MATH 233A: Additive Combinatorics

Sarah Peluse

February 2, 2025

Contents

1 Logistics

1. I've made a canvas page with the syllabus, and I'll also post the assignments there. If you want to be added to the canvas page, send me an email.

2. Email: `speluse@stanford.edu`

3. There will be a final project, which consists of you finding a paper in additive combinatorics, reading it on your own, and then discussing it with me. The meetings will happen during the last week of class, and I'll make an online form to schedule them soon. I'll often mention topics and results that we will not have the time to go into, so that can help you find papers to read, but you don't have to pick something that I mention.

2 Roth's theorem

One of the first ever results in Ramsey theory is van der Waerden's theorem.

Theorem 1 (van der Waerden, 1927)**.** *Let $k, r \in \mathbf{N}$. For any r-coloring of the natural numbers, some color class contains a nontrivial k-term arithmetic progression*

$$x, x+y, \dots, x+(k-1)y \qquad (y \neq 0).$$

Motivated by this result, Erdős and Turán conjectured in 1936 that any subset of $\mathbf{N}$ with positive upper density[1] must contain arbitrarily long nontrivial arithmetic progressions. That is, van der Waerden's theorem should hold because, in any finite coloring of the natural numbers, some color class must have positive density.

Conjecture 1 (Erdős and Turán, 1936)**.** *If $A \subset \mathbf{N}$ has positive upper density, then A contains arbitrarily long nontrivial arithmetic progressions.*

This conjecture can be restated in a more finitary form.

Exercise 1. Prove that the above statement is equivalent to the following: Fix $k \in \mathbf{N}$. If $A \subset [N]$ contains no nontrivial k-term arithmetic progressions, then $|A| = o(N)$.

[1] If $A \subset \mathbf{N}$, its density in the first N integers $[N] := \{1, \dots, N\}$ is $\frac{|A \cap [N]|}{N}$. If the limit of this quantity exists, then the limit is called the *natural density* of A. Its limsup and liminf always exist. We call $\limsup_{N\to\infty} \frac{|A\cap[N]|}{N}$ and $\liminf_{N\to\infty} \frac{|A\cap[N]|}{N}$ the upper and lower densities, respectively, of A. So, A has positive upper density if $\limsup_{N\to\infty} \frac{|A\cap[N]|}{N} > 0$.

In 1975, Szemerédi proved this conjecture via a purely combinatorial argument in which he introduced his regularity lemma for graphs, which is now a fundamental tool in extremal graph theory. The first nontrivial case of Szemerédi's theorem is the case of three-term arithmetic progressions, which was proved more than twenty years earlier by Roth.

Theorem 2 (Roth, 1953)**.** *If $A \subset [N]$ contains no nontrivial three-term arithmetic progressions, then*

$$|A| \ll \frac{N}{\log\log N}.$$

Here and throughout the course, I use Vinogradov's asymptotic notation: for two quantities X and Y, $X \ll Y$, $Y \gg X$, and $X \asymp Y$ mean $X = O(Y)$, $Y = \Omega(X)$, and $X \ll Y \ll X$, respectively.

One of the central problems in additive combinatorics has been to determine the best possible bounds in Roth's theorem, i.e., to determine the size of the largest subset of $[N]$ containing no nontrivial three-term arithmetic progressions. Attempts to improve these bounds have led to the development of numerous new ideas and techniques. Here is a brief history of progress that has been made:

Roth (1953)	$\frac{N}{\log\log N}$
Heath-Brown (1987) and Szemerédi (1990)	$\frac{N}{(\log N)^c}$
Bourgain (1999)	$\frac{N}{(\log N)^{1/2-o(1)}}$
Bourgain (2008)	$\frac{N}{(\log N)^{2/3-o(1)}}$
Sanders (2012)	$\frac{N}{(\log N)^{3/4-o(1)}}$
Sanders (2011)	$\frac{N(\log\log N)^6}{\log N}$
Bloom (2016)	$\frac{N(\log\log N)^4}{\log N}$
Schoen (2021)	$\frac{N(\log\log N)^{3+o(1)}}{\log N}$
Bloom–Sisask (2020)	$\frac{N}{(\log N)^{1+c'}}$
Kelley–Meka (2023)	$\frac{N}{\exp\left(C\sqrt[12]{\log N}\right)}$
Bloom–Sisask (2023)	$\frac{N}{\exp\left(C'\sqrt[9]{\log N}\right)}$

Here, c and c' are small positive constants, C and C' are some positive constants, the $-o(1)$ in the exponent of $\log N$ in the third, fourth, and fifth rows hides bounded powers of $\log\log N$ in the numerator, and the $o(1)$ in the exponent of $\log\log N$ in the last row hides a bounded power of $\log\log\log N$. I'll say a bit about what goes into some of the more important improvements after we prove Roth's theorem.

Note that the subset of integers below N with no twos in their ternary expansion,

$$A_N := \left\{ n \in [N] : n = \sum_{i=0}^{k} a_i 3^i \text{ with } a_i \in \{0,1\} \right\},$$

contains no three-term arithmetic progressions (you can see this by considering the first nonzero digit of the common difference). This set has size $|A_N| \gg N^{\log 2/\log 3}$. In 1942, Salem and Spencer constructed subsets of $[N]$ of size $N\exp(-C\log N/\log\log N)$ containing

no three-term arithmetic progressions for some constant $C > 0$, showing for the first time that there exist three-term progression free sets of size $\gg_\varepsilon N^{1-\varepsilon}$ for all $\varepsilon > 0$; due to this, sets free of three-term arithmetic progressions are sometimes called *Salem–Spencer sets*. A famous construction of Behrend from 1946, which we will see later, gives a three-term progression free subset of $[N]$ of size $\gg N \exp\bigl(-(2\sqrt{2} + o(1))\sqrt{\log_2 N}\bigr)$. Over the past 80 years, there were various minor improvements on Behrend's construction that improved the $o(1)$ term, but no one was able to improve the $2\sqrt{2}$ until very recently, when Elsholtz, Hunter, Proske, and Sauermann introduced a new construction of a three-term progression free set in $[N]$ of size $\gg N \exp\Bigl(-(2\sqrt{\log_2(24/7)} + o(1))\sqrt{\log_2 N}\Bigr)$. One can check that $2\sqrt{\log_2(24/7)} < 2\sqrt{2}$.

2.1 Notation

Before we begin proving Roth's theorem, we will fix some standard notation for the rest of the course. For any real number t and natural number q, we will write $e(t) := e^{2\pi i t}$ and $e_q(t) := e(t/q)$, the latter of which can also naturally be viewed as a function on $\mathbf{Z}/q\mathbf{Z}$. We will use $\|t\|$ to denote the distance from t to the nearest integer.

For any finite nonempty set X and $f : X \to \mathbf{C}$, we denote the average of f over X by

$$\mathbf{E}_{x \in X} f(x) := \frac{1}{|X|} \sum_{x \in X} f(x).$$

Sometimes, when the averaging set is implicitly understood, I will simply write $\mathbf{E}_x$ in place of $\mathbf{E}_{x \in X}$. When G is any finite abelian group, we define the L^p and ℓ^p norms of $g : G \to \mathbf{C}$ with the normalizations

$$\|g\|_{L^p} := (\mathbf{E}_{x \in G} |g(x)|^p)^{1/p} \qquad \text{and} \qquad \|g\|_{\ell^p} := \left(\sum_{x \in G} |g(x)|^p \right)^{1/p}.$$

When $h : G \to \mathbf{C}$ as well, we define the inner product on G by

$$\langle g, h \rangle := \mathbf{E}_{x \in G} g(x) \overline{h(x)}$$

and convolution by

$$(g * h)(x) := \mathbf{E}_{y \in G} g(x - y) h(y).$$

Denote the dual group of G (i.e., the set of characters of G) by $\widehat{G}$. Recall that $\widehat{\mathbf{Z}/q\mathbf{Z}} \cong \mathbf{Z}/q\mathbf{Z}$ for any $q \in \mathbf{N}$ and

$$\widehat{\mathbf{Z}/q\mathbf{Z}} = \{x \mapsto e_q(\xi x) : \xi \in \mathbf{Z}/q\mathbf{Z}\}.$$

For any $\xi \in \widehat{G}$, we define the Fourier coefficient of g at ξ by

$$\hat{g}(\xi) := \mathbf{E}_{x \in G} g(x) \overline{\xi(x)}.$$

With our choice of normalizations, the Fourier inversion formula reads

$$g(x) = \sum_{\xi \in \widehat{G}} \hat{g}(\xi) \xi(x),$$

Plancherel's theorem reads

$$\langle g, h\rangle = \mathbf{E}_{x\in G} g(x)\overline{h(x)} = \sum_{\xi\in\widehat{G}} \hat{g}(\xi)\overline{\hat{h}(\xi)},$$

and we have $\widehat{g * h} = \hat{g}\cdot\hat{h}$.

2.2 Proof of Roth's theorem

Roth's original argument was via a downward induction argument on density. The modern way to phrase this argument (due to Gowers) is in the form of a *density increment argument*. The idea of the argument is to show that if a subset $A \subset [N]$ of density α does not contain many three-term arithmetic progressions, then either N is small in terms of α (precisely, $N \ll \frac{1}{\alpha^{O(1)}}$, which implies that the density of A is very small in terms of N) or A must have density substantially larger than α on a "structured" subset of $[N]$ or N. This structured subset resembles $[N]$ enough that one can repeat the argument relative to it, but now with a set having density greater than α. One iterates this repeatedly, and since density cannot be greater than 1, the iteration must terminate at some point; at this point, the structured set must be small. We can then retrace the steps of the iteration to derive an upper bound for the original density α in terms of N.

Here is our density increment lemma.

Lemma 1. *Let $A \subset [N]$ have density α and assume that A contains no nontrivial three-term arithmetic progressions. Then either*

$$N < \frac{8}{\alpha^2},$$

or there exists an arithmetic progression $P = a + q\cdot[N']$ with $N' \geq 2^{-11}\alpha^2\sqrt{N}$ such that

$$\frac{|A\cap P|}{|P|} \geq \alpha + 2^{-10}\alpha^2.$$

We will show how iterating this lemma proves Roth's theorem.

Theorem 3. *If $A \subset [N]$ contains no nontrivial three-term arithmetic progressions, then*

$$|A| \ll \frac{N}{\log\log N}.$$

Proof assuming Lemma 1. First, note that three-term arithmetic progressions are translation-dilation invariant, so that if A contains no nontrivial arithmetic progressions and $P = a + q\cdot[N']$, then the shifted and rescaled set

$$A' := \{n \in [N'] : a + qn \in A\}$$

also contains no nontrivial three-term arithmetic progressions.

Now, suppose that $A \subset [N]$ has density α and contains no nontrivial three-term arithmetic progressions. Set $A_0 := A$, $N_0 := N$, and $\alpha_0 := \alpha$. Repeatedly applying the density-increment lemma produces a sequence of triples (A_i, N_i, α_i) satisfying

1. $A_{i+1} \subset [N_{i+1}]$ has density α_{i+1},
2. A_{i+1} contains no nontrivial three-term arithmetic progressions,
3. $N_{i+1} \geq 2^{-11}\alpha_i^2\sqrt{N_i}$, and
4. $\alpha_{i+1} \geq \alpha_i + 2^{-10}\alpha_i^2$,

all provided that $N_i \geq 8\alpha_i^{-2}$. Since the density of A_{i+1} in $[N_{i+1}]$ cannot go above 1, this iteration must terminate at some step $i_0 \ll \frac{1}{\alpha}$ by the inequality $\alpha_{i+1} \geq \alpha_i + 2^{-10}\alpha_i^2$. Indeed, this inequality says that

$$\alpha_i \geq \alpha \left(1 + \frac{\alpha}{2^{10}}\right)^i,$$

so that $\alpha_i \geq 2\alpha$ after $i \leq \frac{2^{10}}{\alpha}$ steps of the iteration, and, analogously, α_i doubles again after at most $\frac{2^{10}}{2\alpha}$ more steps of the iteration, and so on.

At the point of termination i_0, the largeness assumption on N_{i_0} must fail, so that we have $N_{i_0} \leq 8\alpha^{-2}$. On the other hand, we have

$$N_{i_0} \geq \left(2^{-11}\alpha^2\right)^{\sum_{i=0}^{i_0} 2^{-i}} N^{2^{-i_0}} \geq 2^{-22}\alpha^4 N^{2^{-O(1/\alpha)}}.$$

Combining these upper and lower bounds yields

$$N^{2^{-O(1/\alpha)}} \ll \frac{1}{\alpha^6}.$$

Taking double logarithms of both sides (when $N \gg 1$) and rearranging yields Roth's theorem. □

Before proving the density increment lemma, we will recall a standard result in Diophantine approximation due to Dirichlet.

Theorem 4. *Let $\gamma \in \mathbf{R}$ and $Q \in \mathbf{N}$. There exist integers a and $1 \leq q \leq Q$ such that*

$$\left|\gamma - \frac{a}{q}\right| < \frac{1}{qQ}.$$

Proof. This is a simple application of the pigeonhole principle. Consider $q\gamma \pmod 1$ for $q = 1, \ldots, Q$. Either $q\gamma \pmod 1$ lies in $[0, 1/Q)$ for some $1 \leq q \leq Q$, in which case certainly $\|q\gamma\| < 1/Q$, or else $q\gamma \pmod 1$ lies in one of the $Q-1$ intervals $[1/Q, 2/Q), \ldots, [(Q-1)/Q, 1)$. In the latter situation, there exist $1 \leq q_1 < q_2 \leq Q$ and $k \in [Q-1]$ such that $q_1\gamma, q_2\gamma \pmod 1$ both lie in $[k/Q, (k+1)/Q)$, in which case $\|(q_2 - q_1)\gamma\| < 1/Q$. Since $1 \leq q_2 - q_1 < Q$, in either case we get that there exists $1 \leq q \leq Q$ such that $\|q\gamma\| < 1/Q$. The conclusion of the theorem follows. □

Now we can prove the density increment lemma.

Proof of Lemma 1. Suppose that $N \geq 8\alpha^{-2}$. First, let $p \in (2N, 4N)$ be prime[2], and note that any three-term arithmetic progression in $\{1, \dots, N\} \subset \mathbf{Z}/p\mathbf{Z}$ corresponds to a genuine three-term arithmetic progression in $[N]$. Thus, the total number of three-term arithmetic progressions in A equals

$$\sum_{x,y \in \mathbf{Z}/p\mathbf{Z}} 1_A(x) 1_A(x+y) 1_A(x+2y). \tag{1}$$

Since A contains only trivial three-term arithmetic progressions, the above equals αN, which satisfies

$$\alpha N = \frac{\alpha^3 N^2}{\alpha^2 N} \leq \frac{\alpha^3 N^2}{8}$$

by our assumption on the size of N.

We define the *balanced function* $f_A : \mathbf{Z}/p\mathbf{Z} \to \mathbf{R}$ of A by $f_A := 1_A - \alpha 1_{[N]}$. Using that $1_A = f_A + \alpha 1_{[N]}$, we get that (1) equals the sum of

$$\sum_{x,y \in \mathbf{Z}/p\mathbf{Z}} 1_A(x) 1_A(x+y) f_A(x+2y), \tag{2}$$

$$\alpha \sum_{x,y \in \mathbf{Z}/p\mathbf{Z}} 1_A(x) f_A(x+y) 1_{[N]}(x+2y), \tag{3}$$

and

$$\alpha^2 \sum_{x,y \in \mathbf{Z}/p\mathbf{Z}} 1_A(x) 1_{[N]}(x+y) 1_{[N]}(x+2y). \tag{4}$$

Note that (counting y in the union of $[\lfloor (N-x)/2 \rfloor] \pmod p$, $-[\lfloor (x-1)/2 \rfloor] \pmod p$, and $\{0\}$) (4) equals

$$\alpha^2 \sum_{x \in \mathbf{Z}/p\mathbf{Z}} 1_A(x) \left(1 + \left\lfloor \frac{x-1}{2} \right\rfloor + \left\lfloor \frac{N-x}{2} \right\rfloor\right) \geq \alpha^2 |A| \frac{N}{4} = \frac{\alpha^3 N^2}{4},$$

where we have used that $N \geq 6$. Recalling that (1) is at most $\alpha^3 N/8$, it follows that at least one of (2) or (3) must have magnitude at least $\alpha^3 N^2/16 \geq 2^{-8}\alpha^3 p^2$. We will assume that we are in the former case; the latter case can be handled similarly.

Now, for any $f, g, h : \mathbf{Z}/p\mathbf{Z} \to \mathbf{C}$, we have the identity

$$\mathbf{E}_{x,y \in \mathbf{Z}/p\mathbf{Z}} f(x) g(x+y) h(x+2y) = \sum_{\xi \in \mathbf{Z}/p\mathbf{Z}} \widehat{f}(\xi) \widehat{g}(-2\xi) \widehat{h}(\xi),$$

which can be seen by plugging in the Fourier inversion formula for each of f, g, and h and then using orthogonality of characters. Thus, we have

$$\begin{aligned}
\frac{\alpha^3}{2^8} &\leq |\mathbf{E}_{x,y} 1_A(x) 1_A(x+y) f_A(x+2y)| \\
&\leq \sum_{\xi} \left| \widehat{1_A}(\xi) \widehat{1_A}(-2\xi) \widehat{f_A}(\xi) \right| \\
&= \max_{\xi \in \mathbf{Z}/p\mathbf{Z}} \left| \widehat{f_A}(\xi) \right| \cdot \left\| \widehat{1_A} \right\|_{\ell^2}^2 = \alpha \max_{\xi \in \mathbf{Z}/p\mathbf{Z}} \left| \widehat{f_A}(\xi) \right|
\end{aligned}$$

[2]Such a p must exist by Bertrand's postulate. It's not actually that important that p is prime–all we will use, besides the size bound on p, is that it is odd.

by the Cauchy–Schwarz inequality and Parseval's identity. Thus, there exists $\xi \in \mathbf{Z}/p\mathbf{Z}$ such that

$$\left| \sum_{x \in \mathbf{Z}/p\mathbf{Z}} f_A(x) e\left(\frac{\xi x}{p}\right) \right| \geq \frac{\alpha^2}{2^8} p.$$

Since f_A has mean zero by definition, we must, in fact, have $\xi \in \{1, \dots, p-1\}$. Further, since f_A is supported on $[N]$, we have

$$\left| \sum_{x \in [N]} f_A(x) e\left(\frac{\xi x}{p}\right) \right| \geq \frac{\alpha^2}{2^7} N.$$

Now, we will apply Dirichlet's theorem with $Q = \left\lceil \sqrt{N} \right\rceil$ to approximate $\frac{\xi}{p}$. This tells us that there exist integers a and $1 \leq q \leq Q$ and a real number $\theta \in [0, 1)$ such that

$$\frac{\xi}{p} = \frac{a}{q} + \frac{\theta}{q\sqrt{N}}.$$

Next, we will partition $[N]$ into arithmetic progressions with common difference q on which the phase $e(\xi x/p)$ is roughly constant. Note that $[N]$ can be partitioned into K arithmetic progressions $P_1, \dots, P_K$ of length $N' := \lceil 2^{-11}\alpha^2\sqrt{N} \rceil$ and common difference q, where $K \geq \lfloor 2^{11}\alpha^{-2}\sqrt{N} \rfloor$, along with q (possibly empty) arithmetic progressions $P_1', \dots, P_q'$ of length at most $N' - 1$ and common difference q. Thus, since $(c + qd)\frac{a}{q} \equiv \frac{ac}{q} \pmod 1$ for all $c, d \in \mathbf{Z}$, we have

$$\sum_{i=1}^{K} \left| \sum_{x \in P_i} f_A(x) e\left(\frac{\theta x}{q\sqrt{N}}\right) \right| + \sum_{j=1}^{q} \left| \sum_{x \in P_j'} f_A(x) e\left(\frac{\theta x}{q\sqrt{N}}\right) \right| \geq \frac{\alpha^2}{2^7} N.$$

Now, note that whenever $x, y \in P_i$ or $x, y \in P_j'$,

$$\begin{aligned} \left| e\left(\frac{\theta x}{q\sqrt{N}}\right) - e\left(\frac{\theta y}{q\sqrt{N}}\right) \right| &= \left| e\left(\frac{\theta(x-y)}{2q\sqrt{N}}\right) - e\left(\frac{\theta(y-x)}{2q\sqrt{N}}\right) \right| \\ &= 2\left| \sin\left(\frac{\pi\theta(x-y)}{q\sqrt{N}}\right) \right| \\ &\leq \frac{\alpha^2}{2^8}, \end{aligned}$$

since $|\sin(\pi\beta)| \leq \pi|\beta|$. Thus,

$$\sum_{i=1}^{K} \left| \sum_{x \in P_i} f_A(x) \right| + \sum_{j=1}^{q} \left| \sum_{x \in P_j'} f_A(x) \right| \geq \frac{\alpha^2}{2^8} N.$$

Since $P_1, \dots, P_K, P_1', \dots, P_q'$ partition $[N]$ and f_A has mean zero and support in $[N]$, we also have

$$\sum_{i=1}^{K} \sum_{x \in P_i} f_A(x) + \sum_{j=1}^{q} \sum_{x \in P_j'} f_A(x) = 0.$$

Summing these two expressions and using that $|t| + t = 2\max(t, 0)$ for all $t \in \mathbf{R}$, we obtain

$$\sum_{i=1}^{K} \max\left(\sum_{x\in P_i} f_A(x), 0\right) + \sum_{j=1}^{q} \max\left(\sum_{x\in P'_j} f_A(x), 0\right) \geq \frac{\alpha^2}{2^9} N.$$

The contribution of the second sum on the left-hand side above is

$$\sum_{j=1}^{q} \max\left(\sum_{x\in P'_j} f_A(x), 0\right) \leq qN' \leq 2^{-10}\alpha^2 N$$

since f_A is 1-bounded. Thus,

$$\sum_{i=1}^{K} \max\left(\sum_{x\in P_i} f_A(x), 0\right) \geq \frac{\alpha^2}{2^{10}} N.$$

So, by the pigeonhole principle, there exists an $i \in [K]$ such that

$$\frac{|A \cap P_i|}{|P_i|} \geq \alpha + \frac{\alpha^2}{2^{10}},$$

as desired. □

Roth's theorem can be bootstrapped from a result giving one three-term arithmetic progression in a set of positive density to many.

Exercise 2. Prove that there exists a function $c : (0, 1] \to (0, 1]$ such that the following holds: If $A \subset [N]$ has density at least α, then A contains at least $c(\alpha)N^2$ three-term arithmetic progressions.

2.3 Improved bounds

We will end the discussion of Roth's theorem by briefly indicating the key ideas going into some of its quantitative improvements. All of these arguments proceed via a density increment argument, with the main differences being the efficiencies of their density increment lemmas (i.e., at each step, the size of the density increment and the size of the structured set on which the increment is obtained).

Note that one source of quantitative inefficiency in our argument was that each step of the density increment iteration reduced the size of the interval on which we worked by a square root. Heath-Brown and Szemerédi instead collected many large nontrivial Fourier coefficients at each step of the iteration, which they used to obtain a much larger (at least $\alpha(1+\Omega(m^{\Omega(1)}))$ for some positive integer $m \ll \alpha^{-O(1)}$) density increment on a long arithmetic progression (of length at least $\gg \alpha N^{1/(m+1)}$). This iteration is more efficient than Roth's, and leads to an improved savings of a power of $\log N$ over the trivial bound.

The key innovation of Bourgain's work was to obtain a density increment on "Bohr sets", which are approximate level sets of the characters $e_p(\xi x)$. These sets have positive density, but have much less additive structure than intervals. Thus, while Bourgain's method obtains

a density increment on a set of positive density, iterating this argument is significantly more difficult. Bourgain's second improvement on the bounds in Roth's theorem comes from a more careful analysis of the set of large nontrivial Fourier coefficients of 1_A.

In 2010, Croot and Sisask proved a variety of theorems that, roughly, say that the convolution of two functions is approximately invariant under translation by a large set of shifts (which one can take to be fairly "additively structured"), and called this phenomenon *almost periodicity*. Sanders's improvement comes not from further analysis on the Fourier side, but by incorporating a "physical side" argument using almost periodicity.

Bloom–Sisask's first bound was proven by (essentially) adapting an argument of Bateman and Katz in the setting of high dimensional vector spaces over finite fields to the integer setting. The main feature of the work of Bateman and Katz was an extremely careful analysis of the set of large nontrivial Fourier coefficients of 1_A. It was a very difficult task to adapt these ideas to the integer setting, and Bloom and Sisask required all of the previously mentioned innovations (and more) in order to make it work.

The argument of Kelley and Meka was a shocking advance, and proceeds almost completely by arguing on the physical side. Aside from its use of an almost periodicity argument, the proof of their density increment lemma is very different than all of the previously mentioned arguments. Good references to learn their methods are the expositions of Bloom–Sisask and Green. Bloom–Sisask's improvement on the root of $\log N$ in the Kelley–Meka bound comes from optimizing the Kelley–Meka argument.

2.4 Two theorems of Sárközy

Answering a question of Lovász, Furstenberg and Sárközy both independently proved in the late 1970s that any subset of the natural numbers having positive upper density must contain two distinct elements that differ by a perfect square. Furstenberg's proof was via ergodic theory (in the same paper in which he gave his ergodic theoretic proof of Szemerédi's theorem), and produced no quantitative bounds, while Sárközy's argument was, like Roth's, via a Fourier-analytic density increment argument, and produced similar explicit quantitative bounds.

Theorem 5 (Sárközy, 1978)**.** *If $A \subset [N]$ contains no two elements $a, a' \in A$ for which $a - a' = b^2$ for some $b \in \mathbf{N}$, then*

$$|A| \ll \frac{N}{(\log N)^{1/3-o(1)}}.$$

You will prove this in your first homework (with a possibly different power of $\log N$ in the denominator). It turns out that one can show that if A as no nontrivial square differences, then $\widehat{1_A}(\xi)$ (where, here, we take the Fourier transform on $\mathbf{Z}$) is large for some ξ close to a rational with small ($\ll \alpha^{-O(1)}$) denominator. This allows one to deduce a density increment on an arithmetic progression contained in $[N]$ of length $\gg \alpha^{O(1)}N$, leading to a more efficient density increment iteration than in our proof of Roth's theorem.

Sárközy's upper bound has since been improved several times, most notably by Pintz–Steiger–Szemerédi in 1988, who proved an upper bound of

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

by a careful analysis of the set of possible large Fourier coefficients of 1_A. An improved version of this analysis was carried out by Bloom and Maynard in 2022, yielding the current best known upper bounds in Sárközy's theorem:

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

In 1984, Ruzsa gave a construction of a square difference-free subset of $[N]$ of size $\asymp N^{.73\ldots}$, and this is still the densest construction known. It is a problem of great interest in additive combinatorics to determine whether the size of the largest square difference-free subset of $[N]$ is asymptotically $\ll N^{1-\delta}$ for some $\delta > 0$.

In his 1978 paper, Sárközy also proved that any subset of the natural numbers having positive upper density must contain two elements that differ by a prime minus one, with explicit quantitative bounds[3].

Theorem 6 (Sárközy, 1978)**.** *If $A \subset [N]$ contains no two elements $a, a' \in A$ for which $a - a' = p - 1$ for some prime p, then*

$$|A| \ll \frac{N}{(\log\log N)^{2-o(1)}}.$$

This theorem has also been improved multiple times, with the current best result due to Green in 2023, who shattered the previous record by proving a power saving bound:

$$|A| \ll N^{1-\delta} \tag{5}$$

for some $\delta > 0$ that can be explicitly computed. This suggests that it's not completely unreasonable to guess that a power saving bound for the size of sets lacking square differences may be the truth. Green's proof does not, in contrast to all others mentioned in this section, proceed via a density increment argument. He instead carefully constructs a special explicit trigonometric polynomial whose existence (by a straightforward Fourier analytic argument) implies the bound (5). The largest known construction of a subset of $[N]$ with no shifted prime common difference is again due to Ruzsa (1984), and has size $\asymp N^{c/\log\log N}$ for some absolute constant $c > 0$.

2.5 Longer progressions

In this subsection, we will work in $\mathbf{Z}/N\mathbf{Z}$ for simplicity, with N odd. Recall the identity

$$\Lambda_3(f, g, h) := \mathbf{E}_{x,y} f(x)g(x + y)h(x + 2y) = \sum_{\xi} \widehat{f}(\xi)\widehat{g}(-2\xi)\widehat{h}(\xi)$$

for any $f, g, h : \mathbf{Z}/N\mathbf{Z} \to \mathbf{C}$. When $A \subset \mathbf{Z}/N\mathbf{Z}$ has density α, the normalized count

$$\Lambda_3(1_A, 1_A, 1_A) = \frac{\#\{(x, y) \in (\mathbf{Z}/N\mathbf{Z})^2 : x, x + y, x + 2y \in A\}}{N^2}$$

[3]Note that $4\mathbf{N}$ contain no two elements that differ by a prime, so it is not interesting to ask about sets avoiding prime differences.

of three-term arithmetic progressions in A equals

$$\sum_{\xi} \widehat{1_A}(\xi)^2 \widehat{1_A}(-2\xi) = \alpha^3 + \sum_{\xi \neq 0} \widehat{1_A}(\xi)^2 \widehat{1_A}(-2\xi)$$

by separating out the contribution of the contribution of the zeroth Fourier coefficient. Thus, by the triangle inequality and Parseval's identity,

$$\left|\Lambda_3(1_A, 1_A, 1_A) - \alpha^3\right| \leq \max_{0 \neq \xi \in \mathbf{Z}/N\mathbf{Z}} \left|\widehat{1_A}(\xi)\right|. \tag{6}$$

For context, if one were to sample a random subset of $\mathbf{Z}/N\mathbf{Z}$ by including each element with probability α independently, then this random subset will almost always have density very close to α and contain very close to $\alpha^3 N^2$ three-term arithmetic progressions. Thus, the inequality (6) says that the difference between the count of three-term arithmetic progressions in A and in a random set of the same density is controlled by the size of the largest trivial Fourier coefficient of 1_A; in this sense, the L^∞-norm of the Fourier transform is a measure of pseudorandomness from the point of view of counting three-term arithmetic progressions.

It is thus natural to ask whether Fourier analysis also controls the count of longer arithmetic progressions in subsets of $\mathbf{Z}/N\mathbf{Z}$. This turns out to not be the case, as can be seen by considering the set

$$\left\{n \in [N] : 0 \leq \left\{\sqrt{2}n^2\right\} \leq \frac{1}{1000}\right\},$$

where $\{t\}$ denotes the fractional part of any $t \in \mathbf{R}$. You will show in your homework that this set has far from the random density of four-term arithmetic progressions, yet has no large nontrivial Fourier coefficients. Given this example, it is maybe then natural to conjecture that if the indicator function of a set A does not correlate with any nontrivial quadratic phase functions $x \mapsto e(\beta x^2 + \gamma x)$, then A has close to the random number $\alpha^4 N^2$ four-term arithmetic progressions; this is again not the case, as you will also show in the homework.

The question remains of whether there is some theory analogous to Fourier analysis that governs the count of arithmetic progressions of length greater than three. Such a theory was first developed by Gowers in the late 1990s and early 2000s, and is now called *higher order Fourier analysis*. Gowers used this to prove the first reasonable quantitative bounds in Szemerédi's theorem. In order to explain a bit about what higher order Fourier analysis is, we will begin by controlling the difference $|\Lambda_3(1_A, 1_A, 1_A) - \alpha^3|$ in an alternative way that is more amenable to generalization.

Let $f_A := 1_A - \alpha$ denote the balanced function of $A \subset \mathbf{Z}/N\mathbf{Z}$ (again, with N odd). Since $\Lambda_3(1_A, 1_A, 1_A) = \Lambda_3(1_A, 1_A, f_A) + \alpha\Lambda_3(1_A, 1_A, 1) = \Lambda_3(1_A, 1_A, f_A) + \alpha^3$, we have

$$\left|\Lambda_3(1_A, 1_A, 1_A) - \alpha^3\right| \leq \left|\Lambda_3(1_A, 1_A, f_A)\right|.$$

Now, to bound the right-hand side, we apply the Cauchy–Schwarz inequality and make a

change of variables to obtain, using that 1_A and f_A are both 1-bounded,

$$\begin{aligned}
|\Lambda_3(1_A, 1_A, f_A)| &= |\mathbf{E}_{x,y} 1_A(x) 1_A(x+y) f_A(x+2y)| \\
&\leq \left(\mathbf{E}_x \left|\mathbf{E}_y 1_A(x+y) f_A(x+2y)\right|^2\right)^{1/2} \\
&= \left(\mathbf{E}_{x,y,z} 1_A(x+y)\overline{1_A(x+z) f_A(x+2y)} f_A(x+2z)\right)^{1/2} \\
&= \left(\mathbf{E}_{x,y,h} 1_A(x+y)\overline{1_A(x+y+h) f_A(x+2y)} f_A(x+2y+2h)\right)^{1/2} \\
&= \left(\mathbf{E}_{x,y,h} 1_A(x)\overline{1_A(x+h) f_A(x+y)} f_A(x+y+2h)\right)^{1/2}.
\end{aligned}$$

A second application of the Cauchy–Schwarz inequality to double the y variable and a change of variables yields

$$\begin{aligned}
|\Lambda_3(1_A, 1_A, f_A)|^4 &\leq \mathbf{E}_{x,h} \left|\mathbf{E}_y f_A(x+y)\overline{f_A(x+y+2h)}\right|^2 \\
&\leq \mathbf{E}_{x,h}\mathbf{E}_{y,z} f_A(x+y)\overline{f_A(x+z) f_A(x+y+2h)} f_A(x+z+2h) \\
&\leq \mathbf{E}_{x,y,h,k} f_A(x)\overline{f_A(x+k) f_A(x+h)} f_A(x+h+k).
\end{aligned}$$

This last expression is the fourth power of the *Gowers U^2-norm* of f_A, $\|f_A\|_{U^2}^4$. Thus, we have shown that

$$\left|\Lambda_3(1_A, 1_A, 1_A) - \alpha^3\right| \leq \|f_A\|_{U^2}.$$

One can show, by plugging in the Fourier inversion formula, that $\|f\|_{U^2} = \|\widehat{f}\|_{\ell^4}$, and from this deduce again from Parseval's identity that $|\Lambda_3(1_A, 1_A, 1_A) - \alpha^3|$ is small whenever 1_A has no large nontrivial Fourier coefficients (though with slightly worse quantitative dependence). Precisely, one has the following (easy to prove) statement, which is called the *inverse theorem for the U^2-norm*:

Lemma 2. *Let $f : \mathbf{Z}/N\mathbf{Z} \to \mathbf{C}$ be 1-bounded. If $\|f\|_{U^2} \geq \delta$, then there exists $\xi \in \mathbf{Z}/N\mathbf{Z}$ such that*

$$\left|\widehat{f}(\xi)\right| \geq \delta^2.$$

One can show, by three applications of the Cauchy–Schwarz inequality followed by a change of variables, that whenever $(N, 6) = 1$ and $f_0, f_1, f_2, f_3 : \mathbf{Z}/N\mathbf{Z} \to \mathbf{C}$ are any 1-bounded functions, then, setting

$$\Lambda_4(f_0, f_1, f_2, f_3) := \mathbf{E}_{x,y} f_0(x) f_1(x+y) f_2(x+2y) f_3(x+3y),$$

we have

$$|\Lambda_4(f_0, f_1, f_2, f_3)| \leq \min_{i \in [4]} \|f_i\|_{U^3},$$

where $\|\cdot\|_{U^3}$ is the *Gowers U^3-norm*, defined by $\|f\|_{U^3}^8$ equaling[4]

$$\mathbf{E}_{x,y,h,k,l} f(x)\overline{f(x+h) f(x+k) f(x+l)} f(x+h+k) f(x+h+l) f(x+k+l)\overline{f(x+h+k+l)}.$$

[4]It is not obvious that $\|\cdot\|_{U^3}$ is a norm; we will discuss this (and other) facts about the U^3-norm later in the course.

