively: Proposition 3 bounds $|CXB|$ by $|CX|$ with X sandwiched between C and B, and to iterate one must feed the output back in as the new C. Commutatively $C = (h-1)B$ turns $C+X+B$ into $X+hB$ and $C+X$ into $X+(h-1)B$; in a general group it only gives $|B^{h-1}XB| \leq K|B^{h-1}X|$, controlling sandwiched products rather than $|XB^h|$. This is no artefact: Plünnecke's theorem genuinely fails in general groups, as $A = H \cup \{x\}$ with H a subgroup and $|HxH| = |H|^2$ shows, for then $|AA| \leq 3|A|$ while $|AAA| \geq (|A|-1)^2$. Sections 4 and 5 of [Pe] restore matters under the extra hypothesis $|AbB| \leq \beta|A|$ for all $b \in B$.

To reach difference sets we use Ruzsa's triangle inequality $|X||Y-Z| \leq |X-Y||X-Z|$, whose proof is a one-line injection: fixing a representation $d = \mathsf{y}(d) - \mathsf{z}(d)$ for each $d \in Y-Z$, the map $(x,d) \mapsto (x-\mathsf{y}(d), x-\mathsf{z}(d))$ is injective, as d is the difference of the coordinates and x is then recovered. Replacing Y by $-Y$ and Z by $-Z$ gives the form we need,

$$|X||Y-Z| \;\leq\; |X+Y||X+Z|. \tag{3}$$

Proof of Theorem 1. If k or l vanishes, Theorem 4 suffices: for instance $|kB| \le |X+kB| \le \alpha^k|X| \le \alpha^k|A|$. Otherwise let X be as in Theorem 4 and apply (3) with $Y = kB$, $Z = lB$: then $|X||kB - lB| \le |X+kB||X+lB| \le \alpha^{k+l}|X|^2 \le \alpha^{k+l}|X||A|$, and we divide by $|X|$. □

12.4 Additive energy and the Balog–Szemerédi–Gowers theorem

Writing $r_A(x) = |\{(a,b) \in A^2 : a+b = x\}|$ we have $E(A) = \sum_x r_A(x)^2$, so the energy is an L^2 quantity. One direction is easy: if $|A+A| \le K|A|$ then Cauchy–Schwarz gives $E(A) \ge |A+A|^{-1}(\sum_{x\in A+A} r_A(x))^2 = |A|^4/|A+A| \ge |A|^3/K$. The converse fails badly, as $A = [N] \cup S$ shows for S a set of N elements in general position: the interval forces $E(A) = \Theta(|A|^3)$ while $S+S$ forces $|A+A| = \Theta(|A|^2)$. Theorem 2 is the repair: nothing follows about A, but a large subset – here $[N]$ – must be structured.

12.4.1 Reduction to two sets, and then to a graph

Put $E(A,B) = |\{(a,b,a',b') \in A\times B\times A\times B : a+b = a'+b'\}|$, so $E(A,A) = E(A)$, and call $A' \subseteq A$, $B' \subseteq B$ *good* if $|A'|, |B'| \ge K^{-O(1)}n$ and $|A'+B'| \le K^{O(1)}n$.

Theorem 5 ([Zh], Theorem 7.13.7)**.** *Let A, B be finite subsets of a commutative group with $|A|, |B| \le n$ and $E(A,B) \ge n^3/K$. Then good $A' \subseteq A$, $B' \subseteq B$ exist.*

Deducing Theorem 2 from Theorem 5 means replacing $A'+B'$ by $A'+A'$, and here the first half of this summary returns. Apply Theorem 5 with $B = A$ and $n = |A|$ to get $A', B' \subseteq A$ with $|A'|, |B'| \ge K^{-O(1)}|A|$ and $|A'+B'| \le K^{O(1)}|A|$. What is needed is

$$|X||Y+Z| \;\le\; |X+Y||X+Z|, \tag{4}$$

applied with $X = B'$ and $Y = Z = A'$, giving $|A'+A'| \le |A'+B'|^2/|B'| \le K^{O(1)}|A| \le K^{O(1)}|A'|$. Now (4) resembles (3) but is genuinely different: reflecting Y or Z turns it into a statement about $|X-Y|$ or $|X-Z|$, so it cannot be obtained from Ruzsa's triangle inequality by the trick used above ([Zh], Remark 7.2.2). It is, however, immediate from Proposition 3: choosing

$X_0 \subseteq X$ minimising $|X_0+Y|/|X_0|$ and applying the proposition with $C = Z$,

$$|Y+Z| \;\le\; |X_0+Y+Z| \;\le\; |X_0+Z|\frac{|X_0+Y|}{|X_0|} \;\le\; |X+Z|\frac{|X+Y|}{|X|}.$$

So the passage to the one-set form runs through exactly the lemma of Section 12.2.

For a bipartite graph G with parts A and B, define the restricted sumset $A +_G B = \{a+b : \; ab \in E(G)\}$.

Theorem 6 (graph BSG; [Zh], Theorem 7.13.9)**.** *Let A, B be finite subsets of a commutative group with $|A|, |B| \le n$ and let G be bipartite with parts A, B satisfying $e(G) \ge n^2/K$ and $|A +_G B| \le Kn$. Then good $A' \subseteq A$, $B' \subseteq B$ exist.*

Theorem 6 implies Theorem 5 by restricting to popular sums. Let $S = \{x : r_{A,B}(x) \ge n/(2K)\}$ and let G join a to b exactly when $a+b \in S$. Unpopular sums contribute $\sum_{x\notin S} r_{A,B}(x)^2 \le \frac{n}{2K}|A||B| \le \frac{n^3}{2K}$, at most half of $E(A,B)$, so $\sum_{x\in S} r_{A,B}(x)^2 \ge n^3/(2K)$; as $r_{A,B} \le n$ this gives $e(G) = \sum_{x\in S} r_{A,B}(x) \ge n^2/(2K)$. Finally $A +_G B \subseteq S$ and each element of S has $\ge n/(2K)$ representations, so $|A +_G B| \le 2Kn$.

12.4.2 Paths of length two and three

The engine is dependent random choice.

Lemma 7 (path of length 2; [Zh], Lemma 7.13.10)**.** *Let $\delta, \varepsilon > 0$ and let G be bipartite with parts A, B and at least $\delta|A||B|$ edges. Then there is $U \subseteq A$ with $|U| \ge \delta|A|/2$ such that at least a $(1-\varepsilon)$-fraction of pairs $(x,y) \in U^2$ have at least $\varepsilon\delta^2|B|/2$ common neighbours.*

Proof. Call $(x,y) \in A^2$ unfriendly if it has fewer than $\varepsilon\delta^2|B|/2$ common neighbours. Pick $v \in B$ uniformly at random and set $U = N(v)$, so $\mathbb{E}|U| = e(G)/|B| \ge \delta|A|$ while $\mathbb{P}(x,y \in U) = \mathrm{codeg}(x,y)/|B| < \varepsilon\delta^2/2$ for unfriendly (x,y). If X counts the unfriendly pairs inside U^2 then $\mathbb{E}X < \varepsilon\delta^2|A|^2/2$, so by Jensen $\mathbb{E}[|U|^2 - X/\varepsilon] \ge (\mathbb{E}|U|)^2 - \mathbb{E}X/\varepsilon > \delta^2|A|^2/2$. Fix v attaining at least this value: then $|U| \ge \delta|A|/2$ and $X \le \varepsilon|U|^2$. □

Lemma 8 (path of length 3; [Zh], Lemma 7.13.11)**.** *Let $\delta > 0$ and let G be bipartite with parts A, B and at least $\delta|A||B|$ edges. Then there are $A' \subseteq A$,*

$B' \subseteq B$ with $|A'| \geq c\delta^C|A|$ and $|B'| \geq c\delta^C|B|$ such that every pair in $A' \times B'$ is joined by at least $c\delta^C|A||B|$ paths of length three, for absolute constants $c, C > 0$.

Sketch. Discarding vertices of A of degree below $\delta|B|/2$ costs at most half the edges, leaving A_1 of size $\geq \delta|A|/2$ and density $\geq \delta/2$. Lemma 7 on (A_1, B) with $\varepsilon = \delta/10$ gives $A_2 \subseteq A_1$ of size $\Omega(\delta^2|A|)$ in which all but a $(\delta/10)$-fraction of pairs are *friendly*, i.e. have $\Omega(\delta^3|B|)$ common neighbours. Take $B' = \{b : \deg(b, A_2) \geq \delta|A_2|/4\}$ and let A' be those $a \in A_2$ friendly to $\geq (1 - \delta/5)|A_2|$ of A_2; counting edges and unfriendly pairs respectively gives $|B'| \geq \delta|B|/4$ and $|A'| \geq |A_2|/2$. For $(a, b) \in A' \times B'$ at least $\delta|A_2|/20$ vertices $a_1 \in A_2$ are both adjacent to b and friendly to a, and each shares $\Omega(\delta^3|B|)$ neighbours b_1 with a, every one giving a path $a\, b_1\, a_1\, b$. □

Proof of Theorem 6. Since $e(G) \geq n^2/K$ and $|A|, |B| \leq n$ we have $|A|, |B| \geq n/K$, and the edge density is at least $\delta = 1/K$. Lemma 8 supplies $A' \subseteq A$, $B' \subseteq B$ of sizes at least $K^{-O(1)}n$ such that every $(a, b) \in A' \times B'$ is joined by at least $K^{-O(1)}n^2$ paths $a\, b_1\, a_1\, b$. For such a path set $x = a + b_1$, $y = a_1 + b_1$, $z = a_1 + b$, each of which lies in $A +_G B$, being a sum along an edge, and note that $a + b = x - y + z$. For fixed (a, b) distinct paths give distinct triples, since b_1 is recovered from x and a, and a_1 from z and b; and triples arising from distinct elements of $A' + B'$ differ, since $x - y + z$ determines the element. Counting triples in $(A +_G B)^3$ gives $K^{-O(1)}n^2|A' + B'| \leq |A +_G B|^3 \leq K^3n^3$, so $|A' + B'| \leq K^{O(1)}n$. With the reductions above, Theorem 2 follows. □

Tracking constants gives explicit but poor exponents; the best known are Schoen's [Sc], namely $|A'| \geq K^{-1/2-o(1)}|A|$ with $|A' - A'| \leq K^{4+o(1)}|A'|$. Both halves above extract a structured subset by opposite means: in Section 12.2 by an extremal principle, in Section 12.4 probabilistically.

References

[BS] Balog, A. and Szemerédi, E., *A statistical theorem of set addition.* Combinatorica **14** (1994), no. 3, 263–268.

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

[Pe] Petridis, G., *New proofs of Plünnecke-type estimates for product sets in groups.* Combinatorica **32** (2012), no. 6, 721–733.

[Pl] Plünnecke, H., *Eine zahlentheoretische Anwendung der Graphentheorie.* J. Reine Angew. Math. **243** (1970), 171–183.

[Ru] Ruzsa, I. Z., *An application of graph theory to additive number theory.* Sci. Ser. A Math. Sci. (N.S.) **3** (1989), 97–109.

[Sc] Schoen, T., *New bounds in Balog–Szemerédi–Gowers theorem.* Combinatorica **35** (2015), no. 6, 695–701.

[Zh] Zhao, Y., *Graph theory and additive combinatorics: exploring structure and randomness.* Cambridge University Press, Cambridge, 2023.

Deepanshu Kush, University of Cambridge
email: deepanshu.kush@cl.cam.ac.uk

13 Roth's theorem

After K. Roth, following Y. Zhao [Z]

A summary written by Stefanos Lappas

Abstract

In this talk, we present a Fourier-analytic proof of Roth's theorem, which states that any subset of the integers containing no three-term arithmetic progression (3-APs) must have density tending to zero. We first study 3-AP-free subsets of the finite vector space $\mathbb{F}_3^n$, where Fourier analysis shows that every such subset has size at most $O(3^n/n)$. We then turn to the integer setting, showing that every 3-AP-free subset of $[N] = \{1, \ldots, N\} \subseteq \mathbb{Z}$ has size at most $O(N/\log\log N)$.

13.1 An introduction to Roth's theorem

A three-term arithmetic progression (3-AP) is a triple $x, x+d, x+2d$ with common difference $d \neq 0$. A set is *3-AP-free* if it contains no such triple. Roth's theorem states that every 3-AP-free subset $A \subseteq [N] = \{1, \ldots, N\}$ has density tending to zero. Quantitatively,

$$|A| = O\left(\frac{N}{\log\log N}\right).$$

Writing $\alpha = |A|/N$ for the density of A in $[N]$, this is equivalent to

$$\alpha = O\left(\frac{1}{\log\log N}\right).$$

The proof follows the same basic scheme throughout:

few 3-APs $\Rightarrow$ large Fourier coefficient $\Rightarrow$ density increment $\Rightarrow$ iteration.

The model case $\mathbb{F}_3^n$ makes this mechanism particularly clear.

13.2 Fourier analysis on finite abelian groups

Let $G = \mathbb{F}_p^n$, with p prime, and set $\omega = e^{2\pi i/p}$. Given $r, x \in \mathbb{F}_p^n$, write $r \cdot x = r_1x_1 + \cdots + r_nx_n$. For $f : G \to \mathbb{C}$, define its Fourier transform $\widehat{f} : G \to \mathbb{C}$ by

$$\widehat{f}(r) =: \mathbb{E}_{x\in G} f(x)\omega^{-r\cdot x} = \frac{1}{p^n}\sum_{x\in G} f(x)\omega^{-r\cdot x}, \quad r \in G.$$

Fourier inversion and Parseval/Plancherel give

$$f(x) = \sum_{r\in G} \widehat{f}(r)\omega^{r\cdot x}, \qquad \mathbb{E}_{x\in G}|f(x)|^2 = \sum_{r\in G} |\widehat{f}(r)|^2.$$

For $f = 1_A$, the zero coefficient is the density:

$$\widehat{1_A}(0) = \frac{|A|}{|G|} =: \alpha.$$

For $f, g, h : G \to \mathbb{C}$, define

$$\Lambda(f, g, h) = \mathbb{E}_{x,y} f(x)g(x + y)h(x + 2y).$$

Fourier inversion gives

$$\Lambda(f, g, h) = \sum_{r\in G} \widehat{f}(r)\widehat{g}(-2r)\widehat{h}(r).$$

In $\mathbb{F}_3^n$, $-2r = r$, so

$$\Lambda_3(f) := \Lambda(f, f, f) = \sum_{r\in\mathbb{F}_3^n} \widehat{f}(r)^3.$$

Thus, the number of 3-APs is controlled by the Fourier coefficients.

13.3 Roth's theorem in $\mathbb{F}_3^n$

Theorem 9 (Roth in $\mathbb{F}_3^n$)**.** *There is an absolute constant $C > 0$ such that every 3-AP-free set $A \subseteq \mathbb{F}_3^n$ satisfies*

$$|A| \le C\frac{3^n}{n}.$$

13.4 Proof sketch of Theorem 9

Let $\alpha = |A|/3^n$. Since A is 3-AP-free, only the trivial progressions contribute, so

$$\Lambda_3(1_A) = \alpha/3^n.$$

But Fourier expansion gives

$$\Lambda_3(1_A) = \sum_r \widehat{1_A}(r)^3.$$

The zero frequency contributes $\widehat{1_A}(0)^3 = \alpha^3$, while Parseval/Plancherel gives

$$\sum_r |\widehat{1_A}(r)|^2 = \alpha.$$

Hence, if $3^n \geq 2\alpha^{-2}$ (if $3^n < 2\alpha^{-2}$, then $|A| < \sqrt{2}3^{n/2} = O(3^n/n)$, so the desired bound already holds in this case), the above Fourier expansion and Parseval/Plancherel imply that some $r \neq 0$ satisfies

$$|\widehat{1_A}(r)| \gtrsim \alpha^2.$$

Consider the hyperplane

$$H = \{x \in \mathbb{F}_3^n : r \cdot x = 0\}.$$

Its three cosets have densities $\alpha_0, \alpha_1, \alpha_2$ satisfying

$$\frac{\alpha_0 + \alpha_1 + \alpha_2}{3} = \alpha.$$

The large Fourier coefficient implies that the densities are not all close to α, and hence one coset has density

$$\alpha' \geq \alpha + c\alpha^2,$$

for some absolute $c > 0$. Restricting A to this denser coset (an affine copy of $\mathbb{F}_3^{n-1}$), the argument can be repeated.

If α_i denotes the density after the i-th iteration and $n_i = n - i$ denotes the dimension of the current affine space at that stage, then

$$\alpha_{i+1} \geq \alpha_i + c\alpha_i^2.$$

Consequently,

$$\frac{1}{\alpha_{i+1}} \leq \frac{1}{\alpha_i} - c'$$

for some absolute constant $c' > 0$. Suppose, for contradiction, that $n > K/\alpha$, where $K > 0$ is a sufficiently large absolute constant. During the first $O(1/\alpha)$ iterations, we still have $n_i \gtrsim 1/\alpha$. Since $\alpha_i \geq \alpha$, it follows that $3^{n_i} \geq 2\alpha_i^{-2}$ for K sufficiently large, so the Fourier argument continues to apply at every stage. Iterating for $O(1/\alpha)$ steps gives a contradiction. Hence $n = O(1/\alpha)$, and since $|A| = \alpha 3^n$, we finally obtain

$$n = O\left(\frac{1}{\alpha}\right) \Rightarrow \alpha = O\left(\frac{1}{n}\right) \Rightarrow |A| = O\left(\frac{3^n}{n}\right).$$

□

13.5 Fourier analysis in the integers

For a finitely supported function $f : \mathbb{Z} \to \mathbb{C}$, define its Fourier transform $\widehat{f} : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ by

$$\widehat{f}(\theta) := \sum_{x \in \mathbb{Z}} f(x) e^{-2\pi i x\theta}, \qquad \theta \in \mathbb{R}/\mathbb{Z}.$$

Fourier inversion and Parseval/Plancherel become

$$f(x) = \int_0^1 \widehat{f}(\theta) e^{2\pi i x\theta} d\theta$$

and

$$\sum_{x \in \mathbb{Z}} |f(x)|^2 = \int_0^1 |\widehat{f}(\theta)|^2 d\theta.$$

The 3-AP counting identity is

$$\Lambda(f, g, h) = \int_0^1 \widehat{f}(\theta)\widehat{g}(-2\theta)\widehat{h}(\theta) d\theta.$$

Thus, the same Fourier principle applies to subsets of the integers.

13.6 Roth's theorem in $\mathbb{Z}$

Theorem 10 (Roth in $\mathbb{Z}$)**.** *Every 3-AP-free subset $A \subseteq [N]$ satisfies*

$$|A| = O\left(\frac{N}{\log\log N}\right).$$

13.7 Proof sketch of Theorem 10

Let $|A| = \alpha N$. Define the de-meaned function

$$f(x) = 1_A(x) - \alpha 1_{[N]}(x).$$

The finite-field argument already shows the main principle: if all nonzero Fourier coefficients were small, then A would have to contain many 3-APs. Since A is 3-AP-free, this cannot happen. A standard Fourier counting estimate therefore yields some $\theta \neq 0$ such that

$$\left|\sum_{x=1}^{N}(1_A - \alpha)(x)e^{2\pi i x\theta}\right| \gtrsim \alpha^2 N.$$

The difference from $\mathbb{F}_3^n$ is that there are no useful hyperplanes in $\mathbb{Z}$. We replace them by long arithmetic subprogressions. Dirichlet's approximation lemma gives, for every $0 < \delta < 1$, there exists an integer $1 \leq d \leq 1/\delta$ such that

$$\|d\theta\|_{\mathbb{R}/\mathbb{Z}} \leq \delta,$$

where $\|d\theta\|_{\mathbb{R}/\mathbb{Z}}$ denotes the distance from $d\theta$ to the nearest integer. On a subprogression $P = \{a, a+d, \ldots, a+(L-1)d\}$, we have

$$e^{2\pi i(a+jd)\theta} = e^{2\pi i a\theta}e^{2\pi i jd\theta}.$$

Thus, when $d\theta$ is close to an integer, the Fourier phase varies slowly on P. Choosing the parameters appropriately gives a subprogression with $|P| \geq N^{1/3}$ on which the phase is sufficiently stable. The large Fourier coefficient above then gives a density increment on P:

$$\frac{|A \cap P|}{|P|} \geq \alpha + c\alpha^2.$$

We can therefore iterate exactly as in the finite-field case. If N_i and α_i denote the length and density after i steps, then

$$N_{i+1} \geq N_i^{1/3} \qquad \text{and} \qquad \alpha_{i+1} \geq \alpha_i + c\alpha_i^2,$$

where $c > 0$ is some absolute constant. The density increment can occur at most $m = O(1/\alpha)$ times. Consequently, $N_m \geq N^{1/3^m}$. On the other hand, the iteration must stop when the subprogression is too short for another density increment, giving

$$N_m \leq \left(\frac{C_1}{\alpha}\right)^{C_2},$$

where $C_1, C_2 > 0$ are some absolute constants. Hence,

$$N^{1/3^m} \leq \left(\frac{C_1}{\alpha}\right)^{C_2}.$$

Taking logarithms and using $m = O(1/\alpha)$ gives

$$\log N \leq e^{O\left(\frac{1}{\alpha}\right)} \Rightarrow \log\log N = O\left(\frac{1}{\alpha}\right).$$

Since $|A| = \alpha N$, we finally obtain

$$\alpha = O\left(\frac{1}{\log\log N}\right) \Rightarrow |A| = O\left(\frac{N}{\log\log N}\right).$$

□

13.8 Concluding remarks

The finite-field and integer proofs thus have the same structure. In $\mathbb{F}_3^n$, a large Fourier coefficient produces a density increment on a hyperplane; in $\mathbb{Z}$, Dirichlet's approximation lemma produces a density increment on a long arithmetic subprogression. The subsequent iteration is the same in both cases.

