arXiv:2302.05537v2 [math.NT] 18 Jun 2023

Strong Bounds for 3-Progressions

Zander Kelley*, Raghu Meka†

Abstract

We show that for some constant $\beta > 0$, any subset A of integers $\{1, \ldots, N\}$ of size at least $2^{-O((\log N)^{\beta})} \cdot N$ contains a non-trivial three-term arithmetic progression. Previously, three-term arithmetic progressions were known to exist only for sets of size at least $N/(\log N)^{1+c}$ for a constant $c > 0$.

Our approach is first to develop new analytic techniques for addressing some related questions in the finite-field setting and then to apply some analogous variants of these same techniques, suitably adapted for the more complicated setting of integers.

1 Introduction

A 3-progression is a triple of integers of the form $(a, a+b, a+2b)$, and it is said to be *nontrivial* if $b \neq 0$. It is straightforward to check that a triple (x, z, y) is of this form if and only if $x + y = 2z$, and that, in this case, the progression is trivial only if $x = y$. We consider the problem of finding a nontrivial 3-progression within a set $A \subseteq [N] \subseteq \mathbb{Z}$, where we assume only that the set A has somewhat large density inside $[N]$. Regarding this problem, we prove the following.

Theorem 1.1. *The following holds for some absolute constant exponent $\beta > 0$. Suppose $A \subseteq [N]$ has density $\delta = |A|/N$. Then, either A contains a nontrivial 3-progression, or else*

$$\delta \leq 2^{-\Omega(\log(N)^{\beta})}.$$

In [Rot53], Roth proved a statement of the same form: that dense enough sets $A \subseteq [N]$ must contain a nontrivial 3-progression, for density threshold $\delta \approx 1/\log\log N$. This was improved by Heath-Brown and Szemerédi to $\delta \approx 1/\log(N)^c$ for some small $c > 0$ [HB87, Sze90]. This bound was further refined in the works of Bourgain, and Sanders [Bou99, Bou08, San12a] where it is shown that one can take $c = 1/2$, $c = 2/3$, and then $c = 3/4$. Sanders then obtained a density-threshold of the form $\delta \approx (\log\log N)^6/\log N$ [San11].[1] This was further sharpened by a factor $(\log\log N)^2$ by Bloom and then again by another factor $\log\log N$ by Schoen [Blo16, Sch21]. The best bound previously available is due to Bloom and Sisask [BS20], who show that a set $A \subseteq [N]$ with no 3-progressions must have density

$$\delta \leq O\left(\frac{1}{\log(N)^{1+c}}\right)$$

*Department of Computer Science, University of Illinois at Urbana-Champaign. Supported by NSF grant CCF-1814788, and CAREER award 2047310. Email: `awk2@illinois.edu`

†Department of Computer Science, University of California, Los Angeles. Supported by NSF AF 2007682. Email: `raghum@cs.ucla.edu.`

[1] See also the work of Bloom and Sisask [BS19], which obtains a comparable result by some different methods which are closely related to the approach taken here.

for some small $c > 0$. We also refer to Appendix A for a more detailed discussion of some previous approaches and how our methods relate to them.

In the other direction, it was shown by Behrend[2] that for infinitely many values N there are indeed subsets $A \subseteq [N]$ of density roughly $\delta \approx 2^{-\log(N)^{1/2}}$ which have no nontrivial 3-progressions.

More specifically, we establish the following.

Theorem 1.2. *Suppose $A \subseteq [N]$ has density at least 2^{-d}. Then the number of triples $(x, y, z) \in A^3$ with $x + y = 2z$ is at least*

$$2^{-O(d^{12})} N^2.$$

Since there are only $|A| \leq N$ trivial 3-progressions, we do indeed obtain a nontrivial 3-progression, unless $\log N \leq O(d^{12})$.

We also consider a similar problem where $A \subseteq \mathbb{F}_q^n$ is a subset of a vector space over some finite field. In this setting, we prove the following.

Theorem 1.3. *Suppose $A \subseteq \mathbb{F}_q^n$ has density at least 2^{-d}. Then the number of triples $(x, y, z) \in A^3$ with $x + y = 2z$ is at least*

$$q^{-O(d^9)} |\mathbb{F}_q^n|^2.$$

As a standalone result, we point out that this theorem is strictly worse than the state-of-the-art bound, which can be obtained from algebraic techniques pioneered by [CLP17] and [EG17], which gave a strong resolution to the cap-set problem. Indeed, Fox and Lovász obtain strong bounds for Green's "arithmetic removal lemma" by combining the algebraic results of [KSS18] with some additional combinatorial arguments. In particular, their results imply the following.

Theorem 1.4 (Special case of Theorem 3 in [FL17])**.** *Suppose $A \subseteq \mathbb{F}_q^n$ has density at least 2^{-d}. Then the number of triples $(x, y, z) \in A^3$ with $x + y = 2z$ is at least $q^{-O(d)} |\mathbb{F}_q^n|^2$.*

Instead, our interest in Theorem 1.3 is that we can obtain it using purely analytic techniques and that these same techniques can then be slightly modified and extended to prove Theorem 1.2. Partially for expository reasons, we devote a substantial portion of this paper to first addressing the finite field case. This will allow us to present most of the key ideas needed to prove our main theorem while ignoring additional complications arising in the setting $A \subseteq \mathbb{Z}$. We also remark that techniques based on additive combinatorics have found several applications within computer science (going all the way back to the paper of [CFL83] who used Behrend's construction for a communication protocol to more recent applications such as [BKM22]). We believe the new analytic techniques will have other applications.

1.1 Structural Results in the Finite Field Setting

This section highlights a structural result in the finite field setting that follows from our analytic techniques but is not known to follow from algebraic methods. To describe it, we recall the notation for sumsets:

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

For a set $A \subseteq \mathbb{F}_q^n$, we also use the notation $\mathrm{Span}^*(A)$ to denote the minimal affine subspace containing A – that is, $\mathrm{Span}^*(A)$ refers to the common intersection of all affine subspaces containing A, which is itself an affine subspace.

[2] [Beh46]. See also the works of Elkin [Elk10], Green and Wolf [GW10], and O'Bryant [O'B11] for some refinements.

The following structural result says that for any dense set $A \subseteq \mathbb{F}_q^n$, there is a reasonably large subset $A' \subseteq A$, which is tightly contained in its own span (in a specific technical sense).

Lemma 1.5 (A large subset tightly contained in its own span)**.** *Fix a parameter $\tau \in \left[0, \frac{1}{2}\right]$. Suppose that $A \subseteq \mathbb{F}_q^n$ has size $|A| \geq 2^{-d}|\mathbb{F}_q^n|$. Then, there is a subset $A' \subseteq A$ with*

i. $\frac{|A'|}{|\mathrm{Span}^*(A')|} \geq \frac{|A|}{|\mathbb{F}_q^n|}$ *and*

ii. $\mathrm{Codim}(\mathrm{Span}^*(A')) \leq O(d^5 \log(1/\tau)^4)$

such that $|A' + A'| \geq (1-\tau)|\mathrm{Span}^*(A')|$.

Taken together, conditions (i) and (ii) roughly say simply that A' is large: for example, they imply that $|A'| \geq q^{-O(d^5 \log(1/\tau)^4)}|\mathbb{F}_q^n|$. However, we are additionally guaranteed (by (i)) that the density of A' inside its span is no worse than the original density of A inside $\mathbb{F}_q^n$. To unpack the conclusion, let us write $\mathrm{Span}^*(A') =: V + \theta$ for some linear subspace V and some shift $\theta \in \mathbb{F}_q^n$. Also let $B := A' - \theta$. Clearly, we have $\mathrm{Span}(B) = V$. That is, B is contained in V, and it also "eventually" generates V: for instance we can surely say that $\{c_1 b_1 + c_2 b_2 \cdots + c_n b_n \,:\, c_i \in \mathbb{F}_q, b_i \in B\} = V$. On the other hand, we can compare this to our conclusion, which says that merely taking

$$B + B = \{b_1 + b_2 \,:\, b_1, b_2 \in B\}$$

is already enough to generate all but a tiny fraction of V: we have $|B + B| \geq (1-\tau)|V|$ and so $|V \setminus (B + B)| \leq \tau |V|$.

We proceed to compare this result with some related results in the literature: Specifically, Sanders' quasipolynomial Bogolyubov–Ruzsa lemma (specialized to the finite field setting) and the critical technical lemma underlying its proof.

Let us first consider what we'll call the *Bogolyubov–Ruzsa Problem*. We recall the notation for sumsets: $2A$ denotes $A + A$ and tA denotes the set of all sums $a_1 + a_2 + \ldots + a_t$ with $a_i \in A$.

Problem 1.6 (Bogolyubov-Ruzsa Problem over $\mathbb{F}_p^n$)**.** *Let $A \subseteq \mathbb{F}_p^n$ be a set of points in a vector space over a prime field. Fix $t \in \mathbb{N}$ and assume that $|A + A| \leq K|A|$. Find (or rather, prove the existence of) an affine subspace V which is as large as possible and contained in tA.*

Note that any sufficiently dense set (i.e., one of size $|A| \geq K^{-1}|\mathbb{F}_p^n|$) trivially satisfies $|A + A| \leq |\mathbb{F}_q^n| \leq K|A|$, so to solve the Bogolyubov-Ruzsa problem one must at least handle the special case of finding a large subspace in tA for all dense sets A. Furthermore, solving the problem in this special case is sufficient to obtain very similar parameters for the general case. The reduction can be summarized as follows: for any $t \leq O(1)$, a set $A \subseteq \mathbb{F}_p^n$ with $|A + A| \leq K|A|$ can be embedded by some linear map ϕ from $\mathbb{F}_p^n$ into $\mathbb{F}_p^m$ in such way that

- ϕ is injective on A (and indeed, even on tA),
- $|\phi(A)| \geq K^{-O(1)}|\mathbb{F}_p^m|$, and
- The set $t\phi(A) = \phi(tA)$ contains an affine subspace $V \subseteq \mathbb{F}_p^m$ only if the preimage $\phi^{-1}(V)$ is an affine subspace in $\mathbb{F}_p^n$.

For a more detailed explanation, see, e.g., the nice description given in [Lov15]. In light of this reduction to the dense case, we restrict our attention to the following.[3]

[3] We note that the reduction from the more general Bogolyubov-Ruzsa problem for vector spaces over prime fields $\mathbb{F}_p$ to the more specific Bogolyubov problem over $\mathbb{F}_p$ uses the fact that subspaces in $\mathbb{F}_p^n$ can be precisely characterized

Problem 1.7 (Bogolyubov Problem over $\mathbb{F}_q^n$)**.** *Let $A \subseteq \mathbb{F}_q^n$. Fix $t \in \mathbb{N}$ and assume that $|A| \geq 2^{-d}|\mathbb{F}_q^n|$. Find an affine subspace V which is as large as possible and contained in tA.*

Regarding this problem, Sanders proves the following.

Theorem 1.8 ([San12b])**.** *Suppose $A \subseteq \mathbb{F}_q^n$ has size $|A| \geq 2^{-d}|\mathbb{F}_q^n|$. Then there is an affine subspace $V \subseteq 4A$ of codimension at most $O(d^4)$ in $\mathbb{F}_q^n$.*

Sanders' approach to the Bogolyubov Problem for $4A$ is first to address what we'll call the *Approximate Bogolyubov Problem* for $2A$:

Problem 1.9 (Approximate Bogolyubov Problem over $\mathbb{F}_q^n$)**.** *Let $A \subseteq \mathbb{F}_q^n$. Fix $t \in \mathbb{N}$ and assume that $|A| \geq 2^{-d}|\mathbb{F}_q^n|$. Find an affine subspace V which is as large as possible and is "mostly" contained in tA, in the sense that*

$$|V \cap (tA)| \geq (1-\tau)|V|$$

for some τ as small as possible.

The connection between the Approximate Bogolyubov Problem for tA and the Exact Bogolyubov Problem for $2tA$ is due to the following simple claim.

Claim 1.10. *Suppose $B \subseteq \mathbb{F}_q^n$ is a set with a large intersection with some linear subspace V: $|B \cap V| > \frac{1}{2}|V|$. Then $B + B$ contains V.*

Proof. Let $x \in V$. We may trivially write $x = (x+y) - y$ for any element $y \in V$. Consider doing so for a uniformly random choice of $y \in V$. By a union bound, the chance that the events $x + y \in B$ and $-y \in B$ occur simultaneously is nonzero. □

Sanders provides the following solution to the Approximate Bogolyubov Problem for sumsets $2A$ over $\mathbb{F}_q^n$.

Theorem 1.11 ([San12b])**.** *Suppose $A \subseteq \mathbb{F}_q^n$ has size $|A| \geq 2^{-d}|\mathbb{F}_q^n|$. Then for any $\tau \geq 2^{-d}$, there is an affine subspace V of codimension at most $O(d^4/\tau^2)$ with*

$$|V \cap (A + A)| \geq (1-\tau)|V|.$$

It is useful to compare this with our structural result above, Lemma 1.5, which accomplishes something similar in spirit. Indeed, given A we may find $A' \subseteq A$ and consider its container $V := \mathrm{Span}^*(A')$, which in our case has codimension $O(d^5 \log(1/\tau)^4)$. We note that $2V$ is again an affine subspace of the same size as V ($2V$ is just a translate of V) which contains $A' + A'$, and that

$$|2V \cap (A + A)| \geq |2V \cap (A' + A')| = |A' + A'| \geq (1-\tau)|2V|.$$

Thus, our result can also be understood as a solution to the Approximate Bogolyubov Problem for $2A$; compared with Sanders' solution, we obtain codimension which is moderately worse in terms of its dependence of the density parameter d in exchange for substantially improved dependence on the error parameter τ. This answers a question posed by Hatami, Hosseini, and Lovett in [HHL18] – they ask whether it is possible to obtain codimension $d^{O(1)} \log(1/\tau)^{O(1)}$ for the Approximate Bogolyubov Problem for $2A$, and we confirm that indeed it is.

as those subsets which are closed under addition – i.e., additive subgroups. However, it is sensible to consider the Bogolyubov problem itself over general finite fields $\mathbb{F}_q$.

We can also answer a related question posed by Schoen and Sisask in Section 9 of [SS16]: they ask whether it is possible to obtain codimension $d^{O(1)}$ for the Exact Bogolyubov Problem for $3A$, as Sanders does for $4A$. The following result confirms this.

Corollary 1.12 (Direct corollary of Lemma 1.5)**.** *Suppose $A \subseteq \mathbb{F}_q^n$ has size $|A| \geq 2^{-d}|\mathbb{F}_q^n|$. Then, there is a subset $A' \subseteq A$ with size*

$$|A'| \geq q^{-O(d^9)}|\mathbb{F}_q^n|$$

such that $A' + A' + A'$ is an affine subspace.

Proof. We apply Lemma 1.5 with parameter τ set slightly smaller than 2^{-d}. We consider the resulting subset A' and affine subspace $\mathrm{Span}^*(A')$, which we write as $V + \theta$ for some linear subspace V and some shift θ. Consider the set $B := A' - \theta \subseteq V$. We argue that $3B = V$, which proves that $3A' = V + 3\theta$. Indeed, if $3B \neq V$, then there is some $x \in V$ such that the set $x - B$ does not intersect with $B + B$. However, $x - B$ is a set in V of size $|x - B| = |B| \geq 2^{-d}|V|$, and we also have $|B + B| \geq (1 - \tau)|V|$, so this is not possible. □

1.2 A robust structural lemma

The structural result given in the previous section is in fact a special case of a more analytically robust variant which we proceed to describe. We recall the notation

$$R_A(x) := \left|\left\{(a_1, a_2) \in A^2 \ : \ a_1 + a_2 = x\right\}\right|,$$

which counts the number of "representations" of x as a sum of two elements of A. We also consider the alternative normalization

$$r_A(x) := \frac{R_A(x)}{|A|^2}$$

which is normalized as a distribution function on $\mathbb{F}_q^n$ – that is, $\sum_{x \in \mathbb{F}_q^n} r_A(x) = 1$. Recall that Lemma 1.5 says that for any dense set $A \subseteq \mathbb{F}_q^n$, there is a reasonably large subset $A' \subseteq A$ which is tightly contained in its own span, in the sense that

$$|A' + A'| \geq (1 - \tau)|\mathrm{Span}^*(A')|.$$

For the sake of clarity let us write $\mathrm{Span}^*(A') = V + \theta$ where V is a linear subspace. We have of course that $|V| = |\mathrm{Span}^*(A')|$. Note that the distribution $r_{A'}(x)$ is supported on the affine subspace $2\mathrm{Span}^*(A') = V + 2\theta$. Our robust variant of Lemma 1.5 states that in fact the distribution $r_{A'}(x)$ is very close to the uniform distribution on $V + 2\theta$. The fact that the support of $r_{A'}$, i.e. $A' + A'$, is large follows as a corollary.

To state this result we need the following notion for measuring closeness of distributions. In what follows, a "distribution" $\pi : \Omega \to \mathbb{R}$ is a nonnegative function on Ω with $\sum_{x \in \Omega} \pi(x) = 1$.

Definition 1.13. *Given $k \geq 1$ and two distributions $\pi, \pi' : \Omega \to \mathbb{R}$, define the k-norm divergence of π from π' as the quantity*

$$\left(\frac{\sum_{x \in \Omega} |\pi(x) - \pi'(x)|^k}{\sum_{x \in \Omega} \pi'(x)^k}\right)^{1/k}.$$

In the case that π' corresponds to some flat distribution $\pi'(x) = \frac{\mathbb{1}_S(x)}{|S|}$, uniform over some subset $S \subseteq \Omega$, and π is supported on S, this can be equivalently expressed as

$$\left(\mathop{\mathbb{E}}_{x \in S} \left|\frac{\pi(x)}{\pi'(x)} - 1\right|^k\right)^{1/k}.$$

Lemma 1.14 (A robust variant of Lemma 1.5)**.** *Fix a constant $\varepsilon \in (0,1)$ and a parameter $k \geq 1$. Suppose that $A \subseteq \mathbb{F}_q^n$ has size $|A| \geq 2^{-d}|\mathbb{F}_q^n|$. Then, there is a subset $A' \subseteq A$ with*

i. $\frac{|A'|}{|\mathrm{Span}^*(A')|} \geq \frac{|A|}{|\mathbb{F}_q^n|}$ *and*

ii. $\mathrm{Codim}(\mathrm{Span}^*(A')) \leq O_\varepsilon(d^5k^4)$

such that the distribution $r_{A'}(x)$ has k-norm divergence at most ε from the uniform distribution on $2\mathrm{Span}^*(A')$.

We note that Lemma 1.5 does indeed follow as a special case by invoking Lemma 1.14 e.g. with $\varepsilon = 1/2$. This is because, if π is a distribution supported on some set $B \subseteq B'$, and $|B' \setminus B| \geq \tau|B'|$, then the k-norm divergence from π to the uniform distribution on B' is at least $\tau^{1/k}$.

2 Technical introduction

The techniques and applications considered in this work are centered around the following question. Let A, B, and C be subsets of a finite abelian group G of size N, each with size at least $2^{-d}N$. We ask: what sort of generic "pseudorandom" conditions on the sets A, B, C are sufficient to ensure that the number of solutions to the equation $a + b = c$, with $(a, b, c) \in A \times B \times C$, is off by only a small multiplicative factor $(1 \pm \varepsilon)$ from the the "expected" number, $|A||B||C|/N$?

We are particularly concerned with the following three conditions on a set $A \subseteq G$, which each attempts to quantify the "additive pseudorandomness" of A in some capacity. For concreteness, we will first state the three conditions – spreadness, regularity, and self-regularity – in the counting measure. However, we will shortly return to restate each of them using alternative normalization conventions which will greatly increase their interpretability. We recall some standard notations for the number of "representations" of x as a sum or difference of elements of some sets A and B: $R_{A,B}(x) := |\{(a,b) \in (A \times B) \,:\, a + b = x\}|$, $R^-_{A,B}(x) := |\{(a,b) \in (A \times B) \,:\, a - b = x\}|$, $R_A(x) := R_{A,A}(x)$, and $R^-_A(x) := R^-_{A,A}(x)$.

Definition 2.1 ((γ, r)-spread)**.** *Let A be a subset of $\mathbb{F}_q^n$. For $\gamma, r \geq 1$, we say A is (γ, r)-spread if,*

$$\frac{|A \cap V|}{|V|} \leq \frac{\gamma|A|}{|\mathbb{F}_q^n|}.$$

for all affine subspaces $V \subseteq \mathbb{F}_q^n$ of codimension at most r.

Our first condition – spreadness – applies only to the case $G = \mathbb{F}_q^n$; however, we will eventually return to consider possible substitutes for when we are working in a different group.[4] Spreadness is important to us because it is *easy to obtain*, for instance, by a density-increment argument.

Definition 2.2 ((γ, k)-regular)**.** *Let A be a subset of a finite abelian group G. For $\gamma, k \geq 1$, We say that A is (γ, k)-regular if*

$$\sum_{a \in A} R_{B,C}(a) \leq \frac{\gamma|A||B||C|}{|G|}$$

for all sets $B, C \subseteq G$ with size at least $|B|, |C| \geq 2^{-k}|G|$.

[4] Specifically, we will consider some alternatives where a Bohr set, or a generalized arithmetic progression, instead plays the role of the subspace.

Definition 2.3 ((γ, k)-self-regular)**.** *We say that $A \subseteq G$ is (γ, k)-self-regular if*

$$\sum_{x \in G} R_A^-(x)^k \leq \frac{\gamma |A|^{2k}}{|G|^{k-1}}.$$

Our latter two conditions – regularity and self-regularity – apply to general finite abelian groups. In contrast to spreadness, regularity is important to us because it is *useful to have*, although it is not clear that it is easy to obtain.

We can summarize the overall argument for our structural result, Lemma 1.14, into three steps:

1. One can use density increments to obtain spreadness. This follows from a simple greedy argument and is formalized in Proposition 2.15.
2. One can (informally speaking) obtain regularity from spreadness. This is a core ingredient of our approach and is formalized in Theorem 2.8.
3. One can obtain strong *two-sided* bounds on the number of solutions to $a + b = c$ from regularity. This is formalized in Theorem 2.11.

Points 2 and 3 above can be considered the two main technical contributions of our work. Taken together, they give the following answer to the central question described above (for the setting $G = \mathbb{F}_q^n$). To control the number of solutions to $a + b + c = 0$ with $a, b, c \in A, B, C$, to within a factor $(1 \pm O(\varepsilon))$ of the expected number $|A||B||C|/|G|$, it suffices that any two of the three sets are $(1 + \varepsilon, \mathrm{poly}(d, 1/\varepsilon))$-spread.

2.1 Density formulation

We take a moment to switch to a more "analytic" language for expressing the three conditions introduced above which both eases their interpretability and clarifies their importance for controlling the number of solutions to $a + b = c$. Here in the introduction, we avoid an extended discussion of details regarding, e.g., normalization conventions (a more detailed description can be found in Section 3). However, we highlight the following key points.

- We use the notation $\|f\|_k := \left(\mathbb{E}_{x \in G} |f(x)|^k\right)^{1/k}$ for $k \geq 1$ and functions $f : G \to \mathbb{R}$, as well as $\|f\|_\infty := \max_{x \in G} |f(x)|$.
- We consider *density functions* on G, which are simply nonnegative functions $F : G \to \mathbb{R}$ with $\|F\|_1 = 1$.
- Given a set $A \subseteq G$, we abuse notation and let $A(x)$ denote the *density function* which corresponds to the uniform density on A. Under this normalization, a set of size $|A| = 2^{-d}|G|$ has corresponding density function $A(x)$ with $\|A\|_\infty = 2^d$.
- Given density functions $A(x)$ and $B(x)$, we consider both the *convolution* $(A * B)(x)$ and the *cross-correlation* $(A \star B)(x)$, both of which are again density functions. Indeed, they are the density functions that are proportional to the representation-counting functions $R_{A,B}(x)$ and $R_{B,A}^-(x)$.