It then follows that

$$\left|\Lambda_4(1_A, 1_A, 1_A, 1_A) - \alpha^4\right| \ll \|f_A\|_{U^3}. \tag{7}$$

The bulk of this course will consist of proving a "local" inverse theorem for the U^3-norm (originally due to Gowers in 1998), which, combined with (7), will allow us to deduce a density increment on a long arithmetic progression whenever $A \subset [N]$ contains no nontrivial four-term arithmetic progressions. This is sufficient to prove Gowers's bound on the size of subsets of $[N]$ lacking four-term arithmetic progressions.

Theorem 7 (Gowers, 1998)**.** *If $A \subset [N]$ contains no nontrivial four-term arithmetic progressions, then*

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

for some absolute constant[5] *$c > 0$.*

Gowers's upper bound has since been improved by Green and Tao to save a small power of $\log N$ over the trivial bound of N.

The proof of the local inverse theorem for the U^3-norm uses multiple foundational results and techniques in additive combinatorics, and is thus a good a result to learn the proof of if one wants an introduction to the area. We will spend the next few weeks on these results, starting with the basic theory of sumsets and product sets.

3 Sumsets and product sets

Let G be any (possibly nonabelian) group and $A, B \subset G$. The *product set* of A and B is defined by

$$AB := \{ab : a \in A \text{ and } b \in B\}.$$

We will also write A^{-1} for $\{a^{-1} : a \in A\}$ and say that A is *symmetric* if $A = A^{-1}$. For any $k \in \mathbf{N}$, we will denote the k-fold product set of A by

$$A^k := \{a_1 \cdots a_k : a_1, \ldots, a_k \in A\}.$$

When G is abelian with operation denoted by $+$, the *sumset* of A and B is defined to be

$$A + B := \{a + b : a \in A \text{ and } b \in B\},$$

and we write $-A$ for $\{-a : a \in A\}$ (so that A is symmetric if $A = -A$) and

$$A - B := \{a - b : a \in A \text{ and } b \in B\}$$

for the *difference set* of A and B. For any $k \in \mathbf{N}$, we will denote the k-fold sumset of A by

$$kA := \{a_1 + \cdots + a_k : a_1, \ldots, a_k \in A\}.$$

Note that this is distinct from the *dilation* $k \cdot A := \{ka : a \in A\}$, though $k \cdot A \subset kA$.

[5] $c = 2^{-2^{13}}$ works

There are a few obvious trivial bounds for the sizes of sumsets and product sets. For any $A, B \subset G$, we have

$$\max\{|A|, |B|\} \leq |AB| \leq |A||B|.$$

When $(G, +)$ is an abelian group and $C \subset G$, we have the stronger trivial upper bound

$$|C + C| \leq |C| + \binom{|C|}{2} = \frac{|C|(|C| + 1)}{2}.$$

These trivial bounds are, in fact, sharp. Indeed, if $A = B = H$ for some finite subgroup H of G, then $|AB| = |A| = |B|$. If, say, $G = F_2 = \langle a, b \rangle$ is the free group on two letters and $A = \{1, a, a^2\}$ and $B = \{1, b, b^2\}$, then $|AB| = |\{1, a, a^2, b, ab, a^2b, b^2, ab^2, a^2b^2\}| = 9 = |A||B|$. Similarly, if $G = \mathbf{Z}$ and $C = \{1, 2, 4\}$, then $|C + C| = |\{1, 2, 4, 3, 5, 6\}| = 6 = \frac{|C|(|C|+1)}{2}$.

3.1 Basic results

Our first nontrivial result on sumsets and product sets will be the following, which is known as *Ruzsa's triangle inequality*.

Lemma 3. *Let $A, B, C \subset G$ be three finite nonempty subsets of G. Then,*

$$|AC^{-1}| \leq \frac{|AB^{-1}||BC^{-1}|}{|B|}.$$

Proof. Note that the desired inequality is equivalent to

$$|AC^{-1}||B| \leq |AB^{-1}||BC^{-1}|.$$

We will prove this by considering the map $\phi : (AC^{-1}) \times B \to (AB^{-1}) \times (BC^{-1})$ defined as follows: fix arbitrarily, for each $x \in AC^{-1}$, $(a_x, c_x) \in A \times C$ such that $x = a_x c_x^{-1}$, and then set

$$\phi(x, b) = \phi(a_x c_x^{-1}, b) = (a_x b^{-1}, b c_x^{-1}).$$

The desired inequality will then follow if we can show that ϕ is injective. To see that ϕ is injective, suppose that $\phi(x, b) = \phi(y, d)$. This means that $(a_x b^{-1}, b c_x^{-1}) = (a_y d^{-1}, d c_y^{-1})$, so, in particular,

$$x = a_x c_x^{-1} = a_x b^{-1} b c_x^{-1} = a_y d^{-1} d c_y^{-1} = a_y c_y^{-1} = y.$$

From this, it also follows that $b c_x^{-1} = d c_x^{-1}$, and thus that $b = d$ as well. Hence, ϕ is injective. □

The reason this lemma is called a "triangle inequality" is due to the existence of the *Ruzsa distance* between two finite nonempty subsets of the same group $A, B \subset G$:

$$d(A, B) := \log \frac{|AB^{-1}|}{\sqrt{|A||B|}}.$$

It's easy to check that $d(A, B)$ is nonnegative (by the trivial lower bound on $|AB^{-1}|$) and symmetric (since $|AB^{-1}| = |BA^{-1}|$, as inversion is a bijection). The Ruzsa triangle inequality is equivalent to the triangle inequality for Ruzsa distance:

$$d(A, C) \leq d(A, B) + d(B, C),$$

as can be seen by exponentiating both sides of the above and multiplying through by $\sqrt{|A||C||B|}$. The Ruzsa distance is not, however, a true distance, since $d(A, A)$ is not always 0 (consider $A = \{0, 1\} \subset \mathbf{Z}/4\mathbf{Z}$) and $d(A, B) = 0$ does not imply that $A = B$ (consider $A = \{0, 2\}, B = \{1, 3\} \subset \mathbf{Z}/4\mathbf{Z}$).

One can deduce from the Ruzsa triangle inequality that if a finite subset of a group has 3-fold product set not much larger than the original set (i.e., the set has *small tripling*), then the size of its n-fold product sets can also be controlled for any positive integer n.

Lemma 4. *Let A be a finite symmetric subset of a group G that contains the identity. If $|A^3| \leq K|A|$, then*

$$|A^n| \leq K^{n-2}|A|$$

for all integers $n \geq 3$.

Proof. We will proceed by induction on n, with the base case $n = 3$ already being covered by assumption. So, assume that $n \geq 4$ and that $|A^{n-1}| \leq K^{n-3}|A|$. Note that, by the definition of Ruzsa distance, we have

$$\frac{|A^n|}{|A|} = \frac{|A^{n-2}A^2|}{|A|} = \frac{\sqrt{|A^{n-2}||A^2|}}{|A|} \exp\left(d(A^{n-2}, A^2)\right).$$

By Ruzsa's triangle inequality, we have

$$\exp\left(d(A^{n-2}, A^2)\right) \leq \exp\left(d(A^{n-2}, A) + d(A, A^2)\right) = \frac{|A^{n-1}||A^3|}{\sqrt{|A^{n-2}||A|} \cdot \sqrt{|A||A^2|}}.$$

Combining this with the above, we deduce that

$$\frac{|A^n|}{|A|} \leq \frac{|A^{n-1}||A^3|}{|A|^2} \leq K^{n-3} \cdot K = K^{n-2}$$

by the induction hypothesis and the initial assumption on $|A^3|$. □

To get a result that applies to all sets with small tripling, one can also use the Ruzsa triangle inequality to show that a set having small tripling implies that its "symmetrization" also has small tripling.

Lemma 5. *Let A be a finite subset of a group G. If $|A^3| \leq K|A|$, then*

$$|(A \cup A^{-1} \cup \{1\})^3| \leq 27K^3|A|.$$

The proof of this lemma is left as an exercise.

It is not, in general, true that a set having small doubling (i.e., $|A^2| \leq K|A|$) implies that its n-fold product sets can also be controlled for any positive integer n. Indeed, let $H \leq G$ be a subgroup of some finite group G and $g \in G \setminus H$, and set $A = H \cup \{g\}$. Then $|A^2| = |H \cup gH \cup Hg| \leq 3|A|$, so A certainly has small doubling, but $A^3 \supset HgH$, which can be very large. For example, if $G = \mathrm{SL}_2(\mathbf{F}_p)$,

$$H = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \in \mathrm{SL}_2(\mathbf{F}_p) \right\},$$

and $g = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, then one can check that $|HgH| = p(p-1)(p-2)$, while $|A| = |H \cup \{g\}| = p(p-1)+1$. So, $|A^3| \gg p|A|$, and there is thus no hope of proving an analogue of Lemma 4 with a small doubling hypothesis for general groups.

Note that the example above crucially relied on G being nonabelian. It turns out that, for abelian groups, small doubling does imply control on the side of all iterated sumsets.

3.2 The Plünnecke–Ruzsa inequality

Next, we will prove the Plünnecke–Ruzsa inequality by following an argument of Petridis from 2014 (in the form of a rephrasing due to Tao). Petridis's proof is significantly shorter and more elegant than all of the previously known proofs. For the entirety of this subsection, G will be abelian with operation $+$.

Theorem 8. *Let $A, B \subset G$ be finite subsets of an abelian group and $n, m \geq 0$ be integers. If $|A+B| \leq K|A|$, then $|nB - mB| \leq K^{n+m}|A|$.*

Note that this gives us our desired result on sets with small doubling when $A = B$. One can ask whether this exponent can be improved to $n+m-1$, since, when $A = B$, it tells us that $|A+A| \leq K|A|$ implies that $|A+A| \leq K^2|A|$. This turns out not to be the case, as you will show in the next problem set.

We will need one new definition before beginning the proof of the Plünnecke–Rusza inequality. A real-valued function f on the set of subsets of some fixed set S is said to be *submodular* if $f(X \cup Y) + f(X \cap Y) \leq f(X) + f(Y)$. Intuitively, a submodular function is one for which the effect of adding additional elements to a set on the value of f has diminishing returns. An equivalent characterization is that, whenever $Y \subset X$, we have

$$f(X \cup \{z\}) - f(X) \leq f(Y \cup \{z\}) - f(Y)$$

whenever $z \notin X$. We begin with a simple lemma about submodular functions.

Lemma 6. *Let S be any set, f be a submodular function on 2^S, and $X_1, \ldots, X_n \subset S$ with $f(X_1) = \cdots = f(X_n) = 0$ and such that $f(Y) \geq 0$ whenever $Y \subset X_i$. Then we must have*

$$f\left(\bigcup_{i=1}^{n} X_i\right) \leq 0.$$

Proof. We proceed by induction on n, the base case $n = 1$ trivially holding by the assumptions of the lemma. Suppose that the desired inequality holds for a general $n-1 \geq 1$. Then, on setting $Y = X_1 \cup \cdots \cup X_{n-1}$, we have

$$f\left(\bigcup_{i=1}^{n} X_i\right) = f(Y \cup X_n) \leq f(Y) + f(X_n) - f(Y \cap X_n) \leq 0$$

since $f(Y) \leq 0$ by the induction hypothesis, $f(X_n) = 0$, and $f(Y \cap X_n) \geq 0$ since $Y \cap X_n \subset X_n$. □

Lemma 7. *Let $\emptyset \neq B \subset G$ and $K \in \mathbf{R}$. The function $f : 2^G \to \mathbf{R}$ defined by*

$$f(X) = |X + B| - K|X|$$

is submodular.

Proof. Let $X, Y \in G$. We first make two simple observations:

$$|X \cup Y| - |X| - |Y| + |X \cap Y| = 0$$

and

$$(X \cap Y) + B \subset (X + B) \cap (Y + B).$$

Thus, we have

$$\begin{aligned} f(X \cup Y) + f(X \cap Y) &= |(X \cup Y) + B| + |(X \cap Y) + B| - K(|X \cup Y| + |X \cap Y|) \\ &\leq |(X + B) \cup (Y + B)| + |(X + B) \cap (Y + B)| - K(|X \cup Y| + |X \cap Y|) \\ &= |X + B| + |Y + B| - K(|X| + |Y|) = f(X) + f(Y), \end{aligned}$$

as desired. □

Now we will apply our lemma on submodular functions to the particular function from the previous lemma.

Lemma 8. *Let $A, B \subset G$ be finite nonempty subsets such that $|A + B| \leq K|A|$. Let $\emptyset \neq X \subset A$ be such that*

$$\frac{|X + B|}{|X|} = \min_{\emptyset \neq Y \subset A} \frac{|Y + B|}{|Y|}.$$

Then, for any $C \subset G$, we have

$$|(X + C) + B| \leq K|X + C|.$$

Proof. Set $K' = |X + B|/|X|$, and define $f(Y) := |Y + B| - K'|Y|$ for $Y \subset G$. Note that $K' \leq K$. Enumerate the elements of C, $C = \{c_1, \ldots, c_n\}$, and set $X_i := X + c_i$, so that $X_1 \cup \cdots \cup X_n = X + C$. By the definition of X, we have

$$f(X_i) = |X + B + c_i| - K'|X + c_i| = |X + B| - K'|X| = 0$$

for all $i = 1, \ldots, n$. Further, for any $Y \subset X_i$ (so that $Y - c_i \subset A$), we have

$$f(Y) = |Y + B| - K'|Y| = |Y - c_i + B| - K'|Y - c_i| \geq 0$$

for all $i = 1, \ldots, n$, again by the definition of X. Thus, we can apply our lemma on submodular functions and obtain that $f(X + C) \leq 0$. That is,

$$|X + C + B| \leq K'|X + C| \leq K|X + C|,$$

as desired. □

Now, we can finally prove the Plünnecke–Ruzsa inequality.

Proof of Theorem 8. First, we will show that there exists $X \subset A$ for which $|X+\ell B| \leq K^\ell |X|$ for all $\ell \in \mathbf{N}$. We will then put two instances of this statement (with $\ell = n$ and $\ell = m$) together using Ruzsa's triangle inequality to complete the proof.

Let $\emptyset \neq X \subset A$ again be such that

$$\frac{|X+B|}{|X|} = \min_{\emptyset \neq Y \subset A} \frac{|Y+B|}{|Y|}.$$

We proceed yet again by induction. The desired inequality automatically holds for this choice of X when $\ell = 1$. Assume that it holds for a general $\ell - 1 \geq 1$. Applying the previous lemma with $C = (\ell-1)B$ yields

$$|X + \ell B| \leq K|X + (\ell-1)B| \leq K^{1+\ell-1}|X| = K^\ell |X|,$$

by the induction hypothesis.

We now apply Ruzsa's triangle inequality with the sets $-nB, X$, and $-mB$, respectively, to obtain that

$$|nB - mB||X| \leq |X + nB||X + mB| \leq K^{n+m}|X|^2.$$

Thus, $|nB - mB| \leq K^{n+m}|X| \leq K^{n+m}|A|$, since $X \subset A$. □

3.3 Approximate groups

Now we return to the general setting where G can possibly be nonabelian. A related notion to a subset having small tripling is that of an approximate group.

Definition 1. *Let $K \geq 1$ be a real number. A subset $A \subset G$ is a K-approximate group if A is symmetric, contains the identity, and is such that A^2 can be covered by at most K left translates of A.*

The last condition is equivalent to there existing $X \subset G$ with $|X| \leq K$ such that $A^2 \subset XA$. The definition of a K-approximate group is very convenient to work with, and is also closely related to having small tripling. Indeed, it's easy to see that a K-approximate group A has tripling at most K^2: letting X with $|X| \leq K$ be such that $A^2 \subset XA$, we have

$$|A^3| \leq |XA^2| \leq |X^2 A| \leq |X|^2|A| \leq K^2|A|.$$

Any dense subset of an approximate group also has small tripling. Indeed, if A is a K-approximate group and $B \subset A$ has density β in A, then

$$|B^3| \leq |A^3| \leq K^2|A| = \frac{K^2}{\beta}|B|.$$

A rough converse also holds–any set with small tripling must be a dense subset of an approximate group. To prove this, we will need another important lemma due to Ruzsa, which is called *Ruzsa's covering lemma.*

Lemma 9. *Let $A, B \subset G$ be finite and nonempty, and assume that $|AB| \leq K|B|$. Then there exists $X \subset A$ with $|X| \leq K$ such that*

$$A \subset XBB^{-1}.$$

Proof. Let $X = \{a_1, \dots, a_n\} \subset A$ be a maximal subset of A such that $a_1B, \dots, a_nB$ are all disjoint. Note that $n \leq |AB|/|B| \leq K$. For every $x \in A$, there exists $i \in [n]$ such that $a_iB \cap xB \neq \emptyset$ by the maximality assumption. That is, $x \in a_iBB^{-1}$. We conclude that $A \subset XBB^{-1}$. □

Now, we can relate the small tripling and approximate group properties.

Theorem 9. *Let $A \subset G$ and $K \geq 1$. The following statements are equivalent, in the sense that if one of them holds for some choice of implied constants, then the other also holds for some choice of implied constants.*

1. $|A^3| \ll K^{O(1)}|A|$
2. *There exists a $O(K^{O(1)})$-approximate group of size $\ll K^{O(1)}|A|$ containing A.*

Proof. To see that the second implication implies the first, just note that, by our observation that a K-approximate group has tripling at most K^2,

$$|A^3| \leq |H^3| \ll K^{O(1)}|H| \ll K^{O(1)}|A|.$$

To see that the first implication implies the second, set $B = A \cup A^{-1} \cup \{1\}$ and $H = B^3$. By Lemma 5, we have $|H| \ll K^{O(1)}|A|$, and H contains A by the inclusion of the identity element in B. Since H is clearly symmetric and contains the identity, it remains to check the last part of the definition of an $O(K^{O(1)})$-approximate group. By Lemmas 4 and 5, we also have that $|H^2B| \ll K^{O(1)}|A| \ll K^{O(1)}|B|$. Thus, by Ruzsa's covering lemma, there exists $X \subset H^2$ of size $\ll K^{O(1)}$ such that

$$H^2 \subset XB^2 \subset XH,$$

since B is symmetric and $H = B^3 \supset B^2$. □

3.4 Product theorems

A natural problem is to classify, given a group G, all K-approximate groups of G (or, equivalently, all sets with small tripling). It's easy to show that any 1-approximate group in G is just a genuine finite subgroup of G, i.e., any finite symmetric subset $A \subset G$ containing the identity for which $|A^2| = |A|$ must be a subgroup. One can, in fact, relax the inequality $|A^2| \leq |A|$ a bit.

Theorem 10. *Let $A \subset G$ be finite and nonempty. If $|A^{-1}A| < \frac{3}{2}|A|$, then there exists a subgroup $H \leq G$ with $|H| < \frac{3}{2}|A|$ such that A is contained in some left coset of H.*

The 3/2 in the above statement cannot be replaced by any larger constant. Indeed, consider $A = \{0, 1\} \subset \mathbf{Z}$. Then $|A - A| = \{-1, 0, 1\}$, so that $|A - A| \leq \frac{3}{2}|A|$, but no coset of a finite subgroup contains A. More generally, for larger K, it is not true that K-approximate groups must always live inside a small number of translates of a not too much larger subgroup. When $G = \mathbf{Z}$, for example, the set $A = [N]$ is a 2-approximate group since

$$A + A = \{2, \dots, 2N\} \subset (1 + [N]) \cup (N + [N]),$$

but the smallest number of cosets of $\{0\}$ (the only finite subgroup of $\mathbf{Z}$) needed to cover $A+A$ is $2N-1$. Examples such as this, which really only appear when G is closely related to a nilpotent group, must be taken into account if one hopes to prove a general classification theorem. Thus, we will postpone discussion of the abelian case (which will play an important role in our proof of the local inverse theorem for the U^3-norm) and general case to later. In the remainder of this subsection, we will prove Theorem 10 and then discuss a result on growth of three-fold product sets in $\mathrm{SL}_n(\mathbf{F}_q)$ and its applications.

We begin by showing that if $|A^{-1}A| < \frac{3}{2}|A|$, then the different intersections $A \cap gA$ satisfy a useful dichotomy.

Lemma 10. *Let $A \subset G$ be a finite subset such that $|A^{-1}A| < \frac{3}{2}|A|$. Then, for all $g \in G$, either $A \cap gA = \emptyset$ or $|A \cap gA| > \frac{|A|}{2}$.*

Proof. Suppose that $A \cap gA \neq \emptyset$, so that there exists $a \in A \cap gA$. This means that $a, g^{-1}a \in A$, and thus that $a^{-1}A, a^{-1}gA \subset A^{-1}A$. But, since $|A^{-1}A| < \frac{3}{2}|A|$, the sets $a^{-1}A$ and $a^{-1}gA$, which both have size $|A|$, must intersect in a set of size greater than $\frac{|A|}{2}$. That is, we have $|a^{-1}A \cap a^{-1}gA| > \frac{|A|}{2}$, which implies that $|A \cap gA| > \frac{|A|}{2}$, as desired. □

We will call any $g \in G$ for which $|A \cap gA| > \frac{|A|}{2}$ in the lemma *involved*. By exploiting the above dichotomy, we will show that the set of involved elements forms a group.

Lemma 11. *Let $A \subset G$ be a finite subset such that $|A^{-1}A| < \frac{3}{2}|A|$. Then, the set $H \subset G$ of involved elements is a finite subgroup of G, and $H = AA^{-1}$.*

Proof. Clearly, $1 \in H$. Also, if $h \in H$, then since $|A \cap hA| = |h^{-1}A \cap A|$, $h^{-1} \in H$ as well. So now suppose that $h_1, h_2 \in H$. Then, since $|A \cap h_1^{-1}A|, |A \cap h_2 A| > \frac{1}{2}|A|$, certainly the intersection $A \cap h_1^{-1}A \cap h_2 A$ is nonempty. Thus,

$$|A \cap h_1 h_2 A| = |h_1^{-1}A \cap h_2 A| \geq |A \cap h_1^{-1}A \cap h_2 A| > 0.$$

By the dichotomy from the previous lemma, this implies that $|A \cap h_1 h_2 A| > \frac{|A|}{2}$, and hence that $h_1 h_2 \in H$. This completes the proof that H is a subgroup. To see that $H = AA^{-1}$, we simply note that if $h = a_1 a_2^{-1} \in AA^{-1}$, then $a_1 \in A \cap hA \neq \emptyset$, so that $h \in H$ by the dichotomy lemma, and if $h \in H$, then there exist $a_1, a_2 \in A$ such that $a_1 = h a_2$, i.e., $h = a_1 a_2^{-1}$, so that $h \in AA^{-1}$. □

Now that we've constructed a group out of the small growth assumption on A, we can prove Theorem 10.

Proof of Theorem 10. Let $H = AA^{-1}$ be as in the previous lemma and $a \in A$, so that $Aa^{-1} \subset H$. This means that $A \subset Ha$. Replacing H with its conjugate $K := a^{-1}Ha$, this implies that $A \subset aK$. By double counting and the dichotomy lemma,

$$|A|^2 = \sum_{h \in H} |A \cap hA| > |H| \frac{|A|}{2}.$$

Thus, $|K| = |H| < 2|A|$. But, for any $k \in K$, $Ak \subset K$, and so since $A \subset K$ and $|Ak| = |A|$, the sets A and Ak have nonempty intersection. It follows that $k \in A^{-1}A$, and hence that $K = A^{-1}A$. We conclude that $|K| < \frac{3}{2}|A|$, as desired. □

We will see the key ideas of this proof appear again later when we discuss the proof of the sum product theorem in finite fields. By a much more elaborate extension of these ideas, one can prove that if a subset of $\mathrm{SL}_n(\mathbf{F}_q)$ isn't contained in a proper subgroup or close to the whole group, then its must have large tripling.

Theorem 11 (Product theorem for special linear groups)**.** *Let $n \geq 2$ be an integer and q be a prime power. There exists an absolute constant $\varepsilon = \varepsilon(n) > 0$ such that the following holds. If $\gamma \leq \varepsilon$, $A \subset \mathrm{SL}_n(\mathbf{F}_q)$ generates $\mathrm{SL}_n(\mathbf{F}_q)$, and $|A| < |G|^{1-O_n(\gamma)}$, then*

$$|A^3| \geq |A|^{1+\gamma}.$$

This theorem was first proven in the case $n = 2$ and q prime in breakthrough work of Helfgott in 2008. Helfgott then extended his argument to the case $n = 3$ in 2011, and the general case (even more generally than above, giving a product theorem for any finite group of Lie type) was proven, independently, by Pyber–Szabo in 2016 and Breuillard–Green–Tao in 2011.

Theorem 11 is relevant to a famous conjecture of Babai from 1979. For any group G and $S \subset G$ a symmetric generating set, we define $\mathrm{Cay}(G, S)$ to be the associated Cayley graph, meaning that $\mathrm{Cay}(G, S)$ is the graph with vertices G with an edge between $g, g' \in G$ whenever $sg = g'$ for some $s \in S$. The assumption that S generates G means that $\mathrm{Cay}(G, S)$ is connected, and the diameter of $\mathrm{Cay}(G, S)$ (i.e., the maximum length of the shortest path from the identity element to each element of G) equals the smallest $d \in \mathbf{N}$ for which $(S \cup \{1\})^d = G$.

Conjecture 2 (Babai's conjecture, 1979)**.** *There exists an absolute constant $C > 0$ such that the following holds. For all nonabelian finite simple groups G and symmetric generating sets $S \subset G$, the diameter of $\mathrm{Cay}(G, S)$ is at most $(\log |G|)^C$.*

By the classification theorem for finite simple groups, it suffices to prove Babai's conjecture for finite simple groups of Lie type and for the alternating group. The product theorems of Pyber–Szabo and Breuillard–Green–Tao have resolved Babai's conjecture for finite simple groups of Lie type of bounded rank. Helfgott and Seress (2011) have obtained very strong bounds when $G = A_n$, proving that the diameter of any $\mathrm{Cay}(A_n, S)$ is bounded by $\exp(O((\log n)^4 \log\log n))$. The current best bounds in the high rank case of finite simple groups of Lie type are based off of ideas from this argument, and have the analogous shape.

Prior Helfgott's work, no nontrivial bounds were known in Babai's conjecture for $\mathrm{PSL}_2(\mathbf{F}_p)$ Cayley graphs aside from very special sets of generators. For example, nothing was known for the generating set

$$S = \left\{ \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} \right\}.$$

Helfgott used his product theorem to prove the following beautiful diameter bound, which is independent of the generating set.

Theorem 12 (Helfgott, 2008)**.** *Let $S \subset \mathrm{SL}_2(\mathbf{F}_p)$ be a symmetric set of generators. Then, $\mathrm{Cay}(\mathrm{SL}_2(\mathbf{F}_p), S)$ has diameter $\ll (\log p)^{O(1)}$.*

This immediately implies the corresponding diameter bound in $\mathrm{PSL}_2(\mathbf{F}_p)$, since $\mathrm{PSL}_2(\mathbf{F}_p)$ is just $\mathrm{SL}_2(\mathbf{F}_p)$ quotiented out by $\pm I$.

3.5 Quasirandom groups and diameter bounds

To deduce diameter bounds for Cayley graphs from Theorem 11, we will need a way to understand product sets of relatively large (i.e., of size $|G|^{1-\delta}$ for $\delta > 0$ smaller than some absolute constant) subsets of $\mathrm{SL}_2(\mathbf{F}_p)$, which will require some basic facts about the representation theory of $\mathrm{SL}_2(\mathbf{F}_p)$.

Definition 2. *Let $D \geq 1$ be an integer and G be a finite group. We say that G is D-quasirandom if all nontrivial irreducible representations of G have dimension at least D.*

This terminology comes from the fact that dense Cayley graphs of quasirandom groups are quasirandom graphs. Intuitively, the more quasirandom a group is, the better its mixing properties. One generally says that a family of finite groups G_n are quasirandom if G_n is D_n-quasirandom for some D_n tending to infinity with $|G_n|$. A standard example of a quasirandom group is A_n, which is $(n-1)$-quasirandom when $n \geq 6$. More relevant to us is the fact that the family $\mathrm{SL}_2(\mathbf{F}_p)$ is quasirandom.

Fact 1. *For all primes p, the group $\mathrm{SL}_2(\mathbf{F}_p)$ is $\frac{p-1}{2}$-quasirandom.*

To prove the desired mixing properties of quasirandom groups, we will need a bit of nonabelian Fourier analysis. Let G be any finite group and $f, f_1, f_2 : G \to \mathbf{R}$ (we will only care about real-valued functions), and define the ℓ^2-norm of f by

$$\|f\|_{\ell^2} = \sqrt{\sum_{x \in G} |f(x)|^2}$$

and the convolution of f_1 and f_2 by

$$(f_1 * f_2)(x) := \sum_{y \in G} f_1(xy^{-1}) f_2(y)$$

for all $x \in G$. We will let $\widehat{G}$ denote the set of irreducible (unitary) representations of G, and for all $\rho \in \widehat{G}$ define the Fourier transform of f at ρ by

$$\widehat{f}(\rho) = \sum_{x \in G} f(x)\rho(x).$$

Then, we have that $\widehat{f_1 * f_2}(\rho) = \widehat{f_1}(\rho)\widehat{f_2}(\rho)$ for all $\rho \in \widehat{G}$. We also have Parseval's identity:

$$\|f\|_{\ell^2}^2 = \frac{1}{|G|} \sum_{\rho \in \widehat{G}} d_\rho \|\widehat{f}(\rho)\|_{HS}^2,$$

where $\|\cdot\|_{HS}$ is the *Hilbert-Schmidt norm* of a matrix, which is defined by $\|M\|_{HS}^2 = \mathrm{tr}(M^*M)$. An important property of the Hilbert–Schmidt norm is that it is *submultiplicative*, i.e., $\|MM'\|_{HS} \leq \|M\|_{HS}\|M'\|_{HS}$. This can be proved by a simple application of the Cauchy–Schwarz inequality.

The following lemma, which will be the key technical ingredient in our forthcoming mixing lemma for large product sets, is due to Babai–Nikolov–Pyber (2007).

Lemma 12. *Let $D \geq 1$ and G be a finite D-quasirandom group. Assume that $f_1, f_2 : G \to \mathbf{R}$ are such that f_1 has mean zero. Then,*

$$\|f_1 * f_2\|_{\ell^2} \leq \frac{|G|^{1/2}}{D^{1/2}} \|f_1\|_{\ell^2} \|f_2\|_{\ell^2}.$$

Proof. By Parseval's identity and the assumption that f_1 has mean zero and the submultiplicativity of the Hilbert–Schmidt norm, we have

$$\|f_1 * f_2\|_{\ell^2}^2 = \frac{1}{|G|} \sum_{1 \neq \rho \in \widehat{G}} d_\rho \|\widehat{f_1}(\rho)\widehat{f_2}(\rho)\|_{HS}^2 \leq \frac{1}{|G|} \sum_{1 \neq \rho \in \widehat{G}} d_\rho \|\widehat{f_1}(\rho)\|_{HS}^2 \|\widehat{f_2}(\rho)\|_{HS}^2.$$

Observe that, again by Parseval's identity,

$$\|f_1\|_{\ell^2}^2 = \frac{1}{|G|} \sum_{\rho \in \widehat{G}} d_\rho \|\widehat{f_1}(\rho)\|_{HS}^2 \geq \frac{d_\rho}{|G|} \|\widehat{f_1}(\rho)\|_{HS}^2$$

for any $\rho \in \widehat{G}$. Rearranging, this says that $d_\rho \|\widehat{f_1}(\rho)\|_{HS}^2 \leq |G| \|f_1\|_{\ell^2}^2$ for all $\rho \in \widehat{G}$. Plugging this back into the first inequality, we obtain that

$$\|f_1 * f_2\|_{\ell^2}^2 \leq \|f_1\|_{\ell^2}^2 \sum_{1 \neq \rho \in \widehat{G}} \|\widehat{f_2}(\rho)\|_{HS}^2 \leq \frac{|G|}{D} \|f_1\|_{\ell^2}^2 \frac{1}{|G|} \sum_{1 \neq \rho \in \widehat{G}} d_\rho \|\widehat{f_2}(\rho)\|_{HS}^2.$$

By Parseval's identity yet again, it follows that

$$\|f_1 * f_2\|_{\ell^2}^2 \leq \frac{|G|}{D} \|f_1\|_{\ell^2}^2 \|f_2\|_{\ell^2}^2.$$

Taking the square root of both sides completes the proof of the lemma. □

We can now prove a general mixing lemma for large subsets of quasirandom groups.

Lemma 13. *Let $D \geq 1$ and G be a finite D-quasirandom group. If $A, B, C \subset G$, then*

$$\left\| 1_A * 1_B - \frac{|A||B|}{|G|} \right\|_{\ell^2} \leq \frac{|G|^{1/2}}{D^{1/2}} \sqrt{|A||B|}.$$

and

$$\left\| 1_A * 1_B * 1_C - \frac{|A||B||C|}{|G|} \right\|_{\ell^\infty} \leq \frac{|G|^{1/2}}{D^{1/2}} \sqrt{|A||B||C|}.$$

Proof. The first inequality follows from writing $f_A := 1_A - \frac{|A|}{|G|}$, and using that $1_A * 1_B = (f_A + |A|/|G|) * 1_B = f_A * 1_B + \frac{|A||B|}{|G|}$ and the previous lemma. To prove the second inequality, we use that, for all $x \in G$,

$$\begin{aligned} \left| \left(1_A * 1_B - \frac{|A||B|}{|G|} \right) * 1_C(x) \right| &= \left| \sum_{y \in G} \left(1_A * 1_B - \frac{|A||B|}{|G|} \right) (xy^{-1}) 1_C(y) \right| \\ &\leq \left\| 1_A * 1_B - \frac{|A||B|}{|G|} \right\|_{\ell^2} \sqrt{|C|}, \end{aligned}$$

by the Cauchy–Schwarz inequality, and combine it with the first inequality and the fact that $[(|A||B|/|G|) * 1_C](x) = |A||B||C|/|G|$. □

This lemma was (essentially) originally due to Gowers in 2008, who used it to answer a question of Babai and Sós. A subset A of a group G is said to be *product free* if there are no $a_1, a_2, a_3 \in A$ for which $a_1 a_2 = a_3$. Babai and Sós asked whether any finite group G has a product free subset of size $\gg |G|$. Gowers showed that the answer to their question is negative for quasirandom G. This is a consequence of the following corollary of our mixing lemma.

Corollary 1. *Let $D \geq 1$ and G be a finite D-quasirandom group. If $A \subset G$ satisfies*

$$|A| > \frac{|G|}{D^{1/3}},$$

then there exist $a_1, a_2, a_3 \in A$ such that $a_1 a_2 = a_3$.