References

[Z] Zhao, Y., *Graph Theory and Additive Combinatorics.* Cambridge University Press, 2023. Available at `https://yufeizhao.com/gtacbook/`.

Stefanos Lappas, Universität Bonn
email: `lappas@math.uni-bonn.de`

14 The Möbius function is strongly orthogonal to nilsequences

After B. Green and T. Tao [GT1]

A summary written by Tiago Moreira

Abstract

We explain the proof of the orthogonality of the Möbius function to nilsequences. The quantitative equidistribution theory of polynomial orbits on nilmanifolds is used as a black box. We also discuss the Heisenberg nilmanifold as an example illustrating the general theory.

14.1 Introduction

The Möbius function $\mu : \mathbb{N} \to \{-1, 0, 1\}$ defined by

$$\mu(n) = \begin{cases} 1, & \text{if } n = 1, \\ (-1)^k, & \text{if } n \text{ is the product of } k \text{ distinct primes}, \\ 0, & \text{if } n \text{ is divisible by the square of a prime.} \end{cases}$$

is one of the basic arithmetic functions in Number Theory. A central question is whether $\mu(n)$ behaves randomly when correlated with structured sequences. A major result of Green and Tao shows that the Möbius function is orthogonal to nilsequences. Nilsequences arise from polynomial orbits on nilmanifolds and generalize classical polynomial phases such as

$$e(\alpha n), \text{ where } e(t) = e^{2\pi i t} \text{ and } \alpha \in \mathbb{R}.$$

The main theorem prescribed in [[GT1], Theorem 1.1] goes as follows:

Theorem 1. *Fix G/Γ to be a nilmanifold. Let $g : \mathbb{Z} \to G$ be a polynomial sequence and $F : G/\Gamma \to \mathbb{R}$ be a Lipschitz function. Then for every $A > 0$,*

$$\frac{1}{N} \sum_{n \leq N} \mu(n) F(g(n)\Gamma) = O_{F,G,\Gamma,A}((\log N)^{-A}).$$

One of the deep parts of the proof relies on the quantitative equidistribution theory of polynomial orbits on nilmanifolds developed in [GT2]. Since this theory is beyond the scope of this summary, we only state the factorization theorem, its relation to quantitative equidistribution, and explain how it is used in the proof. The Heisenberg example will then illustrate the main ideas in a concrete non-abelian setting.

14.2 Nilmanifolds and nilsequences

A Lie group G is called *nilpotent* if its lower central series

$$G_1 = G, \qquad G_{i+1} = [G_i, G], \text{ for } i \geq 1,$$

eventually becomes trivial. Here, the commutator is defined by

$$[g, h] = g^{-1}h^{-1}gh.$$

The simplest examples are abelian groups. For instance, $\mathbb{R}^d$ is nilpotent of step one since all commutators vanish.

A subgroup Γ of a Lie group G is called a *lattice* if it is discrete and the quotient G/Γ is compact. A quotient G/Γ with G nilpotent and Γ a lattice is called a *nilmanifold*. The torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ is the easiest example of a nilmanifold.

A polynomial sequence on a nilpotent group is a sequence $g : \mathbb{Z} \to G$ whose repeated discrete derivatives eventually vanish. The corresponding sequence $(F(g(n)\Gamma))_{n\in\mathbb{Z}}$ is called a *nilsequence*, for F a continuous function.

The theorem therefore states that the Möbius function has negligible correlation with all possible nilsequences on every nilmanifold.

14.3 Factorization of polynomial orbits

The main structural result used in the proof is the factorization theorem for polynomial sequences described in [[GT2], Theorem 1.19].

Theorem 2 (Factorization theorem)**.** *A polynomial sequence can be decomposed in the form*

$$g(n) = \varepsilon(n)g'(n)\gamma(n),$$

where:

- $\varepsilon(n)$ *is a smooth sequence that changes very slowly;*
- $\gamma(n)$ *is periodic;*
- $g'(n)\Gamma$ *is quantitatively equidistributed in a subnilmanifold.*

The proof of this result uses primarily the quantitative equidistribution theory for polynomial orbits on nilmanifolds previously mentioned. Roughly speaking, this theory states that the only obstruction to equidistribution is the existence of a non-trivial horizontal character, which makes the orbit behave almost periodically.

The inherit properties of the above decomposition are part of the key ingredients in the proof of the main theorem, which allow us to reduce the study of general polynomial orbits to the equidistributed case. The idea is that the first two components contain merely simple structures, while the third component contains the random-like behavior that appears under averaging. Hence, it is sufficient to only study the equidistributed component.

14.4 Proof of Theorem 1

We consider the sum

$$S(N) = \sum_{n \leq N} \mu(n) F(g(n)\Gamma).$$

Using Theorem 2, we can write

$$g(n) = \varepsilon(n) g'(n) \gamma(n).$$

The periodic component $\gamma(n)$ takes only finitely many values. For this reason, after splitting the sum into arithmetic progressions, we may assume that $\gamma(n)$ is constant. The smooth component $\varepsilon(n)$ changes slowly. Since F is Lipschitz, replacing $\varepsilon(n)$ by a fixed element produces only a small error. By these observations, the problem reduces to estimating sums of the form

$$\sum_{n \leq N} \mu(n) F_1(g'(n)\Gamma),$$

where $g'(n)\Gamma$ is quantitatively equidistributed and F_1 is another Lipschitz function on the same nilmanifold. The equidistribution property allows one to study the function F_1 through harmonic analysis tools on the nilmanifold. The zero Fourier coefficient gives an average that involves only the Möbius function, which can be handled using a stronger version of the Prime Number Theorem. The remaining Fourier terms lead to oscillatory sums of the form

$$\sum_{n \leq N} \mu(n) e(P(n)), \text{ with } P \text{ a real-valued polynomial.}$$

The estimates for these sums (generalized Vinogradov???s method) imply that for every $A > 0$,

$$S(N) = O_{F,G,\Gamma,A}(N(\log N)^{-A}).$$

To prove the theorem, divide by N on both sides, giving the desired estimate

$$\frac{1}{N}\sum_{n\leq N}\mu(n)F(g(n)\Gamma) = O_{F,G,\Gamma,A}((\log N)^{-A}).$$

14.5 The Heisenberg example

The Heisenberg nilmanifold is the simplest case where the non-commutative structure of a nilmanifold appears. Consider the Heisenberg group

$$G = \left\{ (x,y,z) = \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x,y,z \in \mathbb{R} \right\}.$$

The multiplication law for this particular example is given by

$$(x,y,z)\cdot(x',y',z') = (x+x', y+y', z+z'+xy').$$

The additional term xy' is responsible for the failure of commutativity. In fact, the commutator is

$$[(x,y,z),(x',y',z')] = (0,0,xy'-x'y).$$

Note that all commutators belong to the subgroup $\{(0,0,z) : z \in \mathbb{R}\}$, which is the center of the group. Thus, the Heisenberg group is two-step nilpotent.

We take the lattice

$$\Gamma = \{(a,b,c) : a,b,c \in \mathbb{Z}\},$$

and define the Heisenberg nilmanifold $X = G/\Gamma$. The horizontal factor is obtained by quotienting out the commutator subgroup: $G/(G_2\Gamma) \simeq (G/G_2)/((G_2\Gamma)/G_2) \simeq \mathbb{T}^2$. The corresponding projection is

$$\pi : X \to \mathbb{T}^2, \qquad \pi(x,y,z) = (x,y) \pmod{\mathbb{Z}^2}.$$

Following the exposition presented in Section 5 of [GT2], our focus now shifts to the equidistribution of linear orbits $g(n) = a^n$, where $a \in G$. We first fix $a = (\alpha,\beta,\gamma)$, and notice that an inductive application of the group law gives us

$$g(n) = \left(n\alpha, n\beta, n\gamma + \binom{n}{2}\alpha\beta \right).$$

Indeed, the quadratic term appears from the non-commutative part of the group multiplication, since each multiplication contributes a term $\alpha\beta$ to the central coordinate. The question now is whether the polynomial orbit prescribed by $g(n)\Gamma$ is equidistributed in X. If the orbit is equidistributed in the nilmanifold, then its projection must be equidistributed in the torus.

Therefore, a failure of equidistribution of the torus projection gives an obstruction. A horizontal character is a homomorphism $\eta : G \to \mathbb{R}/\mathbb{Z}$ which is trivial on G_2. Hence, it only depends on the tuple (x, y) and has the form

$$\eta(x, y, z) = mx + \ell y \pmod 1, \quad \text{where } (m, \ell) \in \mathbb{Z}^2.$$

Applying this to the orbit gives the following expression:

$$\eta(g(n)) = n(m\alpha + \ell\beta) \pmod 1.$$

For the torus, we have the classical Weyl's criterion that states: $(x_n)_{n\in\mathbb{N}} \in \mathbb{T}^2$ is equidistributed if and only if for every non-zero character χ, we have

$$\frac{1}{N}\sum_{n\leq N} e(\chi(x_n)) \xrightarrow{N\to+\infty} 0.$$

If $m\alpha+\ell\beta$ is close to a rational number with small denominator, the character changes slowly along the orbit. Thus, the sequence cannot be equidistributed.

This illustrates the general principle introduced when referring to the proof of Theorem 2: failure of equidistribution of polynomial orbits is caused by algebraic obstructions described by non-trivial horizontal characters.

Finally, the Heisenberg example also explains why polynomial phases naturally appear when studying nilsequences in the context of Theorem 1. For a sufficiently regular function $F : X \to \mathbb{R}$, we can use its Fourier expansion in the central coordinate:

$$F(x, y, z) = \sum_{k\in\mathbb{Z}} F_k(x, y)e(kz), \text{ with } F_{-k}(x, y) = \overline{F_k(x, y)}.$$

Applying it to the orbit $g(n)$ produces polynomial phases. These oscillatory sums are then controlled by estimates for M??bius-weighted exponential sums, which provide the cancellation required for the claimed orthogonality.

References

[GT1] B. Green and T. Tao, *The Möbius function is strongly orthogonal to nilsequences*, Annals of Mathematics 175 (2012), no. 2, 541–566.

[GT2] B. Green and T. Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Annals of Mathematics 175 (2012), no. 2, 465–540.

Tiago Moreira, SISSA
email: s59tdias@uni-bonn.de

15 The nonabelian U^2 inverse theorem

After W.T. Gowers and O. Hatami [1]

A summary written by Oissin O'Donnell

Abstract

We state the U^2 inverse theorem for matrix-valued functions on finite groups and nonabelian Fourier analysis necessary for its proof, contrasting these with Fourier analysis on finite abelian groups and the abelian U^2 inverse theorem.

15.1 Introduction

If G is a finite abelian group and $f : G \to \mathbb{C}$, the U^2 norm of f is given by

$$\|f\|_{U^2}^4 = \mathbf{E}_{x,a,b} f(x)\overline{f(x+a)f(x+b)}f(x+a+b).$$

This can be expressed in terms of the Fourier transform as

$$\|f\|_{U^2}^4 = \sum_{\chi \in \widehat{G}} |\hat{f}|^4.$$

Combining this formula for the U^2 norm with Plancherel's theorem we can easily obtain a *inverse theorem* for the U^2 norm: if $|f(x)| \leq 1$ for all $x \in G$ (or in fact if $\|f\|_{L^2} \leq 1$) and $\|f\|_{U^2}$ is large then f must correlate significantly with a character of G.
We may (simultaneously) generalise the U^2 norm to matrix-valued functions on nonabelian groups, where a convenient definition is

$$\|f\|_{U^2}^4 = \mathbf{E}_{xy^{-1}zw^{-1}=e} \operatorname{tr}(f(x)f(y)^* f(z)f(w)^*).$$

In [1] Gowers and Hatami prove an inverse theorem for the matrix valued U^2 norm.

Theorem 1 ([1] Corollary 5.6)**.** *Let G be a finite group, $\theta > 0$ and $f : G \to M_n(\mathbb{C})$ such that $\|f(x)\|_{\mathrm{op}} \leq 1$ for all $x \in G$ and $\|f\|_{U^2}^4 \geq \theta n$. Then there exists $\frac{\theta n}{2-\theta} \leq d \leq \frac{(2-\theta)n}{\theta}$, a unitary representation $\rho : G \to \mathbf{U}(d)$ and $n \times d$ partial unitary matrices U, V such that*

$$\langle f, V\rho U^* \rangle := \mathbf{E}_{x \in G} \operatorname{tr}(f(x)U\rho(x)^* V^*) \geq \tau^4 d$$

where $\tau = \max((\frac{\theta}{2})^{\frac{1}{2}}, (\frac{\theta}{2-\theta})^2)$.

Here we say an $n \times d$ matrix is *partial unitary* if it can be extended to a unitary matrix, i.e. if $n \leq d$ it has orthonormal rows, and if $n \geq d$ it has orthonormal columns. Note that the matrices U and V are necessary in Theorem 1, intuitively G may have too few representations for f to correlate with a representation, but we can always obtain correlation with "part of" a representation.

15.2 Preliminaries

15.2.1 Matrix norms

If A is an $m \times n$ matrix with singular values $\lambda_1, ..., \lambda_{\min(m,n)}$ and $1 < p < \infty$ the *Schatten p-norm* of A is given by

$$\|A\|_p^p = \sum_{i=1}^{\min(m,n)} \lambda_i^p = \operatorname{tr}(|A|^p).$$

We also let $\|A\|_\infty = \|A\|_{\text{op}}$ be the operator norm of A. The Schatten norms have the following properties

Lemma 2. *For all matrices A we have:*

1. $\|A\|_p = \|A^*\|_p$.
2. *(Unitary invariance) If U, V are unitary, then $\|UAV\|_p = \|A\|_p$.*
3. *More generally, $\|BAC\|_p \leq \|B\|_{\text{op}} \|A\|_p \|C\|_{\text{op}}$, wherever B and C are matrices such that BAC is defined.*
4. *(Hölder's inequality) If $\frac{1}{p} + \frac{1}{q} = 1$ then $|\operatorname{tr}(AB^*)| \leq \|A\|_p \|B\|_q$, in particular $|\operatorname{tr} AB^*| \leq \|A\|_{\text{op}} \|B\|_1$.*
5. *(Hilbert-Schmidt norm)*

$$\|A\|_2^2 = \operatorname{tr}(AA^*) = \sum_{i,j} |A_{ij}|^2.$$

6. *$\|\cdot\|_4$ is the "box norm" given by*

$$\|A\|_4^4 = \|A\|_\square^4 := \sum_{i,j,k,l} A_{ik} \overline{A_{il} A_{jk}} A_{jl}.$$

If G is a finite group we define an inner product on functions $f, g : G \to M_n(\mathbb{C})$ by

$$\langle f, g \rangle = \mathbf{E}_{x \in G} \operatorname{tr}(f(x)g(x)^*), \tag{1}$$

and the L^p norm of a matrix-valued function by

$$\|f\|_p = \begin{cases} \left(\mathbf{E}_{x \in G} \|f(x)\|_p^p\right)^{\frac{1}{p}} & \text{if } 1 \leq p < \infty, \\ \sup_{x \in G} \|f(x)\|_{\text{op}} & \text{if } p = \infty. \end{cases}$$

15.2.2 The Fourier transform

If G is a finite, not necessarily abelian group, let $\widehat{G}$ denote the set of equivalence classes of irreducible unitary representations of G. Abusing notation slightly, we fix a representative $\rho : G \to \mathbf{U}(\dim(\rho))$ for each $\rho \in \widehat{G}$. The Fourier transform of a function $f : G \to M_n(\mathbb{C})$ is then given by

$$\hat{f}(\rho) = \mathbf{E}_{x \in G} f(x) \otimes \overline{\rho(x)}. \tag{2}$$

Note that $\hat{f}(\rho)$ is an $nd_\rho \times nd_\rho$ matrix where $d_\rho = \dim(\rho)$. It is also sometimes useful to view it as an operator acting on $n \times d_\rho$ matrices by

$$(\hat{f}(\rho))(A) = \mathbf{E}_x f(x) A \rho(x)^*, \tag{3}$$

explaining the choice of $\overline{\rho(x)}$ rather than $\rho(x)$ or $\rho(x)^*$ in equation 2. Similar to in the abelian, scalar-valued case we have an inversion formula for the Fourier transform, although we do not state it here. The properties we need of the Fourier transform are as follows.

Lemma 3. *Let G be a finite group and $f, g : G \to M_n(\mathbb{C})$.*

1. *(Plancherel theorem)*

$$\|f\|_{L^2}^4 = \sum_{\rho \in \widehat{G}} \dim(\rho) \|\hat{f}(\rho)\|_2^2,$$

and more generally

$$\langle f,\, g \rangle = \sum_{\rho} \dim(\rho) \operatorname{tr}(\hat{f}(\rho)\hat{g}(\rho)^*).$$

2. *(U^2 norm)*

$$\|f\|_{U^2}^4 = \sum_{\rho} \dim(\rho) \|\hat{f}(\rho)\|_4^4.$$

15.3 Remarks on the proof of the inverse theorem

Using only Lemma 3 we can easily prove a nonabelian U^2 inverse theorem for scalar-valued functions. Note that for scalar-valued functions we only need an L^2 rather than a pointwise bound on f.

Theorem 4. *Let G be a finite group, $0 < \theta \leq 1$ and $f : G \to \mathbb{C}$ such that $\|f\|_{L^2} \leq 1$ and $\|f\|_{U^2}^4 \geq \theta$. Then there exists $d \leq 1/\theta$, an irreducible unitary representation $\rho : G \to \mathbf{U}(d)$ and unit vectors $u, v \in \mathbb{C}^d$ such that*

$$\langle f, \overline{u}^\intercal A v \rangle \geq \theta^{\frac{1}{2}}.$$

Proof. (Sketch) Note that by our assumptions and Lemma 3 we have

$$\theta \leq \sum_\rho \dim(\rho) \|\hat{f}(\rho)\|_4^4 \leq \sum_\rho \dim(\rho) \|\hat{f}(\rho)\|_{\text{op}}^4 \|\hat{f}(\rho)\|_2^2 \leq \sup_\rho \|\hat{f}(\rho)\|_{\text{op}}^2.$$

Hence there exists some representation ρ_0 such that $\hat{f}(\rho_0)$ has a singular value λ of size at least $\theta^{\frac{1}{2}}$. Note that, applying Lemma 3 again, $\dim(\rho_0)\theta \leq \|f\|_{L^2}^2 \leq 1$. Singular value decomposition of $\hat{f}(\rho)$ gives us unit vectors $u, v \in \mathbb{C}^d$ such that

$$\lambda = \overline{u}^\intercal \hat{f}(\rho) v = \mathbf{E}_x f(x) \overline{u}^\intercal \overline{\rho(x)} v.$$

Replacing u and v with their complex conjugates gives the result. □

For Theorem 1 using an approach similar to the proof of Theorem 4 selecting all the singular values of the $\hat{f}_\rho$ of size at least $(\frac{\theta}{2})^{\frac{1}{2}}$ and the interpretation of the Fourier transform in equat ion 3 supplies us with singular values $\lambda_1, ..., \lambda_m$, irreducible unitary representations $\rho_1, ..., \rho_m$ and $n \times \dim(\rho_i)$ matrices $U_1, ..., U_m$, $V_1, ..., V_m$ such that:

1. If $d = \dim(\rho_1) + ... + \dim(\rho_m)$ then

$$\frac{\theta n}{2 - \theta} \leq d \leq \frac{(2 - \theta) n}{\theta}.$$

2. $\|U_i\|_2^2 = \|V_i\|_2^2 = \dim(\rho_i)$ and $\text{tr}(U_i U_j^*) = \text{tr}(V_i V_j^*) = 0$ whenever $i \neq j$ but $\rho_i = \rho_j$.