Finally, let us introduce the following notations relevant to capturing the spreadness and regularity conditions. For nonnegative $f : \mathbb{F}_q^n \to \mathbb{R}$, we define

$$\|f\|_{\perp,r} := \max_{\substack{V \subseteq \mathbb{F}_q^n \\ \mathrm{Codim}(V) \leq r}} \langle V, f \rangle, \tag{1}$$

where the maximization is quantified over affine subspaces $V \subseteq \mathbb{F}_q^n$ of codimension at most r. For nonnegative $f : G \to \mathbb{R}$, we define

$$\|f\|_{*,k} := \max_{\substack{B,C \subseteq G \\ \|B\|_\infty, \|C\|_\infty \leq 2^k}} \langle B * C, f \rangle \,, \tag{2}$$

where the maximization is quantified over subsets $B, C \subseteq G$ of sizes $|B|, |C| \geq 2^{-k}|G|$. We can restate each of our three conditions above in the density formulation.

Definition 2.4 (Spreadness – density formulation)**.** *We say that $A \subseteq \mathbb{F}_q^n$ is (γ, r)-spread if*

$$\|A\|_{\perp,r} \leq \gamma.$$

Definition 2.5 (Regularity – density formulation)**.** *We say that A is (γ, k)-regular if*

$$\|A\|_{*,k} \leq \gamma.$$

Definition 2.6 (Self-regularity – density formulation)**.** *We say that A is (γ, k)-self-regular if*

$$\|A \star A\|_k \leq \gamma.$$

Generally, for all three conditions, we are most interested in the regime where $\gamma = 1 + \varepsilon$ and ε is a small constant.

2.2 Regularity from spreadness; Upper-bounds from regularity

We can also restate our central question in the density formulation as follows.

Question 2.7. *Let $A, B, C \subseteq G$. Suppose that the corresponding density functions satisfy the (min-)entropy deficit condition $\|A\|_\infty, \|B\|_\infty, \|C\|_\infty \leq 2^d$. What sort of generic pseudorandom conditions on $A(x), B(x), C(x)$ are sufficient to ensure that*

$$\langle A * B, C \rangle = 1 \pm O(\varepsilon),$$

where ε is some small constant?

We note that (basically by definition) imposing the condition of $(1 + \varepsilon, d)$-regularity on the set A is sufficient to obtain good *upper bounds* for this question (and also necessary if we wish to avoid imposing any further conditions on the sets B and C beyond the given constraints on their sizes) as we have $\langle A * B, C \rangle = \langle A, B \star C \rangle$.

Thus, if this quantity is at least $1 + \varepsilon$ then the fact that $\|A\|_{*,d} \geq 1 + \varepsilon$ is witnessed by the convolution of C with $-B$. Our first main technical result says that spread sets are regular.

Theorem 2.8 (Structure vs. Pseudorandomness in $\mathbb{F}_q^n$ – simplified combination of Theorem 4.6 and Theorem 4.10)**.** *Suppose $A \subseteq \mathbb{F}_q^n$ has size $|A| \geq 2^{-d}|\mathbb{F}_q^n|$. If $\|A\|_{*,k} \geq 1 + \varepsilon$, then $\|A\|_{\perp,r} \geq 1 + \frac{\varepsilon}{8}$ for some $r \leq O_\varepsilon(k^7 d)$. Similarly, if $\|A \star A\|_k \geq 1 + \varepsilon$, then $\|A\|_{\perp,r} \geq 1 + \frac{\varepsilon}{4}$. for some $r \leq O_\varepsilon(k^4 d^4)$.*

We can offer a couple of different interpretations of this result. Firstly we can interpret it analytically as follows. Suppose the density function $A(x)$ has a deviation upwards from the uniform density function, which can be detected, or "witnessed", by some convolution of two large sets (i.e.,

by a density corresponding to a rather weak additive structure). Then, the above theorem says that this deviation can be witnessed almost as well by the uniform density over a reasonably large affine subspace (i.e., a set with a very strict additive structure). Alternatively, we can interpret the contrapositive in the well-known structure vs. pseudorandomness paradigm. Specifically, we have the following dichotomy: a given set A is either *pseudorandom*, in the sense that it is $(1+\varepsilon, k)$-regular, or it has some non-negligible amount of additive structure which can be detected by some large affine subspace V,

$$\langle V, A\rangle \geq 1 + \frac{\varepsilon}{8},$$

and we obtain the resulting *density increment* of A onto V:

$$\frac{|A \cap V|}{|V|} \geq \left(1 + \frac{\varepsilon}{8}\right) \frac{|A|}{|\mathbb{F}_q^n|}.$$

We give the following quick summary of the proof of Theorem 2.8 which is intended for experts. Alternatively, a more detailed overview can be found in Appendix B, which in particular gives some additional discussion of the tools involved, as well as an account of how the proof is arrived at "naturally" as the result of a sequence of observations.

For brevity, we discuss only the most important claim made in Theorem 2.8: that $\|A \star A\|_k \geq 1 + \Omega(1)$ implies $\|A\|_{\perp,r} \geq 1 + \Omega(1)$ for some $r \leq O(k^4 d^4)$. We will establish the stronger claim $\|A \star A\|_{\perp,r} \geq 1 + \Omega(1)$, which is sufficient. Our starting point is the following consequence of Sanders' invariance lemma.

Proposition 2.9. *Let $k \geq 1$ and $\varepsilon \in [2^{-k}, 1]$. For any bounded function $f : \mathbb{F}_q^n \to [0,1]$, we have*

$$\|f\|_{\perp,k^4/\varepsilon^2} \geq \|f\|_{*,k} - O(\varepsilon).$$

Suppose $\|A \star A\|_k \geq 1 + 2\varepsilon$ for some small constant ε, and consider the super-level set indicator

$$f(x) := \mathbb{1}((A \star A)(x) \geq 1 + \varepsilon).$$

Our plan is to argue that $\|f\|_{*,k'} \geq 1 - \frac{\varepsilon}{8}$ for some $k' \leq O(kd)$. From this it follows that

$$\|A \star A\|_{\perp,O(k^4d^4)} \geq (1+\varepsilon)\|f\|_{\perp,O(k^4d^4)} \geq (1+\varepsilon)(1 - \tfrac{\varepsilon}{4}) \geq 1 + \Omega(1),$$

as desired. To argue that we indeed have $\|f\|_{*,k'} \approx 1$ for some value k' which is not too large, we develop the following tool (which is applicable more generally to any finite abelian group G).

Lemma 2.10 (Sifting lemma)**.** *Suppose $A \subseteq G$ has size $|A| = \delta|G|$. Fix a nonnegative function $h : G \to \mathbb{R}_{\geq 0}$ and an integer $k \geq 2$. Let $D(x)$ denote the unique density function which is proportional to $(A \star A)(x)^k$, and suppose that $\langle D, h\rangle = \eta$.*

Then there is a subset $A' \subseteq A$ with $\langle A' \star A', h\rangle \leq 2 \cdot \eta$ and $\frac{|A'|}{|G|} \geq \frac{1}{2} \cdot \delta^k$.

Applying this to $h := 1 - f$ shows that $\|f\|_{*,O(kd)} \approx 1$. This completes the argument; it remains only to prove the sifting lemma and work out the quantitative details – this is done in Section 4.1. We remark briefly that the sifting lemma can be seen as an extension of Schoen's "Pre-BSG" lemma – this is discussed in more detail in Appendix B.3.

2.3 Strong two-sided bounds from self-regularity

We have seen that spreadness is sufficient to give good *upper bounds* for our central question, Question 2.7. However, for applications, typically, one is more interested in *lower bounds* – in particular, in settings where we would like to establish the existence of solutions. For this we turn to the following: our second main technical contribution, which says that if A and B are both self-regular, then $A * B$ is near-uniform.

Theorem 2.11 (Strong two-sided bounds from self-regularity)**.** *Let A, B be subsets of a finite abelian group G. Suppose that A and B are both $(1+\varepsilon, \lceil k/\varepsilon \rceil)$-self-regular for some $\varepsilon \in [0, \frac{1}{4}]$ and some even integer $k \geq 2$. Then*

$$\|A * B - 1\|_k \leq 2\varepsilon.$$

In particular, $\langle A * B, C \rangle = 1 \pm O(\varepsilon)$ for any set C of size $|C| \geq 2^{-k}|G|$.

Proof. Our first step is to apply the following claim, which can be proved without much trouble by a slightly nontrivial application of Cauchy–Schwarz.

Proposition 2.12 (Decoupling inequality)**.** *Let A, B be density functions on a finite abelian group G. For even integers $k \in \mathbb{N}$ we have*

$$\|A * B - 1\|_k \leq \|A \star A - 1\|_k^{1/2} \|B \star B - 1\|_k^{1/2}.$$

So, it suffices to show that both $\|A \star A - 1\|_k \leq 2\varepsilon$ and $\|B \star B - 1\|_k \leq 2\varepsilon$; without loss of generality let us consider A. We argue that the near-uniformity of $A \star A$ follows from an upper bound on $\|A \star A\|_{k'}$ (with $k' := \lceil k/\varepsilon \rceil$). This follows by combining two simple claims.

Proposition 2.13 (Positive correlation for spectrally positive functions – restatement of Corollary 5.5)**.** *Let G be a finite abelian group, and let $f_1, f_2, \ldots, f_t$ be some real-valued functions on G which are "spectrally positive." That is, each of the Fourier coefficients $\widehat{f_i}(\alpha)$ is real and non-negative. Then, for uniformly random $x \in G$,*

$$\mathbb{E}[f_1(x) f_2(x) \cdots f_t(x)] \geq \mathbb{E}[f_1(x)]\, \mathbb{E}[f_2(x)] \cdots \mathbb{E}[f_t(x)] \geq 0.$$

Proposition 2.14 (Odd moments – reformulation of Proposition 5.7)**.** *Let Z be a real-valued random variable. We use the notation $\|Z\|_k := \mathbb{E}\left[|Z|^k\right]^{1/k}$. Suppose Z has non-negative odd moments: that is, $\mathbb{E}\left[Z^t\right] \geq 0$ for all odd integers $t \in \mathbb{N}$. If $\|1 + Z\|_{k'} \leq 1 + \varepsilon$ for some integer k' and some $\varepsilon \in [0, \frac{1}{4}]$, then $\|Z\|_k \leq 2\varepsilon$ for any even integer $k \leq \varepsilon k'$.*

Our density function $(A \star A)(x)$ is indeed spectrally positive, and so is its centering, $F := A \star A - 1$. From this we see that $A \star A$ has non-negative odd central moments:

$$\mathbb{E}\left((A \star A)(x) - 1\right)^t = F(x)^t \geq 0$$

for all odd $t \in \mathbb{N}$. Since we have assumed the upper bound $\|A \star A\|_{k'} \leq 1 + \varepsilon$, we obtain the desired two-sided bound $\|A \star A - 1\|_k \leq 2\varepsilon$. This proves the first claim. For the second claim, we can write

$$\langle A * B, C \rangle = \langle 1, C \rangle + \langle A * B - 1, C \rangle = 1 + \langle A * B - 1, C \rangle,$$

and use a Hölder inequality to estimate

$$|\langle A * B - 1, C\rangle| \leq \|A * B - 1\|_k \|C\|_{1+\frac{1}{k-1}} \leq \|A * B - 1\|_k \|C\|_\infty^{1/k} \|C\|_1$$
$$= \|A * B - 1\|_k \|C\|_\infty^{1/k} \leq 2\|A * B - 1\|_k. \quad \square$$

We remark that it is always possible, for generic density functions D, to infer some kind of bound on $\|D - 1\|_k$ from an upper bound $\|D\|_k \leq 1 + \varepsilon$. However, this generic bound degrades rapidly as k increases: consider, for example, a density function which is uniform over some subset of G of size $(1 - \varepsilon)|G|$. In this case we have

$$\|D\|_{k'} \leq \|D\|_\infty = \frac{1}{1 - \varepsilon} \approx 1 + \varepsilon$$

for any k', and yet

$$\|D - 1\|_k \geq \varepsilon^{1/k}.$$

So, for our arguments above, it was quite important that there are no self-convolutions which are shaped in such a way. More specifically: self-convolutions $A \star A$ which have deviations downwards from 1 must also have deviations upwards from 1 of comparable strength, as measured by the k-norm, for (roughly speaking) any choice of k. This is what allows us to efficiently convert our upper bound from Theorem 2.8 into a lower bound. It is interesting to contrast this situation with the discussion surrounding Question A.3 in the appendix. Specifically, we discuss a formal setting where one can obtain strong upper-control on the quantity $\langle A * B, C\rangle$ from the density-increment method, but it is impossible to obtain any nontrivial lower bound $\langle A * B, C\rangle > 0$.

A review of all the steps. We recall that, as a result of the combination of all of our observations, one can infer from the starting assumption $\langle A * B, C\rangle \leq 1 - \Omega(1)$ that either A or B must have a density increment onto some large affine subspace V. Now that we have seen everything needed for this, we offer the following summary listing all the steps in one place.

- Infer by Hölder that $\|A * B - 1\|_k \geq \Omega(1)$, for $k \geq d$, as witnessed by C.
- Argue that k-norm distance to uniform is maximized by considering self-convolutions: $\|A * B - 1\|_k \leq \|A \star A - 1\|_k^{1/2} \|B \star B - 1\|_k^{1/2}$. Conclude that either $\|A \star A - 1\|_k \geq \Omega(1)$ or $\|B \star B - 1\|_k \geq \Omega(1)$; Suppose $\|A \star A - 1\|_k \geq \Omega(1)$.
- Argue by spectral positivity that the presence of deviations of $A \star A$ downwards from 1 entails the presence of deviations upwards from 1 of comparable strength. More concretely: argue that (upon increasing k slightly) we have $\|A \star A\|_k \geq 1 + \Omega(1)$.
- Use sifting to find a convolution $A' \star A'$ witnessing $\|f\|_{*,O(dk)} \approx 1$, where

$$f := \mathbb{1}(A \star A \geq 1 + \Omega(1)).$$

- Use Sanders' invariance lemma (which is itself powered by the Croot-Sisask lemma and Chang's inequality) to deduce that $\|f\|_{\perp,O(d^4k^4)} \approx 1$.
- Conclude that $\langle V, A \star A\rangle \geq 1 + \Omega(1)$ for some V.
- Conclude that $\langle V', A\rangle \geq 1 + \Omega(1)$ for some $V' = V + a$.

Let us now outline our plan for the remainder of the paper.

In Section 2.4, we discuss the well-known density-increment framework and describe how it can be used to find a large subset $A' \subseteq A$ which is spread relative to its span. We use this to complete the proof our main result for finite field setting: the structural lemma, Lemma 1.14.

In the middle part of the paper, we develop our main tools used to prove Theorem 2.8 and Theorem 2.11: "sifting" and "spectral positivity," respectively. Both tools apply to general finite abelian groups G. We also establish some "local" variants of these techniques in preparation for later when we will consider subsets $A \subseteq \mathbb{Z}$, as well as subsets of cyclic groups.

- In Section 3, we go over some definitions and conventions related to normalization, convolutions, and Fourier analysis.
- In Section 4, we prove the "sifting lemma" (Lemma 4.4) and use is to finish the proof of Theorem 2.8.
- In Section 5, we develop the simple tools needed to prove Theorem 2.11. These include various k-norm inequalities for convolutions and self-convolutions, the latter of which relies on the notion of spectral positivity.

In the final part of the paper, we consider the setting $A \subseteq [N] \subseteq \mathbb{Z}$.

- In Section 6, we give an overview of our plan to establish the existence of many 3-progressions in the integer case, discussing what changes must be made to our approach which handles the finite field case. Most notably, we consider some potential replacements for the notion of "spreadness", where some kind of approximate subgroup, such as a generalized arithmetic progression of bounded rank or a Bohr set of bounded rank, instead plays the role of the subspace of bounded codimension.
- In Section 7, we go over some definitions and basic properties related to generalized progressions, Bohr sets, and Freiman homomorphisms. We also introduce a potentially new device related to the notion of a Freiman homomorphism which we call a "safe" set.
- In Section 8, we complete our proof that dense sets $A \subset [N]$ have many 3-progressions, which relies on tools including sifting, spectral positivity, safe sets, as well as a translation-invariance lemma which is due to Schoen and Sisask.

2.4 The density-increment framework: completing the proof of Lemma 1.14

In this section we prove our main result for the finite field setting by combining Theorem 2.8 and Theorem 2.11 with the well-known density-increment approach.

Let us argue that spreadness is a property that is easy to obtain. Before we begin, we discuss the notion of relative spreadness. For a set $A \subseteq \mathbb{F}_q^n$ contained within some linear subspace V of codimension s, let us say that A is spread relative to V if, upon embedding

$$A \subseteq V \cong \mathbb{F}_q^{n-s},$$

the set is spread in the sense of Definition 2.1. Additionally, suppose that A is contained in some affine subspace V', which can be uniquely described as $V' = V + \theta$ for some linear subspace V and some $\theta \in V^{\perp}$. Let us say that A is spread relative to V' if $A - \theta$ is spread relative to V.

Now, given a set $A \subseteq \mathbb{F}_q^n$ of density at least 2^{-d}, we can find a large set A' which is spread (relative to the ambient space $\mathrm{Span}^*(A')$) by the following simple greedy algorithm. Let $A_0 := A$. We proceed to describe a nested sequence of subsets $A_i \subseteq A_{i-1}$; let $V_i = \mathrm{Span}^*(A_i)$. If the current set A_i is not $(1+\varepsilon, r)$ spread (relative to its container V_i), then there must be an affine subspace $V' \subseteq V_i$, which has codimension at most r inside V_i, such that

$$\frac{|A_i \cap V'|}{|V'|} > (1+\varepsilon)\frac{|A_i|}{|V_i|}.$$

We then pass to the subset $A_{i+1} := A_i \cap V'$. Since we begin with a set with density $\delta_0 \geq 2^{-d}$, and the density of A_i in V_i increases by at least a factor $(1+\varepsilon)$ on every iteration, we must have density

$$\delta_i \geq (1+\varepsilon)^i \cdot 2^{-d}.$$

after i iterations. However, density cannot exceed 1 at any point. Thus, in the case of $\varepsilon \in [0,1]$, this process must terminate within some number $t \leq d/\varepsilon$ iterations; by definition, this means that the final set A_t is $(1+\varepsilon, r)$-spread inside its own span. To summarize, by this simple argument, we have proved the following.

Proposition 2.15. *Let $\varepsilon \in [0,1]$. Suppose $A \subseteq \mathbb{F}_q^n$ has size $|A| \geq 2^{-d}|\mathbb{F}_q^n|$. Then, for some linear subspace $V \subseteq \mathbb{F}_q^n$ of codimension $s \leq rd/\varepsilon$, and some shift $\theta \in V^\perp$, the subset $A' := A \cap (V+\theta)$ satisfies the following:*

1. $\frac{|A'|}{|V|} \geq \frac{|A|}{|\mathbb{F}_q^n|}$, *and*
2. *The set $A' - \theta \subseteq V$ is $(1+\varepsilon, r)$-spread inside $V \cong \mathbb{F}_q^{n-s}$ in the sense of Definition 2.1.*

We remark that it is somewhat more common to frame the structure vs. pseudorandomness approach as a sort of "branching process", where at each step, our current set is either pseudorandom and therefore (possibly by a lengthy argument) satisfactory, or we find a density increment and then repeat this argument recursively. Inspired by [BK12] and [ALWZ20], we find it clarifying to insist that we do all of our density-incrementing up front and then argue that the resulting (spread) set must be satisfactory. Ultimately, though, there does not seem to be a tangible advantage arising from either viewpoint.

Next, we argue that spreadness of A and B implies near-uniformity of $A * B$ (simply by combining Theorem 2.8 with Theorem 2.11).

Proposition 2.16 (Near-uniformity from spreadness)**.** *We have the following for some absolute constant $c \geq 1$. Suppose $A, B \subseteq \mathbb{F}_q^n$ are two sets each of size at least $2^{-d}|\mathbb{F}_q^n|$, where $d \geq 1$. Let $r \in \mathbb{N}$, $\varepsilon \in [0, \frac{1}{4}]$, and $k \geq 1$ be such that*

$$r \geq d^4 k^4/\varepsilon^c.$$

If $\|A\|_{\perp,r} \leq 1+\varepsilon$ and $\|B\|_{\perp,r} \leq 1+\varepsilon$, then

$$\|A * B - 1\|_k \leq O(\varepsilon).$$

Proof. We argue by the contrapositive. Let $\eta \in [0,1]$, and suppose that

$$\|A * B - 1\|_k \geq \eta.$$

By Theorem 2.11, this means that either $\|A \star A\|_{k'} \geq 1 + \Omega(\eta)$ or $\|B \star B\|_{k'} \geq 1 + \Omega(\eta)$ for some $k' \leq O(k/\eta)$; without loss of generality suppose $\|A \star A\|_{k'} \geq 1 + \Omega(\eta)$. We use Theorem 2.8 (or rather, the more specific Theorem 4.6) to infer that

$$\|A\|_{\perp,r} \geq 1 + \Omega(\eta)$$

for some $r \leq M(\eta) \cdot d^4 k^4$, where $M(\eta) \leq \mathrm{poly}(1/\eta)$. Since $\eta \in [0,1]$ was arbitrary, the result follows. □

By combining the two claims above we easily obtain our structural lemma stated in the introduction, Lemma 1.14.

Proof of Lemma 1.14. We are given a constant ε, a parameter $k \geq 1$, and a set $|A| \geq 2^{-d}|\mathbb{F}_q^n|$. Changing ε and d only slightly, we may assume $d \geq 1$ and $\varepsilon \leq 1/4$.

We set a parameter $r \in \mathbb{N}$ to be some constant factor larger than $k^4 d^4/\varepsilon^c$. We invoke Proposition 2.15 and consider the subset $A' = A \cap (V + \theta)$ which is $(1+\varepsilon, r)$-spread relative to its span, $V + \theta$, an affine subspace with codimension at most $rd/\varepsilon = O(d^5 k^4/\varepsilon^{c+1})$ in $\mathbb{F}_q^n$. By our choice of r, we may apply Proposition 2.16. So, the distribution

$$r_{A'-\theta}(x) := \frac{R_{A-\theta}(x)}{|A|^2}$$

has k-norm divergence from the uniform distribution on V bounded by $O(\varepsilon)$. Equivalently, $r_{A'}(x)$ has $O(\varepsilon)$ divergence from the uniform distribution on $V + 2\theta$. Since $\varepsilon \in [0, \frac{1}{2}]$ was arbitrary the result follows. □

Furthermore, our lower bound for 3-progressions in the finite field setting also follows as a direct consequence of Lemma 1.14.

Proof of Theorem 1.3 from Lemma 1.14. Suppose $A \subseteq \mathbb{F}_q^n$ has size $|A| \geq 2^{-d}|\mathbb{F}_q^n|$. If q is even, we clearly have $|A|^2$ solutions to $x + y = 2z$, since $2z = 0$ for all z. So, let us assume that q is odd. We seek to lower bound the number of 3-progressions in A, which is to say the quantity

$$\sum_{z \in A} R_A(2z).$$

Our plan is to invoke Lemma 1.14, with $\varepsilon = 1/4$ and $k = d + 1$, to obtain a nice subset A', and then we will count only the solutions to $x + y = 2z$ with $x, y, z \in A'$, and ignore the rest. For the sake of clarity let us write $\mathrm{Span}^*(A) = V + \theta$ where V is a linear subspace, and consider instead the translated set $B := A' - \theta \subseteq V$, noting that translating an entire set does not change the number of solutions to $x + y = 2z$ with x, y, z in the set.