Proof. Apply the second inequality of the mixing lemma to get that

$$\left| 1_A * 1_A * 1_{A^{-1}}(0) - \frac{|A|^3}{|G|} \right| \leq \frac{|G|^{1/2}|A|^{3/2}}{|D|^{1/2}}.$$

Thus, $1_A * 1_A * 1_{A^{-1}}(0) > 0$ provided that

$$\frac{|A|^3}{|G|} > \frac{|G|^{1/2}|A|^{3/2}}{|D|^{1/2}}.$$

Rearranging gives the conclusion of the lemma. □

By an almost identical argument, if A, B, and C are sufficiently large, then $ABC = G$.

Corollary 2. *Let $D \geq 1$ and G be a finite D-quasirandom group. If $A, B, C \subset G$ with $|A||B||C| > \frac{|G|^3}{D}$, then $ABC = G$.*

We can now combine the above corollary with the fact that $\mathrm{SL}_2(\mathbf{F}_p)$ is $\frac{p-1}{2}$-quasirandom and Helfgott's product theorem to prove his diameter bound for Cayley graphs.

Proof. Since S generates $\mathrm{SL}_2(\mathbf{F}_p)$, we certainly have $|S| \geq 2$. Let $C > 0$ be some constant to be chosen shortly. Applying the product theorem with $\gamma \ll 1$, we obtain that

$$\left|(S \cup \{1\})^d\right| \geq C |\mathrm{SL}_2(\mathbf{F}_p)|^{9/10},$$

say, for $d \ll_C (\log p)^{O(1)}$. Since $|\mathrm{SL}_2(\mathbf{F}_p)| = p^3 - p$ and $\mathrm{SL}_2(\mathbf{F}_p)$ is $\frac{p-1}{2}$-quasirandom, to show that $(S \cup \{1\})^{3d} = \mathrm{SL}_2(\mathbf{F}_p)$, it suffices to check that

$$C |\mathrm{SL}_2(\mathbf{F}_p)|^{9/10} \geq \frac{|\mathrm{SL}_2(\mathbf{F}_p)|}{\left(\frac{p-1}{2}\right)^{1/3}} \gg \frac{p^3}{p^{1/3}} = p^{8/3}.$$

Since $|\mathrm{SL}_2(\mathbf{F}_p)|^{9/10} \asymp p^{27/10}$ and $27/10 > 8/3$, there exists a constant C such that the above always holds, which we will take to be our constant. □

The mixing lemma tells us that quasirandom groups have *product mixing* for large sets, in the sense that if G is D-quasirandom and $A, B, C \subset G$ have densities α, β, and γ, respectively, with α, β, γ sufficiently large (depending on $|G|$ and D), then

$$\#\{(a,b,c) \in A \times B \times C : ab = c\} \sim \alpha\beta\gamma|G|^2.$$

Note that such a result does not hold in abelian groups such as $\mathbf{Z}/N\mathbf{Z}$, since $A = \{0, 1, \ldots, N/10\}$ contains substantially more than $\frac{N^2}{1000}$ solutions to $a_1 + a_2 = a_3$. In the paper in which he introduced the notion of D-quasirandom groups, Gowers asked whether they also exhibit mixing for three-term geometric progressions x, xy, xy^2. Tao proved that this is the case for the family of quasirandom groups $\mathrm{SL}_d(\mathbf{F}_p)$, and then I proved it for all nonabelian finite simple groups, and then Bhangale, Harsha, and Roy proved it in full generality.

One interesting open question is whether quasirandom groups exhibit mixing for four-term geometric progressions x, xy, xy^2, xy^3. It's not even known whether any particular family of groups exhibits mixing for these configurations. Tao has some partial results in this direction in $\mathrm{SL}_2(\mathbf{F}_p)$ for the set of shifts restricted to the subset of y diagonalizable over $\mathbf{F}_p$.

3.6 The sum-product theorem

Consider first finite $A, B \subset \mathbf{Z}$. The simplest example we know of A for which $A + A$ is not much larger than A are intervals like $A = [N]$, and the simplest example we know of B for which BB is not much larger than A are geometric progressions like $B := \{1, 2, 4, \ldots, 2^n\}$. Note that, however $|B+B| = \frac{|B|(|B|+1)}{2} \gg |B|^2$ is maximally large, and $|AA| \gg \frac{N^2}{(\log N)^2}$ (since the number of primes in $[N]$ is asymptotically $\frac{N}{\log N}$) is almost maximally large[6]. Thus, it is natural to ask whether a subset of $\mathbf{Z}$ can have both small sumset and small product set. This is known as the sum-product problem.

Erdős and Szemerédi were the first to prove, in 1983, that there exists an absolute constant $\delta > 0$ such that

$$\max(|A + A|, |AA|) \gg |A|^{1+\delta} \tag{8}$$

for all finite $A \subset \mathbf{Z}$, and conjectured that, in fact

$$\max(|A + A|, |AA|) \geq |A|^{2-o(1)}$$

for all finite $A \subset \mathbf{R}$. There have been a number of quantitative improvements upon the lower bound (8), though obtaining an exponent of $2 - o(1)$ seems out of reach. We will mention some of these improvements later on. Note that the sum-product problem also makes sense in finite fields $\mathbf{F}_p$ (and, more generally, non-prime fields $\mathbf{F}_q$), provided one rules out the obvious issue of A being close to everything $\mathbf{F}_p$.

We will prove the following sum-product theorem that holds in both finite fields and $\mathbf{R}$ simultaneously.

[6]It is not hard to show by elementary methods in analytic number theory that $|[N]^2| = o(N^2)$. Determining the exact order of magnitude of $|[N]^2|$ is known as the *multiplication table problem*, and was fully resolved by Kevin Ford in 2008.

Theorem 13. *There exists an absolute constant $\delta_0 > 0$ such that the following holds. Let $0 < \delta < \delta_0$, $\mathbf{F}$ be a field, and $A \subset \mathbf{F}$ be finite and nonempty. Then either*

$$\max(|A+A|, |AA|) \gg |A|^{1+\delta},$$

or there exists a subfield $\mathbf{F}'$ of $\mathbf{F}$ with $|\mathbf{F}'| \ll |A|^{1+O(\delta)}$ and a nonzero $x \in \mathbf{F}$ such that $c\mathbf{F}'$ contains all but $O\left(|A|^{O(\delta)}\right)$ elements of A.

Note that the latter case cannot occur when $\mathbf{F}$ has characteristic zero, and so we obtain a sum-product theorem in $\mathbf{R}$, in particular. When $\mathbf{F}$ is a prime field, the latter case just says that $|A|$ is large: $|A| \gg |\mathbf{F}|^{1-O(\delta)}$. When $\mathbf{F}$ is finite but not prime, it says that A is close to a dilate of a subfield of $\mathbf{F}$.

Assume that $A \subset \mathbf{F}$ is finite. We showed previously that if $|A+A|$ is not much larger than $|A|$, then the same is true (with at most polynomial losses depending on n and m) for $|nA - mA|$ in general. We also showed that if $|A^2|$ is not much larger than $|A|$ and A doesn't contain 0, then the same holds (again, with at most polynomial losses depending on ℓ and k) for $|A^\ell A^{-k}|$ in general. One may hope that if both $|A+A|$ and $|AA|$ are not much larger than $|A|$, then arbitrary rational combinations of A must also not grow much. It turns out that this is not true in general. To see why, consider the subfield $\mathbf{F}_p$ of $\mathbf{F}_{p^2}$, let $\omega \in \mathbf{F}_{p^2} \setminus \mathbf{F}_p$, and set $A = \mathbf{F}_p \cup \{\omega\}$. Then, $|A+A| = |\mathbf{F}_p \cup (\omega + \mathbf{F}_p) \cup \{2\omega\}| < 2|A|$ and $|AA| = |\mathbf{F}_p \cup \omega\mathbf{F}_p \cup \{\omega^2\}| < 2|A|$, yet $A^2 + A^2 = \mathbf{F}_{p^2}$. One can rectify the situation by taking a large subset B of the set A, namely $\mathbf{F}_p$. Then, all polynomial combinations of B fail to grow the size of the set at all.

A lemma of Katz and Tao (often called the "Katz–Tao Lemma") shows that taking a large subset works in general.

Lemma 14. *Let $\mathbf{F}$ be a field and $A \subset \mathbf{F}^\times$ be finite and nonempty. Assume that $|A+A| \leq K|A|$ and $|AA| \leq K|A|$ for some $K \geq 1$. Then, there exists $B \subset A$ with $|B| \gg K^{-1}|A|$ such that $|B^2 - B^2| \ll K^{O(1)}|A|$.*

Proof. We may as well assume that $|A| \gg K^C$ for some absolute constant $C > 0$ (to be fixed later), or else the lemma is trivial. First of all, by double counting, we have that

$$|A|^2 = \left\| \sum_{a \in A} 1_{aA} \right\|_{\ell^1}.$$

(Note that $\sum_{a \in A} 1_{aA}$ is finitely supported by the assumption that A is finite, so taking this ℓ^1-norm makes sense.) Since $\sum_{a \in A} 1_{aA}$ is supported on AA, by the Cauchy–Schwarz inequality, we have

$$|A|^4 \leq |AA| \left\| \sum_{a \in A} 1_{aA} \right\|_{\ell^2}^2.$$

Using that $|AA| \leq K|A|$ and rearranging, we arrive at

$$\frac{|A|^3}{K} \leq \left\| \sum_{a \in A} 1_{aA} \right\|_{\ell^2}^2.$$

Expanding out the square reveals that the right-hand side above equals

$$\sum_{a,a' \in A} |aA \cap a'A|.$$

Thus, by the pigeonhole principle, there exists some $c \in A$ such that

$$\sum_{a \in A} |aA \cap cA| \geq \frac{|A|^2}{K}.$$

Since $c \neq 0$ by assumption, we may as well assume (by applying a dilation to A) that $c = 1$.

Now set $B := \{a \in A : |aA \cap A| \geq |A|/2K\}$, so that

$$\frac{|A|^2}{2K} \leq \sum_{a \in B} |aA \cap A| \leq |A||B|,$$

which implies that $|B| \geq |A|/2K$. Note that $|aA - A| \leq 2K^3|A|$ whenever $a \in B$. Indeed, if $a \in B$, then $|aA \cap A| \geq |A|/2K$ by definition, and since $|aA + aA| = |A + A| \leq K|A|$, we therefore have

$$|aA + (aA \cap A)|, |A + (aA \cap A)| \leq K|A|.$$

It then follows from Ruzsa's triangle inequality that $|aA - A| \leq 2K^3|A|$.

Similarly, if $b, b' \in B$, then $|bb'A - A| \ll 4K^6|A|$. Indeed, we have by the above that $|bA - A|, |bb'A - bA| \leq 2K^3|A|$, and so by another application of the Ruzsa triangle inequality, we have $|bb'A - A| \leq 4K^6|A|$.

Now, if $b, b' \in B^2$, then $|bA - A|, |b'A - A| \leq 4K^6|A|$, which implies, by the Ruzsa covering lemma, that each of bA and $b'A$ can each be covered by at most $4K^6$ translates of $A - A$, say $bA \subset X + (A - A)$ and $b'A \subset Y + (A - A)$ for $X, Y \subset \mathbf{F}$ each of size at most $4K^6$. Then,

$$|(b-b')A - A| \leq |bA - b'A - A| = |X - Y + (A - A) - (A - A) - A| \leq 16K^{12}|2A - 3A| \ll K^{O(1)}|A|$$

by the Plünneke–Ruzsa inequality.

The upshot is that if $b \in B^2 - B^2$, then $|bA - A| \ll K^{O(1)}|A|$. Thus, by the Cauchy–Schwarz inequality,

$$|A|^2 = \left\| \sum_{a \in A} 1_{bA-a} \right\|_{\ell^1} \leq \sqrt{|bA - A|} \left\| \sum_{a \in A} 1_{bA-a} \right\|_{\ell^2} \ll K^{O(1)}|A|^{1/2} \left\| \sum_{a \in A} 1_{bA-a} \right\|_{\ell^2}.$$

Note that the square of the ℓ^2-norm on the right-hand side equals

$$\sum_{a,a' \in A} |(bA - a) \cap (bA - a')| = \#\left\{(a_1, a_2, a_3, a_4) \in A^4 : ba_1 - a_2 = ba_3 - a_4\right\}.$$

So, for each $b \in B^2 - B^2$, there are $\gg K^{-O(1)}|A|^3$ choices of $(a_1, a_2, a_3, a_4) \in A^4$ for which $ba_1 - a_2 = ba_3 - a_4$. The number of these for which $a_1 = a_3$, which implies that $a_2 = a_4$ as well, is $\ll |A|^2$. Thus, provided that $|A| \gg K^{O(1)}$, there are, in fact, $\gg K^{-O(1)}|A|^3$ choices of $(a_1, a_2, a_3, a_4) \in A^4$ for which $ba_1 - a_2 = ba_3 - a_4$ with $a_1 \neq a_3$ is also $\gg K^{-O(1)}|A|^3$. Note, however, that any such quadruple of a_i's uniquely determines b, since $b = \frac{a_2 - a_4}{a_1 - a_3}$. It follows that the number of possible b's is $\ll K^{O(1)}|A|$, i.e., that $|B^2 - B^2| \ll K^{O(1)}|A| \ll K^{O(1)}|B|$. □

The Katz–Tao lemma can be further upgraded to give control of more polynomial combinations of B. The proof of the following lemma will be left as an exercise.

Lemma 15. *Let $\mathbf{F}$ be a field and $A \subset \mathbf{F}^\times$ be finite and nonempty. Assume that $|A^2 - A^2| \leq K|A|$ for some $K \geq 1$. Then, for all $n \in \mathbf{N}$, we have $|A^n - A^n| \leq K^{O(n)}|A|$.*

Next, we prove a dichotomy-like lemma analogous to the one from last class.

Lemma 16. *Let $\mathbf{F}$ be a field, $A \subset \mathbf{F}$ be finite and nonempty, and $x \in \mathbf{F}$. Then at least one of the following two statements holds:*

1. $|A + xA| = |A|^2$
2. $|A + xA| \leq |(A - A)A + (A - A)A|$

Proof. Suppose that $|A + xA| \neq |A|^2$, which means that $|A + xA| < |A|^2$, and thus that the map $\phi : A \times A \to A + xA$ defined by $\phi(a, a') = a + xa'$ is not injective. So, there exist $a_1, a_2, a_3, a_4 \in A$ with $(a_1, a_2) \neq (a_3, a_4)$ such that $a_1 + xa_2 = a_3 + xa_4$, i.e.

$$x = \frac{a_1 - a_3}{a_4 - a_2},$$

since $a_2 \neq a_4$. Thus,

$$|A + xA| = \left|A + \frac{a_1 - a_3}{a_4 - a_2}A\right| = |(a_4 - a_2)A + (a_1 - a_3)A| \leq |(A - A)A + (A - A)A|,$$

as desired. □

Observe that if B is as in the Katz–Tao lemma, then $|(B - B)B + (B - B)B| \leq |2B^2 - 2B^2| \ll K^{O(1)}|B|$ by the Plünnecke–Ruzsa inequality. Thus, for such B, this lemma gives a true dichotomy (provided $|B|$, and thus $|A|$, is large enough in terms of K).

Now we can prove the sum-product theorem

Proof of Theorem 13. We will assume that $|A + A|, |AA| \leq |A|^{1+\delta} =: K|A|$ for sufficiently small $\delta > 0$. Note that, by taking δ small enough and assuming that $|A|$ is larger than a sufficiently large absolute constant, we can force $|A| \geq C_1 K^{C_2}$ for any fixed $C_1, C_2 > 0$, and also may as well assume that $0 \notin A$. We can then apply the Katz–Tao lemma to obtain $B \subset A$ with $|B| \gg K^{-1}|A|$ such that $|B^2 - B^2| \ll K^{O(1)}|B|$. By applying a dilation, we may assume that $1 \in B$. By Lemma 15, $|B^n - B^n| \ll_n K^{O(n)}|B|$ for all $n \in \mathbf{N}$. It then follows from the Plünnecke–Ruzsa inequality and having $1 \in B$ that any iterated sumset of at most n (dilated) product sets of the form $\pm B^m$ for $m \leq n$ has size at most $\ll_n K^{O(n^2)}|B|$.

By the dichotomy lemma, for all $x \in \mathbf{F}$, either $|B + xB| = |B|^2$ (the "noninvolved" case) or $|B + xB| \leq CK^C|B|$ (the "involved" case) for some fixed absolute constant $C > 0$. Note that all elements of B are involved (provided we take $|A| \gg K^{O(1)}$), since $|B^2 - B^2| \ll K^{O(1)}|B|$ implies that $|B + BB| \leq |B^2 + B^2| \ll K^{O(1)}|B|$ since $1 \in B$. Similarly, by the discussion above, if $x_1, x_2 \in \mathbf{F}$ are involved, then

$$|B + x_1x_2B|, |B + (x_1 + x_2)B|, |B + (x_1 - x_2)B| \ll K^{O(1)}|B|$$

which, provided that $|A| \gg K^{O(1)}$, implies that x_1x_2, x_1+x_2, x_1-x_2 are all involved. Thus, the set of involved elements is closed under mutliplication, addition, and subtraction. Since 0 and 1 are also, clearly, involved elements, it follows that the set R of involved elements form a nontrivial subring of $\mathbf{F}$. Recalling that all elements of B are involved, we get that $|R| \gg K^{-O(1)}|A|$. To see that R is, in fact, a field, it suffices to show that $|R| \ll K^{O(1)}|A|$, which, in particular, implies that R is finite.

To see that $|R| \ll K^{O(1)}|A|$, we argue as in the end of the proof of the Katz–Tao lemma. Let $x \in R$. Then,

$$|B|^2 = \left\| \sum_{b \in B} 1_{b+xB} \right\|_{\ell^1} \leq \sqrt{|B+xB|} \left\| \sum_{b \in B} 1_{b+xB} \right\|_{\ell^2} \ll K^{O(1)}|B|^{1/2} \left\| \sum_{b \in B} 1_{b+xB} \right\|_{\ell^2},$$

so that

$$\# \left\{ (b_1, b_2, b_3, b_4) \in B^4 : b_1 + xb_2 = b_3 + xb_4 \right\} \gg K^{-O(1)}|B|^3,$$

and thus, assuming that $|A| \gg K^{O(1)}$, we have that, for each $x \in R$, there are $\gg K^{-O(1)}|B|^3$ quadruples $(b_1, b_2, b_3, b_4) \in B^4$ with $b_2 \neq b_4$ such that $b_1 + xb_2 = b_3 + xb_4$. Each quadruple uniquely determines x and there are only $|B|^4$ total quadruples, and so we must have $|R| \ll K^{O(1)}|B| \ll K^{O(1)}|A|$, as desired.

To finish (since we may have applied a nontrivial dilation at the beginning of the proof), we will show that R that contains all but $O(K^{O(1)})$ elements of A. We noted above that $B \subset R$, so that, since the B provided by the Katz–Tao lemma is a subset of A, $|A \cap R| \geq |B| \gg K^{-O(1)}|A|$. Thus, by the Ruzsa triangle inequality, we have

$$|A + R| \leq \frac{|A + (A \cap R)||(A \cap R) + R|}{|A \cap R|} \ll K^{O(1)}|R|.$$

It follows from the Ruzsa covering lemma that A is contained in the union of $\ll K^{O(1)}$ translates $s_1 + R, \ldots, s_K + R$ of R. Similarly,

$$|AR| \leq \frac{|A(A \cap R)||(A \cap R)R|}{|A \cap R|} \ll K^{O(1)}|R|,$$

and so, by the Ruzsa covering lemma again, A is contained in the union of $\ll K^{O(1)}$ dilates $x_1R, \ldots, x_LR$, with $x_1, \ldots, x_L \neq 0$, of R. Note that if $x \notin R$, then

$$|(s + R) \cap xR| \leq 1.$$

Indeed, if $s + r_1 = xr_2$ and $s + r_3 = xr_4$ for $r_1, r_2, r_3, r_4 \in R$, then $r_1 - r_3 = x(r_2 - r_4)$. If $r_2 \neq r_4$, then this equation forces $x \in R$, which contradicts our assumption. If $r_2 = r_4$, then $r_1 = r_3$ as well. Thus, all but $\ll K^{O(1)}$ elements of A are contained in R. □

The finite field setting is more difficult to prove sum-product theorems in than the integer or real setting, since in the latter settings one can take advantage of "geometric" properties of $\mathbf{R}^2$. While the first sum-product result in the real setting was proven in the 1980s, it wasn't until work of Bourgain–Glibichuk–Konyagin in 2006 that the first true (i.e., not requiring a nontrivial lower bound on the size of the set) sum-product theorem was proven in the finite

field setting. The world record for δ_0 that can be taken in the sum-product theorem for prime fields when $|A| \leq p^{1/2}$ is $\delta_0 = \frac{1}{4}$, due to Mohammadi and Stevens in 2023. The exponent $\delta_0 = \frac{1}{4}$ was obtained in the real setting in 1997, using a tool from incidence geometry (the Szemerédi–Trotter theorem). Exploiting geometry in $\mathbf{R}^2$ in a different way, Solymosi proved an exponent of $\delta_0 = \frac{1}{3}$ in the sum-product theorem in the reals in 2009. The current world record in the sum-product theorem in the reals is $\delta_0 = \frac{1}{3} + \frac{2}{1167}$, due to Rudnev and Stevens in 2020.

3.7 The Balog–Szemerédi–Gowers theorem

Let $(G, +)$ be an abelian group and $A \subset G$ be finite. The *additive energy* $E(A)$ of A is the number of *additive quadruples*

$$a_1 + a_2 = a_3 + a_4 \qquad a_1, a_2, a_3, a_4 \in A$$

in A, i.e.,

$$E(A) = \#\left\{(a_1, a_2, a_3, a_4) \in A^4 : a_1 + a_2 = a_3 + a_4\right\}.$$

We can also define an asymmetric notion of additive energy. If $B \subset G$ is finite as well, then

$$E(A, B) := \#\left\{(a, a', b, b') \in A^2 \times B^2 : a + b = a' + b'\right\}.$$

Note that we have the trivial bounds

$$|A|^2 \leq E(A) \leq |A|^3$$

and

$$|A||B| \leq E(A, B) \leq \min\left(|A||B|^2, |A|^2|B|\right) \leq |A|^{3/2}|B|^{3/2}.$$

An example of a set with large additive energy is $[N] \subset \mathbf{Z}$, since $E([N]) \gg N^3$. An example of a set with small additive energy is $B = \{1, 2, 4, \ldots, 2^N\}$, since $E(B) = 2\binom{N}{2} + N \gg N^2$. Additive energy and the size of sumsets can be related by a standard application of the Cauchy–Schwarz inequality. Indeed, let $A, B \subset G$ be finite, and note that

$$|A||B| = \sum_{x \in G} (1_A * 1_B)(x) \leq |A + B|^{1/2} \, \|1_A * 1_B\|_{\ell^2},$$

and thus

$$\frac{|A|^2|B|^2}{|A + B|} \leq \|1_A * 1_B\|_{\ell^2}^2 = \sum_{x \in G} (1_A * 1_B)(x)^2 = E(A, B).$$

So, $|A + B|$ being small implies that $E(A, B)$ is large. The converse fails to hold, however. Consider the subset

$$A := [N] \cup \left\{2^N, \ldots, 2^{N-1}\right\}$$

of $\mathbf{Z}$, which has size $2N$. Then $E(A) \geq E([N]) \gg N^3 \gg |A|^3$, but $|A + A| \geq \binom{N}{2} + N \gg N^2 \gg |A|^2$.

The Balog–Szemerédi–Gowers theorem provides a partial converse, saying that a set with large additive energy must contain a large subset with small doubling. We will actually prove a more general, asymmetric version.

Theorem 14. *Let $(G,+)$ be an abelian group and $A, B \subset G$ be finite. Suppose that*

$$E(A,B) \geq \frac{|A|^{3/2}|B|^{3/2}}{K}$$

for some $K \geq 1$. Then, there exist $A' \subset A$ and $B' \subset B$ with $|A'| \gg K^{-O(1)}|A|$ and $|B'| \gg K^{-O(1)}|B|$ such that $|A'+B'| \ll K^{O(1)}|A'|^{1/2}|B'|^{1/2}$.

A version of this result that was far weaker quantitatively was first proven by Balog and Szemerédi, and was improved to polynomial dependence on K by Gowers in his paper proving the first reasonable bounds in Szemerédi's theorem for 4-APs.

Corollary 3. *Let $(G,+)$ be an abelian group and $A \subset G$ be finite. Suppose that*

$$E(A) \geq \frac{|A|^3}{K}$$

for some $K \geq 1$. Then, there exists $A' \subset A$ with $|A'| \gg K^{-O(1)}|A|$ such that $|A'+A'| \ll K^{O(1)}|A'|$.

The above symmetric version of the Balog–Szemerédi–Gowers theorem can be deduced from the asymmetric version using the Ruzsa triangle inequality.

We will deduce Theorem 14 from the following purely graph-theoretic statement.

Lemma 17. *Let A and B be finite sets and $H = (A \sqcup B, E)$ be a bipartite graph with edge set E between A and B. Suppose that*

$$|E| \geq \frac{|A||B|}{K}$$

for some $K \geq 1$. Then, there exist $A' \subset A$ and $B' \subset B$ with $|A'| \gg K^{-O(1)}|A|$ and $|B'| \gg K^{-O(1)}|B|$ such that every $a' \in A'$ and $b' \in B'$ are joined by $\gg K^{-O(1)}|A||B|$ paths of length three.

Note that the maximum possible number of paths of length three between A and B is $|A|^2|B|^2$, and the maximum between any fixed $a \in A$ and $b \in B$ is $|A||B|$. Thus, this lemma says that in any dense bipartite graph one can find a dense bipartite subgraph $A' \times B'$ such that any $a \in A'$ and $b \in B'$ are connected by a positive proportion of the maximum number of paths of length three between a and b.

Proof of Theorem 14 from Lemma 17. Set

$$E := \left\{ (a,b) \in A \times B : (1_A * 1_B)(a+b) > \frac{|A|^{1/2}|B|^{1/2}}{2K} \right\}.$$

Then, since

$$\sum_{(a,b) \in A \times B} (1_A * 1_B)(a+b) = E(A,B) \geq \frac{|A|^{3/2}|B|^{3/2}}{K}$$

and

$$\sum_{(a,b)\in(A\times B)\setminus E}(1_A * 1_B)(a+b) \le \frac{|A|^{3/2}|B|^{3/2}}{2K},$$

it follows that

$$\frac{|A|^{3/2}|B|^{3/2}}{2K} \le \sum_{(a,b)\in E}(1_A * 1_B)(a+b) \le |E|\sqrt{|A||B|}$$

since $|(1_A * 1_B)(x)| \le \|1_A\|_{\ell^2}\|1_B\|_{\ell^2}$ by the Cauchy–Schwarz inequality. It follows that $|E| \ge |A||B|/2K$.

Now, consider the bipartite graph H with vertex sets A and B and edge set E (for each element in $A \cap B$, we create a vertex in both A and B viewed as vertex sets, and view each edge in E as an undirected edge despite the elements of E being ordered pairs). By Lemma 17, there exist $A' \subset A$ and $B' \subset B$ with $|A'| \gg K^{-O(1)}|A|$ and $|B'| \gg K^{-O(1)}|B|$ such that the number of paths of length three between any $a' \in A'$ and $b' \in B'$ is $\gg K^{-O(1)}|A||B|$. That is, for all $a' \in A'$ and $b' \in B'$, there exist $\gg K^{-O(1)}|A||B|$ pairs $a, b \in A \times B$ for which $(a', b), (a, b), (a, b') \in E$. It follows that

$$\sum_{a,b\in G}(1_A * 1_B)(a'+b)(1_A * 1_B)(a+b)(1_A * 1_B)(a+b') \gg K^{-O(1)}|A|^{5/2}|B|^{5/2}.$$

The left-hand side above equals

$$\begin{aligned}
&\sum_{a,b\in G}(1_A * 1_B)(a'+b)(1_{-A} * 1_{-B})(-a-b)(1_A * 1_B)(a+b') \\
&= \sum_{a,b\in G}(1_A * 1_B)(b)(1_{-A} * 1_{-B})(-a-b+(a'+b'))(1_A * 1_B)(a) \\
&= \sum_{\substack{c_1,c_2,c_3\in G \\ c_1+c_2+c_3=a'+b'}}(1_A * 1_B)(c_1)(1_{-A} * 1_{-B})(c_2)(1_A * 1_B)(c_3) \\
&= (1_A * 1_B * 1_{-A} * 1_{-B} * 1_A * 1_B)(a'+b')
\end{aligned}$$

by making the changes of variables $a \mapsto a - b'$ and $b \mapsto b - a'$ and then $c_1 = b$, $c_2 = -a - b + (a' + b')$, and $c_3 = b'$.

Thus, we have

$$(1_A * 1_B * 1_{-A} * 1_{-B} * 1_A * 1_B)(a'+b') \gg K^{-O(1)}|A|^{5/2}|B|^{5/2}$$

for all $a' \in A'$ and $b' \in B'$. On the other hand,

$$\sum_{x\in G}(1_A * 1_B * 1_{-A} * 1_{-B} * 1_A * 1_B)(x) = |A|^3|B|^3.$$

It follows that

$$|A'+B'| \ll \frac{|A|^3|B|^3}{K^{-O(1)}|A|^{5/2}|B|^{5/2}} \ll K^{O(1)}|A|^{1/2}|B|^{1/2},$$

as desired. □

In order to prove Lemma 17, we will need a preparatory lemma about paths of length two.

Lemma 18. *Let A and B be finite sets, $H = (A \sqcup B, E)$ be a bipartite graph with edge set E between A and B, and $\varepsilon > 0$. Suppose that*

$$|E| \geq \frac{|A||B|}{K}$$

for some $K \geq 1$. Then, there exist $A' \subset A$ with $|A'| \geq |A|/\sqrt{2}K$ such that all but at most $\varepsilon|A'|^2$ pairs of vertices $a'_1, a'_2 \in A'$ are joined by at least $\frac{\varepsilon}{2K^2}|B|$ paths of length two.

Proof. The proof proceeds by a technique in extremal combinatorics known as dependent random choice. We will choose A' to be the set of neighbors of a random element $b \in B$; such sets are more likely to be "well connected" than a typical random subset of the same size, since all elements at least share an edge with b.

Let $b \in B$ be chosen uniformly at random and let $A' = A'(b) := \{a \in A : \{a, b\} \in E\}$ be the neighborhood of b. The expected size of A' is

$$\mathbf{E}_{b\in B}|A'| = \mathbf{E}_{b\in B}\#\{a \in A : \{a, b\} \in E\} = \frac{|E|}{|B|} \geq \frac{|A|}{K} \tag{9}$$

by the assumption on the size of $|E|$. Set

$$S := \left\{(a'_1, a'_2) \in (A')^2 : a'_1 \text{ and } a'_2 \text{ are connected by fewer than } \frac{\varepsilon|B|}{2K^2} \text{ paths of length two}\right\}.$$

Then

$$\mathbf{E}_{b\in B}|S| < \frac{|A|^2}{|B|}\frac{\varepsilon|B|}{2K^2} = \frac{\varepsilon|A|^2}{2K^2},$$

since any pair $(a_1, a_2) \in A^2$ that is connected by $< \frac{\varepsilon|B|}{2K^2}$ paths of length two can only lie in $< \frac{\varepsilon|B|}{2K^2}$ different sets $A'(b)^2$.

By (9) and the Cauchy–Schwarz inequality,

$$\frac{|A|^2}{K^2} \leq \mathbf{E}_{b\in B}|A'|^2.$$

Thus, we have

$$\mathbf{E}_{b\in B}\left(|A'|^2 - \frac{|S|}{\varepsilon} - \frac{|A|^2}{2K^2}\right) \geq 0,$$

and so there must exist $b \in B$ for which

$$|A'|^2 - \frac{|S|}{\varepsilon} \geq \frac{|A|^2}{2K^2}.$$

It then follows that $|A'| \geq |A|/\sqrt{2}K$, and also that $\frac{|S|}{\varepsilon} \leq |A'|^2$, so that $|S| \leq \varepsilon|A'|^2$. □

Now, we can finally prove Lemma 17 and complete the proof of the Balog–Szemerédi–Gowers theorem.

Proof of Lemma 17. Let $A_1 \subset A$ be the set of vertices of degree at least $\frac{|B|}{2K}$ and let $E_1 \subset E$ be the set of edges in E connecting A_1 to B. Then,

$$|E_1| > \frac{|A||B|}{K} - \frac{|A||B|}{2K} = \frac{|A||B|}{2K}$$

and thus

$$|A_1| \geq \frac{|E_1|}{|B|} > \frac{|A|}{2K}.$$

Now, we apply Lemma 18 with the bipartite graph (A_1, B, E_1) and $\varepsilon > 0$ to be chosen later. This yields $A_2 \subset A_1$ with $|A_2| \gg K^{-O(1)}|A|$ such that all but $\varepsilon|A_2|^2$ of the pairs of vertices $a, a' \in A_2$ are joined by at least $C^{-1}K^{-C}\varepsilon|B|$ paths of length two for some absolute constant $0 < C \ll 1$. Let $E_2 \subset E$ be the set of edges connecting A_2 with B.

Let $A' \subset A_2$ be the set

$$\left\{a \in A_2 : \left|\left\{a' \in A_2 : a, a' \text{ joined by fewer than } C^{-1}K^{-C}\varepsilon|B| \text{ paths of length } 2\right\}\right| \leq \sqrt{\varepsilon}|A_2|\right\}.$$

Then,

$$|A_2 \setminus A'| \cdot \sqrt{\varepsilon}|A_2| \leq \varepsilon|A_2|^2,$$

and so $|A_2 \setminus A'| \leq \sqrt{\varepsilon}|A_2|$. Also note that since every vertex in $A_2 \subset A_1$ has degree at least $|B|/2K$,

$$|E_2| \geq \frac{|A_2||B|}{2K} \gg \frac{|A||B|}{K^{O(1)}}$$

and thus, by an argument similar to the one at the beginning of the proof, there exist $B' \subset B$ with $|B'| \geq (C')^{-1}K^{-C'}|B|$ such that each $b \in B'$ has degree at least $(C')^{-1}K^{-C'}|A|$ in the bipartite graph (A_2, B, E_2) for some $0 < C' \ll 1$.

Now, by taking $\varepsilon = D^{-1}K^{-D}$ for some fixed $D \gg 1$, we get that any pair of vertices $(a', b') \in A' \times B'$ is connected by $\gg K^{-O(1)}|A||B|$ paths of length three. Indeed, b' has at least $(C')^{-1}K^{-C'}|A|$ neighbors $a \in A_2$, and the number of those for which a and a' are joined by fewer than $C^{-1}K^{-C}\varepsilon|B|$ paths of length two is at most $\sqrt{\varepsilon}|A_2| \leq \sqrt{\varepsilon}|A|$. Taking $D = 2C'$ then yields that b' has at least $(2C')^{-1}K^{-C'}|A|$ neighbors $a \in A_2$ for which a and a' are joined by at least $(2CC')^{-1}K^{-(C+2C')}|B|$ paths of length two. This means that b' and a' are connected by at least

$$(2C')^{-1}K^{-C'}|A| \cdot (2CC')^{-1}K^{-(C+2C')}|B| \gg K^{-O(1)}|A||B|$$

paths of length three, as desired. □

4 An application to bounding exponential sums

Next, we will present a beautiful application of several results developed in the previous section (most notably, the sum-product and Balog–Szemerédi–Gowers theorems) to bounding the size of additive characters of $\mathbf{F}_p$ over multiplicative subgroups, due to Bourgain, Glibichuk, and Konyagin.