3. If $U = (U_1|...|U_n)$ is the $n \times d$ matrix given by concatenating the U_is and similarly $V = (V_1|...|V_n)$, then if $\rho = \rho_1 \oplus ... \oplus \rho_m$

$$\langle f, V \rho U^* \rangle \geq \tau d.$$

The difficulty then is in modifying the matrices U and V so that $\|U\|_{\mathrm{op}} = \|V\|_{\mathrm{op}} = 1$ while maintaining a lower bound on $\langle f, V\rho U^* \rangle$. In [1] this is done by using the (finite-dimensional) Hahn-Banach theorem and the orthogonality of matrix coefficients of irreducible unitary representations (see [2] for an alternative, operator-algebraic approach). Once we have this, using the fact that any matrix A with $\|A\|_{\mathrm{op}} \leq 1$ is a convex combination of partial unitary matrices we can in fact replace U and V by partial unitary matrices, proving Theorem 1.

References

[1] Gowers, W.T. and Hatami, O., *Inverse and stability theorems for approximate representations of finite groups* Sbornik: Mathematics **208** (2017), no.12, 1784–1817;

[2] De Chiffre, M., Ozawa, N. and Thom, A., *Operator algebraic approach to inverse and stability theorems for amenable groups* Mathematika, **65** (2019), no.1, 98–118.

Oissin O'Donnell, Cambridge
email: oissin1994@gmail.com

16 Effective bounds for restricted 3-APs in $\mathbb{F}_3^n$

After A. Bhangale, S. Khot, Y. P. Liu, D. Minzer [BKLMa]

A summary written by Anthony Ostuni

Abstract

We show that any set $A \subseteq \mathbb{F}_3^n$ of density $\gg (\log\log\log n)^{-c}$ contains a 3-AP whose common difference lies in $\{0,1\}^n \setminus \{0^n\}$.

16.1 Introduction

We have already seen that any sufficiently dense set $A \subseteq \mathbb{F}_3^n$ contains a 3-AP. These proofs, however, tell us nothing about the common difference that occurs, and it is natural to ask about what happens if we insist the difference falls into a prescribed subset of $\mathbb{F}_3^n$, say $\{0,1\}^n \setminus \{0^n\}$. Deep results such as the density Hales-Jewett (DHJ) theorem imply that A also contains such a *restricted 3-AP*, but they historically have come with poor quantitative bounds. In this work, we discuss effective bounds for this problem.

Theorem 1. *There are universal constants $C, c > 0$ such that any set $A \subseteq \mathbb{F}_3^n$ of density at least $C(\log\log\log n)^{-c}$ contains a restricted 3-AP.*

Remark 2. *The result in [BKLMa, Theorem 2] is a bit more general, but proving the simpler form above suffices to demonstrate the main ideas.*

Recent work. The authors further developed their machinery in [BKLMb] to prove "reasonable" upper bounds for the DHJ theorem over $[3]^n$. Very recently, they obtained a counting version of Theorem 1 (over $\mathbb{F}_p^n$) where the common difference lies in $\{0,1,2\}^n$ [BKLMc]. Another recent work tweaks the capset proof to give polynomial bounds for subsets of $\mathbb{F}_p^n$ avoiding 3-APs whose common difference lies in some sufficiently large set [CFP].

Notation. Throughout, $A \subseteq \mathbb{F}_3^n$ is a set with density $\alpha \gg (\log\log\log n)^{-c}$ and no restricted 3-APs. In an abuse of notation, we also use $A(\cdot)$ to denote its own indicator function. We use μ to denote the uniform distribution over $\{(a,a,a) : a \in \mathbb{F}_3\} \cup \{(a,a+1,a+2) : a \in \mathbb{F}_3\}$ (i.e., valid coordinate patterns of a restricted 3-AP), and $\mu^{\otimes n}$ to denote the distribution over $(\mathbb{F}_3^n)^3$,

where each coordinate pattern is independently sampled from μ. We write $I \sim_{1-\delta} [n]$ for a random set $I \subseteq \{1, 2, \ldots, n\}$, where each $i \in [n]$ is included independently with probability $1 - \delta$. We use $\mathbb{D}$ for the closed unit disk $\{c \in \mathbb{C} : |c| \leq 1\}$.

16.2 Proof overview

In this section, we provide an overview of the proof of Theorem 1. The high-level idea is to perform a similar density increment argument as in the proof of Roth's theorem, although the basis-dependent constraints complicate matters. We first convert the absence of restricted 3-APs into a noticeable 3-wise correlation.

Lemma 3. *There exist functions $f, g, h : \mathbb{F}_3^n \to [-1, 1]$ such that*

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}} [f(x)g(y)h(z)] \right| \geq \frac{\alpha^3}{8},$$

where either f, g, or h is equal to the balanced indicator function $A - \alpha$.

Proof. If A has no restricted 3-APs, then

$$\mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}} [A(x)A(y)A(z)] \leq \Pr_{(x,y,z)\sim\mu^{\otimes n}} [x = y = z] = 2^{-n}, \tag{1}$$

since triples drawn from μ are constant with probability $1/2$. Expanding each A as $\alpha + (A - \alpha)$, the LHS of (1) becomes a sum of eight terms. One term is α^3, and the other seven include at least one copy of $(A - \alpha)$, which is 1-bounded. The lemma follows by rearrangement. □

We may now invoke (a special case of) one of the main technical results of [BKLMa]: a local inverse theorem for such correlations.

Theorem 4 (Local inverse theorem)**.** *For every $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) \geq \exp(-\varepsilon^{-O(1)})$ such that the following holds. For any functions $f, g, h : \mathbb{F}_3^n \to [-1, 1]$ satisfying*

$$\left| \mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}} [f(x)g(y)h(z)] \right| \geq \varepsilon,$$

we have

$$\Pr_{\substack{I\sim_{1-\delta}[n] \\ z\in\mathbb{F}_3^I}} \left[\exists \{P_i : \mathbb{F}_3 \to \mathbb{D}\}_{i\in\bar{I}} \text{ with } \left| \mathop{\mathbb{E}}_{x\sim\mathbb{F}_3^{\bar{I}}} \left[f_{I\to z}(x) \prod_{i\in\bar{I}} P_i(x_i) \right] \right| \geq \delta \right] \geq \delta.$$

In words, Theorem 4 says that if three 1-bounded functions correlate with respect to $\mu^{\otimes n}$, then after taking a severe random restriction of the ambient space, one such function correlates with a bounded product function with decent probability. We will discuss some of the ideas behind the proof of Theorem 4 in Section 16.3, but let's assume it for the moment.

Our goal is now to obtain a density increment using Theorem 4. We will require the following lemma, where we define $f_A \coloneqq A - \alpha$.

Lemma 5. *Suppose for some δ satisfying $n \gg \frac{1}{\delta}\log(\frac{1}{\delta})$, we have*

$$\Pr_{\substack{I\sim_{1-\delta}[n]\\ z\in\mathbb{F}_3^I}}\left[\exists\{P_i:\mathbb{F}_3\to\mathbb{D}\}_{i\in\bar{I}} \text{ with } \left|\mathop{\mathbb{E}}_{x\sim\mathbb{F}_3^{\bar{I}}}\left[(f_A)_{I\to z}(x)\prod_{i\in\bar{I}}P_i(x_i)\right]\right|\geq\delta\right]\geq\delta.$$

Then $\exists I\subseteq[n]$ of size $|I|\leq(1-\frac{\delta}{2})n$ and $z\in\mathbb{F}_3^I$ such that either

1. *The density of $A_{I\to z}$ in $\mathbb{F}_3^{\bar{I}}$ is $\geq\alpha+\Omega(\delta^3)$, or*
2. *The density of $A_{I\to z}$ in $\mathbb{F}_3^{\bar{I}}$ is $\geq\alpha-O(\delta^2)$ and there exists a 1-bounded product function $P=P_1\cdots P_{\bar{I}}$ such that $|\mathbb{E}_{x\sim\mathbb{F}_3^{\bar{I}}}[(f_A)_{I\to z}(x)P(x)]|\geq\delta$.*

Proof. Let $\mathcal{E}$ be the event that I has size $|I|\leq(1-\frac{\delta}{2})n$ and there exists P such that $|\langle(f_A)_{I\to z},P\rangle|\geq\delta$. By Chernoff's inequality and a union bound, we have $\Pr[\mathcal{E}]\geq\delta/2$. Let α_R denote the density of A under a restriction R. Assume for all restrictions $R\in\mathcal{E}$ that $\alpha_R-\alpha<-\frac{\delta^2}{4}$, as otherwise we have already satisfied conclusion 2. Then,

$$\mathop{\mathbb{E}}_R[(\alpha_R-\alpha)\cdot\mathbb{1}_{\mathcal{E}}]\leq-\frac{\delta^2}{4}\Pr[\mathcal{E}]\leq-\frac{\delta^3}{8}.$$

Since the random restrictions are chosen uniformly, we have $\mathbb{E}_R[(\alpha_R-\alpha)]=0$, so it must be that $\mathbb{E}_R[(\alpha_R-\alpha)\cdot\mathbb{1}_{\mathcal{E}^c}]\geq\frac{\delta^3}{8}$. Thus (again appealing to Chernoff's inequality to ensure the probability of I being too large is much smaller than δ^3) there must exist a restriction R satisfying conclusion 1. □

We can finally sketch the proof of Theorem 1.

Proof sketch of Theorem 1. We will perform a density increment argument. At each step, the density α and the ambient dimension n will change according to

$$\alpha\mapsto\alpha+\exp(-\alpha^{-O(1)})\quad\text{and}\quad n\mapsto(\exp(-\alpha^{-O(1)})\cdot n)^{O(1)}.$$

There can be $\ll \exp(\alpha^{-O(1)})$ iterations before the density exceeds 1, so we can carry out the argument to completion as long as $\alpha \gg (\log\log\log n)^{-c}$.

We apply Lemma 3, Theorem 4, and Lemma 5 with $\delta = \exp(-\alpha^{-O(1)})$. We assume conclusion 2 holds, as otherwise we immediately get the desired density increment. The final step is to perform a different, more technical, restriction. The basic idea is as follows. First randomly choose $n^{O(1)}$ disjoint groups of $n^{O(1)}$ coordinates, and insist that all the coordinates within a group take the same value. Then, randomly fix the values of all coordinates not included in one of the groups. We want that for each group T, the product $\prod_{t\in T} P_t$ should be almost constant, and that the density of A does not drop by much under this restriction. (Morally, this means that $\mathbb{E}[(A-\alpha)_{I\to z}] \approx \mathbb{E}[(f_A)_{I\to z}\cdot P] \geq \delta$.) With enough technical work, one can show that some such restriction does satisfy these properties, giving the desired density increment. We now iterate on the new function, where each group of identified coordinates is considered to be a single coordinate. □

16.3 Sketch of inverse theorem

In this section, we present some of the main ideas underlying the proof of Theorem 4. Much of the analysis centers around the following analog of the U^2-Gowers uniformity norm designed to capture product structure.

Definition 6 (Swap norm)**.** *Let $x, y \in \mathbb{F}_3^n$. We write $(x', y') \sim (x \leftrightarrow y)$ to denote the distribution over $\mathbb{F}_3^n \times \mathbb{F}_3^n$, where independently for each index $i \in [n]$, we set (x'_i, y'_i) to be either (x_i, y_i) or (y_i, x_i) with equal probability.*

For a function $f : \mathbb{F}_3^n \to \mathbb{C}$, we define

$$\mathsf{swap}(f) = \mathbb{E}_{x,y\in\mathbb{F}_3^n (x',y')\leftrightarrow(x,y)}[f(x)f(y)\overline{f(x')f(y')}].$$

The swap norm of f *is* $\|f\|_{\mathsf{sw}} \coloneqq \mathsf{swap}(f)^{1/4}$.

As in Gowers' analysis, one can convert some arithmetic structure into a noticeable swap norm essentially via iterated applications of the Cauchy-Schwarz inequality (although the applications are a bit subtler in this case).

Lemma 7. *If there exist functions $f, g, h : \mathbb{F}_3^n \to [-1, 1]$ such that*

$$\left|\mathop{\mathbb{E}}_{(x,y,z)\sim\mu^{\otimes n}}[f(x)g(y)h(z)]\right| \geq \varepsilon,$$

then $\mathbb{E}_{\substack{I\sim_{1-\Omega(1)}[n]\\ z\sim\mathbb{F}_3^I}}[\mathsf{swap}(f_{I\to z})] \geq \varepsilon^{O(1)}$.

Once equipped with Lemma 7, Theorem 4 follows from an inverse theorem for the swap norm. We unfortunately do not have space to sketch the full proof of this, but we can show such a theorem in the "99% regime", where there is no need to randomly restrict to correlate with a product function.

Lemma 8. *Let $f : \mathbb{F}_3^n \to \mathbb{C}$ satisfy $\|f\|_2 \leq 1$ and $\mathsf{swap}(f) \geq 1 - c \geq 0.95$, then there are functions $P_1, \ldots, P_n : \mathbb{F}_3 \to \mathbb{C}$ with $\|P_i\|_2 = 1$ such that $|\langle f, P_1 \cdots P_n \rangle| \geq 1 - O(c)$.*

Remark 9. *The above P_i have bounded L_2-norm, whereas in Theorem 4 they have bounded L_∞-norm. With a brief foray into the analysis of exponential sums, one can move between these guarantees at the cost of decreasing the correlation polynomially. The details can be found in [BKLMa, Section 6].*

Proof sketch of Lemma 8. Decompose f, viewed as a function on $\mathbb{F}_3 \times \mathbb{F}_3^{n-1}$, according to its singular value decomposition as $f = \sum_i \lambda_i g_i h_i$, where $g_i : \mathbb{F}_3 \to \mathbb{C}$ and $h_i : \mathbb{F}_3^{n-1} \to \mathbb{C}$. Set g_i corresponding to the largest singular value λ_i (which is nearly 1) to be one of the functions in $P = g_i P'$ and iterate on h_i. By orthonormality $\langle f, P \rangle = \lambda_i \langle h_i, P' \rangle$, and one can show that $\mathsf{swap}(h_i)$ is noticeably larger than $\mathsf{swap}(f)$. □

References

[BKLMa] Bhangale, A., Khot, S., Liu, Y. P., and Minzer, D., *On Approximability of Satisfiable k-CSPs: VI.* Merged version appeared in FOCS (2025).

[BKLMb] Bhangale, A., Khot, S., Liu, Y. P., and Minzer, D., *Reasonable Bounds for Combinatorial Lines of Length Three.* Merged version appeared in FOCS (2025).

[BKLMc] Bhangale, A., Khot, S., Liu, Y. P., and Minzer, D., *A Counting Lemma for Somewhat Restricted 3-APs.* Preprint arXiv:2608.17365 (2026).

[CFP] Conlon, D., Fox, J., Pham, H. T., *A note on arithmetic progressions with restricted differences.* Preprint arXiv:2605.13628 (2026).

Anthony Ostuni, UC San Diego
email: aostuni@ucsd.edu

17 The Green–Tao theorem: an exposition

After D. Conlon, J. Fox, and Y. Zhao [CFZ]

A summary written by Felix Pernegger

17.1 Overview

An arithmetic progression of length k is a sequence

$$a,\ a+d,\ a+2d,\dots,\ a+(k-1)d$$

with common difference $d>0$. Green and Tao [GT] proved in 2004 that such progressions can be found inside the primes with no bound on their length.

Theorem 1 (Green–Tao)**.** *The set of prime numbers contains arbitrarily long arithmetic progressions.*

One might hope to apply Szemerédi's theorem directly: every subset $A\subseteq\mathbb{N}$ of positive upper density

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

contains arbitrarily long arithmetic progressions [Sz]. However, the primes have density 0. The main idea of Green–Tao is nevertheless to use Szemerédi's theorem *as a black box*: one transfers it from dense sets to sparse sets which are sufficiently pseudorandom. Roughly, an unbounded sparse weight is replaced by a bounded dense model, and a counting lemma shows that the normalized arithmetic-progression counts are asymptotically the same.

The proof has three main components. First, a *relative Szemerédi theorem* transfers the dense theorem to functions dominated by a pseudorandom measure. Second, after the W-trick removes local biases of the primes modulo small primes, a smoothed sieve weight gives a pseudorandom majorant for the modified primes. Third, one verifies the required pseudorandomness by a linear-forms estimate. Combining these facts gives a k-term progression of primes for every k.

17.2 The relative Szemerédi theorem

Work in $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$. A weighted form of Szemerédi's theorem says that for every $k \geq 3$ and $\delta > 0$ there is $c = c(k, \delta) > 0$ such that every $g : \mathbb{Z}_N \to [0, 1]$ with $\mathbb{E}g \geq \delta$ satisfies

$$\mathbb{E}_{x,d} \prod_{j=0}^{k-1} g(x + jd) \geq c - o_{k,\delta}(1). \tag{1}$$

For the primes, the natural weights are not bounded by 1, so one needs a pseudorandomness hypothesis.

For $j \in [k]$ and $\omega \in \{0, 1\}^{[k]\setminus\{j\}}$, put

$$L_{j,\omega}(x^{(0)}, x^{(1)}) = \sum_{i \neq j} (j - i) x_i^{(\omega_i)}.$$

A function $\nu : \mathbb{Z}_N \to [0, \infty)$ satisfies the *k-linear forms condition* if, for every choice $n_{j,\omega} \in \{0, 1\}$,

$$\mathbb{E} \prod_{j=1}^{k} \prod_{\omega \in \{0,1\}^{[k]\setminus\{j\}}} \nu(L_{j,\omega})^{n_{j,\omega}} = 1 + o(1). \tag{2}$$

Thus all correlations needed for counting k-term progressions behave asymptotically as for the constant measure 1.

Theorem 2 (Relative Szemerédi, Theorem 4.3 of [CFZ]). *If ν satisfies the k-linear forms condition, then for every $\delta > 0$ and every $0 \leq f \leq \nu$ with $\mathbb{E}f \geq \delta$,*

$$\mathbb{E}_{x,d} \prod_{j=0}^{k-1} f(x + jd) \geq c(k, \delta) - o_{k,\delta}(1),$$

with the same constant as in (1).

Sections 5 and 6 provide the two ingredients for the proof. For $h : \mathbb{Z}_N \to \mathbb{R}$, define

$$\|h\|_{\square,r} := \sup_{A_1,\dots,A_r \subseteq \mathbb{Z}_N^{r-1}} \left| \mathbb{E}_{x_1,\dots,x_r} h(x_1 + \cdots + x_r) \prod_{i=1}^{r} \mathbb{1}_{A_i}(x_{-i}) \right|.$$

The dense model theorem says that if $\|\nu - 1\|_{\square,k-1}$ is small, then every $0 \le f \le \nu$ has a model $\widetilde{f} : \mathbb{Z}_N \to [0,1]$ with $\|f - \widetilde{f}\|_{\square,k-1}$ small. The relative counting lemma says that, under the linear forms condition, this approximation preserves the k-AP count up to $o(1)$.