Lemma 1.14 tells us that the distribution $r_B(x)$ has k-norm divergence at most $1/4$ from the uniform distribution on V. This in particular means that the number of points $x \in V$ where

$$r_B(x) \leq \frac{1}{2|V|}$$

is at most $2^{-k}|V|$. Now using the fact that q is odd, we note that the dilation map $z \mapsto 2z$ is a permutation of V, and so

$$\sum_{z \in B} r_B(2z) \geq \left(|B| - 2^{-k}|V|\right) \frac{1}{2|V|} \geq \frac{2^{-d}}{4},$$

where we have used that fact that $|B| \geq 2^{-d}|V|$ and that $k = d+1$. Overall, we see that

$$\begin{aligned}\sum_{z\in A} R_A(2z) &\geq \sum_{z\in B} R_B(2z)\\ &\geq 2^{-O(d)}|B|^2\\ &\geq q^{-O(d^5k^4)}|\mathbb{F}_q^n|^2\\ &= q^{-O(d^9)}|\mathbb{F}_q^n|^2.\end{aligned}$$

□

3 Preliminaries

3.1 Densities and normalization

Definition 3.1. *For an arbitrary finite set Ω, a density function on Ω is simply a non-negative function $D : \Omega \to \mathbb{R}_{\geq 0}$ normalized so that*

$$\mathop{\mathbb{E}}_{x\in\Omega} D(x) = 1$$

.

- Let $A \subseteq \Omega$ be a subset of size $|A| = \delta|\Omega|$. Abusing notation, we also write

  $$A(x) := \frac{\mathbb{1}_A(x)}{\delta} = \begin{cases} 1/\delta & \text{for } x \in A \\ 0 & \text{otherwise}\end{cases}$$

  to denote the (re-normalized) indicator function of A. Under this normalization, $A(x)$ is a density function.

- For any two functions $f, g : \Omega \to \mathbb{R}$ we define $\langle f, g\rangle := \mathbb{E}_{x\in\Omega} f(x)g(x)$. For any set A we have the probabilistic interpretation

  $$\langle A, g\rangle = \mathop{\mathbb{E}}_{x\in\Omega} A(x)g(x) = \mathop{\mathbb{E}}_{a\in A} g(a).$$

  We can also apply this more generally to an arbitrary density function B. Let $b \sim B$ denote a random variable in Ω whose probability distribution is proportional to B. Then $\langle B, g\rangle$ has the interpretation

  $$\langle B, g\rangle = \mathop{\mathbb{E}}_{x\in\Omega} B(x)g(x) = \mathop{\mathbb{E}}_{b\sim B} g(b).$$

  Occasionally it will be useful to allow more generally for complex-valued functions f, g (which can appear e.g. during some intermediate calculations involving Fourier expansions of real-valued functions, even if the resulting quantity must be also real). In this case we insist on the convention

  $$\langle f, g\rangle = \mathop{\mathbb{E}}_{x\in\Omega} \overline{f(x)}g(x)$$

  so that at least when $B(x)$ is real-valued and $g(x)$ is complex-valued we still have

  $$\langle B, g\rangle = \mathop{\mathbb{E}}_{x} B(x)g(x) = \mathop{\mathbb{E}}_{b\sim B} g(b).$$

- For $k \geq 1$, we use the notation

  $$\|f\|_k = \left(\mathop{\mathbb{E}}_{x\in\Omega} |f(x)|^k\right)^{1/k},$$

and

$$\|f\|_\infty = \max_{x\in\Omega} |f(x)|\,.$$

- It is a consequence of Jensen's inequality that for $1 \le k \le k'$ and any function f we have

$$\|f\|_k \le \|f\|_{k'}.$$

3.2 Convolutions

- We specialize the finite set Ω from above to be a finite abelian group G. Given two functions $f, g : G \to \mathbb{R}$, we define their convolution with the following normalization:

$$(f * g)(x) := \mathop{\mathbb{E}}_{y\in G} f(y)g(x-y) = \mathop{\mathbb{E}}_{y\in G} f(x-y)g(y) = \frac{1}{|G|} \sum_{\substack{y,z\in G\\ y+z=x}} f(y)g(z).$$

- For a density function B, we can interpret $B * g$ as

$$(B * g)(x) := \mathop{\mathbb{E}}_{y\in G} B(y)g(x-y) = \mathop{\mathbb{E}}_{b\sim B} g(x-b).$$

- For two real-valued function f, g, we define the *cross-correlation* $f \star g$ as the convolution of $f(-x)$ with g:

$$(f \star g)(x) := \mathop{\mathbb{E}}_{y\in G} f(-y)g(x-y) = \mathop{\mathbb{E}}_{y\in G} f(y)g(x+y),$$

  and we have

$$(B \star g)(x) = \mathop{\mathbb{E}}_{b\sim B} g(x+b).$$

- If A, B are densities corresponding to (independent) random variables a, b, then as a consequence of our normalization conventions, $A * B$ is again a density. Indeed, it is the density corresponding to the random variable $a + b \in G$. Similarly, $A \star B$ is the density corresponding to the random variable $b - a \in G$.
- We note the identity $\langle f * g, h\rangle = \langle f, g \star h\rangle$.

3.3 Fourier analysis on finite abelian groups

- Let $\{e_\alpha(\cdot) \,:\, \alpha \in G\}$ denote the set of characters of the finite abelian group G. Each character is a function from G to the set of complex numbers of modulus 1. The product of two characters is again a character: For $\alpha, \beta \in G$, $e_\alpha \cdot e_\beta = e_{\alpha+\beta}$. Beyond this, the main important properties of the characters (for the development of the Fourier expansion) are

  (i) orthogonality: $\langle e_\alpha, e_\beta\rangle = \mathbb{E}_{x\in G}\, \overline{e_\alpha(x)} e_\beta(x) = \mathbb{E}_{x\in G}\, e_\alpha(-x)e_\beta(x) = \mathbb{1}(\alpha = \beta)$, and

  (ii) symmetry: $e_\alpha(x) = e_x(\alpha)$.

  In the case that the group G is presented explicitly as $G = \mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2} \times \cdots \times \mathbb{Z}_{N_r}$ we can consider the following concrete description:

$$e_\alpha(x) := \exp\left(2\pi i \sum_{j=1}^{r} \frac{\alpha_j x_j}{N_j}\right).$$

- For a set $A \subseteq G$, we define the fourier coefficients

$$\widehat{A}(\alpha) := \langle A, e_\alpha \rangle = \mathop{\mathbb{E}}_{a \in A} e_\alpha(a).$$

- Note that $\widehat{A}(-\alpha) = \langle A, e_{-\alpha} \rangle = \overline{\langle A, e_\alpha \rangle} = \overline{\widehat{A}(\alpha)}$.
- For a function $f : G \to \mathbb{R}$, we define the fourier coefficients

$$\widehat{f}(\alpha) := \langle f, e_\alpha \rangle = \mathop{\mathbb{E}}_{y \in G} f(y) e_\alpha(y).$$

- We can express any function by its fourier expansion

$$f(x) = \sum_{\alpha \in G} \widehat{f}(\alpha) e_\alpha(-x).$$

Proof.

$$\begin{aligned}
\sum_\alpha \left(\mathop{\mathbb{E}}_y f(y) e_\alpha(y) \right) e_\alpha(-x) &= \sum_y \mathop{\mathbb{E}}_\alpha f(y) e_\alpha(y) e_\alpha(-x) \\
&= \sum_y f(y) \mathop{\mathbb{E}}_\alpha e_y(\alpha) e_x(-\alpha) \\
&= \sum_y f(y) \mathbb{1}(y = x) \\
&= f(x).
\end{aligned}$$

□

- We have the following fourier-analytic identities for real-valued functions $A(x), B(x), f(x)$. In general they can all be verified, with no creativity required, by the following process: express any function appearing by its Fourier expansion, expand any products of sums appearing into a sum of products, and then use the identity $\mathbb{E}_{x \in G} e_\alpha(x) = \mathbb{1}(\alpha = 0)$ to evaluate any expectations.
  1. $\langle A, f \rangle = \sum_\alpha \widehat{A}(-\alpha) \widehat{f}(\alpha) = \sum_\alpha \widehat{A}(\alpha) \widehat{f}(-\alpha)$
  2. $(A * B)(x) = \sum_\alpha \widehat{A}(\alpha) \widehat{B}(\alpha) e_\alpha(-x)$
  3. $(A \star B)(x) = \sum_\alpha \widehat{A}(-\alpha) \widehat{B}(\alpha) e_\alpha(-x)$
  4. $(A \star A)(x) = \sum_\alpha |\widehat{A}(\alpha)|^2 e_\alpha(-x)$
  5. $\|A\|_2^2 = \langle A, A \rangle = \sum_\alpha |\widehat{A}(\alpha)|^2$
  6. $\|A \star A\|_2^2 = \sum_\alpha |\widehat{A}(\alpha)|^4$

4 Proof of Theorem 2.8, and Sifting

4.1 Sifting lemma

It will be convenient to set aside the following generic argument.

Proposition 4.1 (Weighted pigeonhole principle, or "first-moment method")**.** *Fix some nonnegative numbers $g_1, g_2, \dots g_m$ and $h_1, h_2, \dots, h_m$, with $\sum_i h_i > 0$. Let*

$$\eta := \frac{\sum_{i=1}^m g_i}{\sum_{i=1}^m h_i}$$

and

$$\mu(h) := \frac{1}{m} \sum_{i=1}^m h_i.$$

(i) There exists a choice of $j \in [m]$ with $\frac{g_j}{h_j} \leq \eta$ (and not of the form $\frac{0}{0}$).

(ii) Furthermore, there exists a choice of j with both

$$\frac{g_j}{h_j} \leq 2\eta \text{ and } h_j \geq \tfrac{1}{2}\mu(h)$$

(iii) More generally, suppose that $H, \tau > 0$ are such that

$$\sum_{i\,:\,h_i \geq H} h_i \geq \tau \sum_i h_i.$$

Then there is a choice of j with both

$$\frac{g_j}{h_j} \leq \frac{\eta}{\tau} \text{ and } h_j \geq H.$$

Proof. The degenerate case $\eta = 0$ is easy to handle: we simply pick the largest value h_j. So suppose $\eta \neq 0$ and consider the first claim. We discard any indices i where both $g_i = h_i = 0$, noting that this does not change the value of $\sum_i g_i$ or $\sum_i h_i$. Consider the equality $\sum_i g_i = \eta \sum_i h_i$. If the desired conclusion does not hold, we have $g_i > \eta h_i$ for all i, and so $\sum_i (g_i - \eta h_i) > 0$, a contradiction. For the second claim, we apply the first claim to the modified sequence

$$h_i' := h_i \cdot \mathbb{1}(h_i \geq \mu/2)\,.$$

For the final claim, we apply the first claim to the modified sequence

$$h_i'' := h_i \cdot \mathbb{1}(h_i \geq H)\,. \qquad \square$$

In preparation to make some combinatorial arguments we briefly switch to the counting measure. This also allows us to also handle finite subsets A of infinite groups G, noting that summations such as $\sum_{x \in G} R_A(x)$ are sensible because R_A is finitely supported.

Lemma 4.2 (Sifting lemma – counting formulation)**.** *Consider*

- *a finite subset A of an abelian group G,*
- *a function $f : G \to \mathbb{R}_{\geq 0}$, and*
- *an integer $k \geq 2$.*

There is a subset $A' \subseteq A$ with

$$\frac{1}{|A'|^2} \sum_{a,b \in A'} f(a-b) \leq 2 \cdot \frac{\sum_x R_A^-(x)^k f(x)}{\sum_x R_A^-(x)^k}$$

and

$$|A'| \geq \tfrac{1}{2} \cdot \frac{\sum_x R_A^-(x)^k}{|A|^k}.$$

Specifically, A' is of the form

$$A'(s) := A \cap (A + s_1) \cap \cdots \cap (A + s_{k-1})$$

for some $s = (s_1, s_2, \cdots s_{k-1}) \in G^{k-1}$.

Proof. For $s \in G^{k-1}$, let

$$A'(s) := A \cap (A + s_1) \cap (A + s_2) \cap \cdots \cap (A + s_{k-1}).$$

We wish to understand the quantity

$$\frac{\sum_{a,b \in A'(s)} f(a-b)}{|A'(s)|^2}$$

for various choices of s. Define

$$g(s) := \sum_{a,b \in A'(s)} f(a-b)$$

and

$$h(s) := |A'(s)|^2.$$

We would like to apply the weighted pigeonhole principle to find a suitable choice of s, and so we need to compute $\sum_s g(s)$ and $\sum_s h(s)$. Before we start the computation we point out ahead-of-time the following combinatorial identities:

$$R_A^-(a-b) = |A \cap (A + a - b)| = |(A - a) \cap (A - b)| = \sum_{t \in G} \mathbb{1}_{A-a}(t) \mathbb{1}_{A-b}(t) = \sum_{t \in G} \mathbb{1}_{A+t}(a) \mathbb{1}_{A+t}(b).$$

We have

$$\begin{aligned}
\sum_x R_A^-(x)^k f(x) &= \sum_{a,b \in A} R_A^-(a-b)^{k-1} f(a-b) \\
&= \sum_{a,b \in A} \left(\sum_t \mathbb{1}_{A+t}(a) \mathbb{1}_{A+t}(b) \right)^{k-1} f(a-b) \\
&= \sum_{a,b \in A} \sum_{s_1, \ldots s_{k-1}} \left(\prod_{j=1}^{k-1} \mathbb{1}_{A+s_j}(a) \right) \cdot \left(\prod_{j=1}^{k-1} \mathbb{1}_{A+s_j}(b) \right) \cdot f(a-b) \\
&= \sum_{s_1, \ldots s_{k-1}} \sum_{a,b \in A} \left(\mathbb{1}_{\bigcap_{j=1}^{k-1}(A+s_j)}(a) \right) \cdot \left(\mathbb{1}_{\bigcap_{j=1}^{k-1}(A+s_j)}(b) \right) \cdot f(a-b) \\
&= \sum_{s_1, \ldots, s_{k-1}} \sum_{a,b \in A'(s)} f(a-b) \\
&= \sum_s g(s).
\end{aligned}$$

A special case of this same calculation (with $f \equiv 1$) also gives

$$\sum_x R_A^-(x)^k = \sum_s h(s) = \sum_s |A'(s)|^2.$$

Thus we have

$$\eta := \frac{\sum_s g(s)}{\sum_s h(s)} = \frac{\sum_x R(x)^k f(x)}{\sum_x R(x)^k}.$$

We remark that in the case that G is finite and A is a dense subset of size $|A| = \delta|G|$, we can obtain a quite satisfactory conclusion already by applying part (ii) of the weighted pigeonhole principle. Indeed, for uniformly random $s \in G^{k-1}$, we have average size

$$\mathop{\mathbb{E}}_s |A'(s)|^2 = \mathop{\mathbb{E}}_s h(s) = \frac{1}{|G|^{k-1}} \sum_x R_A^-(x)^k = \frac{1}{|G|^{k-2}} \mathop{\mathbb{E}}_x R_A^-(x)^k = \delta^{2k}|G|^2 \mathop{\mathbb{E}}_x (A \star A)(x)^k \geq \delta^{2k}|G|^2.$$

However, we consider the following argument which is better in general. We wish to determine a value M which is as large as possible and also satisfies

$$\sum_{s\,:\,|A'(s)| \geq M} |A'(s)|^2 \geq \tfrac{1}{2} \cdot \sum_s |A'(s)|^2.$$

For any choice of M we have

$$\begin{aligned}
\sum_{s\,:\,|A'(s)| \leq M} |A'(s)|^2 &\leq M \cdot \sum_s |A'(s)| \\
&= M \cdot \sum_s \sum_{a \in A} \mathbb{1}_{A+s_1}(a) \cdot \mathbb{1}_{A+s_2}(a) \cdots \mathbb{1}_{A+s_{k-1}}(a) \\
&= M \cdot \sum_{a \in A} \sum_s \mathbb{1}_{A-a}(s_1) \cdot \mathbb{1}_{A-a}(s_2) \cdots \mathbb{1}_{A-a}(s_{k-1}) \\
&= M \cdot |A|^k.
\end{aligned}$$

This shows that we may take

$$M = \tfrac{1}{2} \cdot \frac{\sum_s |A'(s)|^2}{|A|^k} = \tfrac{1}{2} \cdot \frac{\sum_x R_A^-(x)^k}{|A|^k}.$$

We conclude by part (iii) of the weighted pigeonhole principle. □

Remark 4.3. *We note that the combinatorial argument above which shows that*

$$\sum_x R_A^-(x)^k f(x) = \sum_s \sum_{a,b \in A'(s)} f(a-b) = \sum_s \sum_x R_{A'(s)}^-(x) f(x)$$

is in fact valid for any function f. This more plainly means that we have the identity

$$R_A^-(x)^k \equiv \sum_s R_{A'(s)}^-(x)$$

for the k-th power of R_A^-. Given this, we can summarize the remaining points of the argument as follows.

- *For any f, we can interpret the sum $\sum_{s,x} R^-_{A'(s)}(x)f(x)$ in two ways: firstly as*

$$\sum_s \left(\sum_x R^-_{A'(s)}(x)f(x) \right)$$

but also as $\sum_x \left(\sum_s R^-_{A'(s)}(x) \right) f(x) = \sum_x R^-_A(x)^k f(x)$.

- *For the particular case $f \equiv 1$, we may further interpret $\sum_x R^-_{A'(s)}(x) = |A'(s)|^2$.*
- *If we have a reasonable bound on the size of $\{s : |A(s')| \neq 0\}$, we can already infer a reasonable bound on the median value of $|A'(s)|^2$.*
- *We also have the identity $\sum_s |A'(s)| = |A|^k$, which is relevant for obtaining a better estimate for the median value of $|A'(s)|^2$.*

Proof of the extended Pre-BSG Lemma (Lemma B.8). We apply the sifting lemma straightforwardly to the sub-level set indicator

$$f(x) = \mathbb{1}\Big(R^-_A(x) \leq c \cdot \kappa^{\frac{1}{k-1}} \cdot |A| \Big),$$

where

$$\kappa = \frac{\sum_x R^-_A(x)^k}{|A|^{k+1}}.$$

We check that

$$\sum_x R^-_A(x)^k f(x) \leq c^{k-1} \cdot \kappa \cdot |A|^{k-1} \sum_x R^-_A(x) = c^{k-1} \cdot \kappa \cdot |A|^{k+1},$$

so indeed

$$\frac{\sum_x R^-_A(x)^k f(x)}{\sum_x R^-_A(x)^k} = \frac{\sum_x R^-_A(x)^k f(x)}{\kappa |A|^{k+1}} \leq c^{k-1}.$$

We also have

$$|A'| \geq \tfrac{1}{2} \cdot \frac{\sum_x R^-_A(x)^k}{|A|^k} = \tfrac{1}{2} \cdot \frac{\kappa |A|^{k+1}}{|A|^k} = \tfrac{1}{2} \cdot \kappa \cdot |A|. \qquad \square$$

We proceed to state a density formulation of the sifting lemma for subsets $A \subseteq G$ of a finite groups G with of $|A| = \delta |G|$. We use the notation

$$D^{\wedge k}(x) := \left(\frac{D(x)}{\|D\|_k} \right)^k$$

to denote the unique density function proportional to D^k. We also point out the translation

$$\frac{\sum_{x \in G} R^-_A(x)^k}{|A|^k} = \frac{|A|^k}{|G|^k} \sum_{x \in G} (A \star A)(x)^k = \frac{|A|^k}{|G|^{k-1}} \mathop{\mathbb{E}}_x (A \star A)(x)^k = \delta^k \cdot |G| \cdot \|A \star A\|_k^k.$$

Lemma 4.4 (Sifting lemma – density formulation)**.** *Consider*

- *a set $A \subseteq G$ of size $|A| = \delta |G|$,*
- *a function $f : G \to \mathbb{R}_{\geq 0}$, and*

- *an integer* $k \geq 2$.

There is a subset $A' \subseteq A$ *with*

$$\left\langle A' \star A', f \right\rangle \leq 2 \cdot \left\langle (A \star A)^{\wedge k}, f \right\rangle$$

and

$$\frac{|A'|}{|G|} \geq \tfrac{1}{2} \cdot \delta^k \cdot \|A \star A\|_k^k \geq \tfrac{1}{2} \cdot \delta^k.$$

One can easily derive a formulation of the extended Pre-BSG Lemma for the dense setting which corresponds e.g. to either part of Proposition B.4. We consider the following formulation which is tailored specifically to our intended application.

Corollary 4.5 (Sifting a robust witness)**.** *Suppose that* $A \subseteq G$ *has size* $|A| \geq 2^{-d}|G|$*, and*

$$\|A \star A\|_k \geq 1 + \varepsilon$$

for some $k \geq 1$ *and* $\varepsilon > 0$*. Consider*

$$S := \{x \in G \,:\, (A \star A) \leq 1 + \varepsilon/2\}.$$

Let $\overline{\varepsilon} = \min\{1, \varepsilon\}$*. There is a subset* $A' \subseteq A$ *with*

$$\left\langle A' \star A', \mathbb{1}_S \right\rangle \leq \frac{\overline{\varepsilon}}{2^4}$$

and

$$\frac{|A'|}{|G|} \geq \begin{cases} 2^{-O(dk)} \cdot \varepsilon^{O(d/\varepsilon)} & \text{when } \varepsilon < 1/2 \\ 2^{-O(dk)} & \text{when } \varepsilon \geq 1/2. \end{cases}$$

Proof. Apply the sifting lemma with $f := \mathbb{1}_S$ and

$$k' = \begin{cases} \lceil k + 2\lg(32/\varepsilon)/\varepsilon \rceil & \text{when } \varepsilon < 1/2 \\ \lceil k + 20 \rceil & \text{when } \varepsilon \geq 1/2. \end{cases} \quad \square$$

The point of the quantity $\overline{\varepsilon} = \min\{1, \varepsilon\}$ is that for all $\varepsilon \geq 0$, $\lambda \geq 1$ we have

$$\left(1 + \frac{\varepsilon}{\lambda}\right)\left(1 - \frac{\overline{\varepsilon}}{4\lambda}\right) \geq 1 + \frac{\varepsilon}{2\lambda}.$$

So $A' \star A'$ above is a robust witness to $A \star A \gg 1 + \varepsilon/4$ with some room to spare: for

$$f := \mathbb{1}(A \star A \geq 1 + \varepsilon/2)$$

we have

$$\left\langle A' \star A', A \star A \right\rangle \geq \left(1 + \frac{\varepsilon}{2}\right)\left\langle A' \star A', f \right\rangle \geq \left(1 + \frac{\varepsilon}{2}\right)\left(1 - \frac{\overline{\varepsilon}}{2^3}\right) \geq 1 + \frac{\varepsilon}{2^2}.$$

Theorem 4.6. *Let $A \subseteq \mathbb{F}_q^n$ be a set of size $|A| \geq 2^{-d}|\mathbb{F}_q^n|$, where $d \geq 1$. Suppose that*

$$\|A \star A\|_k \geq 1 + \varepsilon$$

for some $k \geq 1$. We have

$$\|A\|_{\perp,r} \geq 1 + \frac{\varepsilon}{4}$$

for some

$$r = \begin{cases} O(d^4k^4) & \text{when } \varepsilon \geq 1/2 \\ O(d^4k^4/\varepsilon^2 + d^4 \lg(1/\varepsilon)^4/\varepsilon^6) & \text{when } \varepsilon < 1/2. \end{cases}$$

In either case we also have the related conclusion

$$\max_{\substack{W \subseteq \mathbb{F}_q^n \\ \dim(W) \leq r}} \|P_W A - 1\|_2^2 \geq \frac{\varepsilon}{4}$$

where $P_W A := W^\perp * A$.

We note that the latter conclusion here is (qualitatively) stronger for small ε. Indeed, we have $\|P_W A\|_2^2 = 1 + \|P_W A - 1\|_2^2$, and in light of the alternative characterization

$$\|A\|_{\perp,r} = \max_{\substack{W \subseteq \mathbb{F}_q^n \\ \dim(W) \leq r}} \|P_W A\|_\infty,$$

we have the lower bound

$$\|A\|_{\perp,r} \geq \|P_W A\|_\infty \geq \|P_W A\|_2 \geq \sqrt{1 + \frac{\varepsilon}{4}} \approx 1 + \frac{\varepsilon}{8}$$

for some small subspace W.