Theorem 15 (Bourgain–Glibichuk–Konyagin, 2006)**.** *For all $\delta > 0$, there exists $\varepsilon = \varepsilon(\delta) > 0$ such that the following holds. Let $p \gg_\delta 1$ and $H \leq \mathbf{F}_p^\times$ be a multiplicative subgroup such that $|H| \geq p^\delta$. Then,*

$$\left|\sum_{x \in H} e_p(\xi x)\right| \leq p^{-\varepsilon}|H|$$

for all nonzero $\xi \in \mathbf{F}_p$.

This theorem says that we can get nontrivial bounds for nontrivial additive character sums over multiplicative subgroups of $\mathbf{F}_p$ of size an arbitrarily small power of p. Prior approaches, which were either purely number-theoretic or incorporated bounds for the number of $\mathbf{F}_p$-points on certain varieties, failed to get below the $p^{1/4}$ barrier. Bourgain and others then extended the ideas going into the proof to bound a wider variety of interesting exponential sums, and also to the more general situation of H having small doubling $|HH| \leq K|H|$.

One should contrast this result with the state of our knowledge of nontrivial multiplicative (i.e., nonprincipal Dirichlet) character sums over short intervals, which are the most "additively structured" subsets of $\mathbf{F}_p$. The classical Pólya–Vinogradov inequality says that

$$\left|\sum_{M < n \leq M+N} \chi(n)\right| \leq 6\sqrt{q}\log q$$

whenever χ is a nonprincipal Dirichlet character modulo q; this inequality is only nontrivial for $N \gg_\varepsilon q^{1/2+\varepsilon}$. Burgess's bound provides nontrivial bounds for such character sums when $N \gg_\varepsilon q^{1/4+\varepsilon}$, and it's a major open problem to obtain nontrivial bounds in general for sums over shorter intervals (any progress would, for example, improve on the best known upper bound for the least quadratic nonresidue modulo a prime). On GRH, we can obtain nontrivial bounds for character sums of length $\gg_\varepsilon q^{O(\varepsilon)}$.

We begin by defining the notion of the large spectrum of a set, which is simply the set of frequencies at which the Fourier transform is large.

Definition 3. *Let $A \subset \mathbf{F}_p$ with density α and $\delta \in [0,1]$. The δ-large spectrum of A is the set*

$$\mathrm{Spec}_\delta(A) := \left\{\xi \in \mathbf{F}_p : \left|\widehat{1_A}(\xi)\right| \geq \delta\alpha\right\}.$$

Observe that $\mathrm{Spec}_1(A) = \{0\}$ and $\mathrm{Spec}_0(A) = \mathbf{F}_p$. Further, from Parseval's identity,

$$\alpha = \mathbf{E}_{x \in \mathbf{F}_p} 1_A(x)^2 = \sum_{\xi \in \mathbf{F}_p} \left|\widehat{1_A}(\xi)\right|^2 \geq (\delta\alpha)^2 \left|\mathrm{Spec}_\delta(A)\right|,$$

so that, by rearranging, we obtain the upper bound

$$|\mathrm{Spec}_\delta(A)| \leq \frac{1}{\alpha\delta^2}.$$

An important fact about the large spectrum of a set is that it posesses some weak additive structure.

Lemma 19. *Let $A \subset \mathbf{F}_p$ with density α and $\delta \in (0,1]$. For any nonempty $S \subset \mathrm{Spec}_\delta(A)$, we have*

$$\#\left\{(\xi_1,\xi_2) \in S^2 : \xi_1 - \xi_2 \in \mathrm{Spec}_{\delta^2/2}(A)\right\} \geq \frac{\delta^2}{2}|S|^2$$

Proof. For each $\xi \in \mathbf{F}_p$, let $\phi(\xi) \in \mathbf{C}$ be such that $\left|\widehat{1_A}(\xi)\right| = \phi(\xi)\widehat{1_A}(\xi)$, so that $|\phi(\xi)| = 1$. We certainly have

$$\delta\alpha|S| \leq \sum_{\xi\in S}\left|\widehat{1_A}(\xi)\right| = \sum_{\xi\in S}\phi(\xi)\mathbf{E}_{x\in\mathbf{F}_p}1_A(x)e_p(-\xi x) = \mathbf{E}_{x\in\mathbf{F}_p}1_A(x)\sum_{\xi\in S}\phi(\xi)e_p(-\xi x).$$

By the triangle inequality and then the Cauchy–Schwarz inequality, we get from the above that

$$\begin{aligned}\alpha^2\delta^2|S|^2 &\leq \alpha\mathbf{E}_{x\in\mathbf{F}_p}1_A(x)\left|\sum_{\xi\in S}\phi(\xi)e_p(-\xi x)\right|^2\\ &= \alpha\mathbf{E}_{x\in\mathbf{F}_p}1_A(x)\sum_{\xi_1,\xi_2\in S}\phi(\xi_1)\overline{\phi(\xi_2)}e_p(-(\xi_1-\xi_1)x)\\ &= \alpha\sum_{\xi_1,\xi_2\in S}\phi(\xi_1)\overline{\phi(\xi_2)}\widehat{1_A}(\xi_1-\xi_2).\end{aligned}$$

Thus, by the triangle inequality,

$$\alpha\delta^2|S|^2 \leq \sum_{\xi_1,\xi_2\in S}\left|\widehat{1_A}(\xi_1-\xi_2)\right|.$$

Since $\left|\widehat{1_A}(\xi_1-\xi_2)\right| \leq \alpha$ and the contribution to the above from pairs (ξ_1,ξ_2) for which $\xi_1-\xi_2 \notin \mathrm{Spec}_{\delta^2/2}(A)$ is at most $\frac{\alpha\delta^2}{2}|S|^2$, we conclude that

$$\#\left\{(\xi_1,\xi_2) \in S^2 : \xi_1 - \xi_2 \in \mathrm{Spec}_{\delta^2/2}(A)\right\} \geq \frac{\delta^2}{2}|S|^2.$$

□

In 2002, Chang proved a more precise and powerful statement about the additive structure of large spectra. To state it, we will need a couple of simple definitions.

Definition 4. *Let $(G,+)$ be any abelian group and $S \subset G$. We say that S is dissociated if the only choice of $(\epsilon_s)_{s\in S} \in \{-1,0,1\}^S$ for which*

$$\sum_{s\in S}\epsilon_s s = 0$$

is the zero vector. For any $A \subset G$, the dimension is the size of the largest dissociated subset of A.

Chang proved a bound on the dimension of the δ-large spectrum of subsets of cyclic groups. In the setting we're focusing on in this section, her result is as follows.

Lemma 20 (Chang, 2002)**.** *Let $A \subset \mathbf{F}_p$ with density α and $\delta \in (0,1]$. Then, the dimension of $\mathrm{Spec}_\delta(A)$ is $\ll \frac{\log \alpha^{-1}}{\delta^2}$.*

One should compare this to our trivial size upper bound $|\mathrm{Spec}_\delta(A)| \leq \frac{1}{\alpha\delta^2}$. Thus, when the density α is small, the dimension of the δ-large spectrum is substantially smaller than its maximum possible size.

Chang used her lemma to improve the best known bounds in the Freiman–Ruzsa theorem, which gives a classification of subsets of the integers with small doubling, and which will be the next major result we prove in this course. Chang's lemma has since found many applications in additive combinatorics, such as to improving the bounds on Roth's theorem (due to Sanders, Bloom, and Bloom–Sisask (in various forms) but now superceded by the Kelley–Meka bounds), in further improving the bounds in the Freiman–Ruzsa theorem (due to Sanders, and which implies quasipolynomial bounds in the inverse theorem for the U^3-norm), and finding long arithmetic progressions in sumsets (due to Green). A version of Chang's lemma for vector spaces over finite fields also has applications in theoretical computer science, in particular, to the analysis of boolean functions.

Returning to the proof of the Bourgain–Glibichuk–Konyagin bound, the basic idea of the argument is that if $H \leq \mathbf{F}_p^\times$ is a multiplicative subgroup and $A := \mathrm{Spec}_\delta(H)$ is more than just the zero frequency, then A is, in fact, 0 along with a union of cosets of H, since

$$\sum_{h \in H} e_p(\xi h) = \sum_{h \in H} e_p(\xi h' h) = \sum_{h \in H} e_p([h'\xi]h)$$

for all $h' \in H$, so that if $\xi \in \mathrm{Spec}_\delta(H)$, then $h'\xi \in \mathrm{Spec}_\delta(H)$. This should translate to A having some amount of nontrivial multiplicative structure. On the other hand, we saw above that A also has some weak additive structure, as long as $\mathrm{Spec}_{\delta^2/2}(H)$ is not too much larger than $A = \mathrm{Spec}_\delta(H)$. The sum-product theorem tells us, morally, that a set cannot simultaneously be both additively structured and multiplicatively structured, and so this should lead to a contradiction unless $A = \{0\}$, which will lead us to a good bound on the original exponential sum. To make all of this talk of "weak additive/multiplicative structure" rigorous, we will use the Balog–Szemerédi–Gowers theorem, and to ensure that $\mathrm{Spec}_{\delta^2/2}(H)$ is not much larger than $\mathrm{Spec}_\delta(H)$, we will use a dyadic pigeonholing argument, which was one of Bourgain's signature techniques.

We begin with deriving a couple of useful consequences of the sum-product theorem. The first is that subsets of $\mathbf{F}_p$ grow rapidly under iterated combined sumsets and product sets

Lemma 21. *Let $A \subset \mathbf{F}_p$ be nonempty, $m \in \mathbf{N}$, and $\delta \in (0,1)$. Then, there exists $k \ll_{m,\delta} 1$ such that*

$$\left|kA^k\right| \gg_{m,\delta} \min\left(|A|^m, p^{1-\delta}\right).$$

Proof. First, observe that if $n \in \mathbf{N}$ with $n \geq 2$, then

$$|nA^n + nA^n|, |(nA^n)(nA^n)| \leq |n^2 A^{n^2}|.$$

Now let $\varepsilon > 0$ be such that, for all $A \subset \mathbf{F}_p$ with $|A| \leq Cp^{1-\delta}$,

$$\max\left(|A+A|, |AA|\right) \geq C|A|^{1+\varepsilon}$$

for some absolute constant $C > 0$. Then, for all $n \in \mathbf{N}$, we have

$$\left|2^{2^n} A^{2^{2^n}}\right| \geq C^{n(1+\varepsilon)^n} |A|^{(1+\varepsilon)^n}$$

unless $\left|2^{2^{n-1}} A^{2^{2^{n-1}}}\right| > Cp^{1-\delta}$, in which case we certainly have $\left|2^{2^n} A^{2^{2^n}}\right| > Cp^{1-\delta}$ as well. Taking $k = 2^{2^n}$ for $n = \log_{(1+\varepsilon)} m$ completes the proof. □

It follows that iterated combined sumsets and product sets expand to fill all of $\mathbf{F}_p$ after few iterations.

Lemma 22. *Let $\delta \in (0,1)$ and $A \subset \mathbf{F}_p$ with $|A| \geq p^\delta$. Then, there exists $k \ll_\delta 1$ such that $kA^k = \mathbf{F}_p$.*

This will be an immediate consequence of the previous lemma and the following one.

Lemma 23. *Let $A \subset \mathbf{F}_p^\times$ with $|A| > p^{3/4}$. Then, $3A^2 = \mathbf{F}_p$.*

Proof. Set $f := \mathbf{E}_{a \in A} 1_{aA}$, and note that, for all $\xi \in \mathbf{F}_p^\times$, we have

$$\left|\widehat{f}(\xi)\right| = \left|\mathbf{E}_{a \in A} \widehat{1_A}(a\xi)\right| \leq \sqrt{\mathbf{E}_{a \in A} \left|\widehat{1_A}(\xi)\right|^2} \leq |A|^{-1/2} \sqrt{\sum_{\eta \in \mathbf{F}_p} \left|\widehat{1_A}(\eta)\right|^2} = \sqrt{\frac{\alpha}{|A|}} = p^{-1/2}$$

by the Cauchy–Schwarz inequality. Now, consider $f * f * f$, which has support $3A^2$. For all $x \in \mathbf{F}_p$, we have, by Fourier inversion, that

$$f(x) = \sum_{\xi \in \mathbf{F}_p} \widehat{f}(\xi)^3 e(\xi x) \geq \alpha^3 - \sum_{0 \neq \xi \in \mathbf{F}_p} \left|\widehat{f}(\xi)\right|^3 \geq \alpha^3 - p^{-1/2}\alpha$$

where we have used that $\widehat{f}(0) = \mathbf{E}_{x \in \mathbf{F}_p} f(x) = \alpha$ and also Parseval's identity again. So, $f(x) > 0$ if $\alpha^3 - p^{-1/2}\alpha$. The latter occurs whenever $\alpha > p^{-1/4}$. Thus, whenever $\alpha > p^{-1/4}$, $f(x) > 0$ for all $x \in \mathbf{F}_p$, i.e., $3A^2 = \operatorname{supp} f * f * f = \mathbf{F}_p$. □

The key consequence of the sum-product theorem that we will require says that if $A \subset \mathbf{F}_p$ is not too large and $B \subset \mathbf{F}_p^\times$ is not too small, then $|A + b \cdot A|$ must be significantly larger than $|A|$ for some $b \in B$.

Proposition 1. *Let $\delta, \delta' \in (0,1)$, $A \subset \mathbf{F}_p$ with $3 \leq |A| \leq p^{1-\delta}$ and $B \subset \mathbf{F}_p^\times$ with $|B| \geq p^{\delta'}$. Then, there exists $b \in B$ such that $|A + b \cdot A| \geq |A|^{1+\Omega_{\delta,\delta'}(1)}$.*

In order to prove this, we will need a simple lemma (originally due to Glibichuk and Konyagin) whose proof is similar to arguments we carried out during our proof of the sum-product theorem.

Lemma 24. *Let $A \subset \mathbf{F}_p$. There exists $x \in \mathbf{F}_p$ such that*

$$|A + x \cdot A| \geq \frac{1}{2} \min\left(|A|^2, p\right).$$

Proof. Observe that

$$\sum_{x\in\mathbf{F}_p} E(A, x\cdot A) = \#\left\{(a_1,a_2,a_3,a_4,x)\in A^4\times\mathbf{F}_p : a_1+xa_2 = a_3+xa_4\right\}$$
$$= |A|^2(|A|-1)^2 + |A|^2(p-1),$$

since any quadruple (a_1,a_2,a_3,a_4) with $a_2\neq a_4$ satisfying $a_1+xa_2 = a_3+xa_4$ uniquely determines x, and this x is nonzero if and only if $a_1\neq a_3$, and any quadruple (a_1,a_2,a_3,a_4) with $a_2 = a_4$ satisfying $a_1+xa_2 = a_3+xa_4$ also requires $a_1 = a_3$ (and there is no restriction on x). Thus, by the pigeonhole principle, there must exist $x\in\mathbf{F}_p$ such that

$$E(A, x\cdot A) \leq \frac{|A|^2(|A|-1)^2 + |A|^2(p-1)}{p-1} \leq \frac{|A|^4}{p} + |A|^2 \leq 2\max\left(\frac{|A|^4}{p}, |A|^2\right)$$

since $\frac{(|A|-1)^2}{p-1}\leq\frac{|A|^2}{p}$ by $|A|\leq p$. Recalling our lower bound

$$E(A, x\cdot A) \geq \frac{|A|^4}{|A+x\cdot A|}$$

and comining it with our upper bound for the additive energy yields

$$|A+x\cdot A| \geq \frac{|A|^4}{2\max\left(\frac{|A|^4}{p}, |A|^2\right)},$$

from which the lemma follows. □

Now we can prove Proposition 1.

Proof of Proposition 1. Analogous to the proof of the sum-product theorem, we have that if $|A+x\cdot A|, |A+y\cdot A|\leq K|A|$, then $|A+(x+y)\cdot A|, |A+(xy)\cdot A|\leq K^9|A|$. Indeed, by Ruzsa's triangle inequality,

$$|x\cdot A - y\cdot A| \leq \frac{|A-x\cdot A||A-y\cdot A|}{|A|}$$

which is at most $K^4|A|$ by the Plünnecke–Ruzsa inequality, and thus, by another application of the Ruzsa triangle inequality, we have

$$|A+(x+y)\cdot A| \leq |A+(x\cdot A+y\cdot A)| \leq \frac{|A+x\cdot A||-x\cdot A+(x\cdot A+y\cdot A)|}{|x\cdot A|} \leq K^9|A|$$

again by the Pünnecke–Ruzsa inequality. For the latter inequality, we simply have

$$|A+(xy)\cdot A| \leq \frac{|A+x\cdot A||-x\cdot A+(xy)\cdot A|}{|x\cdot A|} \leq K^3|A|$$

by the Ruzsa triangle inequality and Plünnecke's inequality again.

Let $m\ll_{\delta'} 1$ be such that $mB^m = \mathbf{F}_p$. If $K\geq 1$ were such that $|A+b\cdot A|\leq K|A|$ for all $b\in B$, then by the above we would have that $|A+x\cdot A|\leq K^{O_m(1)}|A|\leq K^{O_{\delta'}(1)}|A|$ for all $x\in\mathbf{F}_p$. But, by the previous lemma, there exists $x\in\mathbf{F}_p$ such that $|A+x\cdot A|\geq\frac{1}{2}\min(|A|^2,p)\gg|A|^{1+\Omega_{\delta'}(1)}$. Combining these upper and lower bounds leads to a contradiction when $K=|A|^c$ for $c\ll_{\delta,\delta'} 1$. □

Next, we will derive a useful consequence of the Balog–Szemerédi–Gowers theorem concerning sumsets of A with its dilates. During the proof of the Bourgain–Glibichuk–Konyagin bound, we will play this off of Proposition 1 to obtain a contradiction.

Corollary 4. *Let $A \subset \mathbf{F}_p$ and $B \subset \mathbf{F}_p^\times$ be such that*

$$E(A, b \cdot A) \geq K|A|^3$$

for all $b \in B$. Then, there exists $x \in B^{-1}$, $A' \subset A$, and $B' \subset x \cdot B$ with $|A'| \gg K^{-O(1)}|A|$, $|B'| \gg K^{-O(1)}|B|$, and

$$|A' + b' \cdot A'| \ll K^{O(1)}|A|$$

for all $b' \in B'$.

Proof. We begin by applying the Balog–Szemerédi–Gowers theorem to A and $b \cdot A$ for each $b \in B$. This produces $C_b, D_b \subset A$ with $|C_b|, |D_b| \gg K^{-O(1)}|A|$ such that $|C_b + b \cdot D_b| \ll K^{O(1)}|A|$. By the Cauchy–Schwarz inequality,

$$K^{-O(1)}|A|^2|B| \ll \sum_{x,y \in A} \sum_{b \in B} 1_{C_b \times D_b}(x,y) \leq |A| \left(\sum_{x,y \in A} \left| \sum_{b \in B} 1_{C_b \times D_b}(x,y) \right|^2 \right)^{1/2},$$

and so

$$K^{-O(1)}|A|^2|B|^2 \ll \sum_{x,y \in A} \sum_{b,b' \in B} 1_{C_b \times D_b}(x,y) 1_{C_{b'} \times D_{b'}}(x,y) = \sum_{b,b' \in B} |(C_b \times D_b) \cap (C_{b'} \times D_{b'})|.$$

By the pigeonhole principle, there exists $b \in B$ (which we will now fix) for which

$$K^{-O(1)}|A|^2|B| \ll \sum_{b' \in B} |(C_b \times D_b) \cap (C_{b'} \times D_{b'})|,$$

and, thus, also $B' \subset B$ with $|B'| \gg K^{-O(1)}|B|$ such that $|(C_b \times D_b) \cap (C_{b'} \cap D_{b'})| \gg K^{-O(1)}|A|^2$ for all $b' \in B'$. This additionally implies, since $C_{b'}, D_{b'} \subset A$, that, in fact, $|C_b \cap C_{b'}|, |D_b \cap D_{b'}| \gg K^{-O(1)}|A|$ for all $b' \in B'$.

By the Plünnecke–Ruzsa inequality, $|2C_b|, |2D_b| \ll K^{O(1)}|A|$ and also $|2C_{b'}|, |2D_{b'}| \ll K^{O(1)}|A|$ for all $b' \in B'$. By the Ruzsa triangle inequality,

$$|C_b - C_{b'}| \leq \frac{|C_b + (C_b \cap C_{b'})||C_{b'} + (C_b \cap C_{b'})|}{|C_b \cap C_{b'}|} \ll K^{O(1)}|A|$$

for all $b' \in B'$, and, similarly, $|D_b - D_{b'}| \ll K^{O(1)}|A|$ for all $b' \in B'$. Thus, letting d denote the Ruzsa distance, it follows by a couple more applications of the Ruzsa triangle inequality and Plünnecke's inequality that

$$\begin{aligned} d(b \cdot D_b, b' \cdot D_b) &\leq d(b \cdot D_b, -C_b) + d(-C_b, b' \cdot D_b) \\ &\leq d(b \cdot D_b, -C_b) + d(-C_b, -C_{b'}) + d(-C_{b'}, b' \cdot D_b) \\ &\leq d(b \cdot D_b, -C_b) + d(C_b, C_{b'}) + d(-C_{b'}, b' \cdot D_{b'}) + d(b' \cdot D_{b'}, b' \cdot D_b) \end{aligned}$$

for all $b' \in B'$, the upshot being that $|b \cdot D_b - b' \cdot D_b| \ll K^{O(1)}|A|$ for all $b' \in B'$. Setting $A' := D_b$, it thus follows from the Plünnecke–Ruzsa inequality that $|A' + b'' \cdot A'| \ll K^{O(1)}|A|$ for all $b'' \in b^{-1}B' \subset b^{-1}B$, as desired. □

Now we can prove the Bourgain–Glibichuk–Konyagin exponential sum bound.

Proof of Theorem 15. Let $\varepsilon > 0$ be a paramter to be chosen later (depending only on δ, and suppose by way of contradiction that $|\mathbf{E}_{h\in H} e_p(\xi h)| > p^{-\varepsilon}$ for some nonzero $\xi \in \mathbf{F}_p$. By our earlier observation on the H-invariance of large spectra of H, $\mathrm{Spec}_{p^{-\varepsilon}}(H)$ contains at least one coset of H in $\mathbf{F}_p^\times$. Since the $p^{-\varepsilon}$-large spectrum of H also certainly contains zero, we have $\left|\mathrm{Spec}_{p^{-\varepsilon}}(H)\right| \geq |H| + 1 \geq p^\delta + 1$.

We now run a dyadic pigeonholing argument to locate a scale α at which $\mathrm{Spec}_\alpha(H)$ and $\mathrm{Spec}_{\alpha^2/2}(H)$ do not differ too much in size. Define a sequence $\alpha_0 > \cdots > \alpha_M$ of reals in $(0,1)$ by setting $\alpha_0 := p^{-\varepsilon}$ and, for $m = 1, \ldots, M$, setting $\alpha_m := \frac{\alpha_{m-1}^2}{2}$, where $M \in \mathbf{N}$ will be chosen later (also depending only on δ). Note that, certainly, $\mathrm{Spec}_{\alpha_m}(H) \subset \mathrm{Spec}_{\alpha_{m-1}}(H)$ for all $m = 1, \ldots, M$, and that each α_m satisfies $\alpha_m \geq (p^{-\varepsilon}/2)^{2^m}$. By the pigeonhole principle, there must exist $m \in [M]$ for which

$$\left|\mathrm{Spec}_{\alpha_m}(H)\right| \leq p^{1/M} \left|\mathrm{Spec}_{\alpha_{m-1}}(H)\right|.$$

Indeed, if the above inequality failed to hold for all $m \in [M]$, then we would have

$$\left|\mathrm{Spec}_{\alpha_0}(H)\right| < p^{-1/M} \left|\mathrm{Spec}_{\alpha_1}(H)\right| < p^{-2/M} \left|\mathrm{Spec}_{\alpha_2}(H)\right| < \cdots < p^{-1} \left|\mathrm{Spec}_{\alpha_M}(H)\right|,$$

which, since the α_M-large spectrum of H can have at most p elements and the α_0-large spectrum of H is nonempty, would lead to a contradiction. Set $A := \mathrm{Spec}_{\alpha_{m-1}}(H)$ and $B := \mathrm{Spec}_{\alpha_m}(H)$.

Recall that we have

$$\sum_{b\in B} (1_A * 1_{-A})(b) = \left|\left\{(\xi_1, \xi_2) \in A^2 : \xi_1 - \xi_2 \in B\right\}\right| \geq \frac{\alpha_{m-1}^2}{2}|A|^2.$$

Thus, by the Cauchy–Schwarz inequality,

$$\frac{\alpha_{m-1}^2|A|^2}{2} \leq \sqrt{|B|}\#\left\{(\xi_1, \xi_2, \xi_3, \xi_4) \in A^4 : \xi_1 - \xi_2 = \xi_3 - \xi_4\right\}^{1/2},$$

i.e.,

$$E(A,A) \geq \frac{\alpha_{m-1}^4}{4}|A|^3\frac{|A|}{|B|} \geq \frac{\alpha_{m-1}^4}{4}p^{-1/M}|A|^3 \gg_M p^{-2^M\varepsilon - 1/M}|A|^3.$$

We have $|A| \geq p^{-\delta}$ and, by the trivial upper bound for the size of large spectra,

$$|A| \leq \frac{1}{p^{\delta-1}\alpha_{m-1}^2} \ll_M p^{1-\delta+2^M\varepsilon}.$$

Thus, since $E(A,A) = E(A, h\cdot A)$ for all $h \in H$, we can apply our corollary to the Balog–Szemerédi–Gowers theorem to get that there exist $A' \subset A$ and $H' \subset H$ with $|A'| \gg_M p^{-O(2^M\varepsilon + 1/M)}|A|$ and $|H'| \gg_M p^{-O(2^M\varepsilon+1/M)}|H|$ such that

$$|A' + h'\cdot A'| \ll_M p^{-O(2^M\varepsilon+1/M)}|A'|$$

for all $h' \in H'$. Taking M sufficiently large in terms of δ and then ε sufficiently small in terms of M and δ, this contradicts Proposition 1. □

5 The Freiman–Ruzsa theorem

In this section, we will prove a classification of subsets of the integers with small doubling. Let $A \subset \mathbf{Z}$ be finite, and enumerate the elements of A: $a_1 < a_2 < \cdots < a_N$. Then, observe that

$$a_1 + a_1 < a_1 + a_2 < \cdots < a_1 + a_N < a_2 + a_N < \cdots < a_N + a_N,$$

and so $A + A$ contains at least $2|A| - 1$ distinct elements, showing that $|A + A| \geq 2|A| - 1$. We already saw that this lower bound is attained by $[N]$ or any other arithmetic progression in $\mathbf{Z}$ of length N, and it is not hard to show that if $|A + A| = 2|A| - 1$ then A must be an arithmetic progression.

More examples of sets with small doubling are given by *generalized arithmetic progressions*

$$\{a + b_1k_1 + \cdots + b_nk_n : k_i = 0, \ldots, M_i - 1 \text{ for all } i \in [n]\} \tag{10}$$

of *dimension* n and *volume* $M_1 \cdots M_n$. For example, consider

$$A = \{k_1 + 3Mk_2 : k_1, k_2 = 0, \ldots, M - 1\}.$$

Then, $|A| = M^2$ (since if $k_1 + 3Mk_2 = \ell_1 + 3M\ell_2$, then $k_1 - \ell_1 = 3M(\ell_2 - k_2)$, which by the size bounds on k_1 and ℓ_1 forces $\ell_2 = k_2$, and thus $\ell_1 = k_1$ as well), so that the cardinality of A is the same as its volume, and

$$A + A = \{k_1 + 3Mk_2 : k_1, k_2 = 0, \ldots, 2M - 2\}.$$

Thus, $|A + A| \leq (2M - 1)^2 \leq 4M^2 = 4|A|$.

We say that a generalized arithmetic progression (10) is *proper* if all of the elements $a + b_1k_1 + \cdots + b_nk_n$ are distinct, i.e., if the volume of the GAP equals its cardinality. Note that a proper n-dimensional GAP has doubling at most 2^n. If A is a proper n-dimensional GAP and $B \subset A$ has density β, then

$$|B + B| \leq |A + B| \leq 2^n|A| \leq= \frac{2^n}{\beta}|B|,$$

and so dense subsets of generalized arithmetic progressions also have small doubling. The Freiman–Ruzsa theorem says that these are all possible examples of sets with small doubling.

Theorem 16. *Let $A \subset \mathbf{Z}$ be finite. If $|A + A| \leq K|A|$, then A is contained in a generalized arithmetic progression of dimension at most $d(K)$ and volume at most $v(K)|A|$.*

Using her eponymous lemma, Chang was able to obtain the explicit bounds $d(K) \leq K^{2+o(1)}$ and $v(K) \leq e^{K^{2+o(1)}}$, and Sanders was able to improve the exponent of K in Chang's bounds to $7/4$. Schoen then proved that $d(K) \leq K^{1+o(1)}$ and $v(K) \leq e^{K^{1+o(1)}}$ suffices. The proof we will present obtains bounds that are single exponential and double exponential in $K^{O(1)}$ for $d(K)$ and $v(K)$, respectively.

One can also formulate a version of the Freiman–Ruzsa theorem in high dimensional vector spaces over finite fields, where generalized arithmetic progressions are replaced by unions of cosets of a subspace. In a recent spectacular breakthrough, Gowers, Green, Manners, and Tao proved polynomial bounds

Theorem 17 (Gowers–Green–Manners–Tao, 2024)**.** *For each prime p, there exists a constant $c = c_p > 0$ such that the following holds. If $A \subset \mathbf{F}_p^n$ with $|A + A| \leq K|A|$, then there exists a subspace $V \leq \mathbf{F}_p^n$ of size $|V| \leq |A|$ such that A is contained in the union of at most K^c cosets of V.*

It is possible to formulate a version of the Freiman–Ruzsa theorem in $\mathbf{Z}$ for which one can obtain improved bounds by replacing generalized arithmetic progressions by a bounded number of translates of the linear image of an intersections of $\mathbf{Z}^n$ with a centrally symmetric convex body in $\mathbf{R}^n$ (i.e., a *convex progression*) having volume at most $|A|$. In this formulation, Sanders has obtained bounds on the dimension of $(\log K)^{O(1)}$ and number of translates of $e^{(\log K)^{O(1)}}$.

5.1 Background from the geometry of numbers

One of the key ingredients in the proof of the Freiman–Ruzsa theorem is the ability to locate large generalized arithmetic progressions in Bohr sets. In order to do this, we will need some tools from the geometry of numbers.

Recall that a *lattice* in $\mathbf{R}^n$ is a discrete cocompact subgroup. Thus, any lattice $\Lambda \subset \mathbf{R}^n$ takes the form

$$\Lambda = \mathbf{Z}v_1 \oplus \cdots \oplus \mathbf{Z}v_n$$

for some linearly independent vectors $v_1, \ldots, v_n \in \mathbf{R}^n$. Note that a fundamental domain of $\mathbf{R}^n/\Lambda$ is given by

$$\mathcal{F}(\Lambda) := \{x_1v_1 + \cdots + x_nv_n : x_1, \ldots, x_n \in [0, 1)\}.$$

We will denote the volume of the fundamental domain by $\mathrm{Vol}(\Lambda)$. Note that

$$\mathrm{Vol}(\Lambda) = \left|\det \begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix}\right|.$$

The following is a simple consequence of the pigeonhole principle.

Lemma 25 (Blichfeldt's Lemma)**.** *Let $K \subset \mathbf{R}^n$ be measurable and $\Lambda \subset \mathbf{R}^n$ be a lattice. If $\mathrm{Vol}(K) > \mathrm{Vol}(\Lambda)$, then there exist distinct $x, y \in K$ such that $x - y \in \Lambda$ (i.e., $K - K$ intersects Λ in more than just the zero vector).*

Proof. Note that we may, without loss of generality, assume that K is bounded. Consider the tiling of $\mathbf{R}^n$ by translates of $\mathcal{F}(\Lambda)$: $\Lambda + \mathcal{F}(\Lambda) = \mathbf{R}^n$. Set, for each $v \in \Lambda$,

$$K_v := [K \cap (v + \mathcal{F}(\Lambda))] - v \subset \mathcal{F}(\Lambda).$$

Observe that, since K is bounded, there are only finitely many $v \in \Lambda$ for which $K_v \neq \emptyset$, and that

$$\sum_{v \in \Lambda} \mathrm{Vol}(K_v) = \mathrm{Vol}(K).$$

Thus, by the pigeonhole principle, there exist distinct $v, w \in \Lambda$ for which $\mathrm{Vol}(K_v \cap K_w) > 0$. Let $a \in K_v \cap K_w$. Then, there exists $a \in K \cap (v + \mathcal{F}(\Lambda))$ and $b \in K \cap (w + \mathcal{F}(\Lambda))$ for which $a - v = b - w$, i.e., $a - b = v - w \in \Lambda$. Since $v + \mathcal{F}(\Lambda)$ and $w + \mathcal{F}(\Lambda)$ are disjoint, a and b must be distinct. □

From this lemma, we can easily deduce what's known as Minkowski's first theorem.

Theorem 18. *Let $K \subset \mathbf{R}^n$ be a centrally symmetric convex body and $\Lambda \subset \mathbf{R}^n$ be a lattice. If*

$$\mathrm{Vol}(K) > 2^n \,\mathrm{Vol}(\Lambda),$$

then K contains a nonzero vector in Λ.

Proof. Apply Blichfeldt's lemma to the dilation $\frac{1}{2}K$ to get that $\frac{1}{2}K$ contains two distinct points $\frac{1}{2}x$ and $\frac{1}{2}y$, with $x, y \in K$, such that $\frac{x-y}{2} \in \Lambda$. Since K is centrally symmetric, $-y \in K$ as well, and since K is convex, so is $\frac{x-y}{2}$. □

Let $K \subset \mathbf{R}^n$ be a centrally symmetric convex body. For each $i \in [n]$, we define the *i^{th} successive minimum* λ_i of K with respect to Λ by taking λ_i to be the infimum of $\lambda \geq 0$ for which the dilation $\lambda K := \{\lambda x : x \in K\}$ contains i linearly independent vectors in Λ. That is,

$$\lambda_i := \inf \{\lambda \in [0, \infty) : \dim \mathrm{span}(\lambda K \cap \Lambda) \geq i\}.$$

The key result from the geometry of numbers that we will need is Minkowski's second theorem, which concerns successive minima.

Theorem 19. *Let $K \subset \mathbf{R}^n$ be a centrally symmetric convex body and $\Lambda \subset \mathbf{R}^n$ be a lattice. If $\lambda_1 \leq \cdots \leq \lambda_n$ are the successive minima of K with respect to Λ, then we have*

$$\mathrm{Vol}(K) \leq \frac{2^n \,\mathrm{Vol}(\Lambda)}{\lambda_1 \cdots \lambda_n}.$$

This theorem is tight, as can be seen by considering $K = \prod_{i=1}^n [-\lambda_i^{-1}, \lambda_i^{-1}]$ and $\Lambda = \mathbf{Z}^n$. Note that Minkowski's first theorem is an immediate consequence of Minkowski's second theorem, since the hypothesis of the first theorem implies that $\lambda_1 \cdots \lambda_n < 1$, which, since the λ_i are increasing, implies that $\lambda_1 < 1$, i.e., that K contains some nonzero vector in Λ. We will not prove Minkowski's second theorem in this class. A proof can be found in any book on the geometry of numbers.