To prove Theorem 2, assume as in [CFZ] that N is coprime to $(k-1)!$. Repeated Cauchy–Schwarz gives $\|\nu - 1\|_{\square,k-1} = o(1)$, hence a dense model $\widetilde{f}$ with $\mathbb{E}\widetilde{f} \ge \delta - o(1)$. For

$$\psi_j(x_{-j}) = \sum_{i \ne j} (j-i)x_i,$$

set $x = \psi_1(x_{-1})$ and $d = x_1 + \cdots + x_k$; then $\psi_j(x_{-j}) = x + (j-1)d$. The counting lemma therefore gives

$$\mathbb{E}_{x,d} \prod_{j=0}^{k-1} f(x + jd) = \mathbb{E}_{x,d} \prod_{j=0}^{k-1} \widetilde{f}(x + jd) + o(1),$$

and (1) proves the theorem.

17.3 A pseudorandom majorant for the primes

Let $w = w(N) \to \infty$ sufficiently slowly and

$$W = \prod_{p \le w} p.$$

The W-trick restricts attention to $Wn + 1$, which avoids every prime divisor $p \le w$. Define

$$\widetilde{\Lambda}(n) = \begin{cases} \dfrac{\phi(W)}{W} \log(Wn+1), & Wn+1 \text{ prime}, \\ 0, & \text{otherwise}. \end{cases}$$

Its mean is $1 + o(1)$ for sufficiently slow growth of w.

Choose a smooth $\chi : \mathbb{R} \to [0,1]$, supported on $[-1,1]$, with $\chi(0) = 1$, and put

$$\Lambda_{\chi,R}(n) = \log R \sum_{d \mid n} \mu(d) \chi\left(\frac{\log d}{\log R}\right), \qquad c_\chi = \int_0^\infty |\chi'(x)|^2 \, dx.$$

Only $d \le R$ occur, and if n has no prime divisor at most R, then $\Lambda_{\chi,R}(n) = \log R$. Taking $R = N^{k^{-1}2^{-k-3}}$, define

$$\nu(n) = \begin{cases} \dfrac{\phi(W)}{W} \dfrac{\Lambda_{\chi,R}(Wn+1)^2}{c_\chi \log R}, & N/2 \le n < N, \\ 1, & \text{otherwise.} \end{cases} \tag{3}$$

The square makes ν nonnegative, and on primes $Wn+1$ the numerator equals $(\log R)^2$. Hence, for some $\delta_k > 0$, one has $\delta_k \widetilde{\Lambda}(n) \le \nu(n)$ on $[N/2, N)$.

The required pseudorandomness follows from Proposition 8.3 of [CFZ]: for fixed pairwise nonproportional linear forms $\psi_1, \dots, \psi_m : \mathbb{Z}^t \to \mathbb{Z}$, with $\theta_i = W\psi_i + 1$, and for product boxes B of side length at least R^{10m},

$$\mathbb{E}_{x \in B} \prod_{i=1}^{m} \Lambda_{\chi,R}(\theta_i(x))^2 = (1 + o(1)) \left(\frac{W c_\chi \log R}{\phi(W)} \right)^m. \tag{4}$$

Partitioning $\mathbb{Z}_N^t$ into slowly many large boxes and separating those crossing the boundary of $[N/2, N)$ then yields the k-linear forms condition for (3).

17.4 Outline of the linear-forms calculation

Expanding the squares in (4) produces divisor variables $d_i, d_i' \le R$ and congruence conditions $d_i, d_i' \mid \theta_i(x)$. Their least common multiple D is at most R^{2m}, so the average over B may be replaced, with negligible error, by the corresponding density modulo D. The cutoff is not multiplicative, so one applies Fourier inversion to $e^x \chi(x)$, writing

$$\chi\left(\frac{\log d}{\log R} \right) = \int_{\mathbb{R}} d^{-(1+i\xi)/\log R} \varphi(\xi)\, d\xi.$$

After this substitution the divisor weights are multiplicative, and the Chinese remainder theorem turns the sum into an Euler product. The W-trick controls the factors for $p \le w$, while for $p > w$ nonproportionality of the forms makes simultaneous congruences contribute only $O(p^{-2})$. Comparison with zeta factors near $s = 1$, using $\zeta(1+s) \sim s^{-1}$, gives $(W \log R/\phi(W))^m$, while a Plancherel/integration-by-parts identity gives c_χ^m. This proves (4).

Finally, let $f(n) = \delta_k \widetilde{\Lambda}(n)$ for $N/2 \le n < N$ and $f(n) = 0$ otherwise. The prime number theorem in the progression $1 \bmod W$ gives $\sum_{N/2 \le n < N} \widetilde{\Lambda}(n) = (1/2 + o(1))N$, so $\mathbb{E} f \ge \delta_k/3$ for large N, and $0 \le f \le \nu$. Theorem 2 gives

a positive k-AP count. The contribution from $d = 0$ is negligible, and the support $[N/2, N)$ prevents wraparound in $\mathbb{Z}_N$. Undoing the W-trick proves Theorem 1.

17.5 Extensions and conjectural generalizations

Section 10 of Conlon–Fox–Zhao [CFZ] records several extensions of the Green–Tao method. If $A \subseteq \mathbb{P}$ has positive relative upper density, then A itself contains arbitrarily long arithmetic progressions. Tao proved that the Gaussian primes contain arbitrarily shaped finite constellations, and Tao–Ziegler proved polynomial analogues for configurations $x + P_1(t), \dots, x + P_r(t)$ with $P_i(0) = 0$.

The Green-Tao theorem is qualitative: it proves that progressions of every length exist, without producing small explicit examples. As of August 2026, the longest explicitly known progression of primes has length 27 [PG]:

$$60518557631784826 1 + 155368778 \cdot 23\# \cdot n, \qquad n = 0, \dots, 26.$$

It was found by Michael Kwok through PrimeGrid in 2023; no AP28 is currently known.

A broad generalization of the Green-Tao conjectures by Erdos asserts that if $A \subseteq \mathbb{N}$ satisfies

$$\sum_{n \in A} \frac{1}{n} = \infty,$$

then A contains arithmetic progressions of every finite length (Green-Tao theorem would then follow, as $\sum_{p \in \mathbb{P}} 1/p = \infty$).

References

[CFZ] D. Conlon, J. Fox, and Y. Zhao, *The Green–Tao theorem: an exposition*, EMS Surv. Math. Sci. **1** (2014), 249–282.

[GT] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. **167** (2008), 481–547.

[Sz] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.

[PG] PrimeGrid, *AP27 records*, accessed August 2026.

18 On the Bogolyubov–Ruzsa lemma

After Tom Sanders [S12]

A summary written by Antônio Rocha-Neves

Abstract

Sanders established a quasipolynomial bound for the Bogolyubov–Ruzsa lemma, demonstrating that dense sets contain large coset progressions in their iterated sumsets. We sketch a short, self-contained proof of this result for the case of $\mathbb{F}_2^n$.

18.1 Main Theorem

The central result of our exposition is the quasipolynomial bound for Bogolyubov–Ruzsa lemma in $\mathbb{F}_2^n$, established by Sanders [S12] which we prove using some probabilistic ideas from the survey of Lovett [L15].

Theorem 1 (Main Theorem)**.** *Let $A \subset \mathbb{F}_2^n$ be a set with density $\alpha = |A|/2^n > 0$. Then the sumset $4A$ contains a vector subspace $V \leq \mathbb{F}_2^n$ of codimension $O(\log^4(2/\alpha))$.*

To locate this subspace, we combine the Croot–Sisask lemma with Chang's theorem. The Croot–Sisask lemma provides a large set of shifts T that leave the convolution $\phi_A * 1_{2A}$ approximately invariant in L^p. By applying Chang's theorem, we bound the dimension of the large Fourier spectrum of T. The orthogonal complement of the span of these frequencies defines the desired low-codimension subspace V, which is subsequently shown to lie entirely in $4A$ via Parseval's identity and the pigeonhole principle.

18.2 Preliminaries and Fourier analysis in $\mathbb{F}_2^n$

Because we work in $\mathbb{F}_2^n$, we consider only real-valued functions $f : \mathbb{F}_2^n \to \mathbb{R}$. We define the expectation $\mathbb{E}_{x\in\mathbb{F}_2^n}[f(x)] = \frac{1}{2^n}\sum_{x\in\mathbb{F}_2^n} f(x)$, and omit the subscript when the domain is $\mathbb{F}_2^n$. For a subset $S \subseteq \mathbb{F}_2^n$, the expectation is $\mathbb{E}_{x\in S}[f(x)] = \frac{1}{|S|}\sum_{x\in S} f(x)$.

If $A \subseteq \mathbb{F}_2^n$ has density $\alpha = |A|/2^n$, its normalized indicator function is $\phi_A = \alpha^{-1}1_A$. Since $\mathbb{E}[\phi_A] = 1$, it acts as a probability density, satisfying $\mathbb{E}_{x\in A}[f(x)] = \mathbb{E}[\phi_A(x)f(x)]$. We abbreviate $\phi_{\{a\}}$ as ϕ_a.

We define the inner product $\langle f, g\rangle = \mathbb{E}[f(x)g(x)]$ and the convolution $f * g(x) = \mathbb{E}_y[f(y)g(x+y)]$. When applied to density functions, these operators compute probabilities over independent uniform random variables. For example, $\langle \phi_A, f\rangle = \mathbb{E}_{x\in A}[f(x)]$ and $\langle \phi_A * \phi_B, 1_C\rangle = \mathbb{P}_{a\in A, b\in B}[a+b \in C]$.

The dual group $\widehat{\mathbb{F}_2^n}$ consists of characters $\gamma_\xi(x) = (-1)^{\langle \xi, x\rangle}$, which form an $\mathbb{F}_2$-vector space under pointwise multiplication. Evaluating k such characters simultaneously at a uniformly random $x \in \mathbb{F}_2^n$ defines a linear map to $\mathbb{F}_2^k$. If the corresponding vectors $\xi_1, \ldots, \xi_k$ are linearly independent, this map is surjective and by the Rank-Nullity theorem, every target vector has an identical preimage size, meaning the character evaluations act as mutually independent uniform random variables on $\{-1, 1\}$.

The Fourier transform is given by $\widehat{f}(\gamma) = \mathbb{E}_x[f(x)\gamma(x)] = \langle f, \gamma\rangle$. Recall that $\widehat{f * g}(\gamma) = \widehat{f}(\gamma)\widehat{g}(\gamma)$, Parseval's identity gives $\langle f, g\rangle = \sum_\gamma \widehat{f}(\gamma)\widehat{g}(\gamma)$, and that $\widehat{\phi_V} = 1_{V^\perp}$.

Alongside Markov and Hölder inequalities, we will use the following version of Khintchine's inequality:

Lemma 2 (Khintchine's inequality)**.** *Let $X_1, \ldots, X_\ell \in [-1, 1]$ be mutually independent, mean-zero random variables. There is a constant C such that for any $p \geq 1$, the p-th moment of their average is bounded by:*

$$\left(\mathbb{E}\left[\left|\frac{1}{\ell}\sum_{i=1}^{\ell} X_i\right|^p\right]\right)^{1/p} \leq C\sqrt{\frac{p}{\ell}}.$$

18.3 The Croot–Sisask lemma

The Croot–Sisask lemma shows that there is a large set of shifts T that leave the convolution $\phi_A * f$ approximately invariant.

Lemma 3 (Croot–Sisask)**.** *Let $A \subseteq \mathbb{F}_2^n$ be a set of density α, and let $f : \mathbb{F}_2^n \to [0, 1]$. For any $p \geq 1$ and $\varepsilon > 0$, there exists a set $T \subseteq \mathbb{F}_2^n$ with density $\tau \geq \frac{1}{2}\alpha^\ell$, where $\ell = \lceil 16C^2 p\varepsilon^{-2}\rceil$, such that for all $t \in T$:*

$$\|\phi_t * \phi_A * f - \phi_A * f\|_p \leq \varepsilon$$

Proof sketch. We approximate the convolution $\phi_A * f(x)$ with an empirical average over ℓ independent samples $s_1, \ldots, s_\ell \in A$ given by $F_s(x) =$

$\frac{1}{\ell}\sum_{i=1}^{\ell} f(x+s_i)$. For a fixed x, the terms $f(x+s_i) - \phi_A * f(x)$ are independent, mean-zero, and bounded by 1, so applying Khintchine's inequality gives:

$$\mathbb{E}_{s_1,\ldots,s_\ell \in A}\left[\left|\frac{1}{\ell}\sum_{i=1}^{\ell}(f(x+s_i) - \phi_A * f(x))\right|^p\right] \leq \left(C\sqrt{\frac{p}{\ell}}\right)^p.$$

Which implies $\mathbb{E}_{s_1,\ldots,s_\ell \in A}[\|F_s - \phi_A * f\|]_{L^p}^p \leq (C\sqrt{p/\ell})^p$. Setting $\ell = \lceil 16C^2 p\varepsilon^{-2}\rceil$ drives the expected error low enough that at least half of the sequences in A^ℓ yield an approximation with L^p error $\leq \varepsilon/2$. Because translating A by y shifts these good sequences into $(A+y)^\ell$, a double-counting argument over all shifts $y \in \mathbb{F}_2^n$ guarantees there must exist a sequence $s^* \in (\mathbb{F}_2^n)^\ell$ that serves as an $\varepsilon/2$-approximation for a large set of shifts X, where $|X|/2^n \geq \frac{1}{2}\alpha^\ell$.

Because this single sequence s^* serves as a shared approximation for every shift in X, for any $x, y \in X$ the triangle inequality gives $\|\phi_{A+x} * f - \phi_{A+y} * f\|_p \leq \varepsilon/2 + \varepsilon/2 = \varepsilon$. Because the L^p norm is translation-invariant, we can shift the functions by y. Fixing $y \in X$ and setting $T = X + y$, substituting $x = y + t$ yields $\|\phi_t * \phi_A * f - \phi_A * f\|_p \leq \varepsilon$ for all $t \in T$, finishing the proof. □

18.4 Chang's theorem

Chang's theorem bounds the dimension of the large Fourier spectrum.

Lemma 4 (Chang's theorem)**.** *Let $A \subseteq \mathbb{F}_2^n$ be a set of density α. For any $\varepsilon > 0$, the dimension k of the subspace spanned by the large Fourier spectrum $\mathrm{Spec}_\varepsilon(\phi_A)$ is bounded by:*

$$k \leq O(\varepsilon^{-2}\log(2/\alpha)).$$

Proof sketch. Let k be the dimension of the spectrum's span. We select $\gamma_1, \ldots, \gamma_k$ directly from $\mathrm{Spec}_\varepsilon(\phi_A)$ to form an $\mathbb{F}_2$-basis of this span. Define the function $F(x) = \sum_{i=1}^k c_i\gamma_i(x)$ where $c_i = \mathrm{sgn}(\widehat{\phi_A}(\gamma_i))$.

By definition of F, we have $\langle \phi_A, F\rangle \geq k\varepsilon$. Conversely, Hölder's inequality gives $\langle \phi_A, F\rangle \leq \|\phi_A\|_q\|F\|_p$.

Because $\phi_A = \alpha^{-1}1_A$, its L^q norm is $\alpha^{-1/p}$. Because the characters are independent in $\widehat{\mathbb{F}_2^n}$, and therefore independent as random variables, Khintchine's inequality bounds $\|F\|_p \leq C\sqrt{pk}$. Setting $p = 2\ln(2/\alpha)$ ensures that $p > 1$ and $\|\phi_A\|_q \leq \sqrt{e}$, forcing $k\varepsilon \leq C\sqrt{epk}$. Squaring both sides and absorbing the constants yields $k \leq O(\varepsilon^{-2}\log(2/\alpha))$. □

18.5 Extracting the subspace (main proof)

We now extract a subspace from $4A$ of codimension $O(\log^4(2/\alpha))$ proving the main result.

The probability that two random elements of A sum into $2A$ is 1, that is, $\langle \phi_A * \phi_A, 1_{2A} \rangle = 1$. Let $m = \lceil 3 + \log_2(2/\alpha) \rceil$. We apply Croot–Sisask to $f = 1_{2A}$ with $\varepsilon = 1/(4me)$ and $p = 1 + \ln(2/\alpha)$, obtaining a set of invariant shifts T. This choice of p bounds the norm, $\|\phi_A\|_q = \alpha^{-1/p} < e$.

By the triangle inequality, a concatenated shift $x = t_1 + \cdots + t_m$ incurs an L^p error of at most $m\varepsilon$. Hölder's inequality translates this directly into a bounded probability drop:

$$\langle \phi_A, \phi_A * 1_{2A} - \phi_x * \phi_A * 1_{2A} \rangle \leq \|\phi_A\|_q (m\varepsilon) \leq \frac{1}{4}.$$

Thus, drawing a_1, a_2 uniformly from A, and $t_1, \ldots, t_m$ uniformly from T (all mutually independent), the probability remains high:

$$\mathbb{P}_{a_1,a_2 \in A, t_i \in T}[a_1 + a_2 + t_1 + \cdots + t_m \in 2A] \geq \frac{3}{4}.$$

Let $V = \mathrm{Spec}_{1/2}(\phi_T)^\perp$. We evaluate the probability drop when adding a uniform random shift $v \in V$. By Parseval's identity, the probability is the Fourier sum:

$$\begin{aligned} &\mathbb{P}_{a_1,a_2 \in A, t_i \in T, v \in V}[a_1 + a_2 + t_1 + \cdots + t_m + v \in 2A] \\ &\quad = \sum_{\gamma \in \widehat{\mathbb{F}_2^n}} \widehat{\phi_A}(\gamma)^2 \widehat{\phi_T}(\gamma)^m \widehat{\phi_V}(\gamma) \widehat{1_{2A}}(\gamma) \geq \frac{5}{8} > \frac{1}{2}. \end{aligned}$$

Where we used that $\widehat{\phi_V} = 1_{V^\perp}$ so the drop in probability can be bounded by summing the Fourier coefficients along $\gamma \notin V^\perp$. For such frequencies, $|\widehat{\phi_T}(\gamma)| < 1/2$, so we obtain $2^{-m} \sum \widehat{\phi_A}^2 = 2^{-m} \alpha^{-1} \leq 1/8$ by the choice of m.

Because this probability is greater than half, there must exist a choice $b = a_1 + a_2 + t_1 + \cdots + t_m$ such that more than half the elements of V lie in $2A + b$. By the pigeonhole principle, any element of V can be written as the difference of two elements in this intersection, meaning $V \subset (2A + b) + (2A + b) = 4A$.

To find the final codimension bound, we combine our two main lemmas. Croot–Sisask guarantees the density of T is $\tau \geq \frac{1}{2}\alpha^\ell$, which implies $\log(2/\tau) = O(\ell \log(2/\alpha))$.

Applying Chang's theorem to T at threshold $1/2$ bounds the codimension of V by $O(\log(2/\tau))$. Substituting our density bound alongside the chosen parameters $\ell = O(p\varepsilon^{-2})$, $p = O(\log(2/\alpha))$, and $\varepsilon^{-1} = O(\log(2/\alpha))$ yields a codimension of $O(\log^4(2/\alpha))$.

References

[L15] Lovett, S., *An Exposition of Sanders' Quasi-Polynomial Freiman-Ruzsa Theorem.* Theory of Computing Library, Graduate Surveys 6, 1–14 (2015).

[S12] Sanders, T., *On the Bogolyubov-Ruzsa lemma.* Analysis and PDE 5, 627–655 (2012).

António Rocha-Neves, Universiteit van Amsterdam
email: arochaneves@proton.me

19 An equivalence between inverse sumset theorems and inverse conjectures for the U^3 norm

After B. Green and T. Tao [GT]

A summary written by Nihan Tanısalı

Abstract

We explain Green and Tao's equivalence between the Polynomial Freiman–Ruzsa conjecture over $\mathbb{F}_2^\infty$ and the polynomial inverse conjecture for the U^3 norm.

19.1 Introduction

A finite set A in an abelian group is a K-approximate group if $A = -A$ and $A + A$ is contained in the union of at most K translates of A. A finite set B K-controls A if $|B| \leq K|A|$ and $A \subseteq B + X$ for some X with $|X| \leq K$.