Proof. In the case $\varepsilon \leq \frac{1}{2}$, we replace k by $\max\{k, \lg(1/\varepsilon)/\varepsilon\}$ for convenience, noting that $\|A \star A\|_k$ does not decrease as a result. We consider

$$f := \mathbb{1}\big(A \star A \geq 1 + \tfrac{\varepsilon}{2}\big)$$

and (using Corollary 4.5) we sift a robust witness $A' \star A'$ with

$$\big\langle A' \star A', f \big\rangle \geq 1 - \frac{\overline{\varepsilon}}{2^4}.$$

and density $|A'|/|G| \geq 2^{-O(dk)}$. We apply Sanders' invariance lemma (Lemma C.3) to obtain a linear subspace V with

$$\big\langle V * A' \star A', f \big\rangle \geq \big\langle A' \star A', f \big\rangle - \frac{\overline{\varepsilon}}{2^4} \geq 1 - \frac{\overline{\varepsilon}}{2^3}$$

and codimension $r \leq O(d^4k^4/\overline{\varepsilon}^2)$. We conclude that

$$\big\langle V * A' \star A', A \star A \big\rangle \geq \left(1 + \frac{\varepsilon}{2}\right) \big\langle V * A' \star A', f \big\rangle \geq \left(1 + \frac{\varepsilon}{2}\right)\left(1 - \frac{\overline{\varepsilon}}{2^3}\right) \geq 1 + \frac{\varepsilon}{4}.$$

This gives

$$\big\langle V', A \star A \big\rangle \geq 1 + \frac{\varepsilon}{4}$$

for some affine subspace V' and

$$\langle V', A\rangle \geq 1 + \frac{\varepsilon}{4}$$

for some $V'' = V' + a$. This proves the first claim.

To prove the latter claim, we depart from the argument above when we reach

$$\langle V * A' \star A', A \star A\rangle \geq 1 + \frac{\varepsilon}{4}.$$

It is pleasant, although not ultimately crucial, to interpret this quantity Fourier-analytically. Letting $W = V^{\perp}$, we can express

$$\langle V * A' \star A', A \star A\rangle = \sum_{\alpha \in W} |\widehat{A'}(\alpha)|^2 |\widehat{A}(\alpha)|^2,$$

and so we have

$$\sum_{\substack{\alpha \in W \\ \alpha \neq 0}} |\widehat{A'}(\alpha)|^2 |\widehat{A}(\alpha)|^2 \geq \frac{\varepsilon}{4}.$$

Using the trivial bound $|\widehat{A'}(\alpha)| \leq 1$ we infer

$$\sum_{\substack{\alpha \in W \\ \alpha \neq 0}} |\widehat{A}(\alpha)|^2 \geq \frac{\varepsilon}{4},$$

and indeed this quantity is the same as $\|P_W A - 1\|_2^2$. □

4.2 Sifting for general convolutions

We now address the problem of obtaining $\|A\|_{\perp,r} \geq 1 + \frac{\varepsilon}{8}$ from an assumption $\|A\|_{*,k} \geq 1 + \varepsilon$. To begin with we have by definition that $\langle A, B * C\rangle \geq 1 + \varepsilon$ for some large sets $|B|, |C| \geq 2^{-k}|G|$. We argue that

$$\langle A, B * C\rangle = \langle A \star B, C\rangle \leq 2^{k/k'} \|A \star B\|_{k'}$$

for any choice of k'. Choosing $k' = 4\lg(e)k/\overline{\varepsilon}$ gives

$$\|A \star B\|_{k'} \geq 2^{-k/k'}(1+\varepsilon) \geq \left(1 - \frac{\overline{\varepsilon}}{4}\right)(1+\varepsilon) \geq 1 + \frac{\varepsilon}{2}.$$

At this point we argue roughly as before – we just need a variant of the sifting lemma for general convolutions $A \star B$. One other small technical difference is that in this setting we do not know of a way to obtain a good lower bound on the related quantity $\max_{\dim(W) \leq r} \|P_W A\|_2$.

We give the following variant of the sifting lemma for general convolutions. It is a bit more convenient to state it for $R^{-}_{A,B}$ rather than $R_{A,B}$, but this loses no generality.

Lemma 4.7 (Sifting general convolutions)**.** *Consider*

- *an abelian group G,*
- *two finite sets $A, B \subseteq G$,*
- *a function $f : G \to \mathbb{R}_{\geq 0}$, and*

- *an integer* $k \geq 2$.

There are subsets $A' \subseteq A$ *and* $B' \subseteq B$ *with*

$$\frac{1}{|A'||B'|} \sum_{a \in A'} \sum_{b \in B'} f(a-b) \leq 2 \cdot \frac{\sum_x R^-_{A,B}(x)^k f(x)}{\sum_x R^-_{A,B}(x)^k}$$

and

$$|A'| \geq \tfrac{1}{4} \cdot \frac{\sum_x R^-_{A,B}(x)^k}{|B|^k},$$
$$|B'| \geq \tfrac{1}{4} \cdot \frac{\sum_x R^-_{A,B}(x)^k}{|A|^k},$$

Specifically, A' *and* B' *are of the form*

$$A'(s) := A \cap (A+s_1) \cap \cdots \cap (A+s_{k-1}),$$
$$B'(s) := B \cap (B+s_1) \cap \cdots \cap (B+s_{k-1})$$

for some $s = (s_1, s_2, \cdots s_{k-1}) \in G^{k-1}$.

To prove this we will need one more form of the weighted pigeonhole principle.

Proposition 4.8 (Continuation of Proposition 4.1)**.**
Under the hypotheses of Proposition 4.1,

(iv) Suppose that $I \subseteq [m]$ *and* $\tau > 0$ *are such that*

$$\sum_{i \in I} h_i \geq \tau \sum_i h_i.$$

Then there is a choice of j *with both*

$$\frac{g_j}{h_j} \leq \frac{\eta}{\tau} \text{ and } j \in I.$$

Proof. Consider the modified sequence $h'_i := h_i \cdot \mathbb{1}(i \in I)$. □

Proof of Lemma 4.7. For $s \in G^{k-1}$, let

$$A'(s) := A \cap (A+s_1) \cap \cdots \cap (A+s_{k-1}),$$
$$B'(s) := B \cap (B+s_1) \cap \cdots \cap (B+s_{k-1})$$

Note that

$$R^-_{A,B}(a-b) = |A \cap (B+a-b)| = |(A-a) \cap (B-b)| = \sum_{t \in G} \mathbb{1}_{A-a}(t) \mathbb{1}_{B-b}(t) = \sum_{t \in G} \mathbb{1}_{A+t}(a) \mathbb{1}_{B+t}(b).$$

We express

$$\begin{aligned}
\sum_x R^-_{A,B}(x)^k f(x) &= \sum_{a \in A} \sum_{b \in B} \left(\sum_t \mathbb{1}_{A+t}(a) \mathbb{1}_{B+t}(b) \right)^{k-1} f(a-b) \\
&= \sum_s \sum_{a \in A} \sum_{b \in B} \left(\mathbb{1}_{\bigcap_{j=1}^{k-1}(A+s_j)}(a) \right) \cdot \left(\mathbb{1}_{\bigcap_{j=1}^{k-1}(B+s_j)}(b) \right) \cdot f(a-b) \\
&=: \sum_s \Phi_f(s).
\end{aligned}$$

We apply the weighted pigeonhole principle to the ratio

$$\frac{\Phi_f(s)}{\Phi_1(s)}.$$

to find a suitable choice of s. Specifically, we apply variant (iv) with some index-set of the form

$$S = \left\{ s \,:\, |A'(s)| \geq M_A \text{ and } |B'(s)| \geq M_B \right\}.$$

For any choice of M_A, M_B we can estimate

$$\sum_{s\,:\,s\notin S} \Phi_1(s) = \sum_{s\,:\,s\notin S} |A'(s)||B'(s)| \leq M_A \sum_s |B'(s)| + M_B \sum_s |A'(s)| = M_A|B|^k + M_B|A|^k.$$

Choosing

$$M_A := \tfrac{1}{4} \cdot \frac{\sum_s \Phi_s(s)}{|B|^k} = \tfrac{1}{4} \cdot \frac{\sum_x R^-_{A,B}(x)^k}{|B|^k},$$

$$M_B := \tfrac{1}{4} \cdot \frac{\sum_s \Phi_s(s)}{|A|^k} = \tfrac{1}{4} \cdot \frac{\sum_x R^-_{A,B}(x)^k}{|A|^k}$$

is sufficient to ensure that

$$\sum_{s\in S} \Phi_1(s) \geq \frac{1}{2} \cdot \sum_s \Phi_1(s). \qquad \square$$

The size guarantees here for A' and B' have a pleasant interpretation in the density formulation. Let G be a finite group and let $A \subseteq G$ be a set of size $|A| = \delta|G|$. We have the translation

$$\frac{\sum_x R^-_{A,B}(x)^k}{|B|^k} = \frac{|A|^k}{|G|^k} \sum_x (A \star B)(x)^k = \delta^k \cdot \|A \star B\|_k^k \cdot |G| \geq \delta^k \cdot |G|.$$

Lemma 4.9 (Sifting general convolutions – density formulation)**.** *Consider*

- *a finite abelian group G,*
- *two sets $A, B \subseteq G$ of sizes $|A| = \delta_A|G|$, $|B| = \delta_B|G|$.*
- *a function $f : G \to \mathbb{R}_{\geq 0}$, and*
- *an integer $k \geq 2$.*

There are subsets $A' \subseteq A$ and $B' \subseteq B$ with

$$\left\langle A' \star B', f \right\rangle \leq 2 \cdot \left\langle (A \star B)^{\wedge k}, f \right\rangle$$

and

$$\frac{|A'|}{|G|} \geq \tfrac{1}{4} \cdot \delta_A^k \cdot \|A \star B\|_k^k \geq \tfrac{1}{4} \cdot \delta_A^k,$$

$$\frac{|B'|}{|G|} \geq \tfrac{1}{4} \cdot \delta_B^k \cdot \|A \star B\|_k^k \geq \tfrac{1}{4} \cdot \delta_B^k.$$

Theorem 4.10. *Let $A \subseteq \mathbb{F}_q^n$ be a set of size $|A| \geq 2^{-d}|\mathbb{F}_q^n|$, where $d \geq 1$. Suppose that*

$$\|A\|_{*,k} \geq 1 + \varepsilon$$

for some $k \geq 1$. We have

$$\|A\|_{\perp,r} \geq 1 + \frac{\varepsilon}{8}$$

for some

$$r = \begin{cases} O(dk^7) & \text{when } \varepsilon \geq 1 \\ O(dk^7/\varepsilon^6 + dk^3 \lg(2/\varepsilon)^4/\varepsilon^6) & \text{when } \varepsilon < 1. \end{cases}$$

Proof. It follows from $\|A\|_{*,k} \geq 1 + \varepsilon$ that for $p = 4\lg(e)k/\overline{\varepsilon} + \lg(2/\overline{\varepsilon})/\overline{\varepsilon}$ we have $\|A \star B\|_p \geq 1 + \frac{\varepsilon}{2}$ for some set of size $|B| \geq 2^{-k}|\mathbb{F}_q^n|$. From here we argue similarly as in the proof of Theorem 4.6. This time we suppress some quantitative details. Let $f := \mathbb{1}(A \star B \geq 1 + \varepsilon/4)$. We sift $A \star B$ to find a convolution $A' \star B'$ witnessing $\langle A' \star B', f\rangle \approx 1$ with $|A'| \geq 2^{-O(dp)}|\mathbb{F}_q^n|$ and $|B'| \geq 2^{-O(kp)}|\mathbb{F}_q^n|$. We apply Sanders' invariance lemma to find a linear subspace V with $\langle V * A' \star B', f\rangle \approx 1$ and codimension $r \leq O(dp \cdot k^3p^3/\overline{\varepsilon}^2) = O(dk^3p^4/\overline{\varepsilon}^2)$. We conclude that $\langle V', A \star B\rangle \geq 1 + \frac{\varepsilon}{8}$ for some translate V' of V and $\langle V'', A\rangle \geq 1 + \frac{\varepsilon}{8}$ for some translate $V'' = V' + b$. ☐

4.3 Local variants

We proceed to give some local variants of the statements above which will be needed for setting $A \subseteq [N] \subseteq \mathbb{Z}$.

Lemma 4.11 (Sifting lemma – a local variant)**.** *Consider*

- *a finite abelian group G,*
- *a set $A \subseteq G$.*
- *some additional subsets $B, C \subseteq G$,*
- *a function $f : G \to \mathbb{R}_{\geq 0}$, and*
- *an integer $k \geq 1$.*

There is some choice of $s = (s_1, s_2, \ldots, s_k) \in G^k$ giving rise to subsets $B' \subseteq B$ and $C' \subseteq C$ of the form

- $B' = B \cap (A - s_1) \cap (A - s_2) \cap \cdots \cap (A - s_k)$ *and*
- $C' = C \cap (s_1 - A) \cap (s_2 - A) \cap \cdots \cap (s_k - A)$

which satisfy

$$\frac{1}{|B'||C'|} \sum_{b \in B'} \sum_{c \in C'} f(b + c) \leq 2 \cdot \frac{\sum_x R_{B,C}(x) R_A^-(x)^k f(x)}{\sum_x R_{B,C}(x) R_A^-(x)^k}$$

and

$$|B'||C'| \geq \tfrac{1}{2} \cdot \frac{\sum_x R_{B,C}(x) R_A^-(x)^k}{|G|^k}.$$

Proof. We express

$$\begin{aligned}
\sum_{x\in G} R_{B,C}(x)R_A^-(x)^k \cdot f(x) &= \sum_{b\in B}\sum_{c\in C} R_A^-(b+c)^k \cdot f(b+c) \\
&= \sum_{b\in B}\sum_{c\in C} |A\cap(A+b+c)|^k \cdot f(b+c) \\
&= \sum_{b\in B}\sum_{c\in C} |(A-b)\cap(A+c)|^k \cdot f(b+c) \\
&= \sum_{b\in B}\sum_{c\in C} \left(\sum_t \mathbb{1}_{A-b}(t)\mathbb{1}_{A+c}(t)\right)^k \cdot f(b+c) \\
&= \sum_{b\in B}\sum_{c\in C} \left(\sum_t \mathbb{1}_{A-t}(b)\mathbb{1}_{t-A}(c)\right)^k \cdot f(b+c) \\
&= \sum_{s}\sum_{b\in B}\sum_{c\in C} \left(\prod_{j=1}^k \mathbb{1}_{A-s_j}(b)\right)\cdot\left(\prod_{j=1}^k \mathbb{1}_{s_j-A}(c)\right) \cdot f(b+c) \\
&=: \sum_s \Phi_f(s).
\end{aligned}$$

We then apply variant (ii) of the weighted pigeonhole principle to the ratio

$$\frac{\Phi_f(s)}{\Phi_1(s)}$$

to find an appropriate choice of s. □

We need the following notation.

Definition 4.12 (Weighted k-(semi)norm)**.** *For a density D on an arbitrary finite set Ω, and $k \geq 1$, we use the notation*

$$\|f\|_{k,D} := \left\langle D, |f|^k \right\rangle^{1/k} = \left(\mathop{\mathbb{E}}_{x\sim D} |f(x)|^k\right)^{1/k}.$$

Corollary 4.13 (Sifting a local robust witness)**.** *Consider*

- *a finite abelian group G of size N,*
- *a subset $A \subseteq G$ of size $|A| = 2^{-d}N$,*
- *some additional subsets $B, C \subseteq G$, and*
- *$k \geq 1$, $\varepsilon \in [0,1]$.*

Suppose that

$$\|A \star A\|_{k,B*C} \geq 1 + 2\varepsilon,$$

and define the sublevel-set indicator function

$$f(x) := \mathbb{1}\big((A\star A)(x) \leq 1+\varepsilon\big).$$

For any integer $k' \geq k$, there are subsets $B' \subseteq B$ and $C' \subseteq C$ with

$$\langle B' * C', f \rangle \leq 2 \cdot \left(\frac{1+\varepsilon}{1+2\varepsilon}\right)^{k'} \leq 2 \cdot 2^{-\varepsilon k'/2}$$

and

$$\frac{|B'||C'|}{|B||C|} \geq \tfrac{1}{2} \cdot 2^{-2dk'}.$$

In particular,

$$\frac{|B'|}{|B|}, \frac{|C'|}{|C|} \geq \tfrac{1}{2} \cdot 2^{-2dk'}.$$

Proof. We straightforwardly apply Lemma 4.11 – it remains to interpret the result. We have

$$\begin{aligned}
\langle B' * C', f \rangle &\leq 2 \cdot \frac{\sum_x R_{B,C}(x) R_A^-(x)^k f(x)}{\sum_x R_{B,C}(x) R_A^-(x)^k} \\
&= 2 \cdot \frac{\mathbb{E}_x \big(B * C\big)(x) \big(A \star A\big)(x)^{k'} f(x)}{\langle B * C, (A \star A)^{k'} \rangle} \\
&\leq 2 \cdot \frac{\mathbb{E}_x \big(B * C\big)(x) (1+\varepsilon)^{k'}}{(1+2\varepsilon)^{k'}} \\
&= 2 \cdot \left(\frac{1+\varepsilon}{1+2\varepsilon}\right)^{k'}
\end{aligned}$$

and

$$\begin{aligned}
|B'||C'| &\geq \frac{1}{2|G|^{k'}} \sum_x R_{B,C}(x) R_A^-(x)^{k'} \\
&= \frac{1}{2|G|^{k'}} \cdot |G| \cdot \frac{|B||C|}{|G|} \cdot \frac{|A|^{2k'}}{|G|^{k'}} \cdot \left\langle B * C, (A \star A)^{k'} \right\rangle \\
&\geq \frac{1}{2|G|^{k'}} \cdot |G| \cdot \frac{|B||C|}{|G|} \cdot \frac{|A|^{2k'}}{|G|^{k'}} \\
&= \tfrac{1}{2} \cdot 2^{-dk'} |B||C|,
\end{aligned}$$

where we have used that

$$\left\langle B * C, (A \star A)^{k'} \right\rangle = \|A \star A\|_{k', B*C}^{k'} \geq \|A \star A\|_{k, B*C}^{k} \geq 1$$

by assumption. □

5 Spectral positivity and k-norm inequalities for convolutions

5.1 A decoupling inequality for convolutions

Proposition 5.1 (Fourier interpretation of k-norms)**.** *For any function $f : G \to \mathbb{R}$ and any even $k \in \mathbb{N}$,*

$$\|f\|_k^k = \mathop{\mathbb{E}}_{x \in G} f(x)^k = \sum_{\substack{(\alpha_1, \alpha_2, \ldots, \alpha_k) \in G^k \\ \alpha_1 + \alpha_2 + \cdots + \alpha_k = 0}} \prod_{j=1}^{k} \widehat{f}(\alpha_j).$$

Proof. Expand the expression $\left(\sum_\alpha \widehat{f}(\alpha)e_\alpha(-x)\right)^k$ and take the expectation. $\square$

Lemma 5.2 (Decoupling inequality)**.** *Let A, B be density functions on a finite abelian group G. For even integers $k \in \mathbb{N}$ we have*

$$\|A * B - 1\|_k \leq \|A \star A - 1\|_k^{1/2} \|B \star B - 1\|_k^{1/2}.$$

In particular,

$$\|A * A - 1\|_k \leq \|A \star A - 1\|_k.$$

Proof. We note that for any density A, $A * 1 = 1 * A \equiv 1$. So, the claimed inequality is the same as

$$\|(A-1) * (B-1)\|_k \leq \|(A-1) \star (A-1)\|_k^{1/2} \|(B-1) \star (B-1)\|_k^{1/2}.$$

We prove more generally that for any functions f, g,

$$\|f * g\|_k \leq \|f \star f\|_k^{1/2} \|g \star g\|_k^{1/2}.$$

By re-scaling, it suffices[5] to prove this for $\|f \star f\|_k = \|g \star g\|_k = 1$. We have

$$\begin{aligned}
\mathop{\mathbb{E}}_x (f * g)(x)^k &= \mathop{\mathbb{E}}_x \left(\sum_\alpha \widehat{f}(\alpha)\widehat{g}(\alpha)e_\alpha(-x) \right)^k \\
&\leq \sum_{\substack{\alpha_1, \alpha_2, \ldots, \alpha_k \\ \alpha_1 + \alpha_2 + \cdots + \alpha_k = 0}} \left| \prod_{i=1}^k \widehat{f}(\alpha_i) \cdot \prod_{i=1}^k \widehat{g}(\alpha_i) \right| \\
&\leq \sum_\alpha \left(\frac{1}{2} \prod_{i=1}^k |\widehat{f}(\alpha_i)|^2 + \frac{1}{2} \prod_{i=1}^k |\widehat{g}(\alpha_i)|^2 \right) \\
&= \frac{1}{2} \|f \star f\|_k^k + \frac{1}{2} \|g \star g\|_k^k \\
&= 1.
\end{aligned}$$

$\square$

We remark that this inequality has appeared before in other works. For example, it is a special case of Lemma 13 in [Shk17].

5.2 Spectral Positivity

Definition 5.3. *Suppose $f : G \to \mathbb{R}$ is a function such that $\widehat{f}(\alpha)$ is real and nonnegative for all $\alpha \in G$. We say that such functions are "spectrally positive", and we use the notation*

$$f \succeq 0$$

to denote the fact that f is such a function.

[5] One can check the degenerate case $\|f \star f\|_k = 0$ separately. This case occurs only when $f \equiv 0$, which can be seen by the calculation $\|f \star f\|_k^2 \geq \|f \star f\|_2^2 = \sum_\alpha |\widehat{f}(\alpha)|^4$.

We note that any self-convolution $f \star f$ is spectrally positive:

$$(f \star f)(x) = \sum_{\alpha \in G} |\widehat{f}(\alpha)|^2 e_\alpha(-x) \succeq 0,$$

and conversely that any spectrally positive function $f(x) = \sum_\alpha \widehat{f}(\alpha) e_\alpha(-x)$ can be expressed as a self-convolution $g \star g$, where

$$g(x) := \sum_\alpha \sqrt{\widehat{f}(\alpha)} e_\alpha(-x).$$

Proposition 5.4. *The set $\{f : f \succeq 0\}$ enjoys the following two closure properties.*

- *(Closure under multiplication.) Suppose $f, g \succeq 0$. Then*

$$h(x) := f(x) \cdot g(x) \succeq 0.$$

- *(Closure under centering.) Suppose $f \succeq 0$, and that D is a symmetric*[6] *density function. Then*

$$h(x) := f(x) - (D * f)(x) \succeq 0.$$

  In particular, by letting $D \equiv 1$,

$$f - \mathbb{E}[f] \succeq 0.$$

Proof. For the first claim, simply express f and g by their Fourier expansions and expand the product:

$$\left(\sum_\beta \widehat{f}(\beta) e_\beta(-x)\right)\left(\sum_{\beta'} \widehat{g}(\beta') e_{\beta'}(-x)\right) = \sum_\alpha \left(\sum_{\beta+\beta'=\alpha} \widehat{f}(\beta)\widehat{g}(\beta')\right) e_\alpha(-x) \succeq 0.$$

For the second claim we first note that since D is a density, for any α we have

$$|\widehat{D}(\alpha)| = \left|\mathop{\mathbb{E}}_{x \sim D} e_\alpha(x)\right| \leq \mathop{\mathbb{E}}_{x \sim D} |e_\alpha(x))| = 1;$$

in fact, this applies more generally to any function D with $\|D\|_1 \leq 1$. Then we compute

$$f(x) - (D * f)(x) = \sum_\alpha \widehat{f}(\alpha)(1 - \widehat{D}(\alpha)) e_\alpha(-x) \succeq 0,$$

where we have used that $D(-x) = D(x)$ and so D has Fourier coefficients $\widehat{D}(\alpha) \in \mathbb{R}$. ☐

Corollary 5.5 (Positive correlation for spectrally positive functions)**.** *Suppose*

$$f_1, f_2, \ldots, f_k \succeq 0$$

are some spectrally positive functions on G. Then

$$\mathbb{E}[f_1 f_2 \cdots f_k] \geq \mathbb{E}[f_1]\,\mathbb{E}[f_2] \cdots \mathbb{E}[f_k].$$

[6] By this we mean that $D(-x) \equiv D(x)$.