5.2 Bohr sets and generalized arithmetic progressions

Recall that, for any $t \in \mathbf{R}$, we use $\|t\|$ to denote the distance from t to the nearest integer. We will now define the notion of a Bohr set in a cyclic group.

Definition 5. *Let $\Gamma \subset \mathbf{Z}/N\mathbf{Z}$ and $\rho > 0$. We define the (homogeneous) Bohr set* $\mathrm{Bohr}(\Gamma, \rho)$ *by*

$$\mathrm{Bohr}(\Gamma, \rho) := \left\{x \in \mathbf{Z}/N\mathbf{Z} : \left\|\frac{\gamma x}{N}\right\| \leq \rho \text{ for all } \gamma \in \Gamma\right\}.$$

We call Γ the frequency set, $|\Gamma|$ the rank, and ρ the radius of the Bohr set.

A definition of Bohr set can be given in any abelian group, though we will not do so here. In vector spaces over finite fields, Bohr sets of sufficiently small radius are simply subspaces. One should think of Bohr sets in cyclic groups as stand-ins for subspaces, especially in the context of proving bounds in Szemerédi's theorem. Note that any interval of the form

$\{-M, \dots, M\}$ with $M \leq \frac{N-1}{2}$ in can be viewed as a Bohr set of rank 1 in $\mathbf{Z}/N\mathbf{Z}$, but the class of Bohr sets is much larger than just intervals, or even modular arithmetic progressions.

Note that the we have the trivial inclusions

$$\mathrm{Bohr}(\Gamma, \rho) \cap \mathrm{Bohr}(\Gamma', \rho) = \mathrm{Bohr}(\Gamma \cup \Gamma', \rho)$$

and

$$\mathrm{Bohr}(\Gamma, \rho) + \mathrm{Bohr}(\Gamma, \rho') \subset \mathrm{Bohr}(\Gamma, \rho + \rho')$$

the latter of which is a consequence of the triangle inequality. Thus, in particular, $k\,\mathrm{Bohr}(\Gamma, \rho) \subset \mathrm{Bohr}(\Gamma, k\rho)$ for all $k \in \mathbf{N}$.

We begin by proving a basic lower bound on the size of a Bohr set.

Lemma 26. *Let $\Gamma \subset \mathbf{Z}/N\mathbf{Z}$ and $\rho > 0$. We have*

$$|\mathrm{Bohr}(\Gamma, \rho)| \geq \rho^{|\Gamma|} N.$$

Proof. For each $\gamma \in \Gamma$, choose $x_\gamma \in \mathbf{R}/\mathbf{Z}$ uniformly at random independently. Then,

$$\mathbf{E}_{\mathbf{x}} \# \left\{ n \in \mathbf{Z}/N\mathbf{Z} : \left\| \frac{\gamma n}{N} - x_\gamma \right\| \leq \frac{\rho}{2} \text{ for all } \gamma \in \Gamma \right\} = \rho^{|\Gamma|} N.$$

Thus, there exists $\mathbf{x} = (x_\gamma)_{\gamma \in \Gamma}$ for which

$$\# \left\{ n \in \mathbf{Z}/N\mathbf{Z} : \left\| \frac{\gamma n}{N} - x_\gamma \right\| \leq \frac{\rho}{2} \text{ for all } \gamma \in \Gamma \right\} \geq \rho^{|\Gamma|} N.$$

Call the set on the left-hand side above B. Note that if $n, m \in B$, then

$$\left\| \frac{\gamma(n-m)}{N} \right\| = \left\| \frac{\gamma n}{N} - x_\gamma + x_\gamma - \frac{\gamma m}{N} \right\| \leq \rho$$

for all $\gamma \in \Gamma$ by the triangle inequality. Picking $m \in B$ arbitrarily, it follows that $m + B \subset \mathrm{Bohr}(\Gamma, \rho)$, and the first inequality follows. □

Next, we will show that Bohr sets contain large generalized arithmetic progressions.

Theorem 20. *Let $\Gamma \subset \mathbf{Z}/N\mathbf{Z}$ with N prime, $|\Gamma| = d$, and $\rho \in (0, 1/2)$. Then* $\mathrm{Bohr}(\Gamma, \rho)$ *contains a symmetric proper generalized arithmetic progression of dimension d and size at least $(\rho/d)^d N$.*

Proof. Write $\Gamma = \{\gamma_1, \dots, \gamma_d\}$ with each γ_i a representative in $\{0, \dots, N-1\}$, set $\gamma := (\gamma_1/N, \dots, \gamma_d/N)$, and consider the lattice

$$\Lambda := \mathbf{Z}\gamma + \mathbf{Z}^d \subset \mathbf{R}^d$$

(note that this certainly is a lattice, since $\frac{1}{N}\mathbf{Z}^d \subset \Lambda \subset \mathbf{Z}^d$) and the box

$$K := \left\{ (x_1, \dots, x_d) \in \mathbf{R}^d : |x_i| \leq \rho \text{ for all } i \in [d] \right\},$$

which is a centrally symmetric convex body. Let $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$ be the successive minima of K with respect to Λ. There also exists an associated *directional basis* $v_1, \dots, v_d \in$

Λ of K with respect to Λ, i.e., $v_1, \ldots, v_d$ is a basis for $\mathbf{R}^d$ with $v_i \in \partial\lambda_i K \cap \Lambda$ for each $i \in [d]$ and such that $(\lambda_i K \cap \Lambda)^\circ$ does not contain any elements of Λ outside of the span of $v_1, \ldots, v_{i-1}$.

For each $i \in [d]$, there exists $a_i \in \{0, \ldots, N-1\}$ such that $v_i = a_i\gamma + \mathbf{Z}^d$. Since $v_i \in \overline{\lambda_i K}$, each of the components of v_i has size at most $\lambda_i\rho$. Thus,

$$\left\| \frac{a_i\gamma_j}{N} \right\| \leq \lambda_i\rho \tag{11}$$

for all $i, j \in [d]$. Set $M_i := \lceil \frac{1}{d\lambda_i} \rceil$ for each $i \in [d]$, and define the generalized arithmetic progression

$$P := \{a_1k_1 + \cdots + a_dk_d : k_i \in (-M_i, M_i) \cap \mathbf{Z} \text{ for all } i \in [d]\}.$$

P is clearly symmetric and has dimension d. We will show that $P \subset \mathrm{Bohr}(\Gamma, \rho)$ and that P is proper. The size lower bound on P will then follow from Minkowski's second theorem.

First, note that, for any $a_1k_1 + \cdots + a_dk_d \in P$ and any $j \in [d]$, we have

$$\left\| \frac{(a_1k_1 + \cdots + a_dk_d)\gamma_j}{N} \right\| \leq \sum_{i=1}^{d} |k_i| \left\| \frac{a_i\gamma_j}{N} \right\| \leq \rho \sum_{i=1}^{d} \lambda_i|k_i| \leq \rho$$

by the size upper bound on each M_i. Thus, $P \subset \mathrm{Bohr}(\Gamma, \rho)$. That P is proper follows from the fact that the directional basis is, in fact, a basis. Indeed, suppose that

$$a_1k_1 + \cdots + a_dk_d = a_1k_1' + \cdots + a_dk_d'$$

for two distinct vectors $(k_1, \ldots, k_d), (k_1', \ldots, k_d') \in \prod_{i=1}^{d} [(-M_i, M_i) \cap \mathbf{Z}]$. This certainly implies that

$$(k_1 - k_1', \ldots, k_d - k_d') \cdot \left(\frac{a_1}{N}, \ldots, \frac{a_d}{N}\right) = 0,$$

and thus that

$$(k_1 - k_1')v_1 + \cdots + (k_d - k_d')v_d \in \mathbf{Z}^d.$$

But, for each $j \in [d]$, the j^{th} component of $(k_1 - k_1')v_1 + \cdots + (k_d - k_d')v_d$ has size bounded by

$$\sum_{i=1}^{d} |k_i - k_i'|\lambda_i\rho \leq \rho \sum_{i=1}^{d} (2M_i - 1)\lambda_i \leq 2\rho,$$

again by the size upper bound on M_i, since all components of v_i have size at most $\lambda_i\rho$. This forces $(k_1 - k_1')v_1 + \cdots + (k_d - k_d')v_d = 0$ by the assumption that $\rho < 1/2$, contradicting that $(k_1, \ldots, k_d)$ and $(k_1', \ldots, k_d')$ are distinct.

Finally, we will obtain a size lower bound on P using Minkowski's second theorem, which tells us that

$$\mathrm{Vol}(K) \leq \frac{2^d \,\mathrm{Vol}(\Lambda)}{\lambda_1 \cdots \lambda_d}.$$

We have $\mathrm{Vol}(K) = (2\rho)^d$ and

$$|P| = \prod_{i=1}^{d} (2M_i - 1) \geq \prod_{i=1}^{d} \frac{1}{d\lambda_i} = \frac{(1/d)^d}{\lambda_1 \cdots \lambda_d}$$

(as $2\lceil t\rceil - 1 \geq t$ whenever $t > 0$). So, it just remains to compute $\mathrm{Vol}(\Lambda)$. Recall that $\mathrm{Vol}(\Lambda)$ is the volume of the fundamental domain $\mathcal{F}(\Lambda)$, and, if $\Lambda_1 \leq \Lambda_2 \subset \mathbf{R}^d$ are any two lattices, then

$$\mathrm{Vol}(\Lambda_1) = \mathrm{Vol}(\Lambda_2)[\Lambda_1 : \Lambda_2].$$

Applying this with $\Lambda_1 = \mathbf{Z}^d$ and $\Lambda_2 = \Lambda$, we obtain

$$\mathrm{Vol}(\Lambda) = \frac{1}{[\mathbf{Z}^d : \Lambda]}.$$

Since we chose N to be prime, $|\Lambda/\mathbf{Z}^d| = N$, and so we get that $\mathrm{Vol}(\Lambda) = N^{-1}$. Thus,

$$|P| \geq \frac{(1/d)^d}{\lambda_1 \cdots \lambda_d} \geq \frac{(1/d)^d \,\mathrm{Vol}(K)}{2^d \,\mathrm{Vol}(\Lambda)} = \left(\frac{\rho}{d}\right)^d N,$$

as desired. □

5.3 Bogolyubov's lemma

By the Plünnecke–Ruzsa inequality, if A has small doubling, then $nA - mA$ isn't too much larger than A itself. It turns out that iterated sum and difference sets are very structured objects: we will show that $2A - 2A$ contains a large generalized arithmetic progression whenever $A \subset \mathbf{Z}/N\mathbf{Z}$ is dense, and thus, by Ruzsa's covering lemma, A can be covered by a small number of translates of a not too much larger GAP, which can then trivially be seen to be a GAP. It is then a small step to go from a set with small doubling in A to one in $\mathbf{Z}/N\mathbf{Z}$ whose structure can be translated back to A.

Lemma 27 (Bogolyubov's lemma)**.** *Let $A \subset \mathbf{Z}/N\mathbf{Z}$ have density α. Then, $2A-2A$ contains a Bohr set of rank less than $\frac{1}{\alpha^2}$ and radius $\frac{1}{4}$.*

Proof. Observe that the support of $1_A * 1_A * 1_{-A} * 1_{-A}$ is exactly $2A-2A$, and thus $x \in 2A-2A$ if and only if

$$(1_A * 1_A * 1_{-A} * 1_{-A})(x) = \sum_{\xi \in \mathbf{Z}/N\mathbf{Z}} \left|\widehat{1_A}(\xi)\right|^4 e_N(-\xi x)$$

is positive. We split up the right-hand side as

$$\alpha^4 + \sum_{\substack{0 \neq \xi \in \mathbf{Z}/N\mathbf{Z} \\ \left|\widehat{1_A}(\xi)\right| \geq \beta}} \left|\widehat{1_A}(\xi)\right|^4 e_N(-\xi x) + \sum_{\substack{\xi \in \mathbf{Z}/N\mathbf{Z} \\ \left|\widehat{1_A}(\xi)\right| < \beta}} \left|\widehat{1_A}(\xi)\right|^4 e_N(-\xi x) +$$

for some parameter $\alpha > \beta > 0$ to be chosen shortly. By Parseval's identity,

$$\left| \sum_{\substack{\xi \in \mathbf{Z}/N\mathbf{Z} \\ \left|\widehat{1_A}(\xi)\right| < \beta}} \left|\widehat{1_A}(\xi)\right|^4 e_N(-\xi x) \right| < \beta^2 \alpha.$$

We also have (by another application of Parseval's identity a couple of sections ago)

$$\left|\mathrm{Spec}_{\beta/\alpha}(A)\right| \leq \frac{\alpha}{\beta^2}.$$

Set $\Gamma := \mathrm{Spec}_{\beta/\alpha}(A) \setminus \{0\}$, so that $|\Gamma| < \frac{\alpha}{\beta^2}$. Note that if $x \in \mathrm{Bohr}(\Gamma, 1/4)$, then, for all $\gamma \in \Gamma$, since $\|\frac{\gamma x}{N}\| \leq \frac{1}{4}$, $\cos(2\pi\gamma x/N) \geq 0$. It follows that

$$(1_A * 1_A * 1_{-A} * 1_{-A})(x) > \alpha^4 + \sum_{\substack{0 \neq \xi \in \mathbf{Z}/N\mathbf{Z} \\ \left|\widehat{1_A}(\xi)\right| \geq \beta}} \left|\widehat{1_A}(\xi)\right|^4 e_N(-\xi x) - \beta^2\alpha \geq \alpha^4 - \beta^2\alpha$$

whenever $x \in \mathrm{Bohr}(\Gamma, 1/4)$. Since we want $\alpha^4 - \beta^2\alpha \geq 0$, the optimal choice of β is $\beta = \alpha^{3/2}$. Then $|\Gamma| < \frac{\alpha}{\beta^2} = \frac{1}{\alpha^2}$, as desired. □

It follows that $2A - 2A$ also contains a large generalized arithmetic progression of small dimension.

Corollary 5. *Assume that N is prime, and let $A \subset \mathbf{Z}/N\mathbf{Z}$ have density α. Then, $2A - 2A$ contains a generalized arithmetic progression of dimension $d < \frac{1}{\alpha^2}$ and size at least $\left(\frac{1}{4d}\right)^d N$.*

The current best bounds in Bogoylubov's lemma are due to Sanders in 2012, who proved that in $2A - 2A$ one can always find a d dimensional GAP of size at least $\exp(-O(d))N$ with $d \ll \log^{O(1)}(2/\alpha)$.

5.4 Freiman homomorphisms and Ruzsa's modeling lemma

Now that we have proved some useful results about subsets of $\mathbf{Z}/N\mathbf{Z}$, we want to show that we can use them to obtain information about subsets A of $\mathbf{Z}$ with small doubling. One may want to apply Bogolyubov's lemma to A modulo some appropriate integer N. But, if N is too small, then A and its image in $\mathbf{Z}/N\mathbf{Z}$ may not have the same "additive behavior", and if N is too large, then the GAP found using Bogoylubov's lemma will be too small relative to $|A|$. Recall that we do not know anything about A except for its size and that it has small doubling; in particular, we definitely cannot assume that it is contained in some interval not too much larger than $|A|$.

In order to find a useful embedding of A into a cyclic group of prime order, we will need the notions of Freiman homomorphisms and isomorphisms.

Definition 6. *Let $k \geq 2$ be an integer, $(G, +)$ and $(H, +)$ be abelian groups, $A \subset G$, $B \subset H$, and $\phi : A \to B$. We say that ϕ is a Freiman k-homomorphism if*

$$\phi(a_1) + \cdots + \phi(a_k) = \phi(a_1') + \cdots + \phi(a_k')$$

whenever

$$a_1 + \cdots + a_k = a_1' + \cdots + a_k'$$

for $a_1, a_1', \ldots, a_k, a_k' \in A$, i.e., ϕ respects all additive $2k$-tuples in A. We say that ϕ is a Freiman k-isomorphism if it is a bijection and its inverse is also a Freiman k-homomorphism.

The most basic examples of Freiman homomorphisms (of all orders) are restrictions of homomorphisms from G to H. For example, the canonical map $[N] \to \mathbf{Z}/2N\mathbf{Z}$ is a Freiman 2-isomorphism, and, more generally, the reduction modulo N map is a Freiman k-isomorphism when restricted to any interval of integers of the form $(x, x+\frac{N}{k}] \cap \mathbf{Z}$. Freiman homomorphisms form a much wider class of functions, however. For example, any map from $A = \{1, 4, 16\}$ is trivially a Freiman 2-homomorphism because A contains no nontrivial additive quadruples. Another example of a Freiman homomorphism of all orders is the canonical map $\phi : \{0,1\}^n \to (\mathbf{Z}/2\mathbf{Z})^n$, since it is just the homomorphism from $\mathbf{Z}^n$ to $(\mathbf{Z}/2\mathbf{Z})^n$ that takes all components modulo 2. However, its inverse function is not even a Freiman 2-homomorphism, since it does not respect the additive relation $(1,0,\ldots,0)+(1,0,\ldots,0) = (0,\ldots,0)+(0,\ldots,0) \pmod 2$. Thus, ϕ is a Freiman homomorphism of all orders but not a Freiman k-isomorphism for any $k \geq 2$.

We say that two subsets A and B of abelian groups are *Freiman k-isomorphic* if there exists a Freiman k-isomorphism between A and B. It is easy to check that he composition of Freiman k-homomorphisms is again a Freiman k-homomorphism, and that Freiman k-homomorphisms are also Freiman k'-homomorphisms for any $2 \leq k' \leq k$. Freiman k-homomorphisms $A \to B$ also induce Freiman homomorphisms of smaller order from $nA - nA \to nB - nB$.

Lemma 28. *Let $\phi : A \to B$ be a Freiman k-homomorphism and $n \leq \frac{k}{2}$ be an integer. Then ϕ induces a Freiman k'-homomorphism $\phi' : nA - nA \to nB - nB$ for any positive integer $k' \leq \frac{k}{2n}$.*

Proof. For any $a_1 + \cdots + a_n - a'_1 - \cdots - a'_n \in nA - nA$, we define $\phi' : nA - nA \to nB - nB$ by

$$\phi'(a_1 + \cdots + a_n - a'_1 - \cdots - a'_n) := \phi(a_1) + \cdots + \phi(a_n) - \phi(a'_1) - \cdots - \phi(a'_n).$$

That ϕ' is actually well defined follows immediately from the assumption that ϕ is a Freiman k-homomorphism and $n \leq \frac{k}{2}$, as does ϕ' being a Freiman k'-homomorphism as long as $k' \leq \frac{k}{2n}$. □

The particular consequence we will use later is that if A is Freiman 8-isomorphic to B, then $2A - 2A$ is Freiman 2-isomorphic to $2B - 2B$.

Freiman homomorphisms and isomorphisms are useful because they preserve important additive structure of sets.

Lemma 29. *Let $\phi : A \to B$ be a Freiman 2-homomorphism. Then, if $A_1, A_2 \subset A$ are finite, we have*

$$|\phi(A_1) + \phi(A_2)| \leq |A_1 + A_2|,$$

with equality if ϕ is a Freiman 2-isomorphism.

Proof. We define a map $f : \phi(A_1) + \phi(A_2) \to A_1 + A_2$ as follows. For any $x \in \phi(A_1) + \phi(A_2)$, pick $a_1 \in A_1$ and $a_2 \in A_2$ such that $\phi(a_1) + \phi(a_2) = x$, and then set $f(x) := a_1 + a_2$. Then, that ϕ is a Freiman 2-homomorphism implies that f is injective: if $f(x) = a_1 + a_2 = a'_1 + a'_2 = f(y)$ for $a_1, a'_1 \in A_1 \subset A$ and $a_2, a'_2 \in A_2 \subset A$, then $\phi(a_1) + \phi(a_2) = \phi(a'_1) + \phi(a'_2)$, so that $x = y$. Note that, when ϕ is a Freiman 2-isomorphism, we have both $|\phi(A_1) + \phi(A_2)| \leq |A_1 + A_2|$ and $|A_1 + A_2| = |\phi^{-1}(\phi(A_1)) + \phi^{-1}(\phi(A_2))| \leq |\phi(A_1) + \phi(A_2)|$. □

Thus, Freiman homomorphisms with sufficiently large image preseve the property of having small doubling. Freiman homomorphisms also preserve generalized arithmetic progressions.

Lemma 30. *Let $P \subset G$ be a d-dimensional generalized arithmetic progression and $\phi : P \to B$ be a Freiman 2-homomorphism. Then, $\phi(P)$ is also a d-dimensional generalized arithmetic progression. If, additionally, ϕ is a Freiman 2-isomorphism, $\phi(P)$ is proper whenever P is.*

Proof. Observe first that ϕ preserves arithmetic progressions. Indeed, for any $a, b \in G$, one can show that $\phi(a + bM) = \phi(a) + (\phi(a + b) - \phi(a))M$ by induction on integers $M \geq 0$: the base cases $M = 0$ and $M = 1$ are trivial, and if we have that $\phi(a + b(M - 1)) = \phi(a) + (\phi(a + b) - \phi(a))(M - 1)$ for a general $M - 1 \geq 1$, then since

$$(a + bM) + 0 = (a + b(M - 1)) + b$$

we have that

$$\phi(a + bM) + \phi(0) = \phi(a + b(M - 1)) + \phi(b) = \phi(a) + (\phi(a + b) - \phi(a))(M - 1) + \phi(b)$$

and, since,

$$(a + b) + 0 = a + b,$$

we have that

$$\phi(b) - \phi(0) = \phi(a + b) - \phi(a)$$

by the assumption that ϕ is a Freiman 2-homomorphism, from which it follows that

$$\phi(a + bM) = \phi(a) + (\phi(a + b) - \phi(a))\, M,$$

as desired.

Now, write

$$P = \{a + b_1k_1 + \cdots + b_dk_d : k_i = 0, \ldots, M_i - 1 \text{ for all } i \in [d]\}$$

for some $a, b_1, \ldots, b_d \in G$. Iteratively applying the above to induct on d, we can deduce that $\phi(P)$ is also a d-dimensional generalized arithmetic progression. Indeed, suppose that we have

$$\phi\left(a' + \sum_{i=1}^{d-1} b_i'k_i'\right) = \phi(a') + \sum_{i=1}^{d-1} (\phi(a' + b_i') - \phi(a'))\, k_i'$$

for all $a', b_1', \ldots, b_{d-1}' \in G$ and nonnegative integers $k_1', \ldots, k_{d-1}'$. Then, we have that $\phi\left(a + \sum_{i=1}^{d} b_ik_i\right)$ equals

$$\phi\big(a + \sum_{i=1}^{d-1} b_ik_i\big) + \big(\phi\big(a + \sum_{i=1}^{d-1} b_ik_i + b_d\big) - \phi\big(a + \sum_{i=1}^{d-1} b_ik_i\big)\big)k_d,$$

, which equals

$$\begin{aligned}\phi(a)+\sum_{i=1}^{d-1}(\phi(a+b_i)-\phi(b_i))k_i+\big(\phi(a+b_d)-\phi(a)\\ +\sum_{i=1}^{d-1}[\phi(a+b_i+b_d)-\phi(a+b_d)-\phi(a+b_i)+\phi(a)]k_i\big)k_d\\ =\phi(a)+\sum_{i=1}^{d-1}\left(\phi(a+b_i)-\phi(b_i)\right)k_i+\left(\phi(a+b_d)-\phi(a)\right)k_d\end{aligned}$$

for all nonnegative integers $k_1,\dots,k_d$, where we have used that

$$(a+b_i+b_d)+a=(a+b_i)+(a+b_d)$$

and so, since ϕ is a Freiman 2-homomorphism,

$$\phi(a+b_i+b_d)-\phi(a+b_d)-\phi(a+b_i)+\phi(a)=0$$

for all $i\in[d-1]$.

Thus, it follows that $\phi(P)$ is a d-dimensional generalized arithmetic progression whenever P is a d-dimensional generalized arithmetic progression. It immediately follows that when ϕ is a Freiman 2-isomorphism (so that it is, necessarily, a bijection), that $\phi(P)$ is proper whenever P is proper. □

Observe that $A=\{1,2,4,\dots,2^{n-1}\}$ cannot possibly be Freiman isomorphic to any subset of a cyclic group $\mathbf{Z}/N\mathbf{Z}$ with N of comparable size to n. Indeed, if A were Freiman 2-isomorphic to $B\subset\mathbf{Z}/N\mathbf{Z}$, so that $|B|=n$, then, by our lemma on the size of sumsets, we would be forced to have $|B+B|=\frac{n(n-1)}{2}+n=\frac{n(n+1)}{2}$, and thus $N\geq\frac{n(n+1)}{2}=\frac{|A|(|A|+1)}{2}$. So, certainly not every subset of $\mathbf{Z}$ is Freiman isomorphic to a dense subset of a cyclic group. On the other hand, Ruzsa showed that all subsets with small doubling have a large subset that is Freiman isomorphic to a dense subset of a cyclic group.

Lemma 31 (Ruzsa's modeling lemma)**.** *Let $k\geq 2$ be an integer and $A\subset\mathbf{Z}$ be finite. If $|kA-kA|\leq K|A|$, then, for every prime $p>2K|A|$, there exists $A'\subset A$ with $|A'|\geq|A|/k$ such that A' is Freiman k-isomorphic to a subset of $\mathbf{Z}/p\mathbf{Z}$.*

Proof. Let $p>2K|A|$ be prime. We may assume, without loss of generality, that $A\subset\mathbf{N}$ with $1\in A$ (as shifts, which are affine homomorphisms, are Freiman homomorphisms). Pick any prime q greater than the largest element of $kA-kA$, and, for every nonzero $\lambda\in\mathbf{Z}/q\mathbf{Z}$, define a map $\phi_\lambda:\mathbf{Z}\to\{0,1,\dots,q-1\}$ as the composition

$$\mathbf{Z}\xrightarrow{(\mathrm{mod}\ q)}\mathbf{Z}/q\mathbf{Z}\xrightarrow{\cdot\lambda}\mathbf{Z}/q\mathbf{Z}\to\{0,\dots,q-1\},\tag{12}$$

where the last map takes an element of $\mathbf{Z}/q\mathbf{Z}$ to its unique representative in the interval $\{0,\dots,q-1\}$. Observe that if $\lambda\in[q-1]$ is chosen uniformly at random, then each $n\in\mathbf{Z}$ that is not divisible by q is mapped uniformly at random to an element of $[q-1]$. Thus,

each $n \in \mathbf{Z}$ that is not divisible by q has probability at most $\frac{1}{p}$ of being divisible by p, since there are $\lfloor \frac{q-1}{p} \rfloor$ multiples of p in $[q-1]$. Thus, the expected number of nonzero elements of $kA - kA$ whose image is divisible by p under the map ϕ_λ is at most

$$\frac{|kA - kA|}{p} < \frac{K|A|}{2K|A|} = \frac{1}{2}.$$

It follows that there exists $\lambda \in (\mathbf{Z}/q\mathbf{Z})^\times$ such that ϕ_λ maps each nonzero element of $kA - kA$ to an element of $[q-1]$ that is not divisible by p.

Recall that the last map in the composition (12) is a Freiman k-isomorphism when restricted to an interval of length at most $\frac{q}{k}$ in $\{0, \dots, q-1\}$. By the pigeonhole principle, there exists some interval I of length $< \frac{q}{k}$ (since q is prime and $q > k$) in $\{0, \dots, q-1\}$ such that $\#\{a \in A : \phi_\lambda(a) \in I\} \geq |A|/k$. Setting $A' := \{a \in A : \phi_\lambda(a) \in I\}$, we thus have that ϕ_λ is a Freiman k-homomorphism from A' onto its image. Letting ϕ be the composition of ϕ_λ with the reduction modulo p map, we then have that ϕ is a Freiman k-homomorphism of A' onto its image in $\mathbf{Z}/p\mathbf{Z}$. Thus, to show that ϕ is, in fact, injective and a Freiman k-isomorphism, it suffices to show that if

$$\phi(a_1) + \cdots + \phi(a_k) = \phi(a'_1) + \cdots + \phi(a'_k)$$

for some $a_1, a'_1, \dots, a_k, a'_k \in A'$, then we must have

$$a_1 + \cdots + a_k = a'_1 + \cdots + a'_k.$$

The former displayed equation implies that

$$b := \phi_\lambda(a_1) + \cdots + \phi_\lambda(a_k) - \phi_\lambda(a'_1) - \cdots - \phi_\lambda(a'_k)$$

is a multiple of p. Without loss of generality (by relabeling), we may assume that the integer b is nonnegative. By our choice of A', each $\phi_\lambda(a_i)$ and $\phi_\lambda(a'_i)$ lies in the interval I, and thus $b \leq q - 1 < q$. We have

$$b \equiv \phi_\lambda(a_1 + \cdots + a_k - a'_1 - \cdots - a'_k) \pmod{q},$$

and thus, since $\phi_\lambda(a_1 + \cdots + a_k - a'_1 - \cdots - a'_k) \in [q-1]$ by the choice of q and λ, we must have

$$b = \phi_\lambda(a_1 + \cdots + a_k - a'_1 - \cdots - a'_k).$$

But, by our choice of λ, p cannot divide $\phi_\lambda(a_1 + \cdots + a_k - a'_1 - \cdots - a'_k)$ when $a_1 + \cdots + a_k - a'_1 - \cdots - a'_k \neq 0$, since $a_1 + \cdots + a_k - a'_1 - \cdots - a'_k \in kA' - kA' \subset kA - kA$. Thus, since $p \mid b$, we must have $a_1 + \cdots + a_k = a'_1 + \cdots + a'_k$, as desired. □

5.5 The proof of the Freiman–Ruzsa theorem

Now, we can finally prove the Freiman–Ruzsa theorem, which we now recall.

Theorem 21. *Let $A \subset \mathbf{Z}$ be finite. If $|A + A| \leq K|A|$, then A is contained in a generalized arithmetic progression of dimension at most $d(K)$ and volume at most $v(K)|A|$.*

Proof. By the Plünnecke–Ruzsa inequality, we have $|8A - 8A| \ll_K |A|$. Thus, by Ruzsa's modeling lemma, whenever $N \gg_K |A|$ is prime, there exists an $A' \subset A$ with $|A'| \gg_K |A|$ that is Freiman 8-isomorphic to a subset $\phi(A')$ of $\mathbf{Z}/N\mathbf{Z}$. We can use Bertrand's postulate to fix such a prime $N \asymp_K |A|$, so that the $\phi(A')$ obtained has density $\gg_K 1$ in $\mathbf{Z}/N\mathbf{Z}$. This Freiman 8-isomorphism induces a Freiman 2-isomorphism of $2A' - 2A'$ with $2\phi(A') - 2\phi(A')$.

We now apply Bogolyubov's lemma, which tells us that $2\phi(A') - 2\phi(A')$ contains a proper generalized arithmetic progression of dimension $\ll_K 1$ and size $\gg_K |A|$. Since Freiman 2-isomorphisms preserve generalized arithmetic progressions, this means that $2A' - 2A'$ likewise contains a generalized arithmetic progression P of dimension $\ll_K 1$ and size $\gg_K |A|$.

Now, since $P \subset 2A - 2A$, we have that $A + P \subset 3A - 2A$. Thus, by the Plünnecke–Ruzsa inequality, we have $|A + P| \ll_K |A|$. Since $|P| \gg_K |A|$, Ruzsa's covering lemma then tells us that A can be covered by $\ll_K 1$ translates of $P - P$, which is still a generalized arithmetic progression of dimension $\ll_K 1$ and size $\ll_K |A|$. That is,

$$A \subset X + P - P$$

for some $X \subset \mathbf{Z}$ with $|X| \ll_K 1$. But, X is trivially contained in a generalized arithmetic progression of dimension $\ll_K 1$, so that $X + P - P$ is contained in a generalized arithmetic progression of dimension $\ll_K 1$ and size $\ll_K |A|$. □

One can check that this argument does, indeed, give bounds that are single exponential and double exponential in $K^{O(1)}$ for $d(K)$ and $v(K)$, respectively

6 Sets lacking four-term arithmetic progressions

Our next goal in this course is to prove the following quantitative version of Szemerédi's theorem for four-term arithmetic progressions.

Theorem 22 (Gowers, 1998)**.** *If $A \subset [N]$ contains no nontrivial four-term arithmetic progressions, then*

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

for some absolute constant $c > 0$.

One of the key ingredients will be the Freiman–Ruzsa theorem. We will begin by covering some of the basic theory of the Gowers uniformity norms.

6.1 The Gowers U^s-norms

The Gowers uniformity norms can be defined for finitely supported functions on any abelian group, though we will only consider them for finite abelian groups (and then, later, specialize even further to cyclic groups).

Definition 7. *Let $(G, +)$ be a finite abelian group, $s \in \mathbf{N}$, and $f : G \to \mathbf{C}$. For any $h \in G$, we define the multiplicative discrete derivative operator Δ_h by*

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

for all $x \in G$. For any $h_1, \dots, h_s \in G$, we define the s-fold multiplicative discrete derivative operator $\Delta_{h_1,\dots,h_s}$ by

$$\Delta_{h_1,\dots,h_s} f = \Delta_{h_1} \Delta_{h_2} \cdots \Delta_{h_s} f.$$

Observe that, for all $h_1, \dots, h_s \in G$ and permutations τ of $[s]$, we have

$$\Delta_{h_1,\dots,h_s} f = \Delta_{h_{\tau(1)},\dots,h_{\tau(s)}} f$$

for all $f : G \to \mathbf{C}$.

Definition 8. *Let $(G, +)$ be an abelian group, $s \in \mathbf{N}$, and $f : G \to \mathbf{C}$. We define the Gowers U^s-norm of f by*

$$\|f\|_{U^s} = \left(\mathbf{E}_{x,h_1,\dots,h_s \in G} \Delta_{h_1,\dots,h_s} f(x)\right)^{1/2^s},$$

where, on the right-hand side, we are taking the unique nonnegative real 2^s-th root.

We will show shortly that $\mathbf{E}_{x,h_1,\dots,h_s \in G} \Delta_{h_1,\dots,h_s} f(x) \geq 0$, and so this definition makes sense. We will also show that these are, in fact, norms when $s \geq 2$, though the U^1-norm is only a seminorm. Recall from the beginning of the course that $\|f\|_{U^2} = \|\widehat{f}\|_{\ell^4}$, from which these properties of $\|\cdot\|_{U^2}$ immediately follow.

We will also require the notion of the Gowers inner product.

Definition 9. *Let $(G, +)$ be an abelian group, $s \in \mathbf{N}$, and, for each $\omega \in \{0,1\}^s$, $f_\omega : G \to \mathbf{C}$. Then the Gowers U^s-inner product of the f_ω's is*

$$\langle (f_\omega)_{\omega \in \{0,1\}^s} \rangle_{U^s} := \mathbf{E}_{x,h_1,\dots,h_s \in G} \prod_{\omega \in \{0,1\}^s} \mathcal{C}^{|\omega|} f_\omega(x + \omega \cdot (h_1, \dots, h_s)),$$

where $\mathcal{C}$ denotes complex conjugation and $|\omega|$ denotes the number of 1's in ω.