Conjecture 1 (PFR over $\mathbb{F}_2^\infty$)**.** *There is an absolute constant $C > 0$ such that every K-approximate group $A \subseteq \mathbb{F}_2^\infty$ is K^C-controlled by a finite subgroup.*

For $f : G \to \mathbb{C}$ on a finite abelian group, write $\mathbb{E}_x f(x) = |G|^{-1} \sum_x f(x)$. The Gowers norm is

$$\|f\|_{U^k}^{2^k} = \mathbb{E}_{x,h_1,\ldots,h_k} \prod_{\omega \in \{0,1\}^k} \mathcal{C}^{|\omega|} f(x + \omega \cdot h),$$

where $\mathcal{C}^z = \overline{z}$. A map $\psi : \mathbb{F}_2^n \to \mathbb{F}_2$ is quadratic if all its third additive derivatives vanish.

Conjecture 2 (PGI(3) over $\mathbb{F}_2^n$)**.** *There is an absolute constant $C > 0$ such that if $\|f\|_\infty \leq 1$ and $\|f\|_{U^3} \geq K^{-1}$, then*

$$\left|\mathbb{E}_x f(x)(-1)^{\psi(x)}\right| \geq K^{-C}$$

for some quadratic $\psi : \mathbb{F}_2^n \to \mathbb{F}_2$.

Green and Tao prove that these two conjectures are equivalent [GT].

19.2 PFR implies PGI(3)

We use the following equivalent formulation of PFR from Lovett [L, Conjecture 3].

Conjecture 3 (Equivalent small-doubling form of PFR)**.** *If $B \subseteq \mathbb{F}_2^m$ satisfies $|B+B| \le L|B|$, then there is $B' \subseteq B$ such that*

$$|B'| \ge L^{-C}|B|, \qquad |\operatorname{span}(B')| \le L^C|B|.$$

Assume PFR and let $f : \mathbb{F}_2^n \to \mathbb{C}$ satisfy $\|f\|_\infty \le 1$ and $\|f\|_{U^3} \ge K^{-1}$. Put

$$f_h(x) = f(x+h)\overline{f(x)}, \qquad \widehat{g}(\xi) = \mathbb{E}_x g(x)(-1)^{\xi\cdot x}.$$

The following is a consequence of Samorodnitsky's argument [S].

Theorem 4. *There is a map $\phi : \mathbb{F}_2^n \to \mathbb{F}_2^n$ such that the set of pairs (x,y) satisfying $\phi(x+y) = \phi(x) + \phi(y)$ and*

$$|\widehat{f_x}(\phi(x))|, |\widehat{f_y}(\phi(y))|, |\widehat{f_{x+y}}(\phi(x+y))| \ge K^{-C}$$

has density at least K^{-C} in $\mathbb{F}_2^n \times \mathbb{F}_2^n$.

Define $A = \{x : |\widehat{f_x}(\phi(x))| \ge K^{-C}\}$ and let $\mathcal{A} = \{(x, \phi(x)) : x \in A\}$. Theorem 4 implies that A has density at least K^{-C} and that $\mathcal{A}$ contains at least $K^{-C}|\mathcal{A}|^2$ additive triples. By the Balog–Szemerédi–Gowers theorem, there is $\mathcal{A}' \subseteq \mathcal{A}$ such that

$$|\mathcal{A}'| \ge K^{-C}|\mathcal{A}|, \qquad |\mathcal{A}' + \mathcal{A}'| \le K^C|\mathcal{A}'|.$$

Applying Conjecture 3 to $\mathcal{A}'$, we obtain $\mathcal{A}'' \subseteq \mathcal{A}'$ such that

$$|\mathcal{A}''| \ge K^{-C}|\mathcal{A}'|, \qquad |\operatorname{span}(\mathcal{A}'')| \le K^C|\mathcal{A}'|.$$

A linear-algebra decomposition of $\operatorname{span}(\mathcal{A}'')$, followed by a counting argument, shows that some affine graph

$$\{(x, Dx + z) : x \in \mathbb{F}_2^n\}$$

contains at least $K^{-C}|\mathcal{A}''|$ points of $\mathcal{A}''$. Since $|\mathcal{A}''| \ge K^{-C}|\mathcal{A}'|$, $|\mathcal{A}'| \ge K^{-C}|\mathcal{A}|$, and $|\mathcal{A}| \ge K^{-C}2^n$, this affine graph contains $(x, \phi(x))$ for a set of x

of density at least K^{-C} in $\mathbb{F}_2^n$. Hence $\phi(x) = Dx + z$ on this set. Since these x lie in A, the definition of A gives

$$\mathbb{E}_x|\widehat{f_x}(Dx+z)|^2 \geq K^{-C}.$$

The following is the consequence of Lemmas 6.10–6.11 and the subsequent argument in [S]. It will be used to pass from an affine frequency map to a symmetric linear map.

Lemma 5. *Suppose that $D : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is linear and*

$$\mathbb{E}_x|\widehat{f_x}(Dx+z)|^2 \geq \eta$$

for some $z \in \mathbb{F}_2^n$. Then there is a symmetric zero-diagonal linear map B such that

$$\mathbb{E}_x|\widehat{f_x}(Bx)|^2 \geq \eta^C.$$

Applying Lemma 5, there is a symmetric zero-diagonal linear map B such that

$$\mathbb{E}_x|\widehat{f_x}(Bx)|^2 \geq K^{-C}.$$

Choose a matrix M with $B = M + M^t$, set $q(x) = x \cdot Mx$, and define $g(x) = f(x)(-1)^{q(x)}$. Since

$$q(x+h) + q(x) + q(h) = x \cdot Bh,$$

we get $|\mathbb{E}_x g_h(x)| = |\widehat{f_h}(Bh)|$ and therefore

$$\|g\|_{U^2}^4 = \mathbb{E}_h|\widehat{f_h}(Bh)|^2 \geq K^{-C}.$$

The identity $\|g\|_{U^2}^4 = \sum_\xi |\widehat{g}(\xi)|^4$, together with Plancherel, yields some ξ with $|\widehat{g}(\xi)| \geq K^{-C}$. That is to say,

$$\left|\mathbb{E}_x f(x)(-1)^{q(x)+\xi\cdot x}\right| \geq K^{-C},$$

and $x \mapsto q(x) + \xi \cdot x$ is quadratic. This proves PGI(3).

19.3 PGI(3) implies PFR

Assume PGI(3). Let $S \subseteq \mathbb{F}_2^n$ have density σ, and let $\phi : S \to \mathbb{F}_2^\infty$ satisfy

$$x_1 + x_2 = x_3 + x_4 \quad \Longrightarrow \quad \phi(x_1) + \phi(x_2) = \phi(x_3) + \phi(x_4). \tag{1}$$

We first show that ϕ agrees with an affine map on at least $\sigma^C 2^n$ points of S. Choose N such that $\phi(S) \subseteq \mathbb{F}_2^N$ and define

$$F(x, y) = \mathbb{1}_S(x)(-1)^{\phi(x)\cdot y}.$$

On every three-dimensional cube contained in S, relation (1) implies that $\omega \mapsto \phi(x + \omega \cdot h)$ is affine-linear. Thus $\omega \mapsto \phi(x + \omega \cdot h) \cdot (y + \omega \cdot k)$ is quadratic, so the phase cancels in the U^3 cube product. Consequently

$$\|F\|_{U^3(\mathbb{F}_2^{n+N})} = \|\mathbb{1}_S\|_{U^3(\mathbb{F}_2^n)} \geq \sigma.$$

By PGI(3), there is a quadratic $\Psi : \mathbb{F}_2^{n+N} \to \mathbb{F}_2$ such that

$$\left|\mathbb{E}_{x,y} \mathbb{1}_S(x)(-1)^{\phi(x)\cdot y + \Psi(x,y)}\right| \geq \sigma^C.$$

Hence for at least $\sigma^C 2^n$ values of $x \in S$, the inner average in y has magnitude at least σ^C. The mixed part

$$B(x, y) = \Psi(x, y) + \Psi(x, 0) + \Psi(0, y) + \Psi(0, 0)$$

is bilinear, so $B(x, y) = L(x) \cdot y$ for a linear $L : \mathbb{F}_2^n \to \mathbb{F}_2^N$. Therefore the fixed function $y \mapsto (-1)^{\Psi(0,y)}$ has a Fourier coefficient of size at least σ^C at frequency $\phi(x) + L(x)$. Plancherel shows that there are only $\sigma^{-O(1)}$ such frequencies. Pigeonholing gives $r \in \mathbb{F}_2^N$ for which

$$\phi(x) = L(x) + r$$

for at least $\sigma^C 2^n$ points of S.

We now deduce PFR. Let $A \subseteq \mathbb{F}_2^\infty$ be a K-approximate group. Choose n minimal such that some linear map $\pi : \mathbb{F}_2^\infty \to \mathbb{F}_2^n$ restricts to a Freiman isomorphism on A, and put $S = \pi(A)$. Minimality implies

$$4S = \mathbb{F}_2^n.$$

Indeed, if $v \notin 4S$, quotienting by $\langle v \rangle$ would still be Freiman-injective on S, contradicting the choice of n. Since $|S + S| = |A + A| \leq K|S|$, Ruzsa's sumset estimate gives

$$2^n = |4S| \leq K^C |S|,$$

so S has density at least K^{-C}.

Apply the affine-agreement statement to $\phi = \pi^{-1} : S \to A$. For $H = L(\mathbb{F}_2^n)$, some coset $r + H$ satisfies

$$|A \cap (r + H)| \geq K^{-C}|A|, \qquad |H| \leq 2^n \leq K^C|A|.$$

Set $B = A \cap (r + H)$ and choose $X \subseteq A$ maximal such that the translates $x + B$, $x \in X$, are disjoint. Since they lie in $A + A$, one has $|X| \leq K^C$. Maximality gives $A \subseteq H + X$, and hence H K^C-controls A. This proves PFR.

19.4 The higher-order result

We now work over $\mathbb{F} = \mathbb{F}_5$. A map $\phi : S \to \mathbb{F}^\infty$ is a Freiman quadratic if its third additive derivative vanishes whenever the corresponding cube lies in S. The input is the U^4 inverse theorem, which states that a 1-bounded function with large U^4 norm correlates with a cubic phase.

Proposition 6 (Green–Tao, Proposition 4.2)**.** *Let $S \subseteq \mathbb{F}^n$ have density $\sigma \in (0, 1/2)$ and let $\phi : S \to \mathbb{F}^\infty$ be a Freiman quadratic. Then there is a quadratic map $\psi : \mathbb{F}^n \to \mathbb{F}^\infty$ such that*

$$|\{x \in S : \phi(x) = \psi(x)\}| \geq \varepsilon(\sigma)|\mathbb{F}|^n.$$

Choose N with $\phi(S) \subseteq \mathbb{F}^N$ and define

$$F(x, y) = \mathbb{1}_S(x) e_{\mathbb{F}}(\phi(x) \cdot y).$$

On every four-dimensional cube contained in S, the phase in the U^4 cube product cancels because ϕ is Freiman quadratic. Hence $\|F\|_{U^4} \geq \sigma$, and the U^4 inverse theorem gives a cubic phase correlating with F.

Decomposing this cubic according to its degree in y, the Gauss-sum, large-sieve, rank-reduction, and Plancherel arguments in [GT, proof of Proposition 4.2, using Proposition 4.3] give a quadratic map $\widetilde{\psi} : \mathbb{F}^n \to \mathbb{F}^N$ and a set $R \subseteq \mathbb{F}^N$ with $|R| = O_\sigma(1)$ such that

$$\phi(x) - \widetilde{\psi}(x) \in R$$

for at least $c_\sigma|\mathbb{F}|^n$ values of $x \in S$. Pigeonholing a value $r \in R$ gives $\phi(x) = \widetilde{\psi}(x) + r$ on at least $\varepsilon(\sigma)|\mathbb{F}|^n$ points. Since $\widetilde{\psi} + r$ is quadratic, the proposition follows.

References

[GT] B. Green and T. Tao, *An equivalence between inverse sumset theorems and inverse conjectures for the U^3 norm*, Math. Proc. Cambridge Philos. Soc. **149** (2010), 1–19

[L] S. Lovett, *An equivalence between the polynomial Freiman–Ruzsa conjecture and the polynomial inverse conjecture for the Gowers norm*

[S] A. Samorodnitsky, *Low-degree tests at large distances*, STOC 2007, 506–515

Nihan Tanisali, INRIA and ??cole Polytechnique
email: nihan.tanisali@inria.fr

20 An inverse theorem for the Gowers U^4-norm

After B. Green, T. Tao and T. Ziegler [GTZ1]

A summary written by Kit Thomas

Abstract

We give a summary of the proof of the U^4-inverse theorem due to Green, Tao and Ziegler [GTZ1], which shows that functions with a large Gowers U^4-norm correlate with a 3-step nilsequence.

20.1 Introduction

This summary will sketch the proof of the following theorem, due to Green, Tao and Ziegler [GTZ1].

Theorem 1 (U^4-inverse theorem, [GTZ1])**.** *Let $f : [N] \to \mathbb{C}$ be a 1-bounded function. Suppose that $\|f\|_{U^4} \geq \delta$. Then there exists a degree 3 polynomial nilsequence, $(\Psi(n))_{n\in\mathbb{Z}}$ of complexity $O_\delta(1)$ such that*

$$|\mathbb{E}_{[N]} f(n)\overline{\Psi(n)}| \gg_\delta 1.$$

Before commenting on the proof of the theorem, let us define some of the objects occurring in its statement.

Gowers norms: Firstly, recall that the Gowers U^s-norm is defined as follows. For an abelian group G and a function $f : G \to \mathbb{C}$ we have

$$\|f\|_{U^s(G)} \coloneqq \left(\mathbb{E}_{x,h_1,\dots,h_s\in G} \Delta_{h_1} \dots \Delta_{h_s} f\right)^{1/2^s}.$$

Here,

$$\Delta_h f(n) \coloneqq f(n+h)\overline{f(n)}$$

denotes the multiplicative derivative of f. To extend this to functions defined on $[N] \coloneqq \{1, \dots, N\}$, set $G = \mathbb{Z}/\tilde{N}\mathbb{Z}$ for $\tilde{N} \geq 2^s N$ and let

$$\|f\|_{U^s[N]} \coloneqq \|f\|_{U^s(G)}/\|\mathbf{1}_{[N]}\|_{U^s(G)},$$

after extending f to G by setting it to zero outside of $[N]$. Henceforth, U^s will be used to denote $U^s[N]$. A key relationship is the fact that, up to normalisation factors,

$$\|f\|_{U^s} = \left(\mathbb{E}_h \|\Delta_h f\|_{U^{s-1}}^{2^{s-1}}\right)^{1/2^s}. \tag{1}$$

Nilsequences: The outcome of Theorem 1 is that our function f correlates with an object $\Psi(n)$ which is a degree 3 nilsequence. For the purposes of this summary, we will avoid rigorously stating the definition of a nilsequence and hope that the following discussion gives an idea of the kind of objects at play.

A feature of this paper (compared to, say, the proof of the inverse theorem for general U^s-norms [GTZ2]) is that we can work with explicit bracket polynomials. These are functions built up of the usual polynomial operations, as well as allowing bracket operations, $\{\cdot\}$, where $\{x\}$ is the fractional part of x. Let $e(x)$ denote $e^{2\pi ix}$. In place of nilsequences in this summary, we will work with functions $e(p(n))$ where $p(n)$ is a bracket polynomial. The fact that taking fractional parts is discontinuous may prevent these functions from being genuine nilsequences. Sequences which are close to being nilsequences (in a precise sense) are referred to in [GTZ1] as *almost-nilsequences*, and in practice, it is possible to get around these discontinuity issues using quantitative equidistribution results.

The following are some examples of almost-nilsequences:

- $e(\alpha n)$, $e(\alpha\{\beta n\})$ (degree 1)
- $e(\alpha n^2)$, $e(\alpha n\lfloor\beta n\rfloor)$ (degree 2)
- $e(\alpha n^3)$, $e(\{\alpha n\}\{\beta n\}\gamma n)$, $e(\{\alpha n\}\beta n\lfloor\gamma n\rfloor)$, $e(\alpha n^2\lfloor\beta n\rfloor)$ (degree 3).

We will also not comment on the notion of complexity of a nilsequence. Therefore, a rough statement of Theorem 1 that we are trying to prove in this sketch could be showing that

$$|\mathbb{E}_{n\in[N]}f(n)\overline{\Psi(n)}| \gg_\delta 1,$$

for some function $\Psi(n)$ which is a product of degree 1, 2 and 3 almost-nilsequences of the above forms.

20.1.1 Preliminaries

A key input for the proof of the U^4 inverse theorem is the U^3 inverse theorem [GT1].

Theorem 2 (U^3 inverse theorem, rough version)**.** *Let $f : [N] \to \mathbb{C}$ be a 1-bounded function with $\|f\|_{U^3} \geq \delta$. Then there exists a sequence $(\Psi(n))_{n\in\mathbb{Z}}$ such that*

$$|\mathbb{E}_{n\in[N]}f(n)\overline{\Psi(n)}| \gg_\delta 1,$$

where $\Psi(n)$ is a degree 2 almost-nilsequence (of bounded complexity).

For our purposes, we can assume that Ψ is the product of terms of the form:

$$e(\alpha n^2), \;\; e(\beta n) \text{ and } e(\xi n \lfloor \xi' n \rfloor).$$

Additionally, the proof requires quantitative results about the equidistribution of nilsequences in nilmanifolds ([GT2], [GT3]). Very roughly speaking, this allows the distribution of functions such as $e(\xi n \lfloor \xi' n \rfloor)$ to be deduced from the distribution of $(\xi n, \xi' n)$ in the torus.

20.2 Proof outline

The starting point of the proof is to use the identity (1) to show that the assumption $\|f\|_{U^4} \geq \delta$ implies that there is a large set of values $H \subseteq [N]$ such that $\|\Delta_h f\|_{U^3} \gg_\delta 1$ for all $h \in H$. Applying the U^3 inverse theorem, we obtain that

$$|\mathbb{E}_{n \in [N]} \Delta_h f(n) \overline{\chi_h(n)}| \gg_\delta 1$$

for each such h, where

$$\chi_h(n) = e(\alpha_h n^2 + \beta_h n + \sum \xi_{h,i} n \lfloor \xi_{h,i'} n \rfloor). \tag{2}$$

The first few steps reveal the dependence of the coefficients in (2) on h. This will allow us to show that $\Delta_h f$ correlates with what is essentially the derivative of a degree 3 nilsequence. The fact this holds for many h allows us to "integrate" and get the conclusion that f correlates with a degree 3 nilsequence.

20.2.1 Correlation among the χ_h

Following an argument of Gowers, the next step is to deduce some weak linear structure among the $\chi_h(n)$ via applications of the Cauchy-Schwarz inequality.

Proposition 3. *For many additive quadruples $(h_1, \ldots, h_4) \in H^4$ with $h_1 + h_2 = h_3 + h_4$,*

$$|\mathbb{E}_{n \in [N]} \chi_{h_1}(n) \chi_{h_2}(n + h_1 - h_4) \overline{\chi_{h_3}(n) \chi_{h_4}(n + h_1 - h_4)}| \gg_\delta 1.$$

20.2.2 Reducing h-dependence

Let Ξ_h be the set of frequencies $\xi_{h,i}$ coming from (2).

The next step is to show that not too many of these frequencies genuinely depend on h. This is done by iteratively decomposing the sets Ξ_h into a "core" set of frequencies which are the same for each h, and "petal" sets consisting of the remaining frequencies. Furthermore, for the terms of the form $e(\xi_{h,i}n\lfloor\xi_{h,i'}n\rfloor)$ appearing in (2), at most one of the values $\xi_{h,i}$ or $\xi_{h,i'}$ can come from a petal set.

At each step of the iteration, Proposition 3 is used to show there is correlation amongst the χ_{h_i}. Quantitative equidistribution results relate this to the distribution of values in the sets Ξ_{h_i} to get the conclusion.

20.2.3 Approximate linearity of h-dependent frequencies

The next step is to show that there is a further decomposition such that the h-dependent frequencies, $\xi_{h,i}$, are bracket linear, meaning they are of the form

$$\xi_{h,i} = b_0 + a_1\{b_1h\} + \ldots + a_d\{b_dh\} + \theta h.$$

Once again, this is proved inductively and makes crucial use of Proposition 3 and quantitative equidistribution results. This reveals essentially rational relations between remaining frequencies $\xi_{h,i}$, at which point tools regarding approximate homomorphisms upgrade this to show bracket linearity.