Proof. In light of the closure property for multiplication, it suffices to verify this for just two functions. Let $f, g \succeq 0$. Write $g \equiv (g - \mathbb{E}[g]) + \mathbb{E}[g]$, and express

$$\langle f, g \rangle = \langle f, g - \mathbb{E}[g] \rangle + \mathbb{E}[f] \cdot \mathbb{E}[g].$$

We note that the function $F(x) := f(x) \cdot (g(x) - \mathbb{E}[g])$ is spectrally positive. In particular,

$$\langle f, g - \mathbb{E}[g] \rangle = \mathbb{E}[F] = \widehat{F}(0) \geq 0.$$ □

Corollary 5.6 (Odd central moments)**.** *Consider a spectrally positive density D on G. The odd central moments of D are non-negative. That is,*

$$\mathop{\mathbb{E}}_{x \in G} (D(x) - 1)^k \geq 0$$

for all odd integers $k \in \mathbb{N}$.

Proof. $F(x) := D(x) - 1 \succeq 0$. Also, $F(x)^k = F(x)F(x) \cdots F(x) \succeq 0$; in particular $\mathbb{E}[F] \geq 0$. Alternatively, consider this as a special case of Corollary 5.5. □

The utility of this is due to the following.

Proposition 5.7. *For a real-valued random variable X and $k \geq 1$, we use the notation*

$$\|X\|_k := \left(\mathbb{E} \left| X \right|^k \right)^{1/k}.$$

Suppose X is such that

- $\mathbb{E}(X - 1)^k \geq 0$ *for all odd $k \in \mathbb{N}$, and*
- $\|X - 1\|_{k_0} \geq \varepsilon$ *for some even $k_0 \geq 2$ and some $\varepsilon \in [0, \frac{1}{2}]$.*

Then, for any integer $k' \geq 2k_0/\varepsilon$,

$$\|X\|_{k'} \geq 4^{-1/k'}(1 + \varepsilon) \geq 1 + \frac{\varepsilon}{2}.$$

We note that one can infer a statement of this sort already from the following observation: for any odd $k \geq k_0$ we have

$$\mathbb{E}(X - 1)^k = \mathbb{E} \left| (X - 1)_+ \right|^k - \mathbb{E} \left| (X - 1)_- \right|^k \geq 0$$

and

$$\mathbb{E} \left| X - 1 \right|^k = \mathbb{E} \left| (X - 1)_+ \right|^k + \mathbb{E} \left| (X - 1)_- \right|^k \geq \varepsilon^k,$$

so it follows that

$$\mathbb{E} \left| (X - 1)_+ \right|^k \geq \frac{\varepsilon^k}{2}.$$

From here one can get a reasonable lower bound on some $\|X\|_{k'}$ e.g. by a basic pruning argument analogous to the proof of the Paley–Zygmund inequality. However, to obtain nicer constants we give a somewhat different proof which can be found in the appendix (see Appendix D).

5.3 Local variants

We proceed to give some local variants of the statements above which will be needed for setting $A \subseteq [N] \subseteq \mathbb{Z}$. We recall the notation from Definition 4.12,

$$\|f\|_{k,D} := \left\langle D, |f|^k \right\rangle^{1/k} = \left(\mathop{\mathbb{E}}_{x \sim D} |f(x)|^k \right)^{1/k}.$$

Lemma 5.8 (Local decoupling inequality)**.** *Let G be a finite abelian group, and suppose D is a spectrally-positive density function on G. Let $(T_\theta D)(x) := D(x-\theta)$ be some translate of D. For arbitrary real-valued functions f, g on G, and for all even $k \in \mathbb{N}$, we have*

$$\|f * g\|_{k,T_\theta D}^2 \leq \|f \star f\|_{k,D} \|g \star g\|_{k,D}.$$

In particular, if A is a density function then

$$\|A * A - 1\|_{k,T_\theta D} \leq \|A \star A - 1\|_{k,D}.$$

Proof. We have

$$\begin{aligned}
\left\langle T_\theta D, (f*g)^k \right\rangle &= \sum_\gamma \widehat{D}(\gamma) e_\gamma(\theta) \sum_{\beta_1+\beta_2+\cdots+\beta_k=\gamma} \prod_{i=1}^k \widehat{f}(\beta_i)\widehat{g}(\beta_i) \\
&\leq \sum_\gamma \sum_{\beta_1+\beta_2+\cdots+\beta_k=\gamma} \widehat{D}(\gamma) \prod_{i=1}^k |\widehat{f}(\beta_i)|\ |\widehat{g}(\beta_i)| \\
&\leq \sum_\gamma \sum_{\beta_1+\beta_2+\cdots+\beta_k=\gamma} \widehat{D}(\gamma) \left(\frac{\eta}{2} \prod_{i=1}^k \left|\widehat{f}(\beta_i)\right|^2 + \frac{1}{2\eta} \prod_{i=1}^k |\widehat{g}(\beta_i)|^2 \right) \\
&= \frac{\eta}{2} \left\langle D, (f \star f)^k \right\rangle + \frac{1}{2\eta} \left\langle D, (g \star g)^k \right\rangle
\end{aligned}$$

for any choice of $\eta > 0$. Optimizing over η, we conclude that

$$\left\langle T_\theta D, (f*g)^k \right\rangle \leq \sqrt{\langle D, (f \star f)^k \rangle} \sqrt{\langle D, (g \star g)^k \rangle}.$$

□

Lemma 5.9 (Lower bounds from upper bounds – local variant)**.** *If*

$$\|A \star A - 1\|_{k,D} \geq \varepsilon$$

for some $\varepsilon \in [0, \frac{1}{2}]$ and spectrally-positive density D, then

$$\|A \star A\|_{k',D} \geq 1 + \frac{\varepsilon}{2}$$

for some $k' \leq O(k/\varepsilon)$.

Proof. Let k be an integer. We note that

$$D \cdot (A \star A - 1)^k = D \cdot ((A-1) \star (A-1))^k$$

is a spectrally-positive function on G. In particular,

$$\mathop{\mathbb{E}}_{x \sim D} ((A \star A)(x) - 1)^k = \mathop{\mathbb{E}}_{x \in G} D(x)((A \star A)(x) - 1)^k \geq 0.$$

So, we may apply Proposition 5.7 to the random variable $X := (A \star A)(x)$ where $x \sim D$. □

6 Finding a 3-progression in $A \subseteq [N]$ – an overview

We describe how to modify our approach above to address the 3-progression problem in the setting $A \subseteq [N] \subseteq \mathbb{Z}$. We recall our result regarding this problem.

Theorem 6.1 (Theorem 1.2, restated)**.** *Suppose $A \subseteq [N]$ has density $\mu \geq 2^{-d}$. Then the number of triples $(x, y, z) \in A^3$ with $x + y = 2z$ is at least*

$$2^{-O(d^{12})} \cdot N^2.$$

We review the well-known approximate equivalence of the 3-progression problem in the setting $A \subseteq [N]$ with the same problem in the setting $A \subseteq \mathbb{Z}_N$. One may embed $A \subseteq [N]$ naturally inside $\mathbb{Z}_{N'}$, say for some $N' \approx 3N$, so that (i) the density of A in its container decreases only slightly and (ii) we do not obtain any new solutions to

$$x + y = 2z \bmod N'$$

which were not present already. The natural reduction in the other direction is even easier: in this direction, we suffer no loss in density.

So, it suffices to consider the problem of finding many 3-progressions in the setting $A \subseteq \mathbb{Z}_N$, $|A| = 2^{-d}N$. Ultimately we will not take quite this approach, but we consider this setting first as it is more directly comparable with the setting $A \subseteq \mathbb{F}_q^n$. Let us discuss the compact summary of the approach for $\mathbb{F}_q^n$ given at the end of Section 2.3. All of our steps are, in fact, more generally applicable to any finite abelian group G, up until the last point, where we use Sanders' invariance lemma to find an increment onto a large subspace. This is still not a problem: applying a more general form of Sanders' lemma can readily prove a statement of the following sort. Suppose $A \subseteq \mathbb{Z}_N$ has noticeably fewer than the expected number of 3-progressions. In that case, we can obtain a density increment onto a choice of either (i) some (translate of a) large Bohr set B of rank at most $O(d^8)$, or (ii) some large generalized arithmetic progression P of rank $O(d^8)$. The problem comes when we try to iterate this. Unlike the situation before, the container-sets B and P are not structurally isomorphic again to some finite abelian group, so our techniques no longer apply.[7] This is not to say that such an approach (i.e., iteratively seeking density increments onto large approximate subgroups) does not work, only that we cannot analyze it; indeed, it seems likely that it should work. To be concrete, we ask the following technical question.

Question 6.2. *Suppose $A \subseteq B$, where $B \subseteq \mathbb{Z}_N$ is a Bohr set of rank r, and we have local density*

$$\mu = \frac{|A|}{|B|} \geq 2^{-d}.$$

Can it be shown that we must have either (i) at least

$$2^{-\mathrm{poly}(r)} \cdot \mu \cdot |A|^2$$

solutions to $x + y = 2z$ with $x, y, z \in A$, or else (ii) a density increment

$$\frac{|A \cap B'|}{|B'|} \geq (1 + \Omega(1)) \cdot \mu$$

of A onto some affine Bohr set B' with

[7] We could still apply them in a literal sense – everything is still a subset of some finite group $G = \mathbb{Z}_N$ – but quantitatively our techniques become trivial because we are constantly comparing to the uniform density function 1 on G.

- *rank* $r' \leq r + \mathrm{poly}(d)$ *and*
- *size* $|B'| \geq 2^{-\mathrm{poly}(r,d)}|B|$ *?*

Answering this question seems maybe not too far out of reach, but in this work, we were unsuccessful in developing sufficiently strong local variants of our techniques that could address it. A positive answer to this question would surely give the "right" version of our proof.[8] Here instead, we will apply some tricks and cut some corners.

We proceed to discuss the overall structure of the arguments in the earlier works of Roth and Szemerédi and Heath-Brown for addressing the 3-progression problem in the integers. More specifically, we follow the interpretation of these works given in [Gow01] and [Gre99]. In what follows, we refer to the arguments of both works together simply as the "early approach" (as opposed to the "modern approach" developed by Bourgain and refined by Sanders, which emphasizes the use of Bohr sets [Bou99, Bou08, San12a]).

In the early approach, one considers an ad hoc passage back and forth from the setting $A \subseteq \mathbb{Z}$ into various cyclic groups to facilitate the density increment argument. We outline the main points.

- We begin with a set $A \subseteq [N]$ of size $|A| = \mu N$ which we are dissatisfied with: it has much fewer than $\mu|A|^2$ solutions to $x + y = 2z$ with $x, y, z \in A$.
- We consider a sort of "temporary" embedding of A into some cyclic group $\mathbb{Z}_{N'}$ where we can do Fourier analysis. This embedding must be notably more efficient than the simple one described above. Ultimately, we will obtain only a small density increment $(1 + \varepsilon)$ at every iteration, so we cannot afford to lose a factor 3 repeatedly. Additionally, we (roughly) need the embedding to be such that A still has substantially fewer solutions to $x + y = 2z \bmod N'$ than expected, which is in tension with the previous constraint.
- Given such an embedding, we argue via Fourier analysis that we may obtain a density increment onto some large arithmetic progression $P \subseteq \mathbb{Z}_{N'}$.
- We must then argue that P can be pulled back to some "genuine" progression $P' \subseteq [N] \subseteq \mathbb{Z}$, so that we obtain a density increment of A onto P'. This limits the kind of progressions we can allow in the previous step. Finally, we observe that the progression P' is structurally isomorphic to the interval $[|P'|]$, which allows us to iterate the argument.

Let us offer a comment which we find clarifying. There is an alternative interpretation of this approach where the actual algorithm is quite simple and natural. The step where we briefly consider some "virtual" embedding into a cyclic group can be relegated to the analysis. Let us elaborate. For a set $A \subseteq [N]$ of size $|A| = \mu N$, say that a subset $A' \subseteq A$ is satisfactory if there are roughly at least $\mu|A'|^2$ solutions to $x + y = 2z$ with $x, y, z \in A'$. We can make a conceptual distinction between the procedure used to find such a subset and the argument proving that the procedure is successful. In the context of the above, we note that if we are dissatisfied with our current subset, the only recourse which is ultimately available to us is to pass to some restriction $A \cap P$ onto a large progression $P \subseteq \mathbb{Z}$ where we obtain a density increment.

Thus, one can give a quite clear description of the algorithm implicit in the early approach, which need not mention cyclic groups: While there is a density increment onto some large progression

[8] Moreover, an approach that successfully answers this question would likely also produce a pleasant analog to our structural result (Lemma 1.14) in the finite field setting. It is not immediately clear what such an analog should look like, specifically.

$P \subseteq \mathbb{Z}$ available, take it. Throughout, hold on to the current container-set P, starting with $P = [N]$, and measure density relative to it. Once there are no more increments available: conclude – we have found our candidate subset $A' \subset A$. From here, it can be considered a separate matter to argue that A' is satisfactory: if it is not, we argue (e.g., by considering a "virtual" embedding into some cyclic group, if we like) that we could obtain one further increment – a contradiction.

A natural extension of this simple algorithm is to broaden the class of allowed container-sets P to include low-rank generalized progressions; this is quite analogous to the allowance of subspaces of codimension larger than one in the setting of $\mathbb{F}_q^n$. This comes morally very close to the algorithm we will use, and we will also analyze it in much the same way as above: by making some ad hoc reductions to the setting of finite groups where our techniques apply. We give a more detailed overview of our analysis in Section 6.1. For now, we focus on fully specifying our algorithm.

We will need to consider an algorithm stronger than the one described above to make up for a specific weakness of our analysis. Namely, this weakness is related to the last step in the early approach described above, where we would like to pull back a progression in a cyclic group to a "genuine" progression in the integers. This step is already costly for arithmetic progressions; in the early approach, one obtains a progression $P \subseteq [N]$ no larger than $\sqrt{N}$. Indeed, the situation is only worse for generalized progressions, and it will be imperative that we somehow avoid such a substantial size loss.

We will be naturally led to consider the (only mildly) more general problem of establishing the existence of many 3-progressions in a set $A \subseteq P \subseteq \mathbb{Z}^r$, where P is some set of the form $P = \prod_{i=1}^r [N_i] \subseteq \mathbb{Z}^r$, and again we have density

$$\mu = \frac{|A|}{|P|} \geq 2^{-d},$$

and furthermore r is assumed to be no larger than $d^{O(1)}$. We call the set A and its container-set P together a "configuration". We consider some more liberal implementations of the density increment framework which would still suffice for lower bounding the number of 3-progressions in A.

In the broadest sense, whenever we are dissatisfied with $A \subseteq P$, we would like to establish the existence of some related configuration $A' \subseteq P'$, possibly lying in a slightly larger-dimensional space $\mathbb{Z}^{r'}$, where

- A' has no more solutions to $x + y = 2z$ than does A,
- the container set P' is still highly structured – in our setting: of the form $\prod_{i=1}^{r'} [N_i']$,
- the density of A' in its container has increased – in our setting by some constant factor $(1 + \Omega(1))$, and
- The size of A' has not decreased too substantially.

To this end, let us say that $A \subseteq P$ is "spread" roughly when it has no possible "density increments" satisfying the criteria above (we will settle on a precise formulation shortly). Then, our procedure for "locating" a satisfactory configuration $A' \subseteq P'$ is as follows: while there are any good density increments available, take one of them. Otherwise, our resulting configuration is spread, and we are left to argue that this forces it to be satisfactory.

Definition 6.3 (Good increments and spread configurations)**.** *Fix some constants $c, K \in \mathbb{N}$. Suppose that $A \subseteq P \subseteq \mathbb{Z}^r$, where P is of the form $[N_1] \times [N_2] \times \cdots \times [N_r]$, and A has density*

$$\frac{|A|}{|P|} \geq 2^{-d}$$

in its container, for some $d \geq 1$.

Let A' be a subset of A, and let ϕ be a labelling of elements $a \in A'$ by points $p \in P' = \prod_{i=1}^{r'} [N_i'] \subseteq \mathbb{Z}^{r'}$; that is, ϕ is some injection $\phi : A' \to P'$.

For $\varepsilon > 0$, we say that (A', ϕ) is a $(1+\varepsilon)$-good increment (or just a "good increment", suppressing the dependence on ε in addition to the dependence on c and K) if the following conditions are satisfied.

1. *The labelling $\phi : A' \to \mathbb{Z}^{r'}$ is a Freiman homomorphism of order 2.*
2. *We have the density increment*
$$\frac{|A'|}{|P'|} \geq (1+\varepsilon) \frac{|A|}{|P|}.$$
3. *We have bounded dimension growth*
$$r' \leq r + Kd^c.$$
4. *We have bounded size loss*
$$\lg |A'| \geq \lg |A| - Kd^c - Kr^c.$$

If the configuration $A \subseteq P$ has no $(1+\varepsilon)$-good increments, we say that A is $(1+\varepsilon)$-spread (relative to P).

We see that this quite liberal type of density increment (that is, one requiring merely a 2-homomorphism) is indeed still useful for investigation of 3-progressions: if $x, y, z \in A'$ satisfy

$$\phi(x) + \phi(y) = 2 \cdot \phi(z) = \phi(z) + \phi(z) \text{ (in } \mathbb{Z}^{r'}),$$

then

$$x + y = z + z = 2 \cdot z \text{ (in } \mathbb{Z}^r).$$

Thus, the number of 3-progressions does not increase when we pass from A' to $\phi(A')$.

Proposition 6.4 (Passing to a spread configuration)**.** *Fix a choice of constants (c, K, ε) quantifying spreadness.*

Suppose $A \subseteq [N]$ has density at least 2^{-d}. Then there is a subset $A' \subseteq A$ and a Freiman 2-homomorphism

$$\phi : A' \to P' = [N_1] \times [N_2] \times \cdots \times [N_r] \subseteq \mathbb{Z}^r$$

such that

- *$\phi(A')$ is spread relative to its container P',*
- *$r \leq O(d^{c+1})$, and*
- *$\lg |A'| \geq \lg N - O\left(d^{c^2+c+1}\right)$.*

Proof. Let $A_0 := A$ and $P_0 := [N]$. We consider the "greedy algorithm" which at each step passes from the current configuration to some good increment, if one exists. In this way we produce a sequence of

- configurations $A_i \subseteq P_i \subseteq \mathbb{Z}^{r_i}$,
- subsets $A'_{i-1} \subseteq A_{i-1}$, and
- bijections $\phi_i : A'_{i-1} \to A_i$

such that in addition, each of the maps ϕ_i is a 2-homomorphism. The density of the i-th configuration is at least

$$\frac{|A_i|}{|P_i|} \geq (1+\varepsilon)^i 2^{-d} \geq 2^{\varepsilon i - d},$$

and so this algorithm must terminate within $n \leq d/\varepsilon$ iterations. We are left with a configuration $A_n \subseteq P_n$ which is guaranteed to be spread. Taking the composition $\phi := \phi_n \circ \phi_{n-1} \circ \cdots \circ \phi_1$ gives a map which is a bijection between A_n and its preimage $\phi^{-1}(A_n) \subseteq A$, and it is also a 2-homomorphism from the preimage $\phi^{-1}(A_n)$ to $\mathbb{Z}^{r_n}$.

Since the density never decreases, we can bound r_n simply by

$$r_n \leq 1 + n \cdot K(d^c + 1) \leq O(d^{c+1}).$$

Since $r_i \leq r_n$ for all i, we can lower-bound the size of A_n simply by

$$\lg |A_n| \geq |A| - n \cdot O(d^c + r^c) \geq \lg |A| - O(d^{c^2+c+1}) = \lg N - O(d^{c^2+c+1}). \qquad \square$$

With the density-increment framework in place, the task of proving Theorem 1.1 is reduced to the task of proving the following.

Lemma 6.5 (Spread configurations have many 3-progressions)**.** *For some fixed choice of constants (c, K, ε) quantifying spreadness, the following holds.*

Suppose that $A \subseteq [N_1] \times [N_2] \times \cdots \times [N_r] \subseteq \mathbb{Z}^r$ has density μ relative to its container, and that A is spread relative to its container. Then there are at least

$$r^{-O(r)} \cdot \mu \cdot |A|^2$$

solutions to $x + y = 2z$ with $(x, y, z) \in A^3$.

We in fact prove the following more specific formulation.

Lemma 6.6. *Suppose $A \subseteq [N_1] \times [N_2] \times \cdots \times [N_r] \subseteq \mathbb{Z}^r$ has density $\mu \geq 2^{-d}$.*

Either the number of triples $(x, y, z) \in A^3$ with $x + y = 2z$ is at least

$$r^{-O(r)} \cdot \mu \cdot |A|^2,$$

or there is a $(1 + 2^{-10})$-good increment (A', ϕ) mapping into $\mathbb{Z}^{r'}$, specifically with

- *$r' \leq r + O(d^8)$ and*
- *$|A'| \geq r^{-O(r)} \cdot 2^{-O(d^2 r)} \cdot 2^{-O(d^{10})} \cdot |A|$.*

Given these specific parameters, we obtain Theorem 1.2 by using essentially the density-increment framework described above but with some optimizations made to certain details.

Proof of Theorem 1.2. We begin with a set $A \subseteq [N]$ with density at least 2^{-d}. We (repeatedly) take any available $(1+2^{-10})$-increment specified by a subset which is smaller than our current set only by a factor $r^{-O(r)} \cdot 2^{-O(d^{10})}$ and a 2-homomorphism ϕ into $\mathbb{Z}^{r'}$ for some dimension r' exceeding the current dimension by only $O(d^8)$. After some number (say n) of such increments, we arrive at some configuration $A_n \subseteq P_n \subseteq \mathbb{Z}^{r_n}$ for which no further increment is possible, and indeed, we must have $n \leq O(d)$. Thus, also we have

- $r_n \leq O(d^9)$ and, noting the asymptotic bound $r_n \lg(r_n) \leq O(d^{10})$,
- $\lg |A_n| \geq \lg |A| - n \cdot O(r_n \lg r_n) - n \cdot O(d^2 r_n) - n \cdot O(d^{10}) \geq \lg N - O(d^{12})$.

We apply Lemma 6.6 to the final configuration $A_n \subseteq P_n$. Since we are in the "spread" case, we must find at least

$$r_n^{-O(r_n)} \cdot 2^{-d} \cdot |A_n|^2 \geq 2^{-O(d^{12})} \cdot N^2$$

3-progressions. □

6.1 Proof overview for Lemma 6.5 and Lemma 6.6

For this overview, let us focus on the one-dimensional case $A \subseteq [N]$, which is sufficient already to illustrate many of the critical points. Our proof of Lemma 6.5 is by contradiction. We assume that the set A has few 3-progressions (i.e., much fewer than the "expected" number, roughly $|A|^3/N$), and we then show that A cannot be spread by exhibiting a density increment satisfying the four criteria in Definition 6.3: We obtain a new configuration via a (i) Freiman 2-homomorphism with (ii) increased density, (iii) bounded dimension growth, and (iv) bounded size loss.

Our strategy is to make ad hoc reductions to the setting where A is instead a subset of a finite group G, where our techniques apply. For example, it follows readily from our prior arguments that if $A \subseteq \mathbb{Z}_N$ has density $\mu \geq 2^{-d}$, and the number of 3-progressions in A deviates substantially from the expected number, then we get a density increment onto some generalized arithmetic progression $P \subseteq \mathbb{Z}_N$.