Note that $\langle f, \dots, f \rangle_{U^s} = \|f\|_{U^s}^{2^s}$ for all $f : G \to \mathbf{C}$ and $s \in \mathbf{N}$. Also note the recursive identity

$$\|f\|_{U^s}^{2^s} = \mathbf{E}_{h \in G} \|\Delta_h f\|_{U^{s-1}}^{2^{s-1}} \tag{13}$$

for all integers $s \geq 2$.

We will begin by proving the *Gowers–Cauchy–Schwarz inequality*

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

In fact, we will prove that

$$\left|\langle (f_\omega)_{\omega \in \{0,1\}^s} \rangle_{U^s}\right|^{2^s} \leq \prod_{\omega \in \{0,1\}^s} \|f_\omega\|_{U^s}^{2^s}$$

from which it will follow that $\|f\|_{U^s}^{2^s}$ is always real and nonnegative by taking $f_{\mathbf{0}} = f$ and $f_\omega = 1$ for all $\mathbf{0} \neq \omega \in \{0,1\}^s$ in the Gowers–Cauchy–Schwarz inequality.

Proof. We proceed by induction on s. For the case $s = 1$, we have that

$$|\langle f_0, f_1\rangle|^2 = |\mathbf{E}_{x,h\in G} f_0(x) f_1(x+h)|^2 = |\mathbf{E}_{x\in G} f_0(x)\mathbf{E}_{y\in G} f_1(y)|^2 = \|f_0\|_{U^1}^2 \|f_1\|_{U^1}^2,$$

on observing that $|\mathbf{E}_{z\in G} f_i(z)|^2 = \mathbf{E}_{z,z'\in G} f_i(z)\overline{f_i(z')} = \mathbf{E}_{z,h\in G} f_i(z)\overline{f_i(z+h)}$ by a change of variables. Now, before we do the inductive step, observe that

$$\langle (f_\omega)_{\omega\in\{0,1\}^s}\rangle_{U^s} = \mathbf{E}_{x,h_1,h_1',\dots,h_s,h_s'\in G} \prod_{\omega\in\{0,1\}^s} \mathcal{C}^{|\omega|} f_\omega(x + (\mathbf{1}-\omega)\cdot(h_1,\dots,h_s) + \omega\cdot(h_1',\dots,h_s'))$$

by inserting extra averaging and making the change of variables $h_i \mapsto h_i' - h_i$ for each $i \in [s]$ and $x \mapsto x + h_1 + \cdots + h_s$ and in the definition of the Gowers inner product. Suppose that we have proven the Gowers–Cauchy–Schwarz inequality for a general $s \in \mathbf{N}$. We can rearrange $\langle (f_\omega)_{\omega\in\{0,1\}^{s+1}}\rangle_{U^{s+1}}$ as

$$\mathbf{E}_{x,h_1,h_1',\dots,h_s,h_s'\in G} \left(\mathbf{E}_{h_{s+1}\in G} \prod_{\omega\in\{0,1\}^s} \mathcal{C}^{|\omega|} f_{\omega 0}(x + (\mathbf{1}-\omega)\cdot(h_1,\dots,h_s) + \omega\cdot(h_1',\dots,h_s') + h_{s+1})\right)$$
$$\cdot \left(\mathbf{E}_{h_{s+1}'\in G} \prod_{\omega\in\{0,1\}^s} \mathcal{C}^{|\omega|} f_{\omega 1}(x + (\mathbf{1}-\omega)\cdot(h_1,\dots,h_s) + \omega\cdot(h_1',\dots,h_s') + h_{s+1}')\right)$$

and then apply the Cauchy–Schwarz inequality to get that $\left|\langle (f_\omega)_{\omega\in\{0,1\}^{s+1}}\rangle_{U^{s+1}}\right|^2$ is bounded above by

$$\mathbf{E}_{x,h_1,h_1',\dots,h_s,h_s'\in G} \left(\mathbf{E}_{h_{s+1}\in G} \prod_{\omega\in\{0,1\}^s} \mathcal{C}^{|\omega|} f_{\omega 0}(x + (\mathbf{1}-\omega)\cdot(h_1,\dots,h_s) + \omega\cdot(h_1',\dots,h_s') + h_{s+1})\right)$$
$$\cdot \left(\mathbf{E}_{h_{s+1}'\in G} \prod_{\omega\in\{0,1\}^s} \mathcal{C}^{|\omega|} f_{\omega 0}(x + (\mathbf{1}-\omega)\cdot(h_1,\dots,h_s) + \omega\cdot(h_1',\dots,h_s') + h_{s+1}')\right)$$

times

$$\mathbf{E}_{x,h_1,h_1',\dots,h_s,h_s'\in G} \left(\mathbf{E}_{h_{s+1}\in G} \prod_{\omega\in\{0,1\}^s} \mathcal{C}^{|\omega|} f_{\omega 1}(x + (\mathbf{1}-\omega)\cdot(h_1,\dots,h_s) + \omega\cdot(h_1',\dots,h_s') + h_{s+1})\right)$$
$$\cdot \left(\mathbf{E}_{h_{s+1}'\in G} \prod_{\omega\in\{0,1\}^s} \mathcal{C}^{|\omega|} f_{\omega 1}(x + (\mathbf{1}-\omega)\cdot(h_1,\dots,h_s) + \omega\cdot(h_1',\dots,h_s') + h_{s+1}')\right).$$

By making the change of variables $x \mapsto x - h_{s+1}$ and $k_{s+1} = h_{s+1}' - h_{s+1}$, these equal

$$\mathbf{E}_{k_{s+1}\in G}\mathbf{E}_{x,h_1,h_1',\dots,h_s,h_s'\in G} \prod_{\omega\in\{0,1\}^s} \mathcal{C}^{|\omega|} \Delta_{k_{s+1}} f_{\omega i}(x + (\mathbf{1}-\omega)\cdot(h_1,\dots,h_s) + \omega\cdot(h_1',\dots,h_s'))$$

for $i = 0, 1$, so that $\left|\langle (f_\omega)_{\omega\in\{0,1\}^{s+1}}\rangle_{U^{s+1}}\right|^2$ is bounded above by

$$\left(\mathbf{E}_{k_{s+1}\in G}\langle(\Delta_{k_{s+1}}f_{\omega 0})_{\omega\in\{0,1\}^s}\rangle_{U^s}\right)\overline{\left(\mathbf{E}_{k_{s+1}\in G}\langle(\Delta_{k_{s+1}}f_{\omega 1})_{\omega\in\{0,1\}^s}\rangle_{U^s}\right)}.$$

Thus, by the triangle inequality and induction hypothesis,

$$\begin{aligned}\left|\langle (f_\omega)_{\omega\in\{0,1\}^{s+1}}\rangle_{U^{s+1}}\right|^2 &\le \mathbf{E}_{k_{s+1}\in G}\left|\langle(\Delta_{k_{s+1}}f_{\omega 0})_{\omega\in\{0,1\}^s}\rangle_{U^s}\right|\mathbf{E}_{k_{s+1}\in G}\left|\langle(\Delta_{k_{s+1}}f_{\omega 1})_{\omega\in\{0,1\}^s}\rangle_{U^s}\right| \\ &\le \left(\mathbf{E}_{k_{s+1}\in G}\prod_{\omega\in\{0,1\}^s}\|\Delta_{k_{s+1}}f_{\omega 0}\|_{U^s}\right)\left(\mathbf{E}_{k_{s+1}\in G}\prod_{\omega\in\{0,1\}^s}\|\Delta_{k_{s+1}}f_{\omega 1}\|_{U^s}\right)\end{aligned}$$

The desired bound now follows from Hölder's inequality and the identity (13). □

It follows that the Gowers norms are also monotone in s:

$$\|f\|_{U^1} \le \|f\|_{U^2} \le \|f\|_{U^3} \le \dots.$$

This can be seen by taking, for each $s \in \mathbf{N}$ and $\omega \in \{0,1\}^s$, $f_{\omega 0} = f$ and $f_{\omega 1} = 1$, so that

$$\|f\|_{U^s}^{2^s} = \langle (f_\omega)_{\omega\in\{0,1\}^{s+1}}\rangle \le \|f\|_{U^{s+1}}^{2^s},$$

and then taking 2^s-th roots yields $\|f\|_{U^s} \le \|f\|_{U^{s+1}}$.

Now, we can finally show that the Gowers uniformity norms are, in fact, norms when $s \ge 2$. We have already shown that they take values in $[0,\infty)$, and, clearly, $\|0\|_{U^s} = 0$ and $\|tf\|_{U^s} = |t|\|f\|_{U^s}$ for all $t \in \mathbf{R}$. To see that $\|f\|_{U^s} = 0$ only if $f = 0$ when $s \ge 2$ (note that $\|f\|_{U^1} = 0$ whenever f has mean zero), suppose that $f : G \to \mathbf{C}$ is not identically zero, so that $f(a) \ne 0$ for some $a \in G$. Then,

$$\langle f, 1_a, 1_a, 1_a\rangle \ne 0,$$

since if $x+h, x+k, x+h+k = a$, then $x = (x+h)+(x+k)-(x+h+k) = a$ as well. On the other hand,

$$0 < |\langle f, 1_a, 1_a, 1_a\rangle| \le \|f\|_{U^2}$$

by the Gowers–Cauchy–Schwarz inequality. Thus, $\|f\|_{U^2} > 0$, and so $\|f\|_{U^s} > 0$ whenever $s > 2$ as well. It just remains to check that the U^s-norms satisfy the triangle inequality for all $s \ge 1$. If $f_0, f_1 : G \to \mathbf{C}$, then, by the Gowers–Cauchy–Schwarz inequality, we have

$$\|f_0+f_1\|_{U^s}^{2^s} = \sum_{\omega\in\{0,1\}^{2^s}}\langle (f_{\omega(i)})_{i\in[2^s]}\rangle_{U^s} \le \sum_{\omega\in\{0,1\}^{2^s}}\|f_0\|_{U^s}^{2^s-|\omega|}\|f_1\|_{U^s}^{|\omega|} = \sum_{j=0}^{2^s}\binom{2^s}{j}\|f_0\|_{U^s}^{2^s-j}\|f_1\|_{U^s}^{j},$$

which equals $\left(\|f_0\|_{U^s}+\|f_1\|_{U^s}\right)^{2^s}$. Thus, $\|f_0+f_1\|_{U^s} \le \|f_0\|_{U^s}+\|f_1\|_{U^s}$.

6.2 The U^3-norm and four-term arithmetic progressions

Now, we specialize to the case of $G = \mathbf{Z}/p\mathbf{Z}$ for p a large prime. For all $f_0, f_1, f_2, f_3 : G \to \mathbf{C}$, define

$$\Lambda(f_0, f_1, f_2, f_3) := \mathbf{E}_{x,y \in G} f_0(x) f_1(x+y) f_2(x+2y) f_3(x+3y).$$

When $f_0 = f_1 = f_2 = f_3 = f$, we will simply write $\Lambda(f)$. In the beginning of the course, we showed (by iterated appliations of the Cauchy–Schwarz inequality) that

$$|\mathbf{E}_{x,y \in G} f_0(x) f_1(x+y) f_2(x+2y)| \leq \|f_2\|_{U^2} \tag{14}$$

whenever $f_0, f_1, f_2 : G \to \mathbf{C}$ are 1-bounded and $p > 2$. We will now deduce from this the analogous inequality for Λ.

Lemma 32. *Let $p > 3$ be prime and $f_0, f_1, f_2, f_3 : G \to \mathbf{C}$ be 1-bounded. Then,*

$$|\Lambda(f_0, f_1, f_2, f_3)| \leq \|f_3\|_{U^3}.$$

Proof. We apply the Cauchy–Schwarz inequality and a change of variables ($z = y + h$ and then $x \mapsto x - y$) to get that

$$\begin{aligned}
|\Lambda(f_0, f_1, f_2, f_3)|^2 &\leq \mathbf{E}_{x \in G} |\mathbf{E}_{y \in G} f_1(x+y) f_2(x+2y) f_3(x+3y)|^2 \\
&\leq \mathbf{E}_{x,y,z \in G} f_1(x+y) f_2(x+2y) f_3(x+3y) \overline{f_1(x+z) f_2(x+2z) f_3(x+3z)} \\
&= \mathbf{E}_{h \in G} \mathbf{E}_{x,y \in G} \Delta_h f_1(x+y) \Delta_{2h} f_2(x+2y) \Delta_{3h} f_3(x+3y) \\
&= \mathbf{E}_{h \in G} \mathbf{E}_{x,y \in G} \Delta_h f_1(x) \Delta_{2h} f_2(x+y) \Delta_{3h} f_3(x+2y) \\
&\leq \mathbf{E}_{h \in G} \|\Delta_{3h} f_3\|_{U^2}
\end{aligned}$$

by (14) (since $p > 2$). Since p is relatively prime to 3 by assumption, a change of variables yields $|\Lambda(f_0, f_1, f_2)|^2 \leq \mathbf{E}_{h \in G} \|\Delta_h f_3\|_{U^2}$. By Hölder's inequality and (13),

$$\mathbf{E}_{h \in G} \|\Delta_h f_3\|_{U^2} \leq \left(\mathbf{E}_{h \in G} \|\Delta_h f_3\|_{U^2}^4\right)^{1/4} = \|f_3\|_{U^3}^2.$$

We therefore conclude that $|\Lambda(f_0, f_1, f_2, f_3)| \leq \|f_3\|_{U^3}$. □

By making different changes of variables, one can, in fact, deduce that

$$|\Lambda(f_0, f_1, f_2, f_3)| \leq \min_{i=0,1,2,3} \|f_i\|_{U^3}$$

whenever $f_0, f_1, f_2, f_3 : G \to \mathbf{C}$ are 1-bounded.

We now proceed as in the proof of Roth's theorem. Suppose that $A \subset [N]$ has density α and contains no nontrivial four-term arithmetic progressions, let $p \in (6N, 12N)$ be a prime (which, again, must exist by Bertrand's postulate), and let $f_A := 1_A - \alpha 1_{[N]}$ denote the balanced function of A, which can be viewed naturally as a function on $\mathbf{Z}/p\mathbf{Z}$. By our choice of p, the four-term arithmetic progressions in $[N]$ (mod p) are in bijective correspondence with the four-term arithmetic progressions in $[N]$, and so the total number of four-term arithmetic progressions in A equals $p^2 \Lambda(1_A)$, which, since A contains only trivial four-term progressions, must equal αN. If $N \geq 100\alpha^{-3}$, say, then

$$\alpha N = \frac{\alpha^4 N^2}{\alpha^3 N} \leq \frac{\alpha^4 N^2}{100},$$

and so

$$\Lambda(1_A, 1_A, 1_A, 1_A) \leq \frac{\alpha^4 N^2}{100p^2} \leq \frac{\alpha^4}{3600}.$$

On the other hand, using that $1_A = f_A + \alpha 1_{[N]}$ and telescoping, we get that $\Lambda(1_A)$ equals

$$\Lambda(1_A, 1_A, 1_A, f_A) + \alpha\Lambda(1_A, 1_A, f_A, 1_{[N]}) + \alpha^2\Lambda(1_A, f_A, 1_{[N]}, 1_{[N]}) + \alpha^3\Lambda(f_A, 1_{[N]}, 1_{[N]}, 1_{[N]}) + \alpha^4\Lambda(1_{[N]}).$$

The first four quantities are bounded above by $\|f_A\|_{U^3}$, $\alpha\|f_A\|_{U^3}$, $\alpha^2\|f_A\|_{U^3}$, and $\alpha^3\|f_A\|_{U^3}$, respectively, and the last quantity is bounded below by $\frac{\alpha^4}{p^2} \cdot \frac{N}{2} \cdot \frac{N}{6} \geq \frac{\alpha^4}{12^3} = \frac{\alpha^4}{1728}$. It thus follows that

$$\|f_A\|_{U^3} \geq \frac{1}{4}\left(\frac{\alpha^4}{1728} - \frac{\alpha^4}{3600}\right) \geq \frac{\alpha^4}{100000}. \tag{15}$$

In order to deduce a density increment lemma from this, we will prove a local inverse theorem for the U^3-norm.

Theorem 23. *Let $f : \mathbf{Z}/p\mathbf{Z} \to \mathbf{C}$ be 1-bounded with $\|f\|_{U^3} \geq \delta$ for some $\delta \in (0, 1/2]$. Then, there exists an arithmetic progression P of length at least $e^{-\delta^{-O(1)}}p^{\delta^{O(1)}}$ and polynomials $Q_i \in \mathbf{Z}[x]$ of degree at most 2 such that*

$$\mathbf{E}_{i \in G}\,|\mathbf{E}_{x \in i+P} f(x) e_p(Q_i(x))| \geq \delta^{O(1)}.$$

As you showed in one of the bonus problems on the first homework, it is not possible to prove that any 1-bounded function f with large U^3-norm must have large correlation on all of $\mathbf{Z}/p\mathbf{Z}$ with a quadratic phase function $e_p(Q(x))$.

From this, one can deduce the following density-increment lemma, from which Gowers's bound for sets lacking four-term arithmetic progressions follows by the same sort of iteration as in the proof of Roth's theorem.

Lemma 33. *There exist absolute constants $c_1, c_2 > 0$ such that the following holds. Let $A \subset [N]$ have density $\alpha \leq \frac{7}{8}$ and assume that A contains no nontrivial four-term arithmetic progressions. Then either $N < \alpha^{-c_1}$, or there exists an arithmetic progression $P = a + q \cdot [N']$ with $N' \geq N^{\alpha^{c_1}}$ such that*

$$\frac{|A \cap P|}{|P|} \geq \alpha + \alpha^{c_2}.$$

Observe that if $A \subset [N]$ has density $> \frac{7}{8}$, then A contains a nontrivial four-term arithmetic progression (in fact, of step size 1), simply by the pigeonhole principle. Next, we will begin the proof of the local inverse theorem for the U^3-norm.

6.3 The structure of the large Fourier coefficients of $\Delta_h f$

Let $G = \mathbf{Z}/p\mathbf{Z}$ for a large prime p, and assume that $f : G \to \mathbf{C}$ is 1-bounded and satisfies $\|f\|_{U^3} \geq \delta$. Then,

$$\mathbf{E}_{h \in G}\|\Delta_h f\|_{U^2}^4 \geq \delta^8,$$

and thus $\|\Delta_h f\|_{U^2} \geq \frac{\delta^2}{2}$, say, for at least a $\frac{\delta^8}{2}$-proportion of $h \in G$, for otherwise we would have

$$\delta^8 \leq \frac{\delta^8}{2} + \frac{\delta^8}{16} < \delta^8.$$

By the inverse theorem for the U^2-norm (which, recall, says that if $g : G \to \mathbf{C}$ is 1-bounded and $\|g\|_{U^2} \geq \delta$, then $|\widehat{g}(\xi)| \geq \delta^2$ for some $\xi \in G$), it follows that there exists $\phi(h) \in G$ such that

$$\left|\widehat{\Delta_h f}(\phi(h))\right| \geq \frac{\delta^4}{4}$$

for such $h \in G$ (the set of which we will call H). We will show that this function ϕ respects many additive quadruples.

Lemma 34. *Let $f : G \to \mathbf{C}$ be a 1-bounded function, $H \subset G$, and $\phi : H \to G$. Assume that*

$$\sum_{h \in H} \left|\widehat{\Delta_h f}(\phi(h))\right|^2 \geq \delta p.$$

Then, there are at least $\delta^4 p^3$ quadruples $(a_1, a_2, a_3, a_4) \in H^4$ such that

$$a_1 + a_2 = a_3 + a_4$$

and

$$\phi(a_1) + \phi(a_2) = \phi(a_3) + \phi(a_4).$$

Proof. Expanding the definition of the Fourier transform, we have that

$$\sum_{h \in H} \mathbf{E}_{x,y \in G} f(x)\overline{f(x+h)f(y)}f(y+h)e_p(-\phi(h)(x-y)) \geq \delta p,$$

which, by making the change of variables $y = x + k$, becomes

$$\sum_{h \in H} \mathbf{E}_{x,k \in G} f(x)\overline{f(x+h)f(x+k)}f(x+h+k)e_p(\phi(h)k) \geq \delta p.$$

The idea is to now apply the Cauchy–Schwarz inequality a couple of times to obtain information about ϕ. By swapping the order of summation, we have that

$$\begin{aligned}
\delta p &\leq \mathbf{E}_{x,k \in G} f(x)\overline{f(x+k)} \sum_{h \in H} \overline{f(x+h)}f(x+h+k)e_p(\phi(h)k) \\
&\leq \left(\mathbf{E}_{x,k \in G}\left|\sum_{h \in H} \overline{f(x+h)}f(x+h+k)e_p(\phi(h)k)\right|^2\right)^{1/2} \\
&= \left(\mathbf{E}_{x,k \in G} \sum_{h_1,h_2 \in H} \overline{f(x+h_1)}f(x+h_1+k)f(x+h_2)\overline{f(x+h_2+k)}e_p([\phi(h_1)-\phi(h_2)]k)\right)^{1/2}
\end{aligned}$$

by the Cauchy–Schwarz inequality and the assumption that f is 1-bounded. Thus, we have

$$\begin{aligned}
\delta^2 p^2 &\leq \mathbf{E}_{x,k \in G} \sum_{h_1,h_2 \in G} 1_H(h_1)1_H(h_2)\overline{f(x+h_1)}f(x+h_1+k)f(x+h_2)\overline{f(x+h_2+k)}e_p([\phi(h_1)-\phi(h_2)]k) \\
&= \mathbf{E}_{x,k \in G} \sum_{h,\ell \in G} 1_H(h)1_H(h+\ell)\overline{f(x)}f(x+k)f(x+\ell)\overline{f(x+\ell+k)}e_p([\phi(h)-\phi(h+\ell)]k)
\end{aligned}$$

by another change of variables. A final application of the Cauchy–Schwarz inequality (in the variables x, k, and ℓ then yields

$$\delta^4 p^3 \leq \mathbf{E}_{k\in G} \sum_{h,h',\ell\in G} \Delta_\ell 1_H(h) \Delta_\ell 1_H(h') e_p([\phi(h) - \phi(h+\ell) - \phi(h') + \phi(h'+\ell)]k).$$

Making the further change of variables $h' = h + m$ makes the above read

$$\delta^4 p^3 \leq \mathbf{E}_{k\in G} \sum_{h,\ell,m\in G} \Delta_{\ell,m} 1_H(h) e_p([\phi(h) - \phi(h+\ell) - \phi(h+m) + \phi(h+\ell+m)]k).$$

Taking the average over $k \in G$ inside and using orthogonality of characters, we thus obtain that

$$\delta^4 p^3 \leq \sum_{h,\ell,m\in G} \Delta_{\ell,m} 1_H(h) \cdot \begin{cases} 1 & \phi(h) - \phi(h+\ell) - \phi(h+m) + \phi(h+\ell+m) = 0 \\ 0 & \text{otherwise} \end{cases}.$$

Since additive quadruples in H are in bijective correspondence with quadruples $(h, h+\ell, h+m, h+\ell+m) \in H^4$, the conclusion of the lemma follows. □

Note that the conclusion of the lemma says that the graph $\Gamma := (h, \phi(h))_{h\in H}$ contains many additive quadruples, i.e., has large additive energy. We will now use this to deduce, using the Balog–Szemerédi–Gowers theorem and the Freiman–Ruzsa theorem, that $\phi(h)$ must agree with a linear function on a large subset of G. Indeed, by the Balog–Szemerédi–Gowers theorem and the above lemma, there exists a subset $H' \subset H$ of size $\gg \delta^{O(1)} p$ such that the graph $\Gamma' := (h', \phi(h'))_{h'\in H'}$ has doubling $\frac{|\Gamma'+\Gamma'|}{|\Gamma'|} \leq \delta^{-O(1)}$. The following lemma says that functions whose graph has small doubling must agree on a long arithmetic progression with a linear function.

Lemma 35. *Let $H \subset G$ and $\phi : H \to G$, and set $\Gamma = ((h, \phi(h))_{h\in H} \subset G^2$ to be the graph of ϕ. If $|\Gamma + \Gamma| \leq K|\Gamma|$ for $K \geq 2$ and $|H| \geq e^{K^{O(1)}}$, then there exists an arithmetic progression $P = a + q[N] \subset G$ with $N' \geq p^{K^{-O(1)}}$ and a linear function $L(x) = bx + c$ such that $\phi(x) = L(x)$ for at least $K^{-O(1)}N$ elements $x \in P$.*

Before we prove this, we recall from the proof of the Freiman–Ruzsa theorem that if $A \subset \mathbf{Z}$ has small doubling, then $2A - 2A$ contains a large proper GAP of small dimension (we then deduced the conclusion of the Freiman–Ruzsa theorem from an application of Ruzsa's covering lemma, which possibly ruined the properness of the GAP). In particular, we proved the following.

Theorem 24. *Let $A \subset \mathbf{Z}$ be finite. If $|A + A| \leq K|A|$ for $K \geq 2$, then $2A - 2A$ contains a proper generalized arithmetic progression of dimension at most $K^{O(1)}$ and size at least $e^{-K^{O(1)}}|A|$.*

As a corollary, we can find a relatively long arithmetic progression on which a set $A \subset \mathbf{Z}^2$ with small doubling is dense.

Corollary 6. *Let $A \subset \mathbf{Z}^2$ be finite. If $|A+A| \leq K|A|$ for $K \geq 2$, then there exists an arithmetic progression $P = (a_1, a_2) + (q_1, q_2)[N]$ of size at least $e^{-K^{O(1)}}|A|^{K^{-O(1)}}$ such that $\frac{|A \cap P|}{N} \geq K^{-O(1)}$.*

Proof. Note first that any finite $A \subset \mathbf{Z}^2$ is Freiman 8-isomorphic to a subset B of $\mathbf{Z}$. Indeed, this is a special case of a problem on the next homework, but one can also see this by letting $M \in \mathbf{N}$ be such that $A \subset [-M, M]^2$ and considering the map $(x, y) \mapsto x + 8My$. Thus, $|B+B| \leq K|B|$ and $2A - 2A$ and $2B - 2B$ are Freiman 2-isomorphic.

We now apply the Freiman–Ruzsa theorem to locate a proper generalized arithmetic progression Q_0 of dimension at most $K^{O(1)}$ and size at least $e^{-K^{O(1)}}|B|$ in $2B - 2B$. Since Freiman 2-isomorphisms preserve proper generalized arithmetic progressions, it follows that there exists a proper generalized arithmetic progression Q of dimension at most $K^{O(1)}$ and size at least $e^{-K^{O(1)}}|A|$ in $2A - 2A$. Note that

$$|A||Q| = \sum_{x \in \mathbf{Z}^2} |A \cap (x + Q)|,$$

but the sum on the right-hand side is supported in $A - Q$, which satisfies $A - Q \subset A - 2A + 2A = 3A - 2A$, which has size at most $K^{O(1)}|A|$ by the Plünnecke–Ruzsa inequality. Thus, by the pigeonhole principle, there exists $x \in \mathbf{Z}^2$ for which $|A \cap Q'| \geq K^{-O(1)}|Q'|$, where $Q' := x + Q$ is also a proper generalized arithmetic progression of the same size and dimension as Q.

Write

$$Q' = \{\mathbf{a} + \mathbf{b}_1 k_1 + \cdots + \mathbf{b}_n k_n : k_i = 0, \ldots, M_i - 1 \text{ for all } i \in [n]\}$$

for $\mathbf{a}, \mathbf{b}_1, \ldots, \mathbf{b}_n \in \mathbf{Z}^2$, where $n \leq K^{O(1)}$ and $\prod_{i=1}^n M_i \geq e^{-K^{O(1)}}|A|$. Without loss of generality, we may as well assume that $M_n \geq M_i$ for all $i \in [n]$, so that $M_n \geq e^{-K^{O(1)}}|A|^{K^{-O(1)}}$. We can write Q' as the disjoint union of $\prod_{i=1}^{n-1} M_i$ shifts of a fixed arithmetic progression of length M_n:

$$\bigsqcup_{\substack{k_i = 0, \ldots, M_i - 1 \\ i = 1, \ldots, n-1}} \{(\mathbf{a} + \mathbf{b}_1 k_1 + \cdots + \mathbf{b}_{n-1} k_{n-1}) + \mathbf{b}_n k_n : k_n = 0, \ldots, M_n - 1\}.$$

The desired conclusion now follows from $|A \cap Q'| \geq K^{-O(1)}|Q'|$ and the pigeonhole principle again. □

Now we can prove Lemma 35.

Proof. Let $\Gamma' \subset \{0, 1, \ldots, p-1\}^2$ be such that $\Gamma \equiv \Gamma' \pmod{p}$. Then,

$$|\Gamma' + \Gamma'| \leq 4|\Gamma + \Gamma| \leq 4K|\Gamma'|,$$

and so, by the above corollary, there exists an arithmetic progression $Q' = (a_1, a_2) + (q_1, q_2)[N]$ of size $N \geq e^{-K^{O(1)}}|H|^{K^{-O(1)}}$ such that $\frac{|\Gamma' \cap Q'|}{N} \geq K^{-O(1)}$. By the assumption on $|H|$, this, in particular, forces $|\Gamma' \cap Q'| > 1$, so that at least one of q_1 or q_2 is not a multiple of p; since Γ' is a graph, we must, in fact, have q_1 not a multiple of p. Thus, setting $Q = Q' \pmod{p}$, we have $|Q| = |Q'|$ and $|\Gamma \cap Q| \geq K^{-O(1)}N$. This completes the proof with $P = a_1 + q_1[N]$ and $L(x) = a_2 + \frac{q_2}{q_1}(x - a_1)$. □

Summarizing the results of this section, we obtain the following.

Lemma 36. *There exists an absolute constant $c > 0$ such that the following holds. Let $f : \mathbf{Z}/p\mathbf{Z} \to \mathbf{C}$ be 1-bounded with $\|f\|_{U^3} \geq \delta$ for some $\delta \in (0, 1/2]$. Then, there exists an arithmetic progression $P = a + q[N]$ of length $N \geq e^{-\delta^{-O(1)}} p^{\delta^{O(1)}}$ and $b, c \in G$ such that*

$$\mathbf{E}_{h \in P} \left| \widehat{\Delta_h f}(bh + c) \right| \geq \delta^{O(1)}.$$

6.4 Proof of the local inverse theorem for the U^3-norm

We will now use Lemma 36 to prove the local inverse theorem for the U^3-norm. First, by way of illustration, suppose we knew that $\widehat{\Delta_h f}(bh + c)$ were large for all shifts $h \in G$. Then, by the Cauchy–Schwarz inequality,

$$\begin{aligned}\delta^{O(1)} &\leq \mathbf{E}_{h \in G} \left| \widehat{\Delta_h f}(bh + c) \right|^2 = \mathbf{E}_{x,y,h \in G} f(x) \overline{f(x + h) f(y)} f(y + h) e_p(-(bh + c)(x - y)) \\ &= \mathbf{E}_{x,h,k \in G} f(x) \overline{f(x + h) f(x + k)} f(x + h + k) e_p((bh + c)k)\end{aligned}$$

as in the proof of the first lemma in the previous section. Setting

$$f_1(x) = f_3(x) = f(x) e_p\left(\frac{b}{2} x^2\right) \qquad \text{and} \qquad f_2(x) = f_4(x) = f(x) e_p\left(\frac{b}{2} x^2 + cx\right),$$

observe that, since

$$\frac{b}{2} x^2 - \frac{b}{2}(x + h)^2 - \frac{b}{2}(x + k)^2 + \frac{b}{2}(x + h + k)^2 - c(x + h) + c(x + h + k) = bhk + ck,$$

we have

$$\delta^{O(1)} \leq \mathbf{E}_{x,h,k \in G} f_1(x) \overline{f_2(x + h) f_3(x + k)} f_4(x + h + k).$$

Thus, by the Gowers–Cauchy–Schwarz inequality, $\|f_1\|_{U^2} \geq \delta^{O(1)}$. It follows by the inverse theorem for the U^2-norm that there exists $c' \in G$ such that

$$\delta^{O(1)} \leq \left| \mathbf{E}_{x \in G} f(x) e_p\left(\frac{b}{2} x^2 - c' x\right) \right|,$$

as desired. Alternatively, we could have applied Fourier inversion and orthogonality of characters to obtain

$$\sum_{\xi \in G} f_1(\xi) f_2(-\xi) f_3(-\xi) f_4(\xi) \geq \delta^{O(1)},$$

and from this and Parseval's identity deduced that f_1 has a large Fourier coefficient.

In our situation, we merely have

$$\mathbf{E}_{h \in P} \left| \widehat{\Delta_h f}(bh + c) \right|^2 \geq \delta^{O(1)},$$

the left-hand side of which expands to

$$\mathbf{E}_{x,k \in G} \mathbf{E}_{h \in P} f(x) \overline{f(x + h) f(x + k)} f(x + h + k) e_p((bh + c)k).$$

By partitioning G into shifts of P, we obtain

$$\delta^{O(1)} \leq \mathbf{E}_{i,j\in G}\left|\mathbf{E}_{x,h,k\in P} f_1(i+x)\overline{f_2(i+x+h)f_3(i+x+j+k)}f_4(i+x+h+j+k)\right|.$$

with f_1, f_2, f_3, and f_4 as above. For each $i, j \in G$, set

$$f_1^{(i,j)}(x) := 1_P f_1(i+x),\ f_2^{(i,j)}(x) := f_2(i+x),\ f_3^{(i,j)}(x) := f_3(i+j+x),\ \text{and}\ f_4^{(i,j)}(x) := f_4(i+j+x).$$

Thus, the above can be rewritten as

$$\mathbf{E}_{i,j\in G}\left|\mathbf{E}_{x,h,k\in P} f_1^{(i,j)}(x)\overline{f_2^{(i,j)}(x+h)f_3^{(i,j)}(x+k)}f_4^{(i,j)}(x+h+k)\right| \geq \delta^{O(1)}.$$

Our final reduction is to note that, by replacing each $f_1^{(i,j)}(x), \ldots, f_4^{(i,j)}(x)$ with $f_1^{(i,j)}(bx+c)$, $f_2^{(i,j)}(bx+2c)$, $f_3^{(i,j)}(bx+2c)$, or $f_4^{(i,j)}(bx+3c)$, respectively, it suffices to prove the desired local correlation with quadratic phases when $P = [N]$.

We can now, almost, proceed as before, except that our variables x, h, and k range over a (possibly very short) interval in G, instead of the whole group. This issue can be dealt with by some standard Fourier-analytic maneuvering. Recall that the N we obtain is far smaller than p; in particular, it is smaller than $\sqrt{N}$. Thus, we can view the $f_k^{(i,j)}$ above (when restricted to $[3N]$, say) as functions on $\mathbf{Z}$. In particular, it suffices to prove the following lemma.

Lemma 37. *Let $f_1, f_2, f_3, f_4 : \mathbf{Z} \to \mathbf{C}$ be 1-bounded functions with f_1 supported on $[N]$ and f_2, f_3, and f_4 supported on $[3N]$. If*

$$\left|\frac{1}{N}\sum_{x\in\mathbf{Z}} \mathbf{E}_{h,k\in[N]} f_1(x)f_2(x+h)f_3(x+k)f_4(x+h+k)\right| \geq \delta,$$

then there exists $\alpha \in \mathbf{R}/\mathbf{Z}$ such that

$$\left|\mathbf{E}_{x\in[N]} f_1(x)e(\alpha x)\right| \gg \delta^{O(1)}.$$

Indeed, such an $\alpha \in \mathbf{R}/\mathbf{Z}$ can be corrected to one in $\frac{1}{p}\mathbf{Z}$ at the cost of an error that's $\ll \frac{N}{p} \ll \frac{1}{\sqrt{p}}$, and our inverse theorem is trivial when p is small in terms of δ.