This can also be extended to show that the quadratic coefficients, α_h, are bracket linear.

20.2.4 Symmetry argument

Combining the previous steps gives that

$$|\mathbb{E}_{n\in[N]}f(n)\overline{f(n+h)}\chi_h(n)| \gg_\delta 1$$

for many values of h, where the $\chi_h(n)$ can be written as the product of terms of the following form:

$$e(\{\alpha h\}\beta n\lfloor\gamma n\rfloor),\ \ e(\alpha\{\beta h\}n^2) \text{ and } e(\theta_h n).$$

Note that the dependence on h for the quadratic terms has been made explicit. Moreover, considering both h and n as variables, the above terms begin to resemble degree 3 almost-nilsequences.

Write $\chi_h(n) = e(3T(h,n,n))e(\theta_h n)$, where $T(x,y,z)$ is a bracket polynomial consisting of terms $\{\alpha x\}\beta y\{\gamma z\}$ and $\alpha\{\beta x\}yz$.

The main work of this final section is to show that T is trilinear and symmetric (upon restricting its domain to certain Bohr sets).

Having deduced these properties, we can use these and Proposition 3 to build a degree 3 almost-nilsequence $\Psi(n)$ such that $f(n)\overline{\Psi(n)}$ has large U^3 norm. By the U^3 inverse theorem, this correlates with a degree 2 almost-nilsequence:

$$|\mathbb{E}_{n\in[N]} f(n)\overline{\Psi(n)\varphi(n)}| \gg_\delta 1.$$

The product $\Psi(n)\varphi(n)$ is a degree 3 almost-nilsequence as required.

References

[GT1] Green, B. and Tao, T., *An inverse theorem for the Gowers U^3 norm.* Proc. Edinb. Math. Soc. (2) **51** (2008), no. 1, 73–153;

[GT2] Green, B. and Tao, T., *The quantitative behaviour of polynomial orbits on nilmanifolds.* Ann. of Math. (2) **175** (2012), no. 2, 465–540;

[GT3] Green, B. and Tao, T., *The Möbius function is strongly orthogonal to nilsequences.* Ann. of Math. (2) **175** (2012), no. 2, 541–566;

[GTZ1] Green, B., Tao, T. and Ziegler, T., *An inverse theorem for the Gowers U^4-norm.* Glasgow Math. J. **53** (2011), no. 1, 1–50;

[GTZ2] Green, B., Tao, T. and Ziegler, T., *An inverse theorem for the Gowers $U^{s+1}[N]$-norm.* Ann. of Math. (2) **176** (2012), no. 2, 1231–1372;

Kit Thomas
email: magd6057@ox.ac.uk

21 Freiman's theorem in groups with bounded exponent

After Yufei Zhao [YZ]

A summary written by Shota Uka

Abstract

We prove Freiman's theorem in groups with bounded exponent.

21.1 Introduction

We begin by introducing the necessary notions and recalling Freiman's theorem for integers and for general abelian groups. We then focus our attention on the case of groups with bounded exponent.

Definition 1. *Let A and B be finite subsets of an ambient abelian group. We define*

(i) the sumset $A + B := \{a + b : a \in A, b \in B\}$,

(ii) the iterated sumset $kA := A + \cdots + A$ (k times),

(iii) the dilated set $\lambda A := \{\lambda a : a \in A\}$,

(iv) the difference set $A - B := \{a - b : a \in A, b \in B\}$.

At the center of Freiman's theorem is the following question: If we know that $A + A$ is small, what can we infer about the set A itself? Let us first start with the converse question: What can we say about the sumset $A + A$, if we know $|A|$? These answer follows from very simple combinatorial considertations: Let $n = |A|$ and denote by $a_1 < a_2 < \cdots < a_n$ the elements of A. Then we have

$$a_1 + a_1 < a_1 + a_2 < \cdots < a_1 + a_n < a_2 + a_n < \ldots a_n + a_n.$$

Hence, we get $2n - 1$ different elements of $A + A$, so

$$2|A| - 1 \leq |A + A|.$$

Since, obviously, there are $\binom{n+1}{2}$ unordered pairs of elements of A, we have

$$|A+A| \leq \binom{|A|+1}{2}.$$

Combining alltogether, we have proven

Proposition 2 (Easy bounds on sumset size)**.** *Let $A \subset \mathbb{Z}$ be a finite set. Then*

$$2|A|-1 \leq |A+A| \leq \binom{|A|+1}{2}.$$

To study the relation between the sumset $A+A$ and A we introduce:

Definition 3 (Doubling constant)**.** *Let A be a finite subset of an abelian group. We call the ratio $|A+A|/|A|$ the doubling constant of A.*

For arithmetic progressions, the previous proposition immediately gives an upper bound of the doubling constant:

Example 4. *(i) Let $A \subset \mathbb{Z}$ be a finite arithmetic progression. Then $|A+A| = 2|A|-1 \leq 2|A|$. Hence, A has doubling constant at most 2.*

(ii) Let B be a finite arithmetic progression and $A \subset B$ such that $|A| \geq |B|/K$. Then

$$|A+A| \leq |B+B| \leq 2|B| \leq 2K|A|.$$

Therefore, A has doubling constant at most $2K$.

We can generalize arithmetic progressions in multiple dimensions in the following sense:

Definition 5 (GAP – generalized arithmetic progression)**.** *A generalized arithmetic progression (GAP) in an abelian group Γ is an affine map*

$$\phi : [L_1] \times \cdots \times [L_d] \to \Gamma,$$

i.e. for some $a_0, \ldots, a_d \in \Gamma$,

$$\phi(x_1, \ldots, x_d) = a_0 + a_1 x_1 + \cdots + a_d x_d.$$

We call this GAP proper if ϕ is injective. The dimension of this GAP is defined to be d and the volume is given by $L_1 \ldots L_d$.

Remark 6. *(i) Abusing notation, we may sometimes use the term GAP to refer to the image of the affine map ϕ, i.e.*

$$\begin{aligned} P &= a_0 + a_1[L_1] + \cdots + a_d[L_d] \\ &= \{a_0 + a_1x_1 + \cdots + a_dx_d : x_1 \in [L_1], \ldots, x_d \in [L_d]\}. \end{aligned}$$

(ii) Let $P = \{a_0 + a_1x_1 + \cdots + a_dx_d : x_1 \in [L_1], \ldots, x_d \in [L_d]\}$. Then clearly $|P + P|/|P| \leq 2^d$, if ϕ is proper.

(iii) Let P be a proper GAP of dimension d and $A \subset P$ such that $|A| \geq |P|/K$. Then

$$|A + A| \leq |P + P| \leq 2^d|P| \leq 2^dK|A|.$$

Hence, A has doubling constant at most 2^dK.

Considering these examples, we see that for certain sets we can immediately find upper bounds for the doubling constant. So, if we know the structure of the set, we can deduce an upper bound for the doubling constant. The reverse question, however, cannot be answered that easily: What can we say about sets with small doubling constant? Freiman's theorem completely characterizes these sets. In fact, we will see that sets with small doubling constant have the same structure as Remark 6 (iii):

Theorem 7 (Freiman's theorem)**.** *Let $A \subset \mathbb{Z}$ be a finite set such that $|A + A| \leq K|A|$. Then $A \subset P$, where P is a GAP of dimension at most $d(K)$ and volume at most $f(K)|A|$. The constants $d(K)$ and $f(K)$ depend only on K.*

Remark 8. *(i) We can show that $d(K) \leq K - 1$ and $f(K) = e^{O(K)}$.*

(ii) Freiman's theorem is also called Freiman-Ruzsa theorem. Ruzsa (1997) managed to provide a significantly simpler proof than Freiman's original proof in 1973.

Note that we only consider a subset A of $\mathbb{Z}$ in the above version of Freiman's theorem. Green and Ruzsa (2007) proved that Freiman's theorem holds for an arbitrary abelian group. To state that version we need the following

Definition 9 (Coset progression)**.** *A coset progression is a set of the form $P + A$, where P is a GAP and H is a subgroup of the ambient abelian group. The dimension of this coset progression is defined to be the dimension of P and the volume is given by $|H| \operatorname{vol}(P)$.*

Then the following generalization of Freiman's theorem holds:

Theorem 10 (Freiman's theorem for general abelian groups)**.** *Let A be a subset of an abelian group such that $|A+A| \leq K|A|$. Then $A \subset P'$, where P' is a coset progression of dimension at most $d(K)$ and volume at most $f(K)|A|$. The constants $d(K)$ and $f(K)$ depend only on K.*

We can now state Freiman's theorem for groups with bounded exponent. To do so, we need

Definition 11. *Let G be an abelian group. We define the exponent of G as the smallest positive integer r such that $rx = 0$ for all $x \in G$. If no such integer exists, we define the exponent to be infinite.*

Example 12. *1. The finite field $\mathbb{F}_2^n$ has exponent 2.*

2. $\mathbb{Z}/N\mathbb{Z}$ has exponent N.

3. $\mathbb{Z}$ has infinite exponent.

Let G be a group and A a subset of G. We denote by $\langle A \rangle$ the subgroup of G generated by A. Clearly, the exponent of G is given by $\sup_{x \in G} |\langle x \rangle|$, where $|\langle x \rangle|$ is the order of the element x. If G is a vector space, then $\langle A \rangle$ is the smallest subspace containing A.

Theorem 13 (Freiman's theorem for groups with bounded exponent)**.** *Let A be a finite set in an abelian group with exponent $r < \infty$. If $|A+A| \leq K|A|$, then*

$$|\langle A \rangle| \leq K^2 r^{K^4} |A|.$$

We will spend the rest of the talk proving this theorem.

21.2 Plünnecke's inequality and Ruzsa's coverling lemma

The proof of Theorem 13 is very short and essentially relies on applying Plünnecke's inequality and Ruzsa's coverling lemma. This section is devoted to these two lemmas. We start with Plünnecke's inequality which states that sumsets of sets with small doubling constant are also small:

Theorem 14 (Plünnecke's inequality)**.** *Let A be a finite subset of an abelian group such that*

$$|A + A| \leq K|A|.$$

Then for all $m, n \in \mathbb{Z}_+$,

$$|mA - nA| \leq K^{m+n}|A|.$$

Remark 15. *Over the years different proofs of this theorem were presented, first by Plünnecke (1970), then a simpler proof by Ruzsa (1989), and finally a very short proof by Petridis (2012) which we follow here.*

The more general statement holds as well:

Theorem 16. *Let A and B be finite subsets of an abelian group such that*

$$|A + B| \leq K|A|.$$

Then for all $m, n \in \mathbb{Z}_+$,

$$|mB - nB| \leq K^{m+n}|A|.$$

The second essential ingredient in the proof of Freiman's theorem is Ruzsa's coverling lemma.

Theorem 17 (Ruzsa coverling lemma)**.** *Let X and B be finite sets in some abelian group. If*

$$|X + B| \leq K|B|,$$

then there exists a subset $T \subset X$ with $|T| \leq K$ and $X \subset T + B - B$.

21.3 Proof of Freiman's theorem in groups with bounded exponent

We can now prove Freiman's theorem in groups with bounded exponent.

Proof. Plünnecke's inequality, Theorem 16, implies

$$|A + (2A - A)| = |3A - A| \leq K^4|A|.$$

Choosing $X = 2A - A$ and $B = A$ in Ruzsa covering lemma, there exists a subset $T \subset 2A - A$ with $|T| \leq |A + (2A - A)|/|A| \leq K^4$ and $2A - A \subset T + A - A$. Adding A to both sides gives

$$3A - A = 2A - A + A \subset T + \underbrace{2A - A}_{\subset T+A-A} \subset 2T + A - A.$$

By induction over $n \geq 0$, we get

$$(n+1)A - A \subset nT + A - A \subset \langle T \rangle + A - A.$$

Recall that we consider an abelian group with bounded exponent. Hence, for every $x \in \langle A \rangle$ there exists $n \in \mathbb{N}$ such that $x \in nA$. Consequently,

$$\langle A \rangle \subset \bigcup_{n \geq 1} (nA + A - A) \subset \langle T \rangle + A - A. \tag{1}$$

Moreover, since the exponent is at most $r < \infty$ and $|T| \leq K^4$,

$$|\langle T \rangle| \leq r^{|T|} \leq r^{K^4}. \tag{2}$$

Again applying Plünnecke's inequality, Theorem 14, we obtain

$$|A - A| \leq K^2 |A|. \tag{3}$$

Combining (1), (2) and (3) gives

$$|\langle A \rangle| \leq |\langle T \rangle + A - A| \leq |\langle T \rangle||A - A| \leq r^{K^4} K^2 |A|.$$

□

References

[YZ] Zhao Yufei, *Graph Theory and Additive Combinatorics: Exploring Structure and Randomness.* Cambridge University Press, 2023.

Shota Uka, TU Vienna
email: e1211760@student.tuwien.ac.at

22 The sum-product problem over the reals

After G. Elekes [Ele] and L. A. Székely [Sz]

A summary written by Pauwel Van Den Eeckhaut

Abstract

We present Elekes's proof of the sum-product estimate

$$\max\{|A+A|, |AA|\} \gtrsim |A|^{5/4}$$

for finite sets $A \subseteq \mathbb{R}$. The main ingredient in this proof is the Szemerédi–Trotter theorem on the number of incidences between points and lines in the plane. Following Székely, we derive this theorem from the crossing number inequality for simple graphs, using a probabilistic argument to prove the latter.

22.1 The sum-product problem

Let $A \subseteq \mathbb{R}$ be a finite set. Its *sumset* and *product set* are defined as

$$A + A = \{a + b : a, b \in A\}, \qquad AA = \{ab : a, b \in A\}.$$

One can use the sum and product sets to quantify the additive and multiplicative structure of A. Indeed, by counting pairs in A, we obtain the bounds

$$|A| \leq |A + A| \lesssim |A|^2, \quad |A| \leq |AA| \lesssim |A|^2.$$

Here, we write $x \lesssim y$ if $x \leq Cy$ for some absolute constant $C > 0$. The sizes of $A + A$ and AA measure how frequently pairs of elements in A produce the same sum or product. When $|A + A|$ is comparable to $|A|$, many pairs in A produce the same sum, so A has high additive structure. On the other hand, when $|A + A|$ is comparable to $|A|^2$, there are few such coincidences, so A has low additive structure. By the same reasoning, $|AA|$ measures the multiplicative structure of A.

For example, arithmetic progressions have high additive structure and low multiplicative structure, whereas geometric progressions have high multiplicative structure and low additive structure. These examples suggest a tension between additive and multiplicative structure: a set cannot be highly structured with respect to both operations simultaneously. This was made precise by Erdős and Szemerédi [ErSz] in the following theorem.

Theorem 1 (Erdős–Szemerédi theorem)**.** *There exists $\varepsilon > 0$ such that every finite set $A \subseteq \mathbb{R}$ satisfies*

$$\max\{|A+A|, |AA|\} \gtrsim |A|^{1+\varepsilon}.$$

In their paper, Erdős and Szemerédi in fact conjectured the much stronger bound

$$\max\{|A+A|, |AA|\} \gtrsim |A|^{2-o(1)}$$

where $o(1)$ is some quantity going to zero as $|A| \to \infty$. This statement, known as the sum-product conjecture, was a longstanding open problem until it was disproved in a recent preprint of Bloom, Sawin, Schildkraut, and Zhelezov [BSSZ]. The analogous question for finite subsets of the integers, however, remains open to this date.

We will prove the following estimate, due to Elekes [Ele], which improves the exponent in Theorem 1.

Theorem 2. *Every finite set $A \subseteq \mathbb{R}$ satisfies*

$$\max\{|A+A|, |AA|\} \gtrsim |A|^{5/4}.$$

Elekes's argument works by converting the problem to a question about the number of incidences between points and lines in the plane. The main input here is the Szemerédi–Trotter theorem, which provides an optimal bound for such incidences. In the next section, we explain Székely's proof of the Szemerédi–Trotter theorem, which is based on the crossing number inequality for simple graphs [Sz]. The final section shows how this implies Theorem 2.

22.2 Crossing numbers and the Szemerédi–Trotter theorem

All graphs below are understood to be finite and simple. Given a graph G, we consider drawings in the plane, where vertices are represented by points, and edges by continuous curves, such that no edge passes through any vertices except its endpoints. The *crossing number* $\mathrm{cr}(G)$ is the minimal number of edge crossings over all such drawings of G. Here, if multiple edges cross at a single point, we count one crossing for each pair of these edges. Our first goal is to obtain a useful lower bound for the crossing number. In the following argument, we follow the exposition by Zhao [Zh].

We begin with the elementary fact that every planar graph $G = (V, E)$ satisfies

$$|E| \leq 3|V|.$$

Indeed, for a connected graph containing at least one cycle, this is a consequence of Euler's formula. The general case then follows by considering connected components separately, and noting that the claimed inequality is trivial for trees.

By definition, every graph G can be made planar by deleting at most $\mathrm{cr}(G)$ many edges. Combining this with the above inequality, we obtain $|E| - \mathrm{cr}(G) \leq 3|V|$, or equivalently

$$\mathrm{cr}(G) \geq |E| - 3|V|. \tag{1}$$

For graphs with many edges, we can use a probabilistic argument to strengthen this bound and to prove the following theorem, originally due to Ajtai, Chv??tal, Newborn and Szemer??di [ACNS], with an independent proof by Leighton [Lei].

Theorem 3 (Crossing number inequality)**.** *Let $G = (V, E)$ be a graph satisfying $|E| \geq 4|V|$. Then*

$$\mathrm{cr}(G) \gtrsim \frac{|E|^3}{|V|^2}.$$

Proof. Fix a drawing of G with $\mathrm{cr}(G)$ many edge crossings. Choose any $p \in (0, 1]$. Now, independently randomly retain each vertex of G with probability p, and discard all other vertices. In this way, we obtain a random graph $G' = (V', E')$ defined as the induced subgraph on the retained vertices. By taking expectations in (1), we see that

$$\mathbb{E}\ \mathrm{cr}(G') \geq \mathbb{E}|E'| - 3\mathbb{E}|V'| = p^2|E| - 3p|V|.$$

Here, we used that an edge is retained if and only if its two endpoints are retained, which occurs with probability p^2. The chosen drawing of G induces a drawing of the subgraph G'. Each edge crossing in the original graph involves four endpoints, so it remains present in G' with probability p^4. Therefore,

$$\mathbb{E}\ \mathrm{cr}(G') \leq p^4 \mathrm{cr}(G).$$

It follows that

$$\mathrm{cr}(G) \geq p^{-2}|E| - 3p^{-3}|V|.$$

Setting $p = 4|V|/|E| \leq 1$, we see that

$$\mathrm{cr}(G) \geq \frac{|E|^3}{64|V|^2}.$$

□

Following the argument of Székely [Sz], we now apply this inequality to incidences between points and lines in $\mathbb{R}^2$ to obtain the Szemerédi–Trotter theorem. Given finite sets P of points and L of lines in $\mathbb{R}^2$, the number of incidences is

$$\#i = |\{(p, l) \in P \times L : p \in l\}|.$$

Trivially, $\#i \leq |P||L|$. The Szemerédi–Trotter theorem improves this bound significantly.

Theorem 4 (Szemerédi–Trotter theorem)**.** *For every finite set P of points and every finite set L of lines in $\mathbb{R}^2$,*

$$\#i \lesssim |P|^{2/3}|L|^{2/3} + |P| + |L|.$$

Proof. We may assume without loss of generality that every line in L contains at least two points of P, since the lines containing at most one point contribute at most $|L|$ incidences, so we can simply discard them.

Construct a graph $G = (V, E)$ with vertex set $V = P$ by connecting every pair of points that lie consecutively on a line in L by an edge. Drawing each edge as the corresponding line segment, we obtain a drawing of G. The graph G is simple, since two distinct points determine at most one line of L.