Lemma 6.7 (Structure vs. Pseudorandomness in $\mathbb{Z}_N$ – special case of Lemma 8.4)**.** *Consider $A \subseteq \mathbb{Z}_N$ of size $|A| \geq 2^{-d}N$.*

Suppose that

$$\|A * A - 1\|_k \geq \Omega(1).$$

Then there exists a (proper) generalized arithmetic progression $P \subseteq G$, with

- *rank at most $r = O(k^4 d^4)$ and*
- *size $|P| \geq 2^{-O(k^5 d^5)} N$*

such that

$$\langle P, A \rangle \geq 1 + \Omega(1).$$

That is, we get a set $P \subseteq \mathbb{Z}_N$ with density-increment

$$\frac{|A \cap P|}{|P|} \geq (1 + \Omega(1)) \frac{|A|}{N}$$

of the form

$$P = \left\{ a + \sum_{i=1}^{r} c_i \cdot x_i \ : \ x_i \in [N_i] \right\}$$

with size $|P| = N_1 N_2 \cdots N_r \geq 2^{-\mathrm{poly}(d,k)} N$.

So, to describe our tentative plan in detail: we plan to embed $A \subseteq [N]$ into $\mathbb{Z}_N$ via the obvious embedding $\phi : x \mapsto x \mod N$, and then to find a density increment onto some subset $\phi(A) \cap P$. We let $A' = \phi^{-1}(\phi(A) \cap P) = A \cap \phi^{-1}(P) \subseteq [N]$ be the preimage. We then compose with the simple Freiman homomorphism ϕ' which takes points $p \in P$ to their "label" $(x_1, x_2, \ldots, x_r) \in [N_1] \times [N_2] \times \cdots \times [N_r] \subseteq \mathbb{Z}^r$ (see Example 7.12). Then, we hope that the pair $(A', \phi' \circ \phi)$ gives us our desired density increment.

There are two distinct issues to address. Firstly, we need to ensure that upon embedding into $\mathbb{Z}_N$, the density $\phi(A) * \phi(A)$ does indeed deviate substantially from 1 at sufficiently many points that we can apply Lemma 6.7. Secondly, we must ensure that the embedding $\phi : A' \to \mathbb{Z}_N$ is a 2-homomorphism.

Although both issues must be addressed, the first should be considered less serious, as explained next. By our assumption that A has few 3-progressions, we have that the number of representations $R_A(2z)$ of $2z$ is much smaller than $|A|^2/N$ for most points $z \in A$. For simplicity, let us assume the $R_A(2z)$ is very small for *all* points $z \in A$ (and, indeed, it is not hard to reduce to this case). For $x \in [N]$, we have

$$(\phi(A) * \phi(A))(\phi(x)) = \frac{N}{|A|^2} R_{\phi(A)}(\phi(x)) = \frac{N}{|A|^2} R_A(x) + \frac{N}{|A|^2} R_A(x+N).$$

We need to avoid the "unlikely" case that identifying each x with $x+N$ results in close approximation $R_{\phi(A)}(\phi(x)) \approx \frac{|A|^2}{N}$ for all but a very tiny number of points x. One has considerable flexibility in applying various ad hoc tricks to avoid this case. For example, one can choose to instead embed into $\mathbb{Z}_{N'}$ for any choice of $N' \in [N, (1+\delta)N]$ for some small parameter δ – for the sake of discussion, say $\delta \leq 1/100$. This gives up a small amount of density in the short term, but it is acceptable to do so to satisfy the hypothesis of Lemma 6.7 since, in the end, we can still obtain an overall increase in density. Another trick one can consider is first passing to some restriction $A \cap I$ for any reasonably large interval $I \subseteq [N]$ with density, say $|A \cap I|/|I| \geq (1 - 1/100)\mu$.

We briefly sketch some details of the specific ad hoc reduction used here. We assume, at only a negligible-factor loss in the density, that A is in fact entirely contained within the interval $U := [2\delta N, N] \subseteq [N]$. Additionally we assume we are in the "nice" case where a reasonably large fraction of A lies in the interval $M = (\frac{N}{2}, \frac{N}{2} + \delta N)$: say $|A \cap M| \geq \Omega(\delta) \cdot |A|$. If we are not in the nice case, we note that by using ideas in Section 7.4 we can pass to a restriction of A to some fairly large interval such that the restriction becomes "nice" (relative to that interval).[9]

The point of these two intervals (the "upper-portion" U and the "middle slice" M of $[N]$) is that they are designed specifically so that if $x, y \in U$ and $z \in M$ then

$$x + y = 2z \mod N$$

only if

$$x + y = 2z.$$

[9] In the one-dimensional case, it would follow already from the fact that A is spread relative to U that A must be "nice". However, the connection between density upper-bounds on structured sets to density lower-bounds on structured sets degrades when we pass to the general case $A \subseteq [N_1] \times [N_2] \times \cdots \times [N_r]$. In contrast, the idea of using translation-invariance remains quantitatively efficient.

In particular, we obtain $R_{\phi(A)}(\phi(2z)) = R_A(2z)$ for all $z \in M$, so the convolution $\phi(A) * \phi(A)$ is indeed much smaller than 1 on all the points $\phi(2z) \in \mathbb{Z}_N$ with $z \in A \cap M$. Since there are at least $\Omega(2^{-d}N)$ such points, we can conclude that, say,

$$\|\phi(A) * \phi(A) - 1\|_k \geq \frac{1}{2}$$

for some $k \leq O(d)$, as desired. Looking ahead, we note that in the general case of $A \subseteq [N_1] \times [N_2] \times \cdots \times [N_r]$, we intend to essentially apply this same trick independently in each of the r coordinates, but with parameter $\delta \approx 1/100r$.

Now we discuss the second key issue: can we ensure that the embedding $\phi : A' \to \mathbb{Z}_N$ is in fact a 2-homomorphism? A notable feature of our approach is that we do not try to ensure that the embedding $\phi : A \to \mathbb{Z}_N$ is a 2-homomorphism with respect to our original set A – this seems difficult to accomplish without conceding an unacceptable amount of density. Instead, we intend to exploit the fact that we only care about the behavior of ϕ on its restriction to $A' \subseteq A$. Our starting point is the following formalization of a commonly-used trick (see Example 7.13), used to embed problems in $\mathbb{Z}$ into $\mathbb{Z}_N$. Suppose $I \subseteq \mathbb{Z}$ is an interval of size $|I| \leq N/t$. Then the natural embedding $\phi : I \to \mathbb{Z}_N$ is a t-homomorphism.

Usually, this trick is applied before the embedding. Typically, one picks a specific interval I, considers the restriction $A \cap I$, and embeds this restriction into $\mathbb{Z}_N$. In contrast, we would like to delay making a specific choice of interval for as long as possible. We proceed as follows. Consider the natural embedding $\phi : [N] \to \mathbb{Z}_N$. We embed A into $\mathbb{Z}_N$ and invoke Lemma 6.7 to find a density-increment $\mu' = |\phi(A) \cap P|/|P| \geq (1 + \Omega(1))\mu$, where P is a generalized progression of rank at most poly(d). Now consider $A' = \phi^{-1}(\phi(A) \cap P) = A \cap \phi^{-1}(P)$, and also consider some partition $[N] = I_1 \cup I_2 \cup I_3$ of $[N]$ into three intervals each of size roughly $N/3$. It must be the case that one of the densities $\mu'_i := |A' \cap I_i|/|I_i|$ is at least as large as μ'; let us start to modify the basic plan laid out above and set $A'' := A' \cap I_i$.

We now have that the embedding $\phi : A'' \to \mathbb{Z}_N$ is a 2-homomorphism, as desired. As stated, a small issue with this plan is that now the container of $\phi(A'')$, $\phi(I_i) \cap P$, is no longer necessarily a generalized progression. Certainly, this container-set, which is the intersection of a progression of rank 1 and a progression of rank r, still has a large amount of additive structure. One way to continue here is to partition $P \cap \phi(I_i)$ into a small number of (still reasonably low-rank) generalized progressions and further restrict onto one of them – this would suffice to complete the proof (for the one-dimensional case $A \subseteq [N]$). This gives a good idea of how we plan to address the second key issue. In the actual proof, we proceed somewhat differently: we mix the idea described here involving the intervals I_1, I_2, I_3 into the *proof* of Lemma 6.7, which works with a low-rank Bohr set B as an intermediate step before arriving at a low-rank progression P inside some translate of B. Compared with generalized progressions, Bohr sets are nicer because the intersection of two Bohr sets is again a Bohr set, which explains how we avoid the small issue encountered above. We note that for the general case of $A \subseteq [N_1] \times [N_2] \times \cdots \times [N_r]$, our corresponding generalization of the trick here involving the intervals I_1, I_2, I_3 is developed and formalized in Section 7.5. See in particular Definition 7.16, which defines the notion of a "safe" set, and Proposition 7.17, which provides a reasonably large, safe Bohr set in $\mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2} \times \cdots \times \mathbb{Z}_{N_r}$.

7 Preliminaries for 3-progressions in the integers

7.1 Generalized arithmetic progressions

Definition 7.1 (Generalized progression)**.** *In an abelian group G, a generalized arithmetic progression (or just a "progression") is a set of the form*

$$P = \left\{ a + \sum_{i=1}^{r} c_i \cdot x_i \ : \ x_i \in [N_i] \right\}$$

for some elements $a, c_1, \cdots c_r \in G$. We say that the number r is the "rank" of G.

In the case that every point $p = a + \sum_{i=1}^{r} c_i \cdot x_i \in P$ is represented only once in this form (i.e. the case when $|P| = N_1 N_2 \cdots N_r$), we say that P is a "proper" progression. Since we will be interested only in proper progressions in this work, we often omit the qualifier "proper".

7.2 Bohr sets

Definition 7.2 (Bohr set)**.** *Let G be a finite abelian group with character group $\widehat{G}$. A Bohr set of rank 1 in G is a set of the form*

$$\{x \in G \ : \ |\gamma(x) - 1| \leq \rho\}$$

for some character $\gamma \in \widehat{G}$ and some $\rho \geq 0$, which we call the radius. A Bohr set of rank r is a set in G describable as the intersection of at most r such sets.

Given a set $\Gamma \subseteq G$, we use the notation

$$\mathrm{Bohr}(\Gamma, \rho)$$

to denote the rank $|\Gamma|$ Bohr set

$$\{x \in G \ : \ |\gamma(x) - 1| \leq \rho \text{ for all } \gamma \in \Gamma\}.$$

The following is an example of a simple connection between progressions and Bohr sets which we'll make use of at a few points.

Example 7.3 (A centered interval is a Bohr set)**.** *Consider the rank-1 Bohr set in $\mathbb{Z}_N$ corresponding to the character $e_1 = (x \mapsto e^{2\pi i x/N})$:*

$$\mathrm{Bohr}(\{e_1\}, \rho) = \left\{ x \in \mathbb{Z}_N \ : \ |e^{2\pi i \frac{x}{N}} - 1| \leq \rho \right\}.$$

For any given $\rho \in [0, 1]$, this set is simply some interval $[-m, m] \subseteq \mathbb{Z}_N$ with

$$\frac{\rho}{2\pi} \leq \frac{m}{N} \leq \frac{\rho}{4}.$$

7.3 Properties of Bohr sets

Given a Bohr set $B = \mathrm{Bohr}(\Gamma, \rho)$, we denote the *dilation* of B by δ by

$$B_\delta := \mathrm{Bohr}(\Gamma, \delta\rho).$$

We note the straightforward sumset inclusion

$$B + B_\delta \subseteq B_{1+\delta}.$$

A convenient fact about Bohr sets and their dilations is that we can easily give some good approximate bounds on their size.

Proposition 7.4 (Bohr set size estimates [TV06, Section 4.4])**.** *Suppose $B = \mathrm{Bohr}(\Gamma, \rho) \subseteq G$ is a Bohr set of rank $|\Gamma| = r$ and with $\rho \in [0, 2]$. We have the size estimates*

- $|B| \geq \left(\frac{\rho}{2\pi}\right)^r |G|$,
- $|B_2| \leq 6^r |B|$, *and*
- *for any $\delta \in [0, 1]$, $|B_\delta| \geq \left(\frac{\delta}{2}\right)^r |B|$.*

The "doubling" estimate

$$|B + B| \leq |B_2| \leq 6^r |B|$$

shows that B is in some sense an "approximate subgroup" – it is (quantitatively) nearly closed under addition, assuming that one considers the factor 6^r to be small. In settings where this factor cannot be considered small, it can be useful to consider instead a slightly different quantification of approximate closure under addition, which motivates the following definition.

Definition 7.5 (Regular Bohr set)**.** *A Bohr set B of rank r is regular if, for all $\delta \in [0, \frac{1}{12r}]$,*

$$\frac{|B_{1+\delta}|}{|B|} \leq 1 + 12r\delta$$

and

$$\frac{|B_{1-\delta}|}{|B|} \geq 1 - 12r\delta.$$

The point here then is that we have the "doubling" estimate

$$|B + B_\delta| \leq |B_{1+\delta}| \leq 2|B|$$

for $\delta \leq 1/12r$. Compared to the bound above, we have removed the exponential dependence on r in the doubling constant at the cost of an exponential-in-r factor loss in the size of one of the summands, which is more acceptable in certain contexts.

Fortunately, regular Bohr sets are easy to obtain:

Proposition 7.6 (Regularizing a Bohr set [TV06, Section 4.4])**.** *Given a Bohr set B, there is some $\delta \in [\frac{1}{2}, 1]$ so that B_δ is regular.*

Ultimately, our interest in Bohr sets in cyclic groups is due to the fact that large Bohr sets are guaranteed to contain a generalized progression which is still fairly large.

Proposition 7.7 (Large progression in a Bohr set [TV06, Proposition 4.23])**.** *Let G be a cyclic group of size N, and let Let $B = \mathrm{Bohr}(\Gamma, \rho) \subseteq G$ be a Bohr set with*

- *rank $|\Gamma| \leq r$ and*
- *radius $\rho \in [0, 1]$.*

Then B contains a (proper) progression P of rank r and size

$$|P| \geq \left(\frac{\rho}{2\pi r}\right)^r N.$$

7.4 Translation invariance for approximate subgroups

Proposition 7.8 (Smoothing an approximate subgroup)**.** *Let A, B be finite subsets of an abelian group G. Here we would also like to allow for infinite groups, so in the present context we switch to the counting measure: we define the distribution functions*

- $\pi_A(x) := \frac{\mathbb{1}_A(x)}{|A|}$,
- $\pi_B(x) := \frac{\mathbb{1}_B(x)}{|B|}$.

We define the convolution of distributions $\pi_A * \pi_B$ according to the counting measure,

$$(\pi_A * \pi_B)(x) := \sum_{y \in G} \pi_B(y)\pi_A(x-y),$$

so that the convolution of two distributions is again a distribution (i.e. $\sum_x (\pi_A * \pi_B)(x) = 1$).

Suppose S is a finite set in G which contains the difference-set $A - B$. Then, for all $x \in A$, we have the identity

$$(\pi_S * \pi_B)(x) = \frac{1}{|S|}.$$

More generally, if ν is any distribution supported on B, then still for all $x \in A$ we have

$$(\pi_S * \nu)(x) = \frac{1}{|S|}.$$

Proof. We count the number of pairs $(s, b) \in S \times B$ for which $s + b = x$.

$$\sum_{s,b} \mathbb{1}(s + b = x) = \sum_{s \in S} \mathbb{1}(s \in x - B) = |S \cap (x - B)| = |x - B| = |B|,$$

since $x - B \subseteq A - B \subseteq S$ for any $x \in A$. So we do in fact have

$$(\pi_S * \pi_B)(x) = \frac{|B|}{|S \times B|} = \frac{1}{|S|}$$

whenever $x \in A$.

Now we consider the case of general ν supported on B. We note that ν may be expressed as a convex-combination of flat distributions supported on B – that is, we may write

$$\nu = \mathop{\mathbb{E}}_{B'}[\pi_{B'}]$$

with respect to some probability distribution over subsets $B' \subseteq B$. Then the claim follows from our previous argument since S still contains $A - B'$ for every such B':

$$(\pi_S * \nu)(x) = \mathop{\mathbb{E}}_{B'}[(\pi_S * \pi_{B'})(x)] = \frac{1}{|S|}$$

for $x \in A$. □

Corollary 7.9 (Strong one-sided approximation)**.** *Let $A, B, S \subseteq G$, where $S \supseteq A - B$. Suppose that $|S| \leq (1 + \delta)|A|$ for some $\delta \geq 0$. Then for any nonnegative function $f : G \to \mathbb{R}_{\geq 0}$ and any distribution ν supported on B,*

$$\langle \pi_S * \nu, f \rangle \geq (1 + \delta)^{-1} \langle \pi_A, f \rangle .$$

Proof. For every point $x \in A$, we compare $\pi_A(x) = \frac{1}{|A|}$ with $(\pi_S * \nu)(x) = \frac{1}{|S|}$. $\square$

Corollary 7.10 (Smoothing a regular Bohr set)**.** *Suppose B is a regular Bohr set of rank r in a finite abelian group G. Then, for any nonnegative function f on G,*

$$\langle B_{1+\delta} * B_\delta, f\rangle \geq (1 - 12\delta r)\langle B, f\rangle .$$

Proof. We note that $-B_\delta = B_\delta$ and $B_{1+\delta} \supseteq B + B_\delta$. If $\delta \geq 1/12r$ then the claim is trivial. Otherwise, we have $|B_{1+\delta}| \leq (1 + 12\delta r)|B|$, and so $|B|/|B_{1+\delta}| \geq (1 + 12\delta r)^{-1} \geq 1 - 12\delta r$. $\square$

7.5 Freiman homomorphisms

Definition 7.11 (Freiman Homomorphism)**.** *Suppose we have a set $A \subseteq G$ where G is an abelian group, and G' is another abelian group. A map*

$$\phi : A \to G'$$

is said to be a Freiman homomorphism of order t if, for any $x_1, x_2, \ldots, x_t \in A$ and $y_1, y_2, \ldots, y_t \in A$,

$$\phi(x_1) + \cdots + \phi(x_t) = \phi(y_1) + \cdots + \phi(y_t)$$

implies

$$x_1 + \cdots + x_t = y_1 + \cdots + y_t.$$

In particular, a linear map $\phi : G \to G'$ is a t-homomorphism on A if and only if ϕ is injective on the sumset tA.

While not immediate, it is easy to check (by making a translation) that a t-homomorphism ϕ is also a t'-homomorphism for $t' < t$; in particular any Freiman homomorphism ϕ must at least be an injection. We also point out the trivial property that if $\phi : A \to G'$ is a t-homomorphism, and $A' \subseteq A$, then the restriction of ϕ to A' is also a t-homomorphism.

We use the following simple Freiman homomorphism often.

Example 7.12 (Labelling a proper progression)**.** *Let*

$$P = \left\{a + \sum_{i=1}^{r} c_i \cdot x_i \ : \ x_i \in [N_i]\right\} \subseteq G$$

be a (proper) progression, and define the map

$$\phi : P \to [N_1] \times [N_2] \times \cdots \times [N_r] \subseteq \mathbb{Z}^r$$

by

$$\phi(p) = (x_1, x_2, \ldots, x_r).$$

This map is a Freiman homomorphism of all orders.

Proof. The only potentially tricky point is that ϕ is actually well-defined, which is true only because P is proper. Besides this, it is clear that equality of sums of vectors $\sum_{j=1}^{t} x^j = \sum_{j=1}^{t} y^j$ within $\mathbb{Z}^r$ implies the equality

$$a + \sum_{i=1}^{r} c_i \cdot (x_i^1 + x_i^2 + \cdots + x_i^t) = a + \sum_{i=1}^{r} c_i \cdot (y_i^1 + y_i^2 + \cdots + y_i^t)$$

in G. $\square$

The following is a formalization of a standard trick used (e.g. in [Gow01]) to reduce questions regarding t-progressions in the integers to some corresponding questions regarding t-progressions in $\mathbb{Z}_N$. The trick itself is so simple that it is often presented without any corresponding formalization. However, in preparation for some more complicated extensions which can no longer be reasonably handled "by inspection", we work out the details here with some care.

Example 7.13 (Embedding an interval)**.** *Define the natural map*

$$\phi : [N] \to \mathbb{Z}_N$$

by

$$\phi(x) = x \mod N.$$

Let $I = [a, a+m] \subseteq [N]$ *be some interval with*

$$m < \frac{N}{t}.$$

Then the map ϕ *is a Freiman* t*-homomorphism when restricted to* I*.*

Proof. The sumset $tI \subseteq \mathbb{Z}$ is contained in $I' = [ta, ta+tm]$. Let $x, y \in I'$. If $x \equiv y \mod N$, this means that N divides the distance $|x-y|$. However, this distance is at most

$$(ta + tm) - ta < N.$$

So in fact we must have $|x-y| = 0$. □

It will turn out to be convenient for us if we could find a more "intrinsic" formulation of this trick, in the sense that we would like to find a formulation which refers to a nice set $S \subseteq \mathbb{Z}_N$ rather than a nice set $I \subseteq [N]$. For instance, we might wish to say that ϕ is such that if S is any interval in $\mathbb{Z}_N$ with length at most N/t, then the map ϕ is a t-homomorphism on the pullback $\phi^{-1}(S) \subseteq [N]$. This statement is in fact false (already for $t \geq 2$), which is witnessed by the example $S = \{-1, 0, 1\} \subseteq \mathbb{Z}_N$: under our map, this set has preimage

$$\phi^{-1}(S) = \{1\} \cup \{N-1, N\},$$

and now the sumset $\phi^{-1}(S) + \phi^{-1}(S)$ contains two multiples of N: both N and $2N$. This demonstrates alarmingly the importance of certain "implementation details" of linear maps between groups in the present context which can often be safely ignored in other contexts.

We can however make the following simple observation that the behavior of the family of sets in $[N]$ obtainable from preimages of intervals in $\mathbb{Z}_N$ is not unboundedly bad; in the terminology of [Gow01, Gre99], an interval in $\mathbb{Z}_N$ corresponds to a union of at most two "genuine" intervals:

Observation 7.14. *Fix the map* $\phi : [N] \to \mathbb{Z}_N$ *defined by* $\phi(x) = x \mod N$*. Define the family of "intervals of size* m*" in* $\mathbb{Z}_N$ *by*

$$\mathcal{F}_m = \{a + \phi([m]) \ : \ a \in \mathbb{Z}_N\}.$$ [10]

For any $I \in \mathcal{F}_m$*, the pullback* $I' = \phi^{-1}(I)$ *is either*

[10] Note that we recover the same family if we instead consider sets $a + \phi(I) \subseteq \mathbb{Z}_N$ where $I = [a, b] \subseteq [N]$ is any other interval of size $|I| = m$.

- *an interval* $I' = [a, b] \subseteq [N]$ *of size* $|I'| = m$*, or*
- *the union of two intervals: specifically* $I' = [1, b] \cup [a, N]$*, with* $b < m$ *and* $a > N - m$*.*

We record the following immediate consequence of Example 7.13 and Observation 7.14 – this statement shows that it is possible to obtain the "intrinsic" formulation of Example 7.13 we were looking for, so long as we are willing to concede a small fraction of points in $[N]$.

Proposition 7.15 (A safe set in $\mathbb{Z}_N$)**.** *Fix the map* $\phi : [N] \to \mathbb{Z}_N$ *defined by* $\phi(x) = x \mod N$*. Fix a parameter* $\delta \in (0, \frac{1}{2}]$*, and define the "upper portion" of* $[N]$ *by*

$$U = [\delta N, N] \subseteq [N].$$

Suppose $I \subseteq \mathbb{Z}_N$ *is some interval*[11] *of size* $|I| \leq \delta N$*, and that* $t \leq 1/\delta$*.*

Then, for any translation by some $a \in \mathbb{Z}_N$*,* ϕ *is a* t*-homomorphism on the set*

$$\phi^{-1}(I + a) \cap U.$$

Proof. Using the characterization in Observation 7.14, we see that the set $\phi^{-1}(I + a) \cap U$ is in fact equal to some interval[12] in $[N]$ of size at most δN. This is the case considered in Example 7.13, so the claim follows. □

We introduce the following definition to capture this phenomenon in general.