Proof of Lemma 37. Set

$$\mu(x) := \frac{1_{[N]}(x)}{N},$$

and observe that $\|\mu\|_{\ell^2}^2 = N^{-1}$, $\|\mu\|_{\ell^1} = 1$, and $\|\mu \star \mu\|_{\ell^1} = 1$, where we temporarily define

$$(a \star b)(x) = \sum_{y\in\mathbf{Z}} a(y)b(x+y).$$

The inequality assumed in the lemma therefore reads

$$\left|\frac{1}{N}\sum_{x,h,k\in\mathbf{Z}} f_1(x)f_2(x+h)f_3(x+k)f_4(x+h+k)\mu(h)\mu(k)\right| \geq \delta.$$

We first make the change of variables $x \mapsto x-h-k$ and apply the Cauchy–Schwarz inequality to double the k variable to obtain

$$\frac{1}{N} \sum_{x,h,k,k' \in \mathbf{Z}} f_2(x-k)\overline{f_2(x-k')f_1(x-k-h)}f_1(x-k'-h)\mu(h)\mu(k)\mu(k') \gg \delta^2.$$

Making the change of variables $x \mapsto x+k$ and $\ell = k-k'$, the above reads

$$\frac{1}{N} \sum_{x,h,\ell \in \mathbf{Z}} f_2(x)\overline{f_2(x+\ell)f_1(x-h)}f_1(x-h+\ell)\mu(h)(\mu \star \mu)(\ell) \gg \delta^2.$$

Using a similar application of the Cauchy–Schwarz inequality to double the h variable, we obtain

$$\frac{1}{N} \sum_{x,\ell,m \in \mathbf{Z}} \Delta_{\ell,m} f_1(x)(\mu \star \mu)(\ell)(\mu \star \mu)(m) \gg \delta^4$$

after making a change of variables.

We now apply the Fourier inversion formula to $(\mu \star \mu)(n)$ to get that

$$\int_0^1 \int_0^1 |\widehat{\mu}(\xi)|^2 |\widehat{\mu}(\eta)|^2 \frac{1}{N} \sum_{x,\ell,m \in \mathbf{Z}} \Delta_{\ell,m} f_1(x) e(\xi\ell + \eta m) d\eta d\xi \gg \delta^{O(1)}.$$

Taking the maximum absolute value of the inside sum thus yields

$$\left\|\widehat{\mu}^2\right\|_{L^1}^2 \max_{\xi,\eta \in \mathbf{R}/\mathbf{Z}} \left| \frac{1}{N} \sum_{x,\ell,m \in \mathbf{Z}} \Delta_{\ell,m} f_1(x) e(\xi\ell + \eta m) \right| \gg \delta^{O(1)}.$$

Since

$$\left\|\widehat{\mu}^2\right\|_{L^1}^2 = \|\widehat{\mu}\|_{L^2}^4 = \|\mu\|_{\ell^2}^4 = N^{-2},$$

it follows that there exist $\xi, \eta \in \mathbf{R}/\mathbf{Z}$ for which

$$\left| \frac{1}{N^3} \sum_{x,\ell,m \in \mathbf{Z}} \Delta_{\ell,m} f_1(x) e(\xi\ell + \eta m) \right| \gg \delta^{O(1)}.$$

To finish, we set $g_1(x) := f_1(x)e(-(\xi+\eta)x)$, $g_2(x) := \overline{f_1(x)e(-\xi\ell)}$ and $g_3(x) := \overline{f_1(x)e(-\eta m)}$, and then use that the left-hand side above equals

$$\frac{1}{N^3} \sum_{x,h,k \in \mathbf{Z}} g_1(x)g_2(x+h)g_3(x+k)f_1(x+h+k).$$

The conclusion of the lemma now follows by viewing g_1, g_2, g_3, and f_1 as functions on $\mathbf{Z}/3N\mathbf{Z}$, say, and then applying the Gowers–Cauchy–Schwarz inequality and the U^2-inverse theorem. □

This completes the proof of the local inverse theorem for the U^3-norm. Next, we will see how to deduce the density increment lemma from it, thus completing the proof of Gowers's bound for sets lacking four-term arithmetic progressions.

6.5 Deducing the density increment

Recall that we used Dirichlet's Diophantine approximation lemma to deduce a density increment in our proof of Roth's theorem.

Theorem 25. *Let $\gamma \in \mathbf{R}$ and $Q \in \mathbf{N}$. There exists $q \in [Q]$ such that*

$$\|q\gamma\| < \frac{1}{Q}.$$

In order to prove a density increment lemma for sets lacking four-term arithmetic progressions from our local inverse theorem for the U^3-norm, we will require a quadratic version of Dirichlet's diophantine approximation lemma.

Lemma 38. *The exists an absolute constant $c > 0$ such that the following holds. Let $\gamma \in \mathbf{R}$ and $N \in \mathbf{N}$. There exists $n \in [N]$ such that*

$$\|n^2\gamma\| \leq \frac{1}{N^c}$$

.

We will require Weyl's inequality for quadratic polynomials, which you proved on the first problem set.

Lemma 39. *Let $\alpha, \beta \in \mathbf{R}$, and assume that there exist $a, q \in \mathbf{Z}$ with q positive and $(a, q) = 1$ such that $|\alpha - a/q| \leq q^{-2}$. Then,*

$$\left|\mathbf{E}_{n\in[N]} e(\alpha n^2 + \beta n)\right| \ll \log N \left(\frac{1}{q} + \frac{1}{N} + \frac{q}{N^2}\right)^{1/2}.$$

We will also require a bound for exponential sums with linear phases you proved along the way.

Lemma 40. *We have*

$$\left|\sum_{N\leq n\leq N+M} e(\gamma n)\right| \ll \min\left(M, \frac{1}{\|\gamma\|}\right)$$

for all $\gamma, N, M \in \mathbf{R}$ with $M \geq 1$.

We will first prove Lemma 38 in the case that γ is rational, and then use Dirichlet's lemma to reduce to this case.

Lemma 41. *Let $\frac{a}{q} \in \mathbf{Q}$ with $(a, q) = 1$ and $N \in \mathbf{N}$. There exists $n \in [N]$ such that*

$$\left\|n^2\frac{a}{q}\right\| \ll \frac{\sqrt{q}\log^2 q}{N}$$

.

Proof. Observe first that it suffices to prove the result when $N \leq q$, since the case $N > q$ holds trivially by taking $n = q$. Let $M \in \mathbf{N}$ be a parameter to be chosen shortly. By orthogonality of characters, if

$$\sum_{|b| \leq M} \mathbf{E}_{n \in [N]} \mathbf{E}_{\xi \in \mathbf{Z}/q\mathbf{Z}} e_q(\xi(an^2 - b)) > 0, \tag{16}$$

then there exists $n \in [N]$ for which $\|n^2 \frac{a}{q}\| \leq M/q$. The contribution to the above coming from $\xi = 0$ is $\frac{2M+1}{q}$. When $\frac{\xi a}{q}$ is far from a rational with small denominator, Weyl's inequality will give us a good bound for the average $\mathbf{E}_{n \in [N]} e_q(\xi(an^2 - b))$. Otherwise, when $\frac{\xi}{q}$ is far from an integer, we can sum over $|b| \leq M$ to get a savings. At least one of these will hold most of the time, allowing us to bound the contribution of $\xi \neq 0$. Indeed, we have, for any $\xi \neq 0$,

$$\begin{aligned}
\sum_{|b| \leq M} \mathbf{E}_{n \in [N]} e_q(\xi(an^2 - b)) &= \left(\sum_{|b| \leq M} e\left(\frac{-\xi b}{q}\right) \right) \left(\mathbf{E}_{n \in [N]} e\left(\frac{\xi a n^2}{q}\right) \right) \\
&\ll \min\left(M, \|\xi/q\|^{-1}\right) \log N \left(\frac{1}{q} + \frac{1}{N} + \frac{q}{N^2}\right)^{1/2} \\
&\ll \min\left(M, \|\xi/q\|^{-1}\right) \frac{\sqrt{q} \log q}{N}.
\end{aligned}$$

Summing over $\xi \neq 0$, we thus obtain that the contribution of nonzero ξ is bounded by

$$\ll \frac{\log q}{\sqrt{q} N} \sum_{\xi=1}^{q-1} \min\left(M, \|\xi/q\|^{-1}\right) \leq \frac{\sqrt{q} \log q}{N} \sum_{\xi=1}^{q-1} \frac{1}{\xi} \ll \frac{\sqrt{q} \log^2 q}{N}.$$

It follows that (16) holds whenever

$$\frac{M}{q} \gg \frac{\sqrt{q} \log^2 q}{N},$$

completing the proof of the lemma. □

Now we can prove our quadratic Diophantine approximation lemma in full generality.

Proof of Lemma 38. Let Q be a parameter to be chosen shortly (which we will take to be greater than N). By Dirichlet's lemma, there exist $(a, q) = 1$ with $q \in [Q]$ such that $|\gamma - a/q| < (qQ)^{-1}$. If $q \leq N$, then

$$\|q^2 \gamma\| \leq q \|q\gamma\| < \frac{q}{Q} \leq \frac{N}{Q}.$$

If $q > N$, we apply the previous lemma to obtain $n \in [N]$ for which $\|n^2 \frac{a}{q}\| \ll \frac{\sqrt{q} \log^2 q}{N}$, so that

$$\|n^2 \gamma\| \ll \frac{N^2}{qQ} + \frac{\sqrt{q} \log^2 q}{N} \ll \frac{N}{Q} + \frac{\sqrt{Q} \log^2 Q}{N}.$$

The optimal choice of Q is when $Q^{3/2} \log^2 Q \asymp N^2$, so we take $Q = N^{4/3} / \log^{4/3} N$, which gives the desired bound with $c = \frac{1}{3} - \frac{1}{1000}$, say. □

We will need one more technical lemma

Lemma 42. *Any arithmetic progression $P = a + q[N]$ in $\mathbf{Z}/p\mathbf{Z}$ can be partitioned into the union of $\ll \sqrt{N}$ integer arithmetic progressions in $\{0, 1, \ldots, p-1\}$.*

Proof. By Dirichlet's lemma, there exists a positive integer $r \leq \sqrt{N}$ such that $\|rq/p\| \leq N^{-1/2}$, i.e., there exists $(s, r) = 1$ for which $|q/p - s/r| \leq \frac{1}{rN^{1/2}}$. First, divide P into arithmetic progressions $P_0, \ldots, P_{r-1}$ depending on $n \pmod r$, which will produce $r \leq \sqrt{N}$ such progressions, each of length $\leq \lceil N/r \rceil \leq 2\sqrt{N}$. Then, since $P_i \subset \{a + q(rn + i)\}$, we have that the difference between two elements of P_i is at most p, and so each can be split into at most two integer arithmetic progressions. □

Now, deducing the density increment lemma is straightforward.

Lemma 43. *There exist absolute constants $c_1, c_2 > 0$ such that the following holds. Let $A \subset [N]$ have density $\alpha \leq \frac{7}{8}$ and assume that A contains no nontrivial four-term arithmetic progressions. Then either $N < \alpha^{-c_1}$, or there exists an arithmetic progression $P = a + q \cdot [N']$ with $N' \geq N^{\alpha^{c_1}}$ such that*

$$\frac{|A \cap P|}{|P|} \geq \alpha + \alpha^{c_2}.$$

Proof. Let $p \in (6N, 12N)$ be prime. By our manipulations a couple of sections ago, if $N \geq 100\alpha^{-3} \geq \alpha^{-O(1)}$, then $f_A := 1_A - \alpha 1_{[N]}$ (viewed as a function on $G = \mathbf{Z}/p\mathbf{Z}$) satisfies $\|f_A\|_{U^3} \geq \alpha^{O(1)}$. The local inverse theorem for the U^3-norm then says that there exists an arithmetic progression P of length at least $p^{\alpha^{O(1)}} \geq N^{\alpha^{O(1)}}$ and polynomials $Q_i \in \mathbf{Z}[x]$ of degree at most 2 (in fact, with the same coefficient in front of the quadratic term) such that

$$\mathbf{E}_{i \in G} \left| \mathbf{E}_{x \in i+P} f_A(x) e_p(Q_i(x)) \right| \geq \alpha^{O(1)}.$$

Write $Q_i(i + a + qx) = a_i x^2 + b_i$, where $P = a + q[N']$. By mimicing the proof of Roth's theorem, using our linear and quadratic Diophantine approximation lemmas to approximate both $\frac{a_i}{p}$ and $\frac{b_i}{p}$ by rationals with denominator not too large, there is a further partition of each $i + P$ into arithmetic progressions $P_{i,j}$ of length $\asymp p^{\alpha^{O(1)}}$ such that

$$\mathbf{E}_{i \in G} \mathbf{E}_{j \in J_i} \left| \mathbf{E}_{x \in P_{i,j}} f_A(x) \right| \geq \alpha^{O(1)}.$$

We can, using the previous lemma, further partition each $P_{i,j}$ into $\ll |P_{i,j}|^{1/2}$ genuine integer arithmetic progressions contained in $[N]$,

$$\mathbf{E}_{i \in G} \mathbf{E}_{j \in J_i'} \left| \mathbf{E}_{x \in P_{i,j}'} f_A(x) \right| \geq \alpha^{O(1)}.$$

Arguing again like in the proof of Roth's theorem, there exists a $P_{i,j}'$ of length $\gg p^{\alpha^{O(1)}}$ on which A has density at least $\alpha + \alpha^{O(1)}$. □

7 The arithmetic regularity lemma

We will now give a couple of applications of the (Fourier analytic) arithmetic regularity lemma, an arithmetic analogue of Szemerédi's regularity lemma that was introduced by Green in 2003 and extended (in various forms) to regularity lemmas for higher degree Gowers norms in works of Gowers and Wolf on the "true complexity" of linear systems and in work of Green and Tao. We will begin by illustrating the ideas in the finite field model setting, and then work in the integer setting.

7.1 The finite field model arithmetic regularity lemma

For the sake of brevity, we will give a name to the situation where a set has no large Fourier coefficients at nonzero frequencies.

Definition 10. *Let $A \subset \mathbf{F}_p^n$ and $\varepsilon > 0$. We say that A is ε-uniform if*

$$\max_{0 \neq \xi \in \mathbf{F}_p^n} \left| \widehat{1_A}(\xi) \right| \leq \varepsilon.$$

If, in addition, $V \leq \mathbf{F}_p^n$ is a subspace and $W = v + V$ is a coset of V, we say that A is ε-uniform on W if $A \cap W$ is ε-uniform as a subset of W in the natural way, i.e.,

$$\max_{0 \neq \xi \in \widehat{V}} \left| \widehat{1_{(A \cap W) - v}}(\xi) \right| \leq \varepsilon,$$

where the Fourier transform of $1_{(A \cap W) - v}$ is taken on V.

Now we can state Green's arithmetic regularity lemma in the setting of high dimensional vector spaces over finite fields.

Lemma 44. *For all primes p and $\varepsilon > 0$, there exists $M(\varepsilon) = M_p(\varepsilon) \in \mathbf{N}$ such that the following holds. For all $A \subset \mathbf{F}_p^n$, there exists a subspace $V \leq \mathbf{F}_p^n$ of codimension at most $M(\varepsilon)$ such that A is ε-uniform on all but at most $\varepsilon |\mathbf{F}_p^n / V|$ cosets of V in $\mathbf{F}_p^n$.*

Thus, the arithmetic regularity lemma gives a partition of $\mathbf{F}_p^n$ into cosets of a space of bounded codimension such that A is ε-uniform on all but an ε-proportion of parts of the partition. Adapting the ideas from Gowers's lower bound construction for Szemerédi's regularity lemma, Green, showed that there exists $A \subset \mathbf{F}_p^n$ for which any ε-uniform partition into cosets of a subspace V requires $\operatorname{codim} V$ to be at least a tower of p's of height $\gg \varepsilon^{-\Omega(1)}$. The proof of the arithmetic regularity lemma is very similar to the proof of Szemerédi's regularity lemma, and also gives an upper bound for $M(\varepsilon)$ of the same shape.

For any $A, B, C \subset \mathbf{F}_p^n$, a *triangle* in $A \times B \times C$ is a triple $(x, y, z) \in A \times B \times C$ satisfying $x + y + z = 0$. One of the original applications of the arithmetic regularity lemma was to prove an arithmetic version of the triangle removal lemma.

Lemma 45 (Arithmetic triangle removal lemma)**.** *For every $\varepsilon > 0$, there exists a $\delta = \delta(\varepsilon) > 0$ such that the following holds. If $A, B, C \subset \mathbf{F}_p^n$ and $A \times B \times C$ contains δp^{2n} triangles, then $A \times B \times C$ can be made triangle-free by removing at most εp^n elements from A, B, and C.*

Later, Král, Serra, and Vena noticed that the arithmetic triangle removal lemma also just follows from the triangle removal lemma for graphs, and thus Fox's improved bound in the triangle removal lemma (where the tower height depends on $\log(2/\varepsilon)$) also applies to the arithmetic triangle removal lemma. Green asked in 2003 whether a polynomial bound (i.e., polynomial dependence of δ on ε) could hold in the arithmetic triangle removal lemma, and this was finally answered in the affirmative in 2017 by Fox and L. M. Lovász using results on "tricolored sum-free sets", which were proven using the slice-rank polynomial method.

There are still numerous genuine uses of both the Fourier-analytic and higher-order arithmetic regularity lemmas. Another application in Green's original paper was to answer a question of Bergelson, Host, and Kra about popular differences in Roth's theorem. Here is the statement of Green's result in the setting of high dimensional vector spaces over finite fields.

Theorem 26 (Popular difference Roth's theorem in $\mathbf{F}_3^n$)**.** *For all $\varepsilon > 0$, there exists $N = N(\varepsilon) > 0$ such that the following holds. If $A \subset \mathbf{F}_3^n$ has density α and $n \geq N$, then there exists a nonzero $y \in \mathbf{F}_3^n$ such that*

$$\#\left\{x \in \mathbf{F}_3^n : x, x+y, x+2y \in A\right\} \geq \left(\alpha^3 - \varepsilon\right) 3^n.$$

Observe that one would expect very close to $\alpha^3 3^n$ 3-APs with common difference y in a random subset of $\mathbf{F}_3^n$ of density α, and so this theorem says that for any set one can get arbitrarily close to this random density.

One application of higher order arithmetic regularity lemmas is to a conjecture of Gowers and Wolf about the *true complexity* of linear configurations, i.e., the smallest s such that the degree U^{s+1}-norm that controls the average

$$\mathbf{E}_{x_1,\ldots,x_m \in G} \prod_{i=1}^{k} f_i(\psi_i(x_1, \ldots, x_m))$$

for any 1-bounded functions $f_1, \ldots, f_k$. Gowers and Wolf made the conjecture that the true complexity of a collection of linear forms $\psi_1, \ldots, \psi_k$ equals the smallest positive integer ℓ such that the polynomials $\psi_1^{\ell+1}, \ldots, \psi_k^{\ell+1}$ are linearly independent (say over $\mathbf{F}_p$ or, if G is a cyclic group, over $\mathbf{Q}$). The conjecture of Gowers and Wolf was proven in the setting of high dimensional vector spaces over finite fields in a sequence of papers by Gowers and Wolf, for special collections of linear forms in the cyclic group setting by Green and Tao, and in full generality by Altman. All of these proofs use a version of the arithmetic regularity lemma. Manners later gave a proof of a slightly weaker form of the conjecture in an extremely elaborate argument mainly using repeated applications of the Cauchy–Schwarz inequality.

We will now prove the popular difference version of Roth's theorem in the finite field model setting.

Proof of Theorem 26. First, we apply the arithmetic regularity lemma to A with $\frac{\varepsilon}{4}$ in place of ε. So, there exists $M \in \mathbf{N}$ depending only on ε and a subspace $V \leq \mathbf{F}_3^n$ of codimension at most M such that A is $\frac{\varepsilon}{4}$-uniform on all but at most an ε-proportion of cosets of V. The idea now is that we will be able to find a popular common difference $0 \neq y \in V$ since we can use the uniformity of A on most cosets of V to count the number of 3-APs with common

difference in V. For any $v+V \in \mathbf{F}_3^n/V$, let α_v denote the density of $A \cap (v+V)$ in $v+V$. If A is $\frac{\varepsilon}{4}$-uniform on $v+V$, then the number of 3-APs in $A \cap (v+V)$ is

$$\begin{aligned}|V|^2 \mathbf{E}_{x,y \in V} 1_{A-v}(x) 1_{A-v}(x+y) 1_{A-v}(x+2y) &= |V|^2 \sum_{\xi \in \widehat{V}} \widehat{1_{A-v}}(\xi)^2 \widehat{1_{A-v}}(-2\xi) \\ &= \alpha_v^3 |V|^2 + |V|^2 \sum_{0 \neq \xi \in \widehat{V}} \widehat{1_{A-v}}(\xi)^2 \widehat{1_{A-v}}(-2\xi) \\ &\geq \left(\alpha_v^3 - \frac{\varepsilon}{4}\right) |V|^2,\end{aligned}$$

by Parseval's identity. Since all of these 3-APs have common difference in V, summing over all $v+V \in \mathbf{F}_3^n/V$ for which A is $\frac{\varepsilon}{4}$-uniform on $v+V$ gives that the total number of 3-APs with common difference in V is at least

$$\left(\sum_{v \in \mathbf{F}_3^n/V} \alpha_v^3 - 2 \cdot \frac{\varepsilon}{4} 3^{\operatorname{codim} V} \right) |V|^2 \geq \left(\alpha^3 - \frac{\varepsilon}{2}\right) 3^n |V|$$

by Jensen's inequality.

Now, let N be large enough so that $|V| \geq \frac{2}{\varepsilon}$ whenever $n \geq N$. The number of trivial 3-APs in A is $|A| = \alpha 3^n$, and thus, by our choice of n, at most $3^n \leq \frac{\varepsilon}{2} 3^n |V|$, so that the number of nontrivial 3-APs in A with common difference in V is at least $(\alpha^3 - \varepsilon) 3^n |V|$. The conclusion of the theorem now follows by the pigeonhole principle. □

Now, we will, finally, prove the Fourier analytic arithmetic regularity lemma for high dimensional vector spaces over finite fields.

Proof of Lemma 44. The argument is very similar to the proof of Szemerédi's regularity lemma, also using an energy increment argument. So, we begin by defining an appropriate notion of mean square density relative to a partition into cosets of a subspace. For any $B \subset \mathbf{F}_p^n$ and subspace $W \leq \mathbf{F}_p^n$, the *energy* is defined by

$$E(B,W) := \mathbf{E}_{v \in \mathbf{F}_p^n} \frac{|B \cap (v+W)|^2}{|W|^2}.$$

First of all, observe that $E(B,W)$ always lies in $[0,1]$. If we set $\mu_W(x) := p^{\operatorname{codim} W} \cdot 1_W$ to be the indicator function of W weighted so that it has mean 1, then, since $W = -W$,

$$(1_B * \mu_W)(v) = \mathbf{E}_{x \in \mathbf{F}_p^n} 1_B(x) \mu_W(x-v) = \frac{|B \cap (v+W)|}{|W|},$$

and so

$$E(B,W) = \|1_B * \mu_W\|_{L^2}^2 = \|\widehat{1_B * \mu_W}\|_{\ell^2}^2 = \sum_{\xi \in \mathbf{F}_p^n} \left|\widehat{1_B}(\xi)\right|^2 \left|\widehat{1_W}(\xi)\right|^2 = \sum_{\xi \in W^\perp} \left|\widehat{1_B}(\xi)\right|^2.$$

As a consequence, we see that if $W' \leq W$, then $E(B,W) \leq E(B,W')$.

Next, we will see that if $V \leq \mathbf{F}_p^n$ is a subspace such that A is not ε-uniform on more than an ε-proportion of cosets of V, then we can obtain a large energy increment. Indeed, if $B \subset \mathbf{F}_p^m$ were not ε-uniform, then there would exist a nonzero $\xi \in \mathbf{F}_p^m$ for which $|\widehat{1_B}(\xi)| > \varepsilon$. Then,

$$E(B, \langle \xi \rangle^\perp) = \sum_{j=0}^{p-1} |\widehat{1_B}(j\xi)|^2 > |\widehat{1_B}(0)|^2 + \varepsilon^2 = E(B, \mathbf{F}_p^n) + \varepsilon^2.$$

Apply this locally to each nonuniform coset $v+V$ of V to obtain refinements by intersecting with $\langle \xi_v \rangle^\perp$. Set $V' := V \cap \langle (\xi_v)_{v \text{ s.t. } v+V \text{ nonuniform}} \rangle^\perp$, so that the codimension of V' is certainly at most $\operatorname{codim} V + |\mathbf{F}_p^n/V|$. Then, since further refining a partition cannot decrease the energy, we have that

$$E(B, V') = E(B, V) + \varepsilon^3.$$

Using this, we can now run an energy increment iteration. We have a sequence of subspaces (V_i) of $\mathbf{F}_p^n$ constructed as follows: $V_0 = \mathbf{F}_p^n$, and, if A is not ε-uniform on all but at most an ε-proportion of cosets of V_i, then we find a subspace $V_{i+1} \leq V_i$ such that $E(A, V_{i+1}) \geq E(A, V_i) + \varepsilon^3$ as above. The energy cannot exceed 1, and so at some point this iteration must terminate (in at most ε^{-3} steps, in fact), at which point we can find a suitable V. □

There is also a useful "functional decomposition" version of the Fourier-analytic arithmetic regularity lemma, from which our earlier formulation follows. Usually, when one refers to an arithmetic regularity lemma (especially for higher degree Gowers norms) one usually means a result of the following form.

Lemma 46. *For every function $\mathcal{F} : \mathbf{Z}_{\geq 0} \to [0, \infty)$, $\varepsilon > 0$, and $f : \mathbf{F}_p^n \to [0, 1]$, there exists a positive integer $M \ll_{\varepsilon, \mathcal{F}} 1$ such that f can be decomposed as*

$$f = f_{str} + f_{sml} + f_{psd},$$

where $f_{str} = f * \mu_V$ for some subspace $V \leq \mathbf{F}_p^n$ of codimension at most M, $\|f_{sml}\|_{L^2} \leq \varepsilon$, and $\|\widehat{f_{psd}}\|_{\ell^\infty} \leq \mathcal{F}(M)^{-1}$.

7.2 The arithmetic regularity lemma in the integer setting

There is also a popular difference version of Roth's theorem in the integer setting, first proven by Green.

Theorem 27. *For all $\varepsilon > 0$, there exists $N_0 = N_0(\varepsilon) \in \mathbf{N}$ such that the following holds. If $N \geq N_0$, then, for all $A \subset [N]$ of density α, there exists a nonzero $y \in \mathbf{Z}$ such that*

$$\#\{x \in [N] : x, x+y, x+2y \in A\} \geq (\alpha^3 - \varepsilon)N$$

The best known bounds are due to Fox–Pham–Zhao, and give N bounded by a tower of 2's of height $\ll \log(2/\varepsilon)$, which (as in the finite field model case) they also show is basically best possible. I will put the proof of Theorem 27 on the last homework. It can be proven by leveraging another Fourier-analytic arithmetic regularity lemma, which we will state and prove shortly.

First, we will define the U^2-norm of a finitely supported function on $\mathbf{Z}$. For any $f : \mathbf{Z} \to \mathbf{C}$ supported on $[N]$, we define

$$\|f\|_{U^2(N)} := \frac{\|f\|_{U^2(\mathbf{Z}/M\mathbf{Z})}}{\|1_{[N]}\|_{U^2(\mathbf{Z}/M\mathbf{Z})}}$$

for any $M \geq 2N$. Observe that the definition of $\|\cdot\|_{U^2(N)}$ does not depend on M, as long as $M \geq 2N$. The inverse theorem for the $U^2(N)$-norm follows immediately from the inverse theorem for the $U^2(\mathbf{Z}/2N\mathbf{Z})$-norm:

Lemma 47. *If $f : \mathbf{Z} \to \mathbf{C}$ is 1-bounded, supported on $[N]$, and satisfies $\|f\|_{U^2(N)} \geq \delta$, then there exists $\xi \in \mathbf{T}$ such that*

$$\left|\mathbf{E}_{n\in[N]} f(n)e(\xi n)\right| \gg \delta^2.$$

We will need the notion of a Lipschitz function on $\mathbf{T}^d$, and so we should pick a metric on $\mathbf{T}^d$. For each $x, y \in \mathbf{T}^d$, let $d(x, y)$ denote the distance between x and y in the Euclidean metric on $\mathbf{T}^d$. Then, the Lipschitz norm of a function $F : \mathbf{T}^d \to \mathbf{C}$ is

$$\|F\|_{\mathrm{Lip}} := \|F\|_{L^\infty} + \sup_{x\neq y} \frac{|F(x) - F(y)|}{d(x, y)}.$$

A function $f : [N] \to \mathbf{C}$ is said to have *complexity* at most M if, for all $n \in [N]$, $f(n) = F(\theta n)$ for some $\theta \in \mathbf{T}^d$ and function $F : \mathbf{T}^d \to \mathbf{C}$ with $d \leq M$ and $\|F\|_{\mathrm{Lip}} \leq M$.

Let $\mathcal{F} : (0, \infty) \to (0, \infty)$ be increasing. We say that a function $f : [N] \to \mathbf{C}$ is *Fourier measurable with growth* $\mathcal{F}$ if, for all $M > 0$, there exists a function $f_M : [N] \to \mathbf{C}$ of complexity at most $\mathcal{F}(M)$ such that $\|f - f_M\|_{L^2} \leq M^{-1}$ (where, here, the L^2-norm is normalized by dividing through by N). We say that $E \subset [N]$ is *Fourier measurable with growth* $\mathcal{F}$ if 1_E is Fourier measurable with growth $\mathcal{F}$. Note that polynomial combinations of Fourier measurable functions of growth $\mathcal{F}$ are again Fourier measurable with growth depending only on $\mathcal{F}$ and the polynomial. We will write that $\mathcal{F} \ll_\delta 1$ for some parameter δ to mean that there exist functions $\mathcal{G}_\delta : (0, \infty) \to (0, \infty)$ for which $\mathcal{F}(M) \leq \mathcal{G}_\delta(M)$ for all $M > 0$.

Recall that, in the finite field model setting, we proved the arithmetic regularity lemma by iteratively refining a partition of $\mathbf{F}_p^n$ into cosets of a single subspace. The notion of Fourier measurability will be useful for formulating an analogue of this procedure in the integer setting, where the analogue of subspaces (Bohr sets) are much less nice.

To this end, we define a *factor* of $[N]$ to be any subalgebra of subsets of $[N]$ (i.e., a collection of subsets containing $[N]$ and closed under complementation, finite union, and finite intersection). Specifying a factor of $[N]$ is the same as specifying a partition $S_1 \sqcup \cdots \sqcup S_k$ of $[N]$ into nonempty subsets–the factor is then simply the factor generated by complements, finite unions, and finite intersections of the S_i, which are called the *atoms* of the factor. Given a factor $\mathcal{B}$ of $[N]$, we denote the atom containing $n \in [N]$ by $\mathcal{B}(n)$, and a function $f : [N] \to \mathbf{C}$ is said to be $\mathcal{B}$*-measurable* if f is constant on atoms of $\mathcal{B}$. We say that a factor $\mathcal{B}'$ is a *refinement* of a factor $\mathcal{B}$ if every atom of $\mathcal{B}$ is a union of atoms of $\mathcal{B}'$. Given an increasing function $\mathcal{F} : (0, \infty) \to (0, \infty)$ and an $M > 0$, we say that a factor $\mathcal{B}$ of $[N]$ is a *Fourier factor of* $[N]$ *with complexity at most* M *and growth* $\mathcal{F}$ if $\mathcal{B}$ has at most M cells, each of which is a Fourier measurable set with growth $\mathcal{F}$.

The *conditional expectation* of a function $f : [N] \to \mathbf{C}$ relative to a factor $\mathcal{B}$ to be the function $\mathbf{E}(f|\mathcal{B})$ on $[N]$ is defined by

$$\mathbf{E}(f|\mathcal{B})(x) = \mathbf{E}_{y \in \mathcal{B}(x)} f(y).$$

We also define the *energy* by $E(f, \mathcal{B}) := \|\mathbf{E}(f|\mathcal{B})\|_{L^2}^2$. When f is the indicator function of a subset of $[N]$, this is the integer analogue of our definition of energy in the finite field model setting. Note that if $\mathcal{B}'$ refines $\mathcal{B}$, then $E(f, \mathcal{B}') \geq E(f, \mathcal{B})$. Indeed, we certainly have

$$\|\mathbf{E}(f|\mathcal{B}') - \mathbf{E}(f|\mathcal{B})\|_{L^2}^2 \geq 0.$$

Expanding the left-hand side yields

$$E(f, \mathcal{B}') + E(f, \mathcal{B}) - 2\mathbf{E}_{n \in [N]} \left[\mathbf{E}(f|\mathcal{B}')(n)\mathbf{E}(f|\mathcal{B})(n)\right] \geq 0.$$

Since the sum of $\mathbf{E}(f|\mathcal{B}')(n)$ over any fixed atom of $\mathcal{B}$ equals the value of $\mathbf{E}(f|\mathcal{B})$ on that atom, we have that

$$\mathbf{E}_{n \in [N]} \left[\mathbf{E}(f|\mathcal{B}')(n)\mathbf{E}(f|\mathcal{B})(n)\right] = \mathbf{E}_{n \in [N]} \mathbf{E}(f|\mathcal{B})(n)^2 = E(f, \mathcal{B}).$$

It follows that $E(f, \mathcal{B}') - E(f, \mathcal{B}) \geq 0$.

Next, we will prove an analogue of the statement that if a set that is not Fourier uniform on a subspace, then one can find a partition of the subspace that has substantially higher energy with respect to the set. We will require (a version of) the Hardy–Littlewood maximal inequality.

Theorem 28. *Let μ be a Borel probability measure on $\mathbf{R}$, and let*

$$(M\mu)(t) := \sup_{r > 0} \frac{\mu([t - r, t + r])}{2r}$$

be the associated maximal function. Then, for all $\lambda > 0$, we have

$$|\{t \in \mathbf{R} : (M\mu)(t) > \lambda\}| \ll \frac{1}{\lambda}.$$

Now we can prove the promised lemma.