A line in L containing $k \geq 2$ points contributes $k - 1 \geq k/2$ edges to G. Hence

$$|E| \geq \frac{\#i}{2}.$$

If $\#i < 8|P|$, the claimed inequality is immediate, so we may assume that $\#i \geq 8|P|$. In this case, $|E| \geq 4|P| = 4|V|$, so by the crossing number inequality,

$$\mathrm{cr}(G) \gtrsim \frac{|E|^3}{|V|^2} \gtrsim \frac{\#i^3}{|P|^2}. \tag{2}$$

In the drawing constructed above, two edges can cross only if they lie on different lines of L. Consequently,

$$\mathrm{cr}(G) \leq \binom{|L|}{2} \lesssim |L|^2. \tag{3}$$

Combining (2) and (3) yields

$$\#i \lesssim |P|^{2/3}|L|^{2/3}.$$

Adding back the discarded lines and the alternative case $\#i < 8|P|$ proves the theorem. □

22.3 Elekes's sum-product bound

We now apply the Szemerédi–Trotter theorem to prove Theorem 2. In fact, we will prove the stronger estimate

$$|A + A||AA| \gtrsim |A|^{5/2}$$

for every finite set $A \subseteq \mathbb{R}$.

Proof of Theorem 2. Assume without loss of generality that $0 \notin A$. Let

$$P = (A + A) \times AA \subseteq \mathbb{R}^2,$$

and for $a, a' \in A$, consider the line $l_{a,a'} := \{(x, y) \in \mathbb{R}^2 : y = a(x-a')\}$. These lines are distinct: their slopes determine a, and since $a \neq 0$, the intercept determines a'. Let

$$L = \{l_{a,a'} : a, a' \in A\}.$$

Thus

$$|P| = |A + A||AA|, \qquad |L| = |A|^2.$$

For every $b \in A$, the point $(a' + b, ab)$ belongs to $P \cap l_{a,a'}$. These points are distinct as b varies. Every line $l_{a,a'}$ is therefore incident to at least $|A|$ points of P, and hence

$$\#i \geq |A||L| = |A|^3. \tag{4}$$

By the Szemerédi–Trotter theorem,

$$\#i \lesssim |P|^{2/3}|L|^{2/3} + |P| + |A|^2. \tag{5}$$

We can absorb the last two terms into the first one. Indeed, the trivial bounds on the sumset and product set give

$$|A|^2 \leq |P| \lesssim |A|^4,$$

which implies that $|P| \lesssim |P|^{2/3}|A|^{4/3}$ and $|A|^2 \lesssim |P|^{2/3}|A|^{4/3}$.

Combining (4) and (5) therefore gives

$$|A|^3 \lesssim |P|^{2/3}|A|^{4/3}.$$

Rearranging, we obtain

$$|P| = |A+A||AA| \gtrsim |A|^{5/2}.$$

Finally, $\max\{|A+A|, |AA|\}^2 \geq |A+A||AA|$, and so

$$\max\{|A+A|, |AA|\} \gtrsim |A|^{5/4}.$$

□

References

[Ele] G. Elekes, *On the number of sums and products.* Acta Arith. **81** (1997), no. 4, 365–367.

[Sz] L. A. Székely, *Crossing numbers and hard Erdős problems in discrete geometry.* Combin. Probab. Comput. **6** (1997), no. 3, 353–358.

[ErSz] P. Erdős and E. Szemerédi, *On sums and products of integers,* in Studies in pure mathematics. (1983), 213–218.

[BSSZ] T. F. Bloom, W. Sawin, C. Schildkraut, and D. Zhelezov, *The sum-product conjecture is false for real numbers.* arXiv preprint arXiv:2605.28781 (2026).

[Zh] Y. Zhao, *Graph theory and additive combinatorics.* Cambridge University Press (2023).

[ACNS] M. Ajtai, V. Chvátal, M. Newborn, E. Szemerédi, *Crossing-free subgraphs.* Ann. Discrete Math. **12** (1982), 9–12.

[Lei] F. T. Leighton, *New lower bound techniques for VLSI.* Math. Systems Theory **17** (1984), 47–70.

Pauwel Van Den Eeckhaut, University of Bonn
email: pauwelvde@gmail.com

23 Szemerédi's regularity lemma and the triangle removal

After Y. Zhao [Z]

A summary written by Hanaé Vandanjon

Abstract

We present Szemerédi's regularity lemma, which decomposes a finite graph into a bounded number of approximately regular pairs. We give its energy-increment proof and derive the triangle removal lemma as an application.

23.1 Introduction

In what follows, G denotes a finite undirected graph, $V(G)$ its vertex set and $E(G)$ its edge set. Roughly speaking, Szemerédi's regularity lemma states that the vertex set of any finite graph can be partitioned into a bounded number of parts such that almost all pairs of parts have a nearly uniform distribution of edges. The regularity lemma was introduced by Szemerédi in his proof of Szemerédi's theorem on arithmetic progressions and has since found many applications in graph theory, among which is the classical triangle removal lemma.

Theorem 1 (Triangle removal lemma)**.** *For every $\varepsilon > 0$, there exists $\delta > 0$ such that any graph on n vertices with fewer than δn^3 triangles can be made triangle-free by removing fewer than εn^2 edges.*

23.2 Szemerédi's regularity lemma

23.2.1 Preliminary definitions

In order to state Szemerédi's regularity lemma, we first formalize the notion of nearly uniform edge distribution.

Definition 2. *Let $U, W \subseteq V(G)$. We denote by $e_G(U,W)$ the number of edges between U and W, i.e. $e_G(U,W) := |\{(u,w) \in U \times W \mid uw \in E(G)\}|$. We define the edge density of U and W by $d_G(U,W) := \dfrac{e_G(U,W)}{|U||W|}$.*

Definition 3. *Let $\varepsilon > 0$ and $U, W \subseteq V(G)$. We say that (U, W) is an ε-regular pair if, for all $A \subseteq U$ with $|A| \geq \varepsilon|U|$ and all $B \subseteq W$ with $|B| \geq \varepsilon|W|$, we have*

$$|d_G(A, B) - d_G(U, W)| \leq \varepsilon.$$

The intuition one should have is that the edges between an ε-regular pair are evenly distributed. When (U, W) is not ε-regular, we call a pair $A \subseteq U$, $B \subseteq W$ violating the above condition a *witnessing pair*. We also extend the notion of regularity to partitions of the vertex set.

Definition 4. *A partition $\mathcal{P} = \{V_1, \ldots, V_k\}$ of $V(G)$ is said to be ε-regular if*

$$\sum_{\substack{(i,j) \in \{1,\ldots,k\}^2 \\ (V_i, V_j) \text{ not } \varepsilon\text{-reg}}} |V_i||V_j| \leq \varepsilon |V(G)|^2.$$

In other words, in a regular partition almost all pairs are regular.

23.2.2 Regularity lemma

We now have all the tools required to state Szemerédi's regularity lemma.

Theorem 5 (Szemerédi's regularity lemma)**.** *For every $\varepsilon > 0$, there exists a positive integer M such that every graph admits an ε-regular partition into at most M parts.*

The proof proceeds by repeatedly refining the trivial partition $\{V(G)\}$ until it becomes ε-regular. Here, a partition $\mathcal{Q}$ is said to refine a partition $\mathcal{P}$ if every part of $\mathcal{Q}$ is contained in a part of $\mathcal{P}$. To control the number of refinements, we introduce an energy quantity which increases at each step and is bounded.

Definition 6. *Let $U, W \subseteq V(G)$. We define the energy of U and W by*

$$q(U, W) := \frac{|U||W|}{|V(G)|^2} d_G(U, W)^2.$$

For partitions $\mathcal{P}_U = \{U_1, \ldots, U_k\}$ of U and $\mathcal{P}_W = \{W_1, \ldots, W_l\}$ of W, we define

$$q(\mathcal{P}_U, \mathcal{P}_W) := \sum_{i=1}^{k} \sum_{j=1}^{l} q(U_i, W_j).$$

Finally, for a partition $\mathcal{P}$ of $V(G)$, we set $q(\mathcal{P}) := q(\mathcal{P}, \mathcal{P})$.

Since $0 \leq d_G(U,W) \leq 1$, we have $0 \leq q(\mathcal{P}) \leq 1$ for every partition $\mathcal{P}$ of $V(G)$. We now state three lemmas describing the behavior of the energy under refinement.

Lemma 7 (Energy never decreases under refinement)**.** *Let $U, W \subseteq V(G)$, and let $\mathcal{P}_U$ and $\mathcal{P}_W$ be partitions of U and W, respectively. Then*

$$q(\mathcal{P}_U, \mathcal{P}_W) \geq q(U,W).$$

Sketch of proof. Write $\mathcal{P}_U = \{U_1, \ldots, U_k\}$ and $\mathcal{P}_W = \{W_1, \ldots, W_l\}$. Pick uniformly and independently $x \in U$ and $y \in W$, and let i, j be such that $x \in U_i$ and $y \in W_j$. Setting $Z := d_G(U_i, W_j)$, we have

$$\mathbb{E}[Z] = d_G(U,W), \qquad \mathbb{E}[Z^2] = \frac{|V(G)|^2}{|U||W|} q(\mathcal{P}_U, \mathcal{P}_W).$$

The result follows from Jensen's inequality. □

This lemma implies in particular that if $\mathcal{Q}_U$ refines $\mathcal{P}_U$, then

$$q(\mathcal{Q}_U, \mathcal{P}_W) \geq q(\mathcal{P}_U, \mathcal{P}_W).$$

Lemma 8 (Energy boost for an irregular pair)**.** *If (U,W) is not ε-regular, witnessed by $A \subseteq U$ and $B \subseteq W$, then*

$$q(\{A, U\backslash A\}, \{B, W\backslash B\}) > q(U,W) + \varepsilon^4 \frac{|U||W|}{|V(G)|^2}.$$

The proof follows by estimating $\mathrm{Var}(Z)$ from above and below, with Z defined as before. We now combine the two preceding lemmas to obtain the refinement needed in the proof of the regularity lemma.

Lemma 9 (Energy boost for an irregular partition)**.** *Let $\mathcal{P} = \{V_1, \ldots, V_k\}$ be a partition of $V(G)$ which is not ε-regular. Then there exists a refinement $\mathcal{Q}$ of $\mathcal{P}$ such that every V_i is partitioned into at most 2^{k+1} parts and*

$$q(\mathcal{Q}) > q(\mathcal{P}) + \varepsilon^5.$$

In particular, $\mathcal{Q}$ has at most $k2^{k+1}$ parts.

Sketch of proof. For every pair (V_i, V_j) which is not ε-regular, choose a witnessing pair $(A^{i,j}, B^{i,j})$. By symmetry, we may choose $A^{i,j} = B^{j,i}$. We define $\mathcal{Q}$ by subdividing each V_i according to these subsets: two vertices belong to the same part of $\mathcal{Q}$ if they belong to exactly the same subsets $A^{i,j}$. Thus $\mathcal{Q}$ is a refinement of $\mathcal{P}$. Inside each V_i, at most $k+1$ such subsets are involved, so $\mathcal{Q}$ divides each V_i into at most 2^{k+1} parts. The energy estimate follows from the two preceding lemmas. □

We are now ready to prove the regularity lemma.

Proof of the regularity lemma 5. Start with the trivial partition $\mathcal{P}_0 = \{V(G)\}$ and apply the energy boost lemma whenever the current partition is not ε-regular. Since the energy increases by at least ε^5 at each step and is bounded above by 1, the procedure terminates after at most ε^{-5} steps. Moreover, a partition with k parts is refined into at most $k2^{k+1} \leq 2^{2^k}$ parts. Since we start the trivial partition of size 1, the final ε-regular partition has at most

$$\text{tower}(\lceil 2\varepsilon^{-5} \rceil)$$

parts, where tower(k) denotes the iterated power of 2 of height k. □

23.3 Triangle removal lemma

The proof of the triangle removal lemma proceeds as follows. We first apply the regularity lemma to obtain a regular partition, and then remove the edges between unsuitable pairs, in a sense to be made precised below. A counting lemma will then imply that the resulting graph is triangle-free. We begin with a simple consequence of regularity.

Lemma 10 (Most vertices have roughly the same degree)**.** *Let (U, W) be an ε-regular pair. Then fewer than $\varepsilon|U|$ vertices in U have fewer than $(d_G(U, W) - \varepsilon)|W|$ neighbors in W.*

Theorem 11 (Triangle counting lemma)**.** *Let G be a graph and $U, W, Z \subseteq V(G)$ such that (U, W), (W, Z) and (Z, U) are all ε-regular pairs for some $\varepsilon > 0$. If $d_G(U, W),\ d_G(W, Z),\ d_G(Z, U) \geq 2\varepsilon$, then*

$$|\{(u, w, z) \in U \times W \times Z : \ uwz \text{ triangle in } G\}| \\ \geq (1 - 2\varepsilon)(d_G(U, W) - \varepsilon) \times (d_G(W, Z) - \varepsilon)(d_G(Z, U) - \varepsilon)|U||W||Z|.$$

The proof follows by applying the previous lemma to count common neighbors. In particular, three pairwise ε-regular sets with sufficiently large edge densities contain many triangles. With this in mind, we are ready to prove the triangle removal lemma.

Proof of the triangle removal lemma 1. Let $\varepsilon > 0$ and let G be a graph on n vertices with fewer than δn^3 triangles, with δ to be defined later. Apply the regularity lemma to obtain an $\frac{\varepsilon}{4}$-regular partition $\{V_1, \ldots, V_m\}$.
For every $(i, j) \in \{1, \ldots, m\}^2$, remove all edges between V_i and V_j whenever

$$\text{(a) } (V_i, V_j) \text{ is not } \frac{\varepsilon}{4}\text{-regular, (b) } d_G(V_i, V_j) < \frac{\varepsilon}{2}, \text{ or (c) } \min(|V_i|, |V_j|) < \frac{\varepsilon n}{4m}.$$

Since the partition is $\frac{\varepsilon}{4}$-regular, condition (a) removes at most $\frac{\varepsilon}{4}n^2$ edges. By definition of the edge density, condition (b) removes at most $\frac{\varepsilon}{2}n^2$ edges. Finally, for condition (c), the union of the parts of size less than $\frac{\varepsilon n}{4m}$ has size at most $\frac{\varepsilon n}{4}$, so at most $\frac{\varepsilon}{4}n^2$ edges are removed. Thus, the process removes at most εn^2 edges in total.
Assume that the resulting graph still contains a triangle with vertices in V_i, V_j and V_k. Since none of the three corresponding pairs was removed, the triangle counting lemma gives:

$$|\{(x, y, z) \in V_i \times V_j \times V_k : \ xyz \text{ triangle in } G\}| \geq \left(1 - \frac{\varepsilon}{2}\right)\left(\frac{\varepsilon}{4}\right)^3\left(\frac{\varepsilon n}{4m}\right)^3.$$

Since every triangle is counted at most 6 times as an ordered triple, this contradicts the assumption that G contains fewer than δn^3 triangles whenever

$$\delta < \frac{1}{6}\left(1 - \frac{\varepsilon}{2}\right)\left(\frac{\varepsilon}{4}\right)^3\left(\frac{\varepsilon}{4m}\right)^3.$$

Since m depends only on ε, we may choose $\delta = \delta(\varepsilon) > 0$ satisfying this inequality. Hence the resulting graph is triangle-free, completing the proof. □

References

[Z] Zhao Y., *Graph Theory and Additive Combinatorics: Exploring Structure and Randomness.* Cambridge University Press; 2023.

Hanaé Vandanjon, Nantes Université
email: hanae.vandanjon@univ-nantes.fr

24 On a Conjecture of Marton

After Gowers, Green, Manners, and Tao [GGMT1]

A summary written by Jianghao Zhang

Abstract

The polynomial Freiman-Ruzsa conjecture for bounded-torsion groups was recently resolved in [GGMT1, GGMT2]. Here we give a brief summary of the main proof ideas in the simplest case $G = \mathbb{F}_2^n$. This exposition closely follows [Green, Tao].

24.1 Introduction

The polynomial Freiman-Ruzsa conjecture in $\mathbb{F}_2^n$ is stated as follows:

Theorem 1. *Suppose that $A \subset \mathbb{F}_2^n$ is non-empty with $|A+A| \leq K|A|$. Then there is a subgroup $H \leq \mathbb{F}_2^n$ with $|H| \leq |A|$ such that A can be covered by at most $2K^{O(1)}$ translates of H.*

This result, known as Polynomial Freiman-Ruzsa (or Marton conjecture), was established in [GGMT1]. The point here is that the number of translates we need to cover A is polynomial in the doubling constant K, which is a strong quantitative dependence.

24.2 The key inductive strategy

The Rusza distance between two subsets A, A' of some abelian group (in particular, $\mathbb{F}_2^n$) is defined by

$$d_{\mathrm{com}}[A, A'] := \log \frac{|A - A'|}{|A|^{\frac{1}{2}}|A'|^{\frac{1}{2}}}$$

One may think of $d_{\mathrm{com}}[A, A']$ as a measure of how the additive structure of A close to the additive structure of A'. By constructing a natural injection, one can show the triangle inequality $d_{\mathrm{com}}[A, C] \leq d_{\mathrm{com}}[A, B] + d_{\mathrm{com}}[B, C]$ for Ruzsa distance.

The condition of Theorem 1 says $d_{\mathrm{com}}[A, A]$ is small. (Here we used $A = -A$ in $\mathbb{F}_2^n$, but even in general $d_{\mathrm{com}}[A, A]$ and $d_{\mathrm{com}}[A, -A]$ are not so

different at the heuristic level.) While the conclusion of Theorem 1 may be informally stated as finding a subgroup H such that $d_{\mathrm{com}}[A, H]$ (i.e. close to A in the additive-structure sense).

The basic proof strategy is induction on the doubling constant. The base case is $|A+A| \approx |A|$. In this case, we can deduce that A is indeed a subspace of $\mathbb{F}_2^n$, so the conclusion holds.

In each induction step, we hope to find some A' such that

1. The additive structure of A' is close to that of A. (e.g. $d_{\mathrm{com}}[A, A'] = O(\log K)$.)

2. A' has a smaller doubling constant than A. (e.g. $|A'+A'| \leq K^{0.99}|A'|$.)

Suppose we can do so. Then iterating this process gives us a sequence of sets A_n such that $A_{n+1} = A_n'$. Since the doubling constant $K_{n+1} \leq K_n^{0.99}$, the base case implies that $A_N = H$ for some large N. Then for $A = A_0$ we may estimate

$$\begin{aligned} d_{\mathrm{com}}[A, H] &\leq d_{\mathrm{com}}[A, A_1] + d_{\mathrm{com}}[A_1, A_2] + \cdots \\ &\lesssim \log K + \log K^{0.99} + \log(K^{0.99})^{0.99} + \cdots \lesssim \log K. \end{aligned}$$

Then one can show (via the above inequality) H is roughly what we want.

How to find the 'improving' A' from A? We have the following two motivating examples:

1. Suppose A is a random subset of a subspace H with $|H| \approx K^2$ such that each element of H is included with probability K^{-1}. In this case, $|A|$ is typically K. Moreover, $P(x \in A + A) \approx \sum_{(y,z):y+z=x} P(y \in A, z \in A) = |H|K^{-2} \approx 1$, so $A + A \approx H$. Thus A has doubling constant roughly K, but $A + A$ has doubling constant roughly 1. It is a good idea to using $A + A$ as the 'improving'.

2. Suppose $A = \bigsqcup_{j=1}^{K}(a_j + H)$ for **suitablely selected** $(a_j)_{1\leq j\leq K}$. Then $A + A = \bigsqcup_{1\leq j\leq j'\leq K}^{K}(a_j + a_{j'} + H)$, so A has doubling constant $\sim K$. However, $A \cap (A + y)$ with $y = a_1 - a_2$ (say) is $a_1 + H$. And $a_1 + H$ has doubling constant 1. Hence is a good idea to using $A \cap (A + y)$ as the 'improving' in this case.

It is possible to build more complicated examples. For example, we may take a subspace H with $|H| \approx K$ and let A be a random subset of $\bigsqcup_{j=1}^{K^{\frac{1}{2}}}(a_j +$

H) such that each element is included with probability $K^{-\frac{1}{2}}$. Then A has doubling constant roughly K. Moreover, $A+A \approx \bigsqcup_{1\leq j\leq j'\leq K^{\frac{1}{2}}}^{K}(a_j+a_{j'}+H)$, so that $A+A$ has doubling constant roughly K as well. We see replacing A with $A+A$ does not improve the doubling constant. However, this replacement changes the shape of A into the form of the second example above, so further replacing $A+A$ with $(A+A)\cap(A+A+y)$ can improve the doubling constant. These examples suggest our induction strategy: Show one of the above two operations or a more complicated combination of them can give us the 'improving' A'.