Definition 7.16 (t-safe set)**.** *Consider two abelian groups* G, G' *and a subset* $A \subseteq G$*. Fix an injection* $\phi : A \to G$*. We say that a set* $B \subseteq G'$ *is "*t*-safe" with respect to* ϕ *if, for every translation* $\theta \in G'$*, the restriction*

$$\phi : A \cap \phi^{-1}(B + \theta) \to G'$$

is a Freiman homomorphism of order t*.*

By a simple generalization of Proposition 7.15, we can describe a highly structured, reasonably large set which is safe with respect to the natural embedding of $[\delta N_1, N_1] \times [\delta N_2, N_2] \times \cdots \times [\delta N_r, N_r]$ into $\mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2} \times \cdots \times \mathbb{Z}_{N_r}$.

Proposition 7.17 (A safe set in $\mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2} \times \cdots \times \mathbb{Z}_{N_r}$)**.** *Fix some natural numbers* $N_1, N_2, \ldots N_r$ *and a parameter* $\delta \in (0, \frac{1}{2}]$*. Consider the set* $[N_1] \times [N_2] \times \cdots \times [N_r] \subseteq \mathbb{Z}^r$*, and define the "upper portion" of this set:*

$$U := \prod_{i=1}^{r} [\delta N_i, N_i].$$

Fix the map $\phi : [N_1] \times [N_2] \times \cdots \times [N_r] \to G = \mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2} \times \cdots \times \mathbb{Z}_{N_r}$ *defined by*

$$\phi(x) = x \mod (N_1, N_2, \ldots, N_r).$$

For $i = 1, 2, \ldots, r$*, let* $m_i \leq \delta N_i$*,* $I_i := [m_i]$*,* $B_i := I_i \bmod N_i \subseteq \mathbb{Z}_{N_i}$*, and let*

$$B := B_1 \times B_2 \times \cdots \times B_r.$$

[11] See Observation 7.14 for what is precisely meant by this.

[12] It may be equal so the "trivial" interval $\emptyset$ – this case is also fine.

If $t \leq 1/\delta$, then B (and hence also any translate of B) is t-safe with respect to the restriction $\phi : U \to G$. That is, the natural embedding $\phi(x) = x \mod (N_1, N_2 \ldots, N_r)$ is a t-homomorphism on the set

$$\phi^{-1}(B + \theta) \cap U$$

for any translation $a \in G$.

Proof. We note that $\phi^{-1}(B+\theta) \cap U = \prod_{i=1}^{r} \phi_i^{-1}(B_i + \theta_i) \cap [\delta N_i, N_i]$, and we apply Proposition 7.15 on each coordinate. To finish, we use the easily verifiable fact that if $\phi_1 : A_1 \to G_1$ and $\phi_2 : A_2 \to G_2$ are both Freiman t-homomorphisms, then the map from $A_1 \times A_2$ to $G_1 \times G_2$ given by

$$(x_1, x_2) \mapsto (\phi_1(x_1), \phi_2(x_2))$$

is also a Freiman t-homomorphism. □

8 Proof of Lemma 6.5 and Lemma 6.6

Lemma 6.5 follows straightforwardly from the following three claims, which are proved in this section. Here, for easier reading we present a slightly informal formulation of each one, suppressing some minor details. The more specific Lemma 6.6 follows from the details found in the corresponding complete formulations.

Proposition 8.1 (Informal version of Proposition 8.10: passing to a "nice" configuration)**.** *Suppose $A \subseteq [N_1] \times [N_2] \times \cdots \times [N_r]$ is a spread configuration. Then there is a large subset of A which is Freiman 2-isomorphic to a "nice" configuration $A' \subseteq P' = [N_1'] \times [N_2'] \times \cdots \times [N_r']$ with the assurance that*

- *$A' \subseteq P'$ is only slightly less spread, and*
- *$|A'| \geq r^{-O(r)}|A|$.*

Proposition 8.2 (Informal version of Proposition 8.9: embedding a "nice" configuration)**.** *Suppose $A \subseteq [N_1] \times [N_2] \times \cdots \times [N_r]$ is a "nice" configuration with density $\mu \geq 2^{-d}$. Then (firstly) the group $G = \mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2} \times \cdots \times \mathbb{Z}_{N_r}$ must be cyclic. Let ϕ be the natural embedding $\phi : A \to G$.*

Either A has at least

$$r^{-O(r)} \cdot \frac{|A|^3}{|G|}$$

3-progressions, or

$$\|\phi(A) \star \phi(A)\|_{k, B \star B} \geq 1 + \Omega(1)$$

for some $k \leq O(d)$ and some regular Bohr set B which

- *has rank r and radius $\rho \approx 1/r$, and*
- *is 2-safe with respect to the embedding $\phi : A \to G$.*

Lemma 8.3 (Informal version of Lemma 8.4: obtaining a density increment)**.** *Suppose A is a subset of a cyclic group G with density $\mu \geq 2^{-d}$, and let B be a regular Bohr set with rank r and radius $\rho \approx 1/r$. If*

$$\|A \star A\|_{k, B \star B} \geq 1 + \Omega(1),$$

then

$$\frac{|A \cap P|}{|P|} \geq (1 + \Omega(1))\, \mu$$

for some generalized progression P which

- *has rank $r' \leq r + \mathrm{poly}(k, d)$,*
- *has size $|P| \geq 2^{-\mathrm{poly}(k,d,r)}|G|$, and*
- *is contained in some translate $B + \theta$.*

Proof of Lemma 6.6. Let $A_0 = A$ and $r_0 = r$. We apply Proposition 8.10 with $\varepsilon = 2^{-9}$. Either this immediately supplies us with the desired density increment, or else we obtain a subset $A_0' \subseteq A_0$ and a 2-homomorphic bijection $\phi_1 : A_0' \to A_1 \subseteq P_1$ where

- $P_1 = [p_1] \times [p_2] \times \cdots \times [p_{r_1}] \subseteq \mathbb{Z}^{r_1}$,
- $|A_1| \geq r^{-O(r)}|A_0|$ and $r_1 \leq r_0$,
- $A_1 \subseteq P_1$ is δ-nice with $\frac{1}{10} \geq \delta \geq \Omega(1/r)$, and
- $\mu_1 := \frac{|A_1|}{|P_1|} \geq (1 - 5\varepsilon)\mu$.

We apply Proposition 8.9. In the case that A_1 has at least

$$r^{-O(r)} \frac{|A_1|^3}{|P|} \geq r^{-O(r)} \frac{|A|^3}{|P|}$$

3-progressions, we are done. Otherwise, we let $\phi_2 : A_1 \to A_2 \subseteq G$ be the natural embedding into the group $G = \mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_{r_1}}$, and we have

$$\|A_2 \star A_2\|_{k, B \star B} \geq 1 + \tfrac{1}{4}$$

for some $k \leq O(d)$ and some large regular Bohr set B which is 2-safe with respect to ϕ_2. Applying Lemma 8.4, we obtain a subset

$$A_2' = A_2 \cap P_3$$

where P_3 is some proper progression which

- has rank $r_3 \leq r + O(d^8)$,
- has size $|P_3| \geq r^{-O(r)} \cdot 2^{-O(d^2 r)} \cdot 2^{-O(d^{10})} \cdot |G|$,
- is contained in some translate $B + \theta$, and
- provides the density increment

$$\mu_3 = \frac{|A_2 \cap P_3|}{|P_3|} \geq \left(1 + \tfrac{1}{32}\right) \mu_1 \geq \left(1 + \tfrac{1}{32}\right) \left(1 - \tfrac{5}{2^9}\right) \mu \geq \left(1 + \tfrac{1}{50}\right) \mu.$$

As the density does not decrease, we note also that

$$|A_2'| \geq r^{-O(r)} \cdot 2^{-O(d^2 r)} \cdot 2^{-O(d^{10})} |A_2|.$$

Since $A_2' \subseteq P_3 \subseteq B + \theta$, the map ϕ_2 is a 2-homomorphism when restricted to $\phi_2^{-1}(A_2')$.

Finally, we let $\phi_3 : P_3 \to \mathbb{Z}^{r_3}$ be the simple Freiman homomorphism described in Example 7.12 and we arrive at our final configuration

$$A_3 = \phi_3(A_2') \subseteq \phi_3(P_3) \subseteq \mathbb{Z}^{r_3}.$$

Letting $\phi := \phi_3 \circ \phi_2 \circ \phi_1$ and $A' := \phi^{-1}(A_3)$, we obtain the desired increment (A', ϕ). □

8.1 Obtaining a density increment

Lemma 8.4 (Structure vs. Pseudorandomness in $\mathbb{Z}_N$ – local variant)**.** *Consider*

- *a cyclic group G.*
- *a subset $A \subseteq G$ of size $|A| \geq 2^{-d}|G|$,*
- *a regular Bohr set $B = \mathrm{Bohr}(\Gamma, \rho)$ of rank $|\Gamma| = r$,*
- *$k \in \mathbb{N}$, and*
- *a constant*[13] *$\varepsilon_0 \in [0, 1]$.*

If

$$\|A \star A\|_{k, B \star B} \geq 1 + \varepsilon_0,$$

then there is a (proper) generalized progression P, contained in some translate $B + \theta$, with

- *rank $r' \leq r + O(k^4 d^4)$ and*
- *size $|P| \geq \rho^{r'} \cdot r^{-O(r)} \cdot 2^{-O(dkr)} \cdot 2^{-O(d^5 k^5)} \cdot N$,*

and such that

$$\frac{|A \cap P|}{|P|} \geq \left(1 + \frac{\varepsilon_0}{8}\right) \frac{|A|}{|G|}.$$

We will need the following translation-invariance lemma due to Schoen and Sisask.

Lemma 8.5 (Special case of [SS16, Theorem 5.4])**.** *Consider*

- *a finite abelian group G,*
- *subsets $X, Y \subseteq G$,*
- *a regular Bohr set $B = \mathrm{Bohr}(\Gamma, \rho) \subseteq G$ with rank $|\Gamma| = r$,*
- *an indicator function $f : G \to \{0, 1\}$, and*
- *$\varepsilon \in [0, 1]$.*

If

$$|Y + B| \leq 2^d |Y|$$

and

$$|X + Y + B| \leq 2^s |X|,$$

then there is a Bohr set $B' = \mathrm{Bohr}(\Gamma', \rho') \subseteq B$ such that

- *$\big|\big((X * Y) \star f\big)(b) - \big((X * Y) \star f\big)(0)\big| \leq \varepsilon$ for all $b \in B'$,*

[13] We allow the implied constants in the O-notation here to depend on ε_0.

- *B' has rank $r' \le r + O\left(ds^3/\varepsilon^2 + ds\log(1/\varepsilon)^2/\varepsilon^2\right)$, and*
- *B' has radius $\rho' \ge \rho \cdot \varepsilon \cdot 2^{-s/2}/(r^2 r')$.*

In particular,

$$|\langle D * X * Y, f\rangle - \langle X * Y, f\rangle| \le \varepsilon$$

for any density function D supported on B'.

It is maybe not immediately clear that this is indeed a special case of Lemma 5.4 of [SS16], so we elaborate. We assume familiarity with the statement as it appears in [SS16]. Let $L \subseteq G$ denote the support of our indicator function $f = \mathbb{1}_L$. The key point is that for our formulation here, one can observe that the only values of f which can possibly play any role are those values $f(z)$ for which z is "near" zero – more accurately: $z \in X + Y + B$. This assertion makes use of the fact that we indeed have $0 \in B$, so that $X + Y \subseteq X + Y + B$. Thus, it makes no difference for us to insist that $L \subseteq X + Y + B$ – that is, to replace f by $f \cdot \mathbb{1}_{X+Y+B}$. Given this, the remaining translation between the statements is straightforward: we apply their Lemma 5.4 with $S = B$. Their sets A, M, L correspond to our sets $Y, X, -L$. Their parameters η, K correspond to our parameters $2^{-s}, 2^d$. Their lemma provides a bound on

$$\left| \mathop{\mathbb{E}}_{\substack{x\in X\\ y\in Y}} f(x+y+z+b) - \mathop{\mathbb{E}}_{\substack{x\in X\\ y\in Y}} f(x+y+z) \right|$$

for all $b \in B'$ and all $z \in G$, whereas we seek to control this difference only at the single point $z = 0$.

Proof of Lemma 8.4. By assumption we have

$$\left\langle B \star B, |A \star A|^k \right\rangle = \left\langle B, B * (A \star A)^k \right\rangle \ge (1+\varepsilon_0)^k .$$

We set $\delta_0 = \varepsilon_0/100r$, and pick $\delta \in [\delta_0/2, \delta_0]$ so that B_δ is regular. We apply Corollary 7.10 to get

$$\left\langle B_{1+\delta} * B_\delta, B * (A \star A)^k \right\rangle \gtrapprox (1+\varepsilon_0)^k ,$$

yielding

$$\left\langle B_{1+\delta} * B_\delta, B * (A \star A)^k \right\rangle^{1/k} \ge 1 + \frac{\varepsilon_0}{2}.$$

It follows that for some fixed $\theta \in B_{1+\delta}$, the translate $C := -B + \theta = B + \theta$ satisfies

$$\left\langle B_\delta * C, (A \star A)^k \right\rangle^{1/k} \ge 1 + \frac{\varepsilon_0}{2}.$$

By (possibly) increasing k only slightly we may assume that $2^{1-\varepsilon_0 k/4} \le 2^{-5}\varepsilon_0$. Let $\varepsilon = \varepsilon_0/4$ and apply Corollary 4.13 to get dense subsets $X \subseteq C$ and $Y \subseteq B_\delta$, both with relative density at least $2^{-O(kd)}$, such that

$$\langle X * Y, f\rangle \le \frac{\varepsilon}{8},$$

where

$$f(x) := \mathbb{1}\big((A \star A)(x) \le 1 + \varepsilon\big) .$$

Now, we invoke Lemma 8.5 with sets X, Y and with Bohr set B_η, for some $\eta \in [\frac{\delta^2}{2}, \delta^2]$, to get a final Bohr set

$$B' = \mathrm{Bohr}(\Gamma', \rho') \subseteq B_\eta$$

full of translations which leave $X * Y$ approximately fixed, in the limited sense that they have little effect on the value $\langle X * Y, f\rangle$. We then pass to a large progression $P \subseteq B'$ with

$$\langle P * X * Y, f\rangle \leq \frac{\varepsilon}{4}.$$

Since

$$|Y + B_\eta| \leq |B_\delta + B_\eta| \leq |B_{\delta+\delta^2}| \leq 2|B_\delta| \leq 2^{O(dk)}|Y|,$$

and

$$|X + Y + B_\eta| \leq |(B + \theta) + B_{\delta+\delta^2}| = |B + B_{\delta+\delta^2}| \leq |B_{1+\delta+\delta^2}| \leq 2|B| \leq 2^{O(dk)}|X|,$$

we obtain a Bohr set B' with parameters

- rank $r' \leq r + O(d^4k^4)$ and
- radius $\rho' \geq \rho \cdot r^{-O(1)} \cdot 2^{-O(dk)}$,

and a progression $P \subseteq B' \subseteq B$ with parameters

- rank $r' = r + O(d^4k^4)$ and
- size $|P| \geq \rho^{r'} \cdot r^{-O(r)} \cdot 2^{-O(dkr)} \cdot 2^{-O(d^5k^5)} \cdot N$,

Since $\langle P * X * Y, f\rangle \leq \varepsilon/4$, it must be the case that $\langle P', f\rangle \leq \varepsilon/4$ for some translate P' of P. Recalling the definition of f, we obtain the lower-bound

$$\left\langle P', A \star A\right\rangle \geq \left(1 - \frac{\varepsilon}{4}\right)(1 + \varepsilon) \geq 1 + \frac{\varepsilon}{2}.$$

Again, it follows that

$$\left\langle P'', A\right\rangle \geq 1 + \frac{\varepsilon}{2} = 1 + \frac{\varepsilon_0}{8},$$

for some further translate P'' of P', which concludes the proof since P'' is itself a progression with the same size and rank as P. □

8.2 Embedding a nice configuration

Definition 8.6. *Fix a parameter $\delta > 0$. For an interval $I = [a, b] \subseteq \mathbb{Z}$, the "upper portion" of I is the subset*

$$U := [a + 2\delta(b - a), b],$$

and the "middle slice" of I is the subset

$$M := \left[a + \tfrac{1}{2}(b - a) + 1, a + (\tfrac{1}{2} + \delta)(b - a)\right].$$

The point of these ad hoc definitions is that if we consider an interval $I = [N] \subseteq \mathbb{Z}$, if $x, y \in U$ and $z \in M$, then

$$x + y = 2z \mod N$$

only if

$$x + y = 2z.$$

This can be checked by inspection: since $x + y, 2z \in [2N]$, the only way we can have N dividing $|x + y - 2z|$ while $|x + y - 2z| \neq 0$ is that $|x + y - 2z| = N$. The sets U and M are designed specifically so that (for any choice of $\delta \in [0, \frac{1}{2}]$) this is not possible.

Definition 8.7. *Fix a parameter $\delta > 0$. For some intervals $I_i \subseteq \mathbb{Z}$, and a cube $P = \prod_{i=1}^r I_i \subseteq \mathbb{Z}^r$, the "upper portion" of P is the subset*

$$U := \prod_{i=1}^r U_i$$

and the "middle slice" of P is the subset

$$M := \prod_{i=1}^r M_i.$$

We note the density lower-bounds

$$\frac{|U|}{|P|} \geq (1-2\delta)^r \geq 1 - 2\delta r,$$

$$\frac{|M|}{|P|} \geq \Omega(\delta)^r.$$

Definition 8.8 (Nice configuration)**.** *Consider a configuration $A \subseteq [N_1] \times [N_2] \times \cdots \times [N_r] \subseteq \mathbb{Z}^r$ with density*

$$\mu = \frac{|A|}{N_1 N_2 \cdots N_r}.$$

Consider a parameter $\delta \leq 1/2r$, and let v, m denote the vectors

- $v = \left(\lfloor(\frac{1}{2}+\frac{\delta}{2})N_1\rfloor, \lfloor(\frac{1}{2}+\frac{\delta}{2})N_2\rfloor, \ldots, \lfloor(\frac{1}{2}+\frac{\delta}{2})N_r\rfloor\right)$,
- $m = \left(\lfloor\frac{\delta}{3}N_1\rfloor, \lfloor\frac{\delta}{3}N_2\rfloor, \ldots, \lfloor\frac{\delta}{3}N_r\rfloor\right)$.

We say the configuration is (δ)-nice if

1. *$N_1, N_2, \ldots, N_r$ are some distinct prime numbers not including 2,*
2. *A is in fact contained within the upper-portion $U = \prod_{i=1}^r [2\delta N_i, N_i]$,*
3. *Given $b, b' \in \mathbb{Z}^r$ drawn uniformly at random from $[-m_1, m_1] \times [-m_2, m_2] \times \cdots \times [-m_r, m_r]$, we have the following "weighted" density lower-bound in the middle-slice:*

$$\mathop{\mathbb{E}}_{b,b'}\left[\mathbb{1}_A(v+b-b')\right] \geq \frac{\mu}{2}.$$

4. *We have a uniform bound on the number of representations of any point $2z$:*

$$\max_{z\in A} R_A(2z) \leq r^{O(r)} \cdot \mathop{\mathbb{E}}_{z\in A} R_A(2z).$$

Proposition 8.9 (Embedding a nice configuration)**.** *Suppose $A \subseteq [N_1] \times [N_2] \times \cdots \times [N_r] \subseteq \mathbb{Z}^r$ is a δ-nice configuration with density*

$$\frac{|A|}{N_1 N_2 \cdots N_r} \geq 2^{-d}.$$

Let ϕ be the natural map from $\mathbb{Z}^r$ to the group $G = \mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2} \times \cdots \times \mathbb{Z}_{N_r}$ given by $\phi(x) = x \mod (N_1, N_2, \ldots, N_r)$. We must have either (i) that the number of triples $(x, y, z) \in A^3$ with $x + y = 2z$ is at least

$$r^{-O(r)}\frac{|A|^3}{|G|},$$

or (ii) that there is an even integer $k \le O(d)$ and a Bohr set $B = \mathrm{Bohr}(\Gamma, \rho) \subseteq G$ such that

$$\|\phi(A) \star \phi(A)\|_{k, B\star B} \ge 1 + \tfrac{1}{4}$$

where the Bohr set B

- *has rank $|\Gamma| = r$,*
- *has radius $\rho \ge \Omega(\delta)$*
- *is regular, and*
- *is 2-safe with respect to the restricted map $\phi : A \to G$.*

Proof of Proposition 8.9. We argue by a case analysis on the value

$$\max_{z \in A} R_A(2z).$$

If this value is at least $\frac{|A|^2}{4|G|}$ then (by the fourth criterion for niceness) we indeed have at least

$$r^{-O(r)} \frac{|A|^3}{|G|}$$

3-progressions in A. We now consider the alternative case.

In light of Lemma 5.8 and Lemma 5.9, it suffices to establish the following: the existence of a 2-safe regular Bohr set B (with adequate rank and radius) such that, for some translate $D = T_\theta(B \star B)$,

$$\|\phi(A) * \phi(A) - 1\|_{k,D} \ge \tfrac{1}{2}$$

for some even integer $k \le O(d)$. For any $z \in A \cap M$, where

$$M := \prod_{i=1}^{r} \left[\frac{N_i}{2}, \frac{N_i}{2} + \delta N_i \right] \subseteq \mathbb{Z}^r,$$

we have by construction that

$$R_{\phi(A)}(2\phi(z)) = R_A(2z) \le \frac{|A|^2}{4|G|}.$$

That is, we have not increased the number of representations of $2z$ as a sum of two elements in A by allowing equality mod $(N_1, N_2, \ldots, N_r)$. After re-normalizing to a density function we obtain

$$(\phi(A) * \phi(A))\,(2\phi(z)) = \frac{|G|}{|A|^2} R_{\phi(A)}(2\phi(z)) \le \frac{1}{4}$$

for such z. Let $v, m \in \mathbb{Z}^r$ be the vectors in Definition 8.8, and let b, b' be uniformly random points drawn from the cube

$$C_m := [-m_1, m_1] \times [-m_2, m_2] \times \cdots \times [-m_r, m_r] \subseteq \mathbb{Z}^r.$$

We note that by construction we always have $v + b - b' \in M$. It follows (from this and the third criterion for niceness in Definition 8.8) that $2\phi(v + b - b')$ has a non-negligible chance of being a point where $\phi(A) * \phi(A)$ differs substantially from uniform, so

$$\mathop{\mathbb{E}}_{b, b' \in C_m} \left| \Big(\phi(A) * \phi(A)\Big) \big(2\phi(v) + 2\phi(b) - 2\phi(b')\big) - 1 \right|^k \ge \tfrac{1}{2} \cdot 2^{-d} \cdot \left(\tfrac{3}{4}\right)^k,$$

still for any choice of k. Define the set

$$B := \{2\phi(b) \ : \ b \in C_m\} \subseteq \phi(2C_m) \subseteq G.$$

By Proposition 7.17, this set is 2-safe with respect to the restriction

$$\phi : U \to G,$$

and since $A \subseteq U$ it is also 2-safe with respect to the further restriction $\phi : A \to G$.

We apply essentially the observation in Example 7.3 to say that B is in fact describable as a Bohr set of the group $G = \mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2} \times \cdots \times \mathbb{Z}_{N_r}$. Indeed, for $i = 1, 2, \ldots, r$ let $\alpha_i \in \mathbb{Z}_{N_i}$ be the multiplicative inverse of $2 \mod N_i$. Now consider the set of characters Γ consisting of

$$\gamma_i : x \mapsto e^{2\pi i \alpha_i x_i / N_i}$$

for $i = 1, 2, \ldots, r$. As in Example 7.3, we can say that

$$B = \mathrm{Bohr}(\Gamma, \rho)$$

for some choice of $\rho \geq \Omega(\delta)$. Additionally, due to the simple structure of B we can verify directly that it is regular:

$$|B_{1+\eta}| \leq \left(1 + \tfrac{2}{\pi}\eta\right)^r |B| \leq (1 + \eta r)|B|$$

and

$$|B_{1-\eta}| \geq \left(1 - \tfrac{2}{\pi}\eta\right)^r |B| \geq (1 - \eta r)\, |B|$$

for $\eta \leq 1/r$.