Lemma 48. *Let $f : [N] \to \mathbf{C}$ be a 1-bounded function. If $\|f\|_{U^2(N)} \geq \delta$, then there exists a Fourier measurable $E \subset [N]$ with growth $\mathcal{F} \ll_\delta 1$ such that*

$$\left|\mathbf{E}_{n \in [N]} f(n) 1_E(n)\right| \gg_\delta 1.$$

Proof. By the $U^2(N)$-inverse theorem, there exists $\alpha \in \mathbf{T}$ such that

$$\left|\mathbf{E}_{n \in [N]} f(n) e(\alpha n)\right| \gg \delta^2.$$

The remainder of the proof is just a sequence of maneuvers to turn this linear phase into the indicator function of a low complexity set. Writing $e(\alpha n) = \cos(2\pi\alpha n) + i \sin(2\pi\alpha n)$, we have by the triangle inequality and pigeonhole principle that

$$\left|\mathbf{E}_{n \in [N]} f(n) g(n)\right| \gg \delta^2$$

for g equal to either $\cos(2\pi\alpha\cdot)$ or $\sin(2\pi\alpha\cdot)$. Writing $g = g_+ - g_-$, where $g_+ = \max(g, 0)$ and $g_- = -\min(g, 0)$ and arguing similarly, we conclude that there exists $h : \mathbf{R} \to [0, 1]$ such that

$$\left|\mathbf{E}_{n\in[N]} f(n)h(n)\right| \gg \delta^2,$$

where, since $\max(g, 0) = \frac{x+|x|}{2}$ and $\min(g, 0) = \frac{x-|x|}{2}$, the function h can be written as $H(\alpha n)$ where $\|H\|_{\mathrm{Lip}} \ll 1$.

We now apply the layercake decomposition to h. For each $t \in [0, 1]$, set

$$E_t := \{n \in [N] : h(n) \geq t\}.$$

Then,

$$h(n) = \int_0^1 1_{E_t}(n)dt.$$

Plugging this in for h and using the triangle inequality yields

$$\int_0^1 \left|\mathbf{E}_{n\in[N]} f(n)1_{E_t}(n)\right| dt \geq C\delta^2$$

for some $C > 0$. Let $S \subset [0, 1]$ be the set of $t \in [0, 1]$ for which $\left|\mathbf{E}_{n\in[N]} f(n)1_{E_t}(n)\right| \geq \frac{C}{2}\delta^2$, say. Then S is a Borel set, and has measure $|S| \geq \frac{C}{2}\delta^2$ (or else it would contradict the inequality above).

We now apply the maximal inequality with the Borel probability measure

$$\mu(A) := \frac{\#\{n \in [N] : h(n) \in A\}}{N}$$

on $\mathbf{R}$, where we extend h to be zero outside of $[0, 1]$. This tells us that

$$|\{t \in [0, 1] : (M\mu)(t) > \lambda\}| \ll \frac{1}{\lambda}$$

for all $\lambda > 0$. Taking $\lambda \asymp \delta^{-2}$, it follows that there must exist some $t \in S$ with $(M\mu)(t) \ll \delta^{-2}$. That is,

$$\#\{n \in [N] : h(n) \in [t - r, t + r]\} \ll \delta^{-2}rN$$

for any $r > 0$. Fix this t.

For each $r > 0$, let $\phi_r : \mathbf{R} \to [0, \infty)$ be a 1-bounded function with $\|\phi_r\|_{\mathrm{Lip}} \ll r^{-1}$ such that $\phi_r(x)$ is zero on $(-\infty, t - r)$ and is 1 on $(t + r, \infty)$. Then, since $\phi_r \circ h$ differs from 1_{E_t} only when $h(n) \in [t - r, t + r]$, we have

$$\|1_{E_t} - \phi_r \circ h\|_{L^2}^2 \ll_\delta r,$$

and so, taking $r = C_\delta M^{-2}$ for C_δ the implied constant above and setting $f_M = \phi_{C_\delta M^{-2}} \circ h$, we have that $\|1_{E_t} - f_M\|_{L^2} \leq M^{-1}$ and that f_M has complexity $\ll_\delta M^2$, since $h(n) = H(\alpha n)$ with $\|H\|_{\mathrm{Lip}} \ll 1$ and $\|\phi_{C_\delta M^{-2}}\|_{\mathrm{Lip}} \ll_\delta M^2$. This means that $E = E_t$ is Fourier measurable with growth $\mathcal{F} \ll_\delta 1$, and since $t \in S$,

$$\left|\mathbf{E}_{n\in[N]} f(n)1_E(n)\right| \geq \frac{C}{2}\delta^2 \gg_\delta 1$$

as desired. □

Now we can prove an analogue of the result from last time saying that if a set is nonuniform on many cosets of a subspace, then one can refine the partition of $\mathbf{F}_p^n$ into a union of cosets of a smaller subspace to obtain a substantial energy increment.

Lemma 49. *Let $\mathcal{F} : (0, \infty) \to (0, \infty)$ be an increasing function, $M > 0$, and $f : [N] \to \mathbf{R}$ be 1-bounded. Suppose that $\mathcal{B}$ is a Fourier factor of $[N]$ with complexity at most M and growth $\mathcal{F}$. If*

$$\|f - \mathbf{E}(f|\mathcal{B})\|_{U^2(N)} \geq \delta,$$

then there exists a refinement $\mathcal{B}'$ of $\mathcal{B}$ of complexity at most $2M$ and growth $\ll_{M,\delta,\mathcal{F}} 1$ such that

$$E(f, \mathcal{B}') - E(f, \mathcal{B}) \gg_\delta 1.$$

Proof. By the above lemma, there exists a Fourier measurable $E \subset [N]$ with growth $\ll_\delta 1$ such that

$$\left|\mathbf{E}_{n\in[N]}(f - \mathbf{E}(f|\mathcal{B}))(n)1_E(n)\right| \gg_\delta 1.$$

Let $\mathcal{B}'$ be the factor generated by $\mathcal{B}$ and E, so that $\mathcal{B}'$ has complexity at most $2M$ (by intersecting each atom of $\mathcal{B}$ with E or $[N] \setminus E$ to obtain atoms for $\mathcal{B}'$) with growth $\ll_{M,\delta,\mathcal{F}} 1$. Then, since 1_E is $\mathcal{B}'$ measurable and thus constant on atoms of $\mathcal{B}'$, we can replace f by the conditional expectation $\mathbf{E}(f|\mathcal{B}')$ above to obtain

$$\left|\mathbf{E}_{n\in[N]}(\mathbf{E}(f|\mathcal{B}') - \mathbf{E}(f|\mathcal{B}))(n)1_E(n)\right| \gg_\delta 1.$$

Applying the Cauchy–Schwarz inequality yields

$$\|\mathbf{E}(f|\mathcal{B}') - \mathbf{E}(f|\mathcal{B})\|_{L^2}^2 \gg_\delta 1.$$

Expanding the left-hand side, we get that

$$1 \ll_\delta E(f, \mathcal{B}') + E(f, \mathcal{B}) - 2\mathbf{E}_{n\in[N]}\mathbf{E}(f|\mathcal{B})\mathbf{E}(f|\mathcal{B}') = E(f, \mathcal{B}') + E(f, \mathcal{B}) - 2\mathbf{E}_{n\in[N]}\mathbf{E}(f|\mathcal{B})^2$$

since the sum of $\mathbf{E}(f|\mathcal{B}')$ over each atoms of $\mathcal{B}$ equals the sum of $\mathbf{E}(f|\mathcal{B})$ over that atom. Thus,

$$E(f, \mathcal{B}') - E(f, \mathcal{B}) \gg_\delta 1.$$

as desired. □

Now, by running an energy increment iteration just like in the finite field model case, we can deduce what's known as the weak arithmetic regularity lemma, which gives a decomposition of any 1-bounded function into a structured part and a U^2-pseudorandom part.

Lemma 50 (Weak Fourier analytic arithmetic regularity lemma)**.** *Let $\mathcal{F} : (0, \infty) \to (0, \infty)$ be an increasing function, $M > 0$, $\mathcal{B}$ be a Fourier factor of $[N]$ of complexity at most M and growth $\mathcal{F}$, and $f : [N] \to \mathbf{R}$ be 1-bounded. There exists a refinement $\mathcal{B}'$ of $\mathcal{B}$ of complexity $\ll_{\delta,M} 1$ and growth $\ll_{\delta,M,\mathcal{F}} 1$ such that*

$$\|f - \mathbf{E}(f|\mathcal{B}')\|_{U^2(N)} \leq \delta.$$

Proof. We repeatedly apply the previous lemma to obtain a sequence $(\mathcal{B}_i)$ of refinements of $\mathcal{B}$ such that $E(f, \mathcal{B}_{i+1}) - E(f, \mathcal{B}_i) \gg_\delta 1$ if $\|f - \mathbf{E}(f|\mathcal{B}_i)\|_{U^2(N)} > \delta$. Since energy cannot exceed 1, this iteration must terminate in $\ll_\delta 1$ steps, at which point we must have $\|f - \mathbf{E}(f|\mathcal{B}_i)\|_{U^2(N)} \leq \delta$ and that $\mathcal{B}' = \mathcal{B}_i$ is a refinement of $\mathcal{B}$ of complexity $\ll_{\delta,M} 1$ and growth $\ll_{\delta,M,\mathcal{F}} 1$. □

Now we can prove the full Fourier analytic arithmetic regularity lemma.

Theorem 29. *Let $\varepsilon > 0$, $\mathcal{F} : (0, \infty) \to (0, \infty)$ be increasing, and $f : [N] \to [0, 1]$. There exists $M \ll_{\varepsilon,\mathcal{F}} 1$ and a decomposition*

$$f = f_{str} + f_{sml} + f_{psd}$$

into 1-bounded functions $[N] \to \mathbf{R}$ such that

1. *f_{str} has complexity at most M,*
2. *$\|f_{sml}\|_{L^2} \leq \varepsilon$,*
3. *$\|f_{psd}\|_{U^2(N)} \leq \mathcal{F}(M)^{-1}$, and*
4. *both f_{str} and $f_{str} + f_{sml}$ take values in $[0, 1]$.*

Proof. The proof proceeds by another iteration. We construct a sequence of increasing real numbers, factors of $[N]$, and 1-bounded real valued functions on $[N]$ as follows. Set $M_0 = 1$ and $\mathcal{B}_0 = \{\emptyset, [N]\}$ to be the trivial factor (which, trivially, has complexity $\ll 1$ and growth $\ll 1$), and f_0 to be the constant function $\mathbf{E}_{n\in[N]} f(n)$. If $(M_i, \mathcal{B}_i)$ has been given and $\mathcal{B}_i$ has complexity and growth $\ll_{i,M,\mathcal{F}} 1$, then, by definition, there exists a function $f_i : [N] \to [0, 1]$ of complexity $M_{i+1} \ll_{\varepsilon,i,\mathcal{F}} 1$ such that $M_{i+1} \geq M_i$ for which

$$\|\mathbf{E}(f|\mathcal{B}_i) - f_i\|_{L^2} \leq \frac{\varepsilon}{2}.$$

Further, by the weak regularity lemma, there exists a refinement $\mathcal{B}_{i+1}$ of $\mathcal{B}_i$ of complexity and growth $\ll_{i,M_{i+1},\mathcal{F}} 1$ such that

$$\|f - \mathbf{E}(f|\mathcal{B}_{i+1})\|_{U^2(N)} \leq \frac{1}{\mathcal{F}(M_{i+1})}.$$

Now, since energy can only increase under refinement, $E(f, \mathcal{B}_i)$ is an increasing sequence of reals between $[0, 1]$. By the pigeonhole principle, there must exist $i \leq \frac{4}{\varepsilon^2} \ll_\varepsilon 1$ such that

$$E(f, \mathcal{B}_{i+1}) - E(f, \mathcal{B}_i) \leq \frac{\varepsilon^2}{4}.$$

For this choice of i, set $M = M_{i+1} \ll_{\varepsilon,\mathcal{F}} 1$, $f_{str} = f_i$, $f_{sml} = \mathbf{E}(f|\mathcal{B}_{i+1}) - f_i$, and $f_{psd} = f - \mathbf{E}(f|\mathcal{B}_{i+1})$. Then, since

$$\|f_{sml}\|_{L^2} \leq \|\mathbf{E}(f|\mathcal{B}_{i+1}) - \mathbf{E}(f|\mathcal{B}_i)\|_{L^2} + \|\mathbf{E}(f|\mathcal{B}_i) - f_i\|_{L^2} \leq 2\frac{\varepsilon}{2} = \varepsilon$$

because $E(f, \mathcal{B}_{i+1}) - E(f, \mathcal{B}_i) = \|\mathbf{E}(f|\mathcal{B}') - \mathbf{E}(f|\mathcal{B})\|_{L^2}^2$, this gives the desired decomposition. □

8 The transference principle

The last topic we will cover in this class is (an instance of) the Fourier analyic transference principle, which allows one to transfer certain results about dense subsets of intervals of integers or abelian groups into statements about dense subsets of sparse pseudorandom subsets. This first appeared in work of Ben Green in 2003, in which he proved Roth's theorem relative to the primes.

Theorem 30. *If A is a subset of the primes that has positive upper density in the primes, then A contains a nontrivial three-term arithmetic progression.*

The ideas in this work of Green were later extended on by Green and Tao to prove their famous theorem about arithmetic progressions in the primes.

We will illustrate the transference principle via an argument due to Prendiville concerning subsets of Sidon sets.

Definition 11. *A subset A of an abelian group is a Sidon set if it contains no nontrivial additive quadruples, i.e., if*

$$a_1 + a_2 = a_3 + a_4$$

for $a_1, a_2, a_3, a_4 \in S$ only when $\{a_1, a_2\} = \{a_3, a_4\}$.

For example, the powers of 2, $\{2^i : i \in \mathbf{Z} \geq 0\}$, are a Sidon set in $\mathbf{Z}$, the logarithms of primes, $\{\log p : p \text{ prime}\}$, are a Sidon set in $\mathbf{R}$ (since the primes form a Sidon set in $\mathbf{Q}^\times$), and the graph $\{(x, x^2) : x \in \mathbf{F}_p\}$ is a Sidon set in $\mathbf{F}_p^2$ when $p > 2$ (as one can check quickly by hand).

As observed by Erdős and Turán in the 1940s, it is not hard to show that if $A \subset [N]$ is a Sidon set, then $|A| \leq \sqrt{N}(1 + o(1))$, and that there exists a Sidon set in $[N]$ of size at least $\sqrt{N}(1 - o(1))$. Similarly, any Sidon set in a finite abelian group of size n has size at most $\sqrt{n}(1 + o(1))$. There are, essentially, three main open problems concerning Sidon sets, the last of which is probably hopeless:

1. Determine, asymptotically, the size of the largest Sidon set contained in $[N]$. Erdős famously offered \$500 to prove an upper bound of $\leq \sqrt{N} + O_\varepsilon(N^\varepsilon)$. The current best bound is $\leq \sqrt{N} + O(N^{1/4})$.

2. Find the densest infinite Sidon set in $\mathbf{N}$. The current world record is due to Ruzsa, who constructed a Sidon set $A \subset \mathbf{N}$ such that $|A \cap [N]| \gg_\varepsilon N^{\sqrt{2}-1-\varepsilon}$ for all $N \in \mathbf{N}$.

3. Classify large (i.e., maximal or close to maximal) Sidon sets in finite abelian groups. Eberhard and Manners have an interesting short paper in which they put many known examples into a single framework.

For the purpose of this section, the most important quality of Sidon sets is that they are very Fourier uniform. We will use the transference principle to give an alternative proof due to Prendiville of the following result originally due to Conlon, Fox, Sudakov, and Zhao:

Theorem 31 (Conlon–Fox–Sudakov–Zhao, 2021)**.** *Let $\alpha > 0$. If $N \gg_\alpha 1$, then any Sidon set $S \subset [N]$ with $|S| \geq \alpha\sqrt{N}$ contains a solution to*

$$x_1 + x_2 + x_3 + x_4 = 4x_5 \tag{17}$$

with $x_1, \dots, x_5$ all distinct.

This is the "relative" version of the following result, which can be proven either directly by using the arithmetic regularity lemma, or by combining a density increment argument with an averaging argument as in the first homework to prove a supersaturation result.

Theorem 32. *For any $A \subset [N]$ with $|A| \geq \alpha N$, we have*

$$\#\left\{(x_1, \dots, x_5) \in A^5 : x_1 + x_2 + x_3 + x_4 = 4x_5\right\} \gg_\alpha N^4.$$

The most common incarnation of the transference principle is via a "dense model lemma", which allows, via relatively standard arguments, one to transfer statements about dense sets (like Theorem 32) into statements about dense subsets of pseudorandom sets (like Theorem 31). Here is the dense model lemma that we will use.

Lemma 51. *Let $S \subset [N]$ be a Sidon set and $\varepsilon > 0$. If $N \gg_\varepsilon 1$, then there exists a dense model $f : [N] \to [0, \infty)$ such that*

$$\|f\|_{L^2} \ll 1$$

and

$$\left\|\widehat{f} - \sqrt{N} \cdot \widehat{1_S}\right\|_{L^\infty} \leq \varepsilon.$$

Here, as in the previous section, we take the normalized Fourier transform of a function g supported on $[N]$: $\widehat{g}(\xi) := \mathbf{E}_{n \in [N]} f(n) e(-\xi n)$ and the normalized L^2-norm $\|g\|_{L^2} := \left(\mathbf{E}_{n \in [N]} |g(n)|^2\right)^{1/2}$.

It is not hard to show that the count of solutions to $x_1 + x_2 + x_3 + x_4 = 4x_5$ in a set is controlled by Fourier analysis. The function f from the lemma can be considered a model for the scaled indicator function $\sqrt{N} \cdot 1_S$ of S (so normalized to have mean $\asymp 1$) in the sense that its Fourier transform is very close to the Fourier transform of $\sqrt{N} \cdot 1_S$. This function f can further be considered a "dense" model because it typically resembles a bounded function with large mean when S is a dense Sidon set, meaning that $|S| \geq \delta\sqrt{N}$. Indeed, the L^∞ bound on $\widehat{f} - \sqrt{N} \cdot \widehat{1_S}$ tells us that $\mathbf{E}_{n \in [N]} f(n)$ is within ε of $\sqrt{N}\mathbf{E}_{n \in [N]} 1_S(n) \geq \delta$, and f's L^2-norm is the same order of magnitude as a 1-bounded function.

We will begin by deriving Theorem 31 from Theorem 32 using the dense model lemma. The first step is to deduce a version of Theorem 32 in which indicator functions of dense sets are replaced by functions with large mean and bounded L^2-norm, as in the dense model lemma.

Corollary 7. *For any $f : [N] \to [0, \infty)$ with $\mathbf{E}_{n \in [N]} f(x) \geq \alpha$ and $\|f\|_{L^2} \leq 1$, we have*

$$\sum_{x_1 + x_2 + x_3 + x_4 = 4x_5} f(x_1) f(x_2) f(x_3) f(x_4) f(x_5) \gg_\alpha N^4.$$

Proof. Set $A := \{n \in [N] : f(n) \geq \alpha/2\}$. Then,

$$\alpha N \leq \sum_{n\in[N]} f(n) = \sum_{n\in A} f(n) + \sum_{n\in[N]\setminus A} f(n) \leq \sum_{n\in A} f(n) + \frac{\alpha N}{2},$$

and so, by the Cauchy–Schwarz inequality,

$$\frac{\alpha N}{2} \leq \sum_{n\in A} f(n) \leq \sqrt{|A|}\sqrt{N}\|f\|_{L^2} \leq \sqrt{|A|N}.$$

Rearranging and squaring both sides yields $|A| \geq \frac{\alpha^2 N}{4}$. By Theorem 32,

$$\sum_{x_1+x_2+x_3+x_4=4x_5} 1_A(x_1)1_A(x_2)1_A(x_3)1_A(x_4)1_A(x_5) \gg_\alpha N^4.$$

Since $\frac{2}{\alpha} f(n) \geq 1_A(n)$ for all $n \in A$, the desired conclusion immediately follows. □

Next, we will prove a lemma that will help us compare the count of solutions to (17) in a dense Sidon set with the number weighted by a dense model.

Lemma 52. *Let $\nu : [N] \to [0, \infty)$ be a weight satisfying*

$$\sum_{x,h,k\in\mathbf{Z}} \Delta_{h,k}\nu(x) \leq N^3.$$

If $f_1, f_2, f_3, f_4, f_5 : [N] \to \mathbf{R}$ satisfy $|f_i| \leq \nu$ for all $i = 1, \ldots, 5$, then

$$\left| \sum_{x_1+x_2+x_3+x_4=4x_5} \prod_{i=1}^{5} f_i(x_i) \right| \leq N^4 \min_{i\in[5]} \|\widehat{f_i}\|_{L^\infty}.$$

Proof. By plugging in the definition of the Fourier transform and using orthogonality of characters, we have

$$N^5 \int_{\mathbf{T}} \widehat{f_1}(\xi)\cdots\widehat{f_4}(\xi)\widehat{f_5}(-4\xi)d\xi = \sum_{x_1+x_2+x_3+x_4=4x_5} \prod_{i=1}^{5} f_i(x_i).$$

For each $i \in [4]$,

$$\begin{aligned}\left| \int_{\mathbf{T}} \widehat{f_1}(\xi)\cdots\widehat{f_4}(\xi)\widehat{f_5}(-4\xi)d\xi \right| &\leq \|\widehat{f_i}\|_{L^\infty} \int_{\mathbf{T}} |\widehat{f_5}(-4\xi)| \prod_{i\neq j\leq 4} |\widehat{f_j}(\xi)|d\xi \\ &\leq \|\widehat{f_i}\|_{L^\infty} \prod_{i\neq j\leq 5} \|\widehat{f_i}\|_{L^4}\end{aligned}$$

by Hölder's inequality and a change of variables, and, similarly,

$$\left| \int_{\mathbf{T}} \widehat{f_1}(\xi)\cdots\widehat{f_4}(\xi)\widehat{f_5}(-4\xi)d\xi \right| \leq \|\widehat{f_5}\|_{L^\infty} \prod_{j=1}^{4} \|\widehat{f_i}\|_{L^4}.$$

For each $i \in [5]$, we have

$$N^4 \|\widehat{f_i}\|_{L^4}^4 = \sum_{x,h,k \in \mathbf{Z}} \Delta_{h,k} f_i(x) \leq \sum_{x,h,k \in \mathbf{Z}} \Delta_{h,k} \nu(x) \leq N^3.$$

Thus,

$$\left| \sum_{x_1+x_2+x_3+x_4=4x_5} \prod_{i=1}^{5} f_i(x_i) \right| \leq N^4 \min_{i \in [5]} \|\widehat{f_i}\|_{L^\infty}.$$

□

Now, assuming the dense model lemma, we can show that dense Sidon sets contain many solutions to (17).

Theorem 33. *Let $S \subset [N]$ be a Sidon set with $|S| \geq \alpha\sqrt{N}$. Then,*

$$\#\{(x_1, \dots, x_5) \in S^5 : x_1 + x_2 + x_3 + x_4 = 4x_5\} \gg_\alpha N^{3/2}.$$

Proof. Let $\varepsilon > 0$ be a parameter to be chosen later, depending only on α. We may as well assume that $N \gg_\varepsilon 1 \gg_\alpha 1$, or else the result follows trivially just by considering the diagonal solutions. Thus, we can apply the dense model lemma to obtain $f : [N] \to [0, \infty)$ such that $\|f\|_{L^2} \ll 1$ and $\|\widehat{f} - \sqrt{N}\widehat{1_S}\|_{L^\infty} \leq \varepsilon$.

Observe first that, since $\widehat{1_S}(0) = |S|/N \geq \alpha$, we have $\mathbf{E}_{n \in N} f(n) \geq \alpha - \varepsilon \gg \alpha$ by ensuring that $\varepsilon \leq \alpha/2$, say. Thus, by Corollary 7,

$$\sum_{x_1+x_2+x_3+x_4=4x_5} \prod_{i=1}^{5} f(x_i) \gg_\alpha N^4.$$

Now, set

$$\Lambda(f_1, \dots, f_5) := \sum_{x_1+x_2+x_3+x_4=4x_5} \prod_{i=1}^{5} f_i(x_i)$$

for all $f_1, \dots, f_5 : [N] \to \mathbf{C}$. Then, writing $f = f - \sqrt{N}1_S + \sqrt{N}1_S$ and using the multilinearity of Λ, we have

$$\begin{aligned}
\Lambda(f, \dots, f) &= \Lambda(f, \dots, f, f - \sqrt{N}1_S) + \Lambda(f, \dots, f, \sqrt{N}1_S) \\
&= \Lambda(f, \dots, f, f - \sqrt{N}1_S) + \Lambda(f, f, f, f - \sqrt{N}1_S, \sqrt{N}1_S) + \Lambda(f, f, f, \sqrt{N}1_S, \sqrt{N}1_S) \\
&= N^{5/2}\Lambda(1_S, \dots, 1_S) + \sum_{j=1}^{5} \Lambda(\overbrace{f, \dots, f}^{j-1 \text{ times}}, f - \sqrt{N}1_S, \overbrace{\sqrt{N}1_S, \dots, \sqrt{N}1_S}^{5-j \text{ times}}).
\end{aligned}$$

Thus,

$$\left|\Lambda(f, \dots, f) - N^{5/2}\Lambda(1_S, \dots, 1_S)\right| \leq \sum_{j=1}^{5} \left| \Lambda(\overbrace{f, \dots, f}^{j-1 \text{ times}}, f - \sqrt{N}1_S, \overbrace{\sqrt{N}1_S, \dots, \sqrt{N}1_S}^{5-j \text{ times}}) \right|$$

and it suffices to bound each of the five terms on the right-hand side.

To do this, we will apply the previous lemma with $C\nu = f + \sqrt{N}1_S$ for some sufficiently large absolute constant $C \geq 1$. Observe that, since $f \geq 0$, we indeed have $\nu : [N] \to [0, \infty)$. Further,

$$\left(\sum_{x,h,k \in \mathbf{Z}} \Delta_{h,k} C\nu(x) \right)^{1/4} = N \|\widehat{C\nu}\|_{L^4} \leq N \left(\|\widehat{f}\|_{L^4} + \sqrt{N} \|\widehat{1_S}\|_{L^4} \right).$$

We have

$$\|\widehat{f}\|_{L^4}^4 \leq \|\widehat{f}\|_{L^\infty}^2 \|\widehat{f}\|_{L^2}^2 \ll \frac{1}{N}$$

where we have used Parseval's identity and that, again assuming $N \gg_\varepsilon 1$, we have $\|\widehat{f}\|_{L^\infty} \ll \|\sqrt{N}\widehat{1_S}\|_{L^\infty} \ll 1$, and, since S is a Sidon set,

$$N^4 \|\widehat{1_S}\|_{L^4}^4 = |S| + 2\binom{|S|}{2} = |S|^2 \ll N.$$

Thus,

$$\left(\sum_{x,h,k \in \mathbf{Z}} \Delta_{h,k} C\nu(x) \right)^{1/4} \ll N \left(N^{-1/4} + N^{1/2-3/4} \right) \ll N^{3/4},$$

and hence $\sum_{x,h,k \in \mathbf{Z}} \Delta_{h,k} \nu(x) \ll C^{-1} N^3$. Fixing $C \gg 1$ then makes ν satisfy $\sum_{x,h,k \in \mathbf{Z}} \Delta_{h,k} \nu(x) \leq N^3$, as needed to apply the previous lemma. Since $|f|, |\sqrt{N}1_S|, |f - \sqrt{N}1_S| \leq C\nu$ and $\|\widehat{f - \sqrt{N}1_S}\|_{L^\infty} \leq \varepsilon$, it thus follows that

$$\left| \Lambda(\overbrace{f, \ldots, f}^{j-1 \text{ times}}, f - \sqrt{N}1_S, \overbrace{\sqrt{N}1_S, \ldots, \sqrt{N}1_S}^{5-j \text{ times}}) \right| \ll \varepsilon N^4$$

for all $j \in [5]$. We conclude, fixing $0 < \varepsilon \ll_\alpha 1$ sufficiently small, that

$$N^{5/2} \Lambda(1_S, \ldots, 1_S) \gg_\alpha N^4.$$

Thus, $\Lambda(1_S, \ldots, 1_S) \gg_\alpha N^{3/2}$, as desired. □

The theorem of Conlon–Fox–Sudakov–Zhao now follows by bounding the number of trivial solutions to (17) in Sidon sets.

Proof of Theorem 31. We will bound the number of solutions to (17) in S with $x_4 = x_5$, say; the analogous arguments for the remaining nine pairs proceed in the same way. The number of such solutions is

$$\sum_{x_1 + x_2 + x_3 = 3x_4} 1_S(x_1) 1_S(x_2) 1_S(x_3) 1_S(x_4) = N^4 \int_{\mathbf{T}} \widehat{1_S}(\xi)^3 \widehat{1_S}(-3\xi) d\xi$$

by orthogonality of characters. But, by the triangle inequality and Hölder's inequality,

$$\left| \int_{\mathbf{T}} \widehat{1_S}(\xi)^3 \widehat{1_S}(-3\xi) d\xi \right| \leq \|\widehat{1_S}\|_{L^4}^4 \ll \frac{1}{N}$$

since S is a Sidon set. Thus, the number of trivial solutions is $\ll N$. Since the total number of solutions is $\gg_\alpha N^{3/2}$, the conclusion of the theorem follows provided that $N \gg_\alpha 1$. □

It just remains to prove the dense model lemma.

Proof. For this proof, since not every function appearing will be supported on $[N]$, we will un-normalize the Fourier transform and L^2-norm, so that when $g : \mathbf{Z} \to \mathbf{C}$ is finitely supported,

$$\widehat{g}(\xi) := \sum_{n \in \mathbf{Z}} g(n) e(-\xi n),$$

and we will want to prove that our dense model satisfies $\|f\|_{L^2} \ll \sqrt{N}$ and

$$\left\| \widehat{f} - \sqrt{N} \cdot \widehat{1_S} \right\|_{L^\infty} \leq \varepsilon N.$$

We have already defined the notion of the large spectrum of a set in the cyclic group setting, and can make the analogous definition in the integer setting. For any $\delta > 0$, we set

$$\mathrm{Spec}_\delta(S) := \left\{ \xi \in \mathbf{T} : \left| \widehat{1_S}(\xi) \right| \geq \delta |S| \right\}.$$

Let $0 < \delta \ll \varepsilon$ be a parameter to be chosen shortly depending only on ε. Set

$$B := \{ n \in [-\delta N, \delta N] \cap \mathbf{Z} : \|\xi n\| \leq \delta \text{ for all } \xi \in \mathrm{Spec}_\delta(S) \},$$

which is a Bohr set in the interval $[-\delta N, \delta N] \cap \mathbf{Z}$, and define

$$g(x) := \frac{1}{|B|} \sum_{y \in \mathbf{Z}} 1_S(x - y) 1_B(y),$$

a weighted convolution of 1_S and 1_B (note that $0 \in B$, so $|B| > 0$). Since $S \subset [N]$, g is supported on $(-\delta N, (1+\delta)N] \cap \mathbf{Z}$. After a bit of massaging, g will become our desired dense model f.

First, note that if $\xi \notin \mathrm{Spec}_\delta(S)$, then

$$\left| \widehat{1_S}(\xi) - \widehat{g}(\xi) \right| = \frac{1}{|B|} \left| \widehat{1_S}(\xi) \right| \left| |B| - \widehat{1_B}(\xi) \right| \leq 2\delta |S|,$$

and if $\xi \in \mathrm{Spec}_\delta(S)$, then, since $e(\xi n) = 1 + O(\delta)$ whenever $n \in B$, we have $\widehat{1_B}(\xi) = |B| + O(\delta |B|)$

$$\left| \widehat{1_S}(\xi) - \widehat{g}(\xi) \right| = \frac{1}{|B|} \left| \widehat{1_S}(\xi) \right| \left| |B| - \widehat{1_B}(\xi) \right| \ll \delta |S|$$

by bounding $\widehat{1_S}(\xi)$ trivially by $|S|$. Setting $g' := \sqrt{N} g$, we thus have

$$\left\| \widehat{g} - \sqrt{N} \cdot \widehat{1_S} \right\|_{L^\infty} \ll \delta N.$$

Now, expanding the definition of the convolution, note that

$$\begin{aligned} \sum_{n \in \mathbf{Z}} g'(n)^2 &= \frac{N}{|B|^2} \sum_{a_1 + a_2 = a_3 + a_4} 1_S(a_1) 1_B(a_2) 1_S(a_3) 1_B(a_4) \\ &\leq \frac{N}{|B|^2} \left(|S||B| + |B|^2 \right) \leq N \left(\frac{|S|}{|B|} + 1 \right), \end{aligned}$$

where we have used that, since S is Sidon, there is at most one representation $a_1 - a_3 = a_4 - a_2$ of any nonzero integer with $a_1, a_3 \in S$. Thus, $\|g'\|_{L^2} \ll N$ provided that $|B| \gg |S|$. We will verify that this is the case whenever N is sufficiently large in terms of δ (and thus, ε).

We can obtain a lower bound on $|B|$ in a similar manner to how we obtained a lower bound on the size of Bohr sets in cyclic groups. Let $\xi_1, \dots, \xi_R$ be a maximal collection of $\frac{1}{N}$-separated frequencies in $\mathrm{Spec}_\delta(S)$. By maximality, every element of $\mathrm{Spec}_\delta(S)$ is within a distance of $\frac{1}{N}$ of some ξ_i, and thus B contains the Bohr set

$$B' := \{[-\delta N/2, \delta N/2] : \|\xi_i n\| \leq \delta/2 \text{ for all } i \in [R]\}$$

by the triangle inequality. By the same argument we used to obtain a lower bound on the size of Bohr sets in cyclic groups, we have $|B'| \geq (\delta/2)^R (\delta N + 1) \geq (\delta/2)^{R+1} N$, say. To bound R, note that, since $|e(\alpha) - e(\beta)| \leq 2\pi \|\alpha - \beta\|$ for all $\alpha, \beta \in \mathbf{T}$, we have

$$\bigcup_{i=1}^{R} \left(\xi_i - \frac{\delta}{4\pi N}, \xi_i + \frac{\delta}{4\pi N} \right) \subset \mathrm{Spec}_{\delta/2}(S),$$

and so, since the ξ_i's are $\frac{1}{N}$-separated,

$$R \frac{\delta}{2\pi N} \leq m(\mathrm{Spec}_\delta(S)) \leq \frac{1}{(\delta|S|)^4} \int_{\mathbf{T}} \left| \widehat{1_S}(\xi) \right|^4 d\xi \leq \frac{1}{\delta^4 |S|^2} \ll \frac{\delta^{-O(1)}}{N}$$

since S is Sidon. Rearranging yields $R \ll \delta^{-O(1)}$, and so

$$|B| \geq (\delta/2)^{O(\delta^{-O(1)})+1} N \gg_\delta N.$$

To finish, we just need to correct g' to a function supported on $[N]$ by setting $f := 1_{[N]} g'$. This cannot possibly increase the L^2-norm, and so it remains to check that $\widehat{f}$ is still a good approximation for $\sqrt{N}\widehat{1_S}$. Note that, by the Cauchy–Schwarz inequality,

$$\begin{aligned} \left| \widehat{g'}(\xi) - \widehat{f}(\xi) \right| &\leq \left(\sum_{-\delta N < n \leq 0} + \sum_{N < n \leq (1+\delta)N} \right) |g'(n)| \\ &\ll \sqrt{\delta N} \|g'\|_{L^2} \ll \sqrt{\delta} N \end{aligned}$$

for all $\xi \in \mathbf{T}$ provided that $N \gg_\delta 1$. Thus, fixing $0 < \delta \ll \varepsilon^2$,

$$\|\widehat{f} - \sqrt{N}\widehat{1_S}\|_{L^\infty} \leq \varepsilon N$$

and $\|f\|_{L^2} \leq \sqrt{N}$, again provided that $N \gg_\varepsilon 1$. □