24.3 Entropy notion

However, this strategy is not easy to made rigorous using the above traditional doubling constant and Ruzsa distance. This is because these traditional notions do not behave well under homomorphism projections. By homomorphism projection we mean some map like $\pi : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2^n$ defined by $\pi(x,y) := x+y$. We need such π to construct our 'improving' candidates. The solution is to work with the entropic analogue of doubling constant/Ruzsa distance.

Definition 2. *Let X, Y be $\mathbb{F}_2^n$-valued random variables. Define their entropic Ruzsa distance as*

$$d[X,Y] := H(X'-Y') - \frac{1}{2}H(X') - \frac{1}{2}H(Y'),$$

*where X', Y' are **independent** copies of X, Y respectively.*

One can reformulate Theorem 1 and the above induction strategy using the language of entropic Ruzsa distance. In particular, our goal becomes: Given $\mathbb{F}_2^n$-valued random variable X, find X' such that $d[X',X'] \leq (1-c_0)d[X,X]$ and $d[X,X'] \leq C_0 d[X,X]$. Here c_0 is a small positive constant, and C_0 is a large positive constant.

The advantage of using entropic Ruzsa distance lies in the following (corollary of) fibring identity:

Lemma 3. *Let Z_1, Z_2, Z_3, Z_4 be independent copies of X. Then*

$$\begin{aligned} 2d[Z_1,Z_2] = d[Z_1+Z_2, Z_1+Z_2] + d[Z_1|Z_1+Z_2, Z_1|Z_1+Z_2] \\ + I(Z_1+Z_2 : Z_2+Z_4 | Z_1+Z_2+Z_3+Z_4). \end{aligned}$$

Here we would like to set $X' := Z_1 + Z_2$ or $X' := (Z_1|Z_1 + Z_2)$ as our 'improving'. If neither of them satisfies our requirements, then Lemma 3 implies $I(Z_1 + Z_2 : Z_2 + Z_4|Z_1 + Z_2 + Z_3 + Z_4)$ must be small. In this case, we enter the endgame.

24.4 Endgame

Smallness of $I(Z_1 + Z_2 : Z_2 + Z_4|Z_1 + Z_2 + Z_3 + Z_4)$ means $Z_1 + Z_2, Z_2 + Z_4$ are 'almost independent' conditioned on $Z_1 + Z_2 + Z_3 + Z_4$. Suppose they are independent for this heuristic discussion. We then make the crucial use of the fact that our group has a bounded characteristic (actually equals to 2): $(Z_1 + Z_2) + (Z_2 + Z_4) = Z_1 + Z_4$. This means the independent sum of two copies of $X' := Z_1 + Z_2$ has the same distribution as X' itself. (The independent sum is $(Z_1 + Z_2) + (Z_2 + Z_4) = Z_1 + Z_4$. Note that Z_1, Z_2, Z_3, Z_4 are independent copies of X.) Thus $d[X', X'] = 0$, which is definitely a satisfactory 'improving'.

To made this discussion rigorous, we need the following entropic analogue of Balog–Szemeredi–Gowers.

Lemma 4. *Let A, B be (not necessarily independent) G-valued random variables. Let $Z := A + B, A'_z := (A|Z = z), B'_z := (B|Z = z)$. Then*

$$\sum_z p_Z(z) d[A'_z, B'_z] \leq 3I(A : B) + 2\big(H(A + B) - \frac{1}{2}H(A) - \frac{1}{2}H(B)\big).$$

Note that the second term of RHS may not equal to d[A, B] since A, B may not be independent.

Lemma 4 roughly says we can find A', B' with small entropic Ruzsa distance from A, B with small $I(A : B)$ and small $H(A+B) - \frac{1}{2}H(A) - \frac{1}{2}H(B)$. To see how this lemma relates to the ordinary BSG, let's recall the following form of ordinary BSG [TV].

Lemma 5. *Let A, B be subsets of an abelian group G. Let E be a subset of $A \times B$ (i.e. a subset of all edges between A and B). Define $A \overset{E}{+} B := \{a + b : (a, b) \in E\}$. Suppose*

$$|E| \geq \frac{|A||B|}{K}, \quad |A \overset{E}{+} B| \leq K|A|^{\frac{1}{2}}|B|^{\frac{1}{2}}.$$

Then we can find $A' \subset A, B' \subset B$ *with* $|A'| \gtrsim K^{-1}|A|, |B'| \gtrsim K^{-1}|B|$ *such that*

$$|A' + B'| \lesssim K^{O(1)}|A'|^{\frac{1}{2}}|B'|^{\frac{1}{2}}.$$

Compare Lemma 4 and Lemma 5. One may regard $|E| \geq \frac{|A||B|}{K}$ as smallness of $I(A : B)$ (almost independence), $|A \overset{E}{+} B| \leq K|A|^{\frac{1}{2}}|B|^{\frac{1}{2}}$ as smallness of $H(A + B) - \frac{1}{2}H(A) - \frac{1}{2}H(B)$ (small 'dependent doubling'), and $|A' + B'| \lesssim K^{O(1)}|A'|^{\frac{1}{2}}|B'|^{\frac{1}{2}}$ as smallness of some $d[A'_z, B'_z]$ (small doubling).

Let's end this summary with one final remark: We take all the distributions to be the same as X in Lemma 3 and many heuristic discussions. In the rigorous proof, one has to work with X, Y with different distributions due to some polarization argument to remove conditioning.

References

[GGMT1] W. T. Gowers, Ben Green, Freddie Manners, and Terence Tao, *On a conjecture of Marton.* Ann. of Math. **201** (2), 515–549, 2025.

[GGMT2] W. Timothy Gowers, Ben J. Green, Freddie Manners, and Terence Tao, *Marton's conjecture in abelian groups with bounded torsion.* Ann. Fac. Sci. Toulouse Math. **6** (2000), 35(1), 1-33, 2026.

[Green] Ben Green, *Additive Combinatorics* , 2025

[Tao] Terence Tao, *On a conjecture of Marton* , 2023

[TV] Terence Tao and Van H. Vu, *Additive combinatorics,* volume 105 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, paperback edition, 2010.

Jianghao Zhang, University of Edinburgh
email: J.Zhang-395@sms.ed.ac.uk

25 Quasipolynomial inverse theorem for the $U^4(\mathbb{F}_p^n)$ norm

After L. Milićević [Mil24+]

A summary written by Daniel Zhu

Abstract

We summarize the proof of the quasipolynomial inverse theorem for the U^4 Gowers norm over finite field vector spaces.

25.1 Background

For an integer $k \geq 2$, a finite abelian group G, and a function $f\colon G \to \mathbb{C}$, the U^k *Gowers norm* is defined as

$$\|f\|_{U^k} = \left(\mathbb{E}_{x,h_1,\dots,h_k \in G}[\Delta_{h_1}^* \cdots \Delta_{h_k}^* f(x)]\right)^{1/2^k},$$

where $\Delta_h^* f(x) = f(x+h)\overline{f(x)}$ is the multiplicative (discrete) derivative. A central question of higher-order Fourier analysis asks what structure we can find in a function f with $|f(x)| \leq 1$ (call such f 1-*bounded*) and $\|f\|_{U^k} \geq \delta$. Generally speaking, we wish to find a family of 1-bounded functions $g(x)$ such that $\|g\|_{U^k}$ is large for "well-understood reasons" and such that f correlates with g, i.e. $|\mathbb{E}_{x \in G} f(x)\overline{g(x)}| \geq \delta'$ for some δ' depending on δ. A result of this form of this kind is known as a "Gowers inverse theorem" and lies at the core of most applications of higher-order Fourier analysis.

Henceforth, we let G be an $\mathbb{F}_p$-vector space, where p is a fixed prime that all constants may depend on. In this case, work of Bergelson, Tao, and Ziegler [BTZ10, TZ12] showed that that we may take g to be a *non-classical degree-*$(k-1)$ *polynomial phase*, a 1-bounded function with $\|g\|_{U^k} = 1$. Unraveling definitions, such $g(x)$ are exactly those of the form $e^{2\pi i P(x)}$ where $P\colon G \to \mathbb{R}/\mathbb{Z}$ satisfies $\Delta_{h_1} \cdots \Delta_{h_k} P(x) = 0$, where $\Delta_h f(x) = f(x+h) - f(x)$ is the discrete derivative. Such P are well-understood, and the name comes from the fact that if $p > k-1$, all such P are of the form $\alpha + Q(x)/p$ where $\alpha \in \mathbb{R}/\mathbb{Z}$ is a constant and $Q\colon G \to \mathbb{F}_p$ is a polynomial of degree at most $k-1$. If $p \leq k-1$ there are extra "non-classical" solutions.

There has been considerable work towards proving quantitative bounds for δ' in terms of δ (absent in [BTZ10, TZ12]), with the ultimate conjecture

that δ' should be bounded by a polynomial in δ. The first nontrivial case is the U^3 norm, where polynomial bounds were recently shown by Gowers, Green, Manners, and Tao in 2023 [GGMT25, GGMT26]. For the U^4 norm, the best-known bound, due to Milićević [Mil24+], is quasipolynomial and the subject of these notes. (Here, δ' is *quasipolynomial* in δ if $\delta' = \exp(-\log(1/\delta)^{O(1)})$, or alternatively, if $\log(1/\delta')$ is bounded by a polynomial in $\log(1/\delta)$.) For higher norms, bounds with a bounded number of exponentials are known in the high-characteristic case [GM20+].

25.2 Philosophical remarks

A common starting point in the proofs of Gowers inverse theorems is the observation that if $f\colon G \to \mathbb{C}$ is 1-bounded and $\|f\|_{U^k} \geq \delta$, we must have

$$\delta^{O(1)} \leq \|f\|_{U^k}^{2^k} = \mathbb{E}_{a\in G}\|\Delta_a^* f\|_{U^{k-1}}^{2^{k-1}}, \tag{1}$$

which implies that $\Delta_a^* f$ must have high U^{k-1} norm for many values of a. Applying the U^{k-1} inverse theorem as an inductive hypothesis, we conclude that for many a there must exist a function $g_a(x)$ that correlates with $\Delta_a^* f(x)$. The relations between $\Delta_a^* f$ for different values of a give constraints on g_a; namely, a Cauchy-Schwarz argument shows that $|\mathbb{E}_{x\in G} g_a(x) g_b(x) \overline{g_c(x) g_d(x)}|$ is large for many quadruples (a,b,c,d) with $a+b=c+d$. The remaining task is to "integrate" this information into a global function g that correlates with f.

In the U^3 case, these g_a are Fourier characters, elements of the Pontryagin dual $\widehat{G}$ of G. As $\widehat{G}$ is just an abelian group, this integration step falls in the realm of classical additive combinatorics.

For the U^4 norm, the g_a are given by quadratic phases, which, roughly speaking, are the same as matrices. Thus, the proof of the U^4 theorem can be viewed as the development and application of an "additive combinatorics of matrices". To accomplish this, the fundamental tools of additive combinatorics are used both as logical inputs and to motivate certain arguments.

25.3 Proof overview

In what follows, we let δ' be shorthand for a quantity that is quasipolynomial in δ. We aim to show the following theorem:

Theorem 1 (Milićević [Mil24+]). *Let $f\colon G \to \mathbb{C}$ be a 1-bounded function with $\|f\|_{U^4} \geq \delta$. Then there is a non-classical cubic polynomial $P\colon G \to \mathbb{R}/\mathbb{Z}$ with*

$$\left|\mathbb{E}_{x\in G}[f(x)e^{-2\pi iP(x)}]\right| \geq \delta'.$$

An important ingredient will be a robust version of Sanders's Bogolyubov-Rusza lemma.

Lemma 2 (Schoen-Sisask [SS16]). *Let G be a finite-dimensional $\mathbb{F}_p$-vector space and let $A \subseteq G$ have size $\delta|G|$. Then there is a subspace $U \subseteq G$ of size at least $\delta'|G|$ such that for each $u \in U$, there are at least $\delta'|G|^3$ quadruples $(a, b, c, d) \in A^4$ with $a + b - c - d = u$.*

25.3.1 Part I: Inductive hypothesis and Cauchy-Schwarz

Suppose $f\colon G \to \mathbb{C}$ is 1-bounded and satisfies $\|f\|_{U^4} \geq \delta$. By (1) and the U^3 inverse theorem, we find that there exists a set $A \subseteq G$ with size at least $\delta'|G|$ such that $\Delta_a^* f$ correlates with a quadratic phase $g_a(x)$. For such a, let $\phi_a\colon G \to \widehat{G}$ be the linear map such that $\Delta_b^* g_a(x)$ is a multiple of $\phi_a(b)$. A Cauchy-Schwarz argument shows that for at least $\delta'|G|^3$ quadruples $(a, b, c, d) \in A^4$ with $a + b = c + d$, we have

$$\operatorname{rank}(\phi_a + \phi_b - \phi_c - \phi_d) \leq \log(1/\delta').$$

25.3.2 Part II: Some quadruples to all quadruples in a subspace

This part follows from an adaptation of the Balog-Szemerédi-Gowers theorem to the case of matrices, which Milićević calls an "abstract Balog-Szemerédi-Gowers argument". One key difference, however, is that while in the original theorem one passes immediately to a set of popular differences, the fact that $\phi_a + \phi_b - \phi_c - \phi_d$ is not exactly zero means that the proof must keep track of raw multiplicities for longer. The end result of a series of graph-theoretic arguments is that, after possibly shrinking A and δ', the set

$$\{\phi_{a_1} - \phi_{a_2} + \cdots - \phi_{a_{16}} \mid (a_1, \ldots, a_{16}) \in A^{16}, a_1 - a_2 + \cdots - a_{16} = 0\}$$

can be covered by at most $1/\delta'$ "balls" of the form $\{\phi \mid \operatorname{rank}(\phi - \psi) \leq \log(1/\delta')\}$.

A probabilistic argument allows us to shrink A further in a way that throws out all the balls that do not consist of low-rank matrices, meaning

that we may assume that $\operatorname{rank}(\phi_{a_1} - \phi_{a_2} + \cdots - \phi_{a_{16}}) \leq \log(1/\delta')$. Finally, applying Lemma 2 allows us to replace A with a subspace H of size at least $\delta'|G|$ such that $\operatorname{rank}(\phi_a + \phi_b - \phi_c - \phi_d) \leq \log(1/\delta')$ for all $(a, b, c, d) \in H^4$ with $a + b - c - d = 0$.

25.3.3 Part III: Low rank to sometimes zero on a bilinear variety

The next step is to remove noise from the ϕ_a. First, for some $r = \log(1/\delta')$, pick a random map $\pi\colon \widehat{G} \to \mathbb{F}_p^r$ and set $U_a = \ker(\pi \circ \phi_a)$. Under some pseudorandomness assumptions (which if false would lead to a quick finish), π can be chosen so that

$$\phi_{a_1} - \phi_{a_2} + \cdots - \phi_{a_{2k}}|_{U_{a_1} \cap \cdots \cap U_{a_{2k}}} = 0 \tag{2}$$

for at least $(1 - \varepsilon)|H|^{2k-1}$ choices of $(a_1, \ldots, a_{2k}) \in H^{2k}$ with $a_1 - a_2 + \cdots - a_{2k} = 0$, where ε is some absolute constant (2^{-24} in the paper) and k is fixed.

Then, for each $a \in H$, pick $x_a, y_a \in H$ at random. By (2) for $k = 2$, the map $\phi_{x_a+a} - \phi_{x_a}$ will likely be equal to $\phi_{y_a+a} - \phi_{y_a}$ on $U_{x_a+a} \cap U_{x_a} \cap U_{y_a+a} \cap U_{y_a}$, in which case they can be "glued together" to yield a map ψ_a on $W_a = (U_{x_a+a} \cap U_{x_a}) + (U_{y_a+a} \cap U_{y_a})$. Moreover, by (2) for $k = 4$, it is also likely that $\psi_a + \psi_b - \psi_c - \psi_d|_{W_a \cap W_b \cap W_c \cap W_d} = 0$ for $(a, b, c, d) \in H^4$ with $a + b = c + d$.

The effect of this maneuver is that the only possible W_a that can result with non-negligible probability are highly structured; specifically, in expectation, there exists a form $\beta\colon H \times G \to \mathbb{F}_p^d$, affine in H and linear in G, for $d = \log(1/\delta')$, for which $B_a = \ker \beta(a, -) \subseteq W_a$ for at least $\delta'|H|$ values of $a \in H$. We can then fix a choice $(x_a, y_a)_{a \in H}$ such that $\psi_a + \psi_b - \psi_c - \psi_d|_{B_a \cap B_b \cap B_c \cap B_d} = 0$ for $(a, b, c, d) \in H^4$ for at least $\delta'|H|^3$ choices of $(a, b, c, d) \in H^4$ with $a + b = c + d$.

25.3.4 Part IV: Some quadruples to all quadruples again

This part largely mirrors part II, running the abstract Balog-Szemerédi-Gowers argument again. One new difficulty, however, is that we are now concerned with a non-transitive relation; namely

$$\psi_a - \psi_b - \psi_c + \psi_d|_{B_a \cap B_b \cap B_c \cap B_d} = 0 \text{ and } \psi_c - \psi_d - \psi_e + \psi_f|_{B_c \cap B_d \cap B_e \cap B_f} = 0$$

does not imply that

$$\psi_a - \psi_b - \psi_e + \psi_f|_{B_a \cap B_b \cap B_e \cap B_f} = 0.$$

To solution to this issue is a notion of "robust transitivity"; indeed, the previous implication is true if the conditions are met for enough (c, d). This argument requires an algebraic regularity lemma to split β into "high-rank" and "low-rank" parts, as well as a counting lemma. This part of the proof is also where the robust part of Lemma 2 is used.

25.3.5 Part V: The finish

The remainder of the proof follows previously known arguments. Once we understand that ψ_a is affine, we can go backwards and show that there is some affine Φ such that $\operatorname{rank}(\phi_a - \Phi(a)) \leq \log(1/\delta')$ for $\delta'|G|$ values of $a \in A$. An application of the U^3 inverse theorem again yields a trilinear $\alpha\colon G \times G \times G \to \mathbb{F}_p$ such that $|\mathbb{E}_{a,b,c,x\in G}[\Delta_a^*\Delta_b^*\Delta_c^* f(x)e^{-2\pi i\alpha(a,b,c)/p}]| \geq \delta'$. A symmetrization argument of Tidor [Tid22] bounds this by a product of expressions of the form $\|f(x)e^{-2\pi iP(x)/p}\|_{U^3}$ where P is a non-classical cubic polynomial. A final application of the U^3 inverse theorem finishes.

References

[BTZ10] V. Bergelson, T. Tao and T. Ziegler, *An inverse theorem for the uniformity seminorms associated with the action of $\mathbb{F}_p^\infty$.* Geom. Funct. Anal. 19 (2010), pp. 1539–1596.

[GGMT25] W. T. Gowers, B. Green, F. Manners, and T. Tao, *On a conjecture of Marton.* Ann. Math. 201 (2025), pp. 515–549.

[GGMT26] W. T. Gowers, B. G. Green, F. Manners, and T. Tao, *Marton's conjecture in abelian groups with bounded torsion.* Ann. Fac. Sci. Toulouse Math. 35 (2026), pp. 1–33.

[GM20+] W. T. Gowers and L. Milićević, *An inverse theorem for Freiman multi-homomorphisms.* arXiv: 2002.11667.

[Mil24+] L. Milićević, *Quasipolynomial inverse theorem for the $\mathsf{U}^4(\mathbb{F}_p^n)$ norm.* arXiv: 2410.08966.

[SS16] T. Schoen and O. Sisask, *Roth's theorem for four variables and additive structures in sums of sparse sets.* Forum Math. Sigma 4 (2016), e5.

[TZ12] T. Tao and T. Ziegler, *The inverse conjecture for the Gowers norm over finite fields in low characteristic.* Ann. Comb. 16 (2012), pp. 121–188.

[Tid22] J. Tidor, *Quantitative bounds for the U^4-inverse theorem over low characteristic finite fields.* Discrete Anal. 2022:14.

Daniel Zhu, Princeton University
email: zhd@princeton.edu