Thus, we have that for the density $D = T_{-2\phi(v)}(B \star B)$,

$$\left\langle D, |\phi(A) * \phi(A) - 1|^k \right\rangle \geq 2^{-(d+1)} \cdot \left(\tfrac{3}{4}\right)^k.$$

Now choose k large enough that $2^{(d+1)/k} \leq \frac{3}{2}$, so that we get

$$\|\phi(A) * \phi(A) - 1\|_{k,D} = \left\langle D, |\phi(A) * \phi(A) - 1|^k \right\rangle^{1/k} \geq 2^{-(d+1)/k} \cdot \tfrac{3}{4} \geq \tfrac{1}{2}. \qquad \square$$

8.3 Obtaining a nice configuration

Proposition 8.10. *Fix a constant*[14] $\varepsilon \in (0, \frac{1}{10}]$ *and let* $\delta := \varepsilon/2r$. *Suppose* $A \subseteq [N_1] \times [N_2] \times \cdots \times [N_r]$ *is a configuration with density* μ. *Then can obtain a subset* $A' \subseteq A$ *and a Freiman homomorphism (of all orders)* $\phi : A' \to P' \subseteq \mathbb{Z}^{r'}$ *with*

- *size* $|A'| \geq r^{-O(r)}|A|$,
- *rank* $r' \leq r$,

and such that we obtain either: (i) a (δ)*-nice configuration with density*

$$\mu' := \frac{|A'|}{|P'|} \geq (1 - 5\varepsilon)\, \mu,$$

or (ii) a configuration which is not necessarily nice but has increased density:

$$\mu' \geq \left(1 + \frac{\varepsilon}{2}\right) \mu.$$

[14] We allow the implied constants in the O-notation here to depend on ε.

Proof. Essentially the idea is to consider the restriction of A to some subcube of the form $P' = \prod_{i=1}^r [a_i, a_i + p_i - 1]$ for some distinct primes p_i each roughly of size δN_i. From here it is easy enough to just translate the whole configuration $(A \cap P') \subseteq P'$ down to $\prod_{i=1}^r [p_i]$ – such a translation is clearly a Freiman homomorphism of all orders.

To ensure that we can make such a choice of primes we use the fact that the number of primes in any interval $[n/2, n]$ is at least $\Omega(n/\log n)$ – this fact follows from the prime number theorem, but such a lower bound also follows from substantially more elementary arguments such as Erdös' proof of Bertrand's postulate.

Without loss of generality suppose that $N_1 \geq N_2 \geq \cdots \geq N_r$. We wish to make a choice of distinct odd primes $p_i \in [\delta N_i/2, \delta N_i]$. Let n_0 be the smallest natural number such that, for any $n \geq n_0$, the number of primes $p \in [n/2, n]$ is at least $r + 1$. It is the case that n_0 is at most $O(\frac{r \log r}{\log \log r})$, although we note it would be fine for us even just to say that $n_0 \leq r^{O(1)}$. If $\delta N_r \geq n_0$ then clearly we can make such choice of primes p_i – otherwise, we will need to "fix" A in the coordinates i corresponding to size-lengths N_i which are too small. Specifically, we pass from A to the subset

$$A' = \{a \in A \,:\, a_r = x\},$$

where we choose $x \in [N_r]$ so that the size of A' is maximized. This results in no loss in density:

$$\frac{|A'|}{N_1 N_2 \cdots N_{r-1}} \geq \frac{|A|}{N_1 N_2 \cdots N_r}$$

and only a small loss in size:

$$|A'| \geq \left(\frac{\delta}{n_0}\right)|A| \geq r^{-O(1)}|A|.$$

With the last coordinate fixed, the projection map $\phi : \mathbb{Z}^r \to \mathbb{Z}^{r-1}$ is clearly a Freiman homomorphism of all orders on the set A'. Continuing in this way, we may assume (at the cost of only a factor $r^{-O(r)}$ loss in size) that we have a configuration $A \subseteq \prod_{i=1}^r [N_i]$ where r has only decreased and that there are some distinct prime numbers $p_i \in [\delta N_i/2, \delta N_i]$ for $i = 1, 2, \ldots, r$.

At this point we stop to apply a pruning step to address niceness criterion number four. Let $A' \subseteq A$ be the subset of points $x \in A$ with

$$R_A(2x) \leq \frac{4}{\varepsilon} \cdot \mathop{\mathbb{E}}_{z \in A} R_A(2z).$$

By a Markov inequality, $|A'| \geq (1 - \varepsilon/4)|A|$, and so A' has density $\mu' \geq (1 - \varepsilon/4)\mu$.

We now consider the restriction of A' to some subcube

$$P_a := a + [p_1] \times [p_2] \times \cdots \times [p_r] \subseteq \mathbb{Z}^r.$$

Let U_a denote the corresponding upper-portion of this cube. Our plan is to set

$$A'' := A' \cap U_a$$

for some choice of $a \in \mathbb{Z}^r$, and our final configuration will be (a translation of) $A'' \subseteq P_a$. Let

- $v = \left(\lfloor(\frac{1}{2} + \frac{\delta}{2})p_1\rfloor, \lfloor(\frac{1}{2} + \frac{\delta}{2})p_2\rfloor, \ldots, \lfloor(\frac{1}{2} + \frac{\delta}{2})p_r\rfloor\right),$
- $m = \left(\lfloor\frac{\delta}{3}p_1\rfloor, \lfloor\frac{\delta}{3}p_2\rfloor, \ldots, \lfloor\frac{\delta}{3}p_r\rfloor\right),$

and define

$$B = [-m_1, m_1] \times [-m_2, m_2] \times \cdots \times [-m_r, m_r] \subseteq \mathbb{Z}^r.$$

We consider the quantities

$$\mu_1''(a) := \frac{|A''|}{|U_a|} = \mathop{\mathbb{E}}_{u \in U_a} \mathbb{1}_{A'}(u)$$

and

$$\mu_2''(a) := \mathop{\mathbb{E}}_{b,b' \in B} \mathbb{1}_{A'}(a + v + b - b').$$

We note that we may assume that

$$\mu_1''(a), \mu_2''(a) \leq (1+\varepsilon)\mu'$$

for all translations $a \in \mathbb{Z}^r$ – otherwise we are done immediately as we would obtain a density increment either onto some U_a or onto some translate of B:

$$\mu'' \geq (1+\varepsilon)(1-\varepsilon/4)\mu \geq (1+\varepsilon/2)\mu.$$

We consider a uniformly-random translation a drawn from the set

$$[-p_1, N_1] \times [-p_2, N_2] \times \cdots \times [-p_r, N_r] \subseteq \mathbb{Z}^r.$$

Applying Proposition 7.8 (or more specifically, its Corollary 7.9), it follows that both of the expectations $\mathbb{E}[\mu_1''(a)]$ and $\mathbb{E}[\mu_2''(a)]$ are at least

$$(1+\delta)^{-r}\mu' \geq (1-\delta r)\mu' \geq (1-\varepsilon/2)\mu.$$

We can now apply a Markov inequality to the (nonnegative) random variables

$$X_1(a) := (1+\varepsilon) - \frac{\mu_1''(a)}{\mu'}$$

and

$$X_2(a) := (1+\varepsilon) - \frac{\mu_2''(a)}{\mu'}$$

to conclude that there is a nonzero probability that we choose a translation a simultaneously satisfying

$$X_1(a), X_2(a) \leq 3\varepsilon,$$

or rather

$$\mu_1''(a), \mu_2''(a) \geq (1-4\varepsilon)\mu'.$$

Finally, we note that

$$\mu'' := \frac{|A''|}{|P_a|} = \frac{|U_a|}{|P_a|}\frac{|A''|}{|U_a|} \geq (1-\varepsilon/2)\mu_1''(a) \geq (1-\varepsilon/2)(1-4\varepsilon)(1-\varepsilon/4)\mu \geq (1-5\varepsilon)\mu$$

and

$$R_{A''}(2x) \leq R_{A'}(2x) \leq \frac{4}{\varepsilon}\frac{1}{|A|}\sum_{z \in A} R_A(2z) \leq \frac{4}{\varepsilon}\frac{1}{|A|}\sum_{z \in A''} R_{A''}(2z) \leq r^{O(r)} \mathop{\mathbb{E}}_{z \in A''} R_{A''}(2z)$$

for all $x \in A''$. □

References

[ALWZ20] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. Improved bounds for the sunflower lemma. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 624–630, 2020.

[Beh46] Felix A Behrend. On sets of integers which contain no three terms in arithmetical progression. *Proceedings of the National Academy of Sciences*, 32(12):331–332, 1946.

[BK12] Michael Bateman and Nets Katz. New bounds on cap sets. *Journal of the American Mathematical Society*, 25(2):585–613, 2012.

[BKM22] Mark Braverman, Subhash Khot, and Dor Minzer. Parallel repetition for the GHZ game: Exponential decay. *CoRR*, abs/2211.13741, 2022.

[Blo16] Thomas F Bloom. A quantitative improvement for Roth's theorem on arithmetic progressions. *Journal of the London Mathematical Society*, 93(3):643–663, 2016.

[Bou99] Jean Bourgain. On triples in arithmetic progression. *Geometric and Functional Analysis*, 9(5):968–984, 1999.

[Bou08] Jean Bourgain. Roth's theorem on progressions revisited. *Journal d'Analyse Mathématique*, 104(1):155–192, 2008.

[BS19] Thomas F Bloom and Olof Sisask. Logarithmic bounds for Roth's theorem via almost-periodicity. *DISCRETE ANALYSIS*, 2019.

[BS20] Thomas F Bloom and Olof Sisask. Breaking the logarithmic barrier in Roth's theorem on arithmetic progressions. *arXiv preprint arXiv:2007.03528*, 2020.

[CFL83] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In David S. Johnson, Ronald Fagin, Michael L. Fredman, David Harel, Richard M. Karp, Nancy A. Lynch, Christos H. Papadimitriou, Ronald L. Rivest, Walter L. Ruzzo, and Joel I. Seiferas, editors, *Proceedings of the 15th Annual ACM Symposium on Theory of Computing, 25-27 April, 1983, Boston, Massachusetts, USA*, pages 94–99. ACM, 1983.

[Cha02] Mei-Chu Chang. A polynomial bound in Freiman's theorem. *Duke mathematical journal*, 113(3):399–419, 2002.

[CLP17] Ernie Croot, Vsevolod F Lev, and Péter Pál Pach. Progression-free sets in are exponentially small. *Annals of Mathematics*, pages 331–337, 2017.

[CS10] Ernie Croot and Olof Sisask. A probabilistic technique for finding almost-periods of convolutions. *Geometric and functional analysis*, 20(6):1367–1396, 2010.

[EG17] Jordan S Ellenberg and Dion Gijswijt. On large subsets of with no three-term arithmetic progression. *Annals of Mathematics*, pages 339–343, 2017.

[Elk10] Michael Elkin. An improved construction of progression-free sets. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 886–905. SIAM, 2010.

[FL17] Jacob Fox and László Miklós Lovász. A tight bound for Green's arithmetic triangle removal lemma in vector spaces. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1612–1617. SIAM, 2017.

[Gow01] William T Gowers. A new proof of Szemerédi's theorem. *Geometric & Functional Analysis GAFA*, 11(3):465–588, 2001.

[Gow11] William T Gowers. What is difficult about the cap set problem? http://gowers.wordpress.com/2011/01/11/what-is-difficult-about-the-cap-set-problem/, 2011. Accessed: 1-5-2023.

[Gre99] Ben Green. Progressions of length 3 following Szemerédi. https://people.maths.ox.ac.uk/greenbj/papers/szemeredi-roth.pdf, 1999.

[GW10] Ben Green and Julia Wolf. A note on Elkin's improvement of Behrend's construction. In *Additive number theory*, pages 141–144. Springer, 2010.

[HB87] David Rodney Heath-Brown. Integer sets containing no arithmetic progressions. *Journal of the London Mathematical Society*, 2(3):385–394, 1987.

[HHL18] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for xor functions. *SIAM Journal on Computing*, 47(1):208–217, 2018.

[KB80] Rob Kaas and Jan M Buhrman. Mean, median and mode in binomial distributions. *Statistica Neerlandica*, 34(1):13–18, 1980.

[KSS18] Robert Kleinberg, Will Sawin, and David Speyer. The growth rate of tri-colored sum-free sets. *Discrete Analysis*, page 3734, 2018.

[Lov15] Shachar Lovett. An exposition of Sanders' quasi-polynomial Freiman-Ruzsa theorem. *Theory of Computing*, pages 1–14, 2015.

[Mes95] Roy Meshulam. On subsets of finite abelian groups with no 3-term arithmetic progressions. *Journal of Combinatorial Theory, Series A*, 71(1):168–172, 1995.

[O'B11] Kevin O'Bryant. Sets of integers that do not contain long arithmetic progressions. *the electronic journal of combinatorics*, 18(P59):1, 2011.

[Rot53] Klaus F Roth. On certain sets of integers. *J. London Math. Soc*, 28(104-109):3, 1953.

[San10] Tom Sanders. Popular difference sets. *Online J. Anal. Comb.*, 5(5):4, 2010.

[San11] Tom Sanders. On Roth's theorem on progressions. *Annals of Mathematics*, pages 619–636, 2011.

[San12a] Tom Sanders. On certain other sets of integers. *Journal d'Analyse Mathématique*, 1(116):53–82, 2012.

[San12b] Tom Sanders. On the Bogolyubov–Ruzsa lemma. *Analysis & PDE*, 5(3):627–655, 2012.

[Sch15] Tomasz Schoen. New bounds in Balog-Szemerédi-Gowers theorem. *Combinatorica*, 35(6):695–701, 2015.

[Sch21] Tomasz Schoen. Improved bound in Roth's theorem on arithmetic progressions. *Advances in Mathematics*, 386:107801, 2021.

[Shk17] ID Shkredov. Some remarks on the Balog–Wooley decomposition theorem and quantities D+, D×. *Proceedings of the Steklov Institute of Mathematics*, 298:74–90, 2017.

[SS13] Tomasz Schoen and Ilya D. Shkredov. Higher moments of convolutions. *Journal of Number Theory*, 133(5):1693–1737, 2013.

[SS16] Tomasz Schoen and Olof Sisask. Roth's theorem for four variables and additive structures in sums of sparse sets. In *Forum of Mathematics, Sigma*, volume 4. Cambridge University Press, 2016.

[Sze90] Endre Szemerédi. Integer sets containing no arithmetic progressions. *Acta Mathematica Hungarica*, 56(1):155–158, 1990.

[TV06] Terence Tao and Van H Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.

A Comparison with prior approaches and other measures of additive pseudorandomness

In this section, we compare our techniques for controlling the number of solutions to $a + b = c$ with some previous approaches, and we compare our conditions for quantifying "additive pseudorandomness" with some analogous conditions considered by other works.

Additive energy. We begin by discussing the notion of "additive energy", which is a good central measure of additive structure to keep in mind as it can be easily compared with everything else we discuss. For a set $A \subset G$ the additive energy of A is the quantity $E(A) = \sum_{x \in A} R_A^-(x)^2$, which is also the same as the number of solutions to $a_1 - a_2 = a_3 - a_4$ with $a_i \in A$. It is always between $|A|^2$ and $|A|^3$. It can be expressed in the density formulation as the quantity

$$E(A) = \frac{|A|^4}{|G|} \|A \star A\|_2^2.$$

Thus, for $k = 2$ our "self-regularity" condition corresponds simply to some bound on the additive energy.[15] For sufficiently dense sets, there is a different lower bound which improves on the trivial bound $E(A) \geq |A|^2$: it is always the case that $E(A)$ is at least $|A|^4/|G|$. The bounds

$$|A|^4/|G| \leq E(A) \leq |A|^3$$

translate gracefully to the following equivalent bounds on $\|A \star A\|_2^2$. By Jensen's inequality and Young's inequality,

$$1 = \|A \star A\|_1^2 \leq \|A \star A\|_2^2 \leq \|A\|_2^2.$$

Note that set of size $|A| = 2^{-d}|G|$ corresponds to a density function with $\|A\|_2^2 = 2^d$. More generally, the bounds above apply to any density function $A(x)$.

It is a typical sort of problem in additive combinatorics to prove something nontrivial about the "additive structure" of A when $E(A)$ is "large". In some contexts this means that $E(A)$ only slightly less than maximal: for instance if $E(A) \geq \frac{1}{10}|A|^3$, or perhaps $E(A) \geq |A|^{3-1/10}$. In the present work, we are interested in what can be said about A given that the additive energy is only slightly greater than minimal: for instance if $\|A \star A\|_2^2 \geq 1 + \frac{1}{10}$. Thus, the alternative normalization $\|A \star A\|_2^2$ of the additive energy is a natural choice for this setting.

The Roth-Meshulam argument. We review the classical Roth-Meshulam argument ([Rot53, Mes95]) in the case of $\mathbb{F}_q^n$, and discuss its relation to the present work. For an additive character

$$e_\alpha : \mathbb{F}_q^n \to \mathbb{C},$$

[15] Similarly, for $k \geq 2$ our self-regularity condition is simply a bound on the higher-order energy $\sum_x R_A^-(x)^k$ which is also a well-studied concept; see e.g. [SS13].

we consider the corresponding Fourier coefficient[16]

$$\widehat{A}(\alpha) := \mathop{\mathbb{E}}_{x\in A} e_\alpha(x) = \langle A, e_\alpha \rangle .$$

We note that, for any set A, $\widehat{A}(0) = 1$, and generally that $|\widehat{A}(\alpha)| \leq 1$.

In the Roth-Meshulam argument, the key quantity for measuring the additive pseudorandomness of a set A is the size of its largest nontrivial Fourier coefficient:

$$\max_{\substack{\alpha\in\mathbb{F}_q^n \\ \alpha\neq 0}} |\widehat{A}(\alpha)|.$$

The overall argument can be summarized compactly as follows.

Proposition A.1. *Let $A, B, C \subseteq \mathbb{F}_q^n$ be sets of size at least $2^{-d}|\mathbb{F}_q^n|$.*

1. *If $\|A\|_{\perp,1} \leq 1+\varepsilon$ then*
$$\max_{\alpha\neq 0} |\widehat{A}(\alpha)| \leq 2\varepsilon.$$

2. $$|\langle A, B * C\rangle - 1| \leq \max_{\alpha\neq 0} |\widehat{A}(\alpha)| \cdot \|B\|_2 \|C\|_2 \leq \max_{\alpha\neq 0} |\widehat{A}(\alpha)| \cdot 2^d.$$

Proof. For a linear subspace W, let P_W denote the projection operator

$$P_W : f \mapsto W^\perp * f.$$

We note that the quantity $\|A\|_{\perp,r}$ can be equivalently characterized as

$$\|A\|_{\perp,r} = \max_{\substack{W\subseteq\mathbb{F}_q^n \\ \dim(W)\leq r}} \|P_W A\|_\infty. \tag{3}$$

We also note that for a function

$$A(x) = \sum_{\alpha\in\mathbb{F}_q^n} \widehat{A}(\alpha) e_\alpha(-x)$$

we have

$$(P_W A)(x) = \sum_{\alpha\in W} \widehat{A}(\alpha) e_\alpha(-x).$$

Fix some $\alpha \neq 0$, and let W be some one-dimensional subspace containing it. We have

$$|\widehat{A}(\alpha)| = |\langle P_W A - 1, e_\alpha\rangle| \leq \|P_W A - 1\|_1.$$

To finish, we use the basic fact that for any density function F,

$$\|F-1\|_1 = 2\|(F-1)_+\|_1 \leq 2\|(F-1)_+\|_\infty = 2(\|F\|_\infty - 1).$$

[16] See Section 3 for details regarding definitions and normalization conventions related to the Fourier expansion.

This proves the first claim. For the second claim we make the following calculation.

$$\begin{aligned}
|\langle A, B * C\rangle - 1| &\leq \sum_{\alpha\neq 0} |\widehat{A}(\alpha)||\widehat{B}(\alpha)||\widehat{C}(\alpha)| \\
&\leq \max_{\alpha\neq 0} |\widehat{A}(\alpha)| \cdot \sum_{\beta} |\widehat{B}(\beta)||\widehat{C}(\beta)| \\
&\leq \max_{\alpha\neq 0} |\widehat{A}(\alpha)| \cdot \sqrt{\sum_{\beta} |\widehat{B}(\beta)|^2}\sqrt{\sum_{\beta} |\widehat{C}(\beta)|^2} \\
&= \max_{\alpha\neq 0} |\widehat{A}(\alpha)| \cdot \|B\|_2\|C\|_2.
\end{aligned}$$

□

Thus, the overall structure of the argument very much resembles our own: to begin, there is a "primitive" pseudorandom condition (spreadness with respect to subspaces of codimension one), which can be ensured by a density increment argument. Then there is an intermediate pseudorandom condition (i.e. a bound on $\max_{\alpha\neq 0} |\widehat{A}(\alpha)|$), which firstly can be derived from spreadness, and secondly is sufficient to directly control the quantity $\langle A, B * C\rangle$ when B and C are large.

We note another similarity to the present work. Consider the following Fourier-analytic interpretation of the additive energy of A:

$$\|A \star A\|_2^2 = 1 + \sum_{\alpha\neq 0} |\widehat{A}(\alpha)|^4 = 1 + \|A \star A - 1\|_2^2.$$

It is well known that (for very large sets A) a bound on $\max_{\alpha\neq 0} |\widehat{A}(\alpha)|$ is roughly equivalent to a bound on $\|A \star A - 1\|_2^2$:

$$\max_{\alpha\neq 0} |\widehat{A}(\alpha)|^4 \leq \sum_{\alpha\neq 0} |\widehat{A}(\alpha)|^4 \leq \max_{\alpha\neq 0} |\widehat{A}(\alpha)|^2 \cdot \sum_{\beta} |\widehat{A}(\beta)|^2 = \max_{\alpha\neq 0} |\widehat{A}(\alpha)|^2 \cdot \|A\|_2^2.$$

Thus, another reasonable summary of the Roth-Meshulam argument is that for very large sets A, B, C,

- $\|A \star A\|_2$ can be controlled by $\|A\|_{\perp,1}$, and
- $\langle A, B * C\rangle$ can be controlled by $\|A \star A\|_2$.

The Bloom-Sisask physical space argument. In [BS19], Bloom and Sisask show (in particular) how to obtain parameters for the cap-set problem very near to those given by the Roth-Meshulam approach, but with arguments working almost exclusively in physical space (rather than Fourier space). This is a notable similarity to the present work, where we make very little use of any "quantitative" Fourier analysis.[17] In particular, the Bloom-Sisask physical space argument more closely resembles our approach than any other prior work. We offer the following interpretation of their argument (as it applies to the problem of controlling the quantity $\langle A * B, C\rangle$, for sets $A, B, C \subseteq \mathbb{F}_q^n$ of density 2^{-d}, by a spreadness assumption).

There are three steps. Firstly: the plan is to control $\langle A * B, C\rangle$ by obtaining a bound on the key quantity $\|A * B - 1\|_k$, for some $k \approx d$. This is precisely what is done also in our work, and already it

[17] We implicitly rely on a small amount of quantitative Fourier analysis through our use of Lemma C.3, whose proof relies on Chang's inequality. Besides this, our remaining Fourier-analytic arguments (i.e., regarding the decoupling inequality and spectral positivity) are "qualitative".

distinguishes the approach from many others, which consider instead some key quantity describable in Fourier space. Secondly, Bloom and Sisask argue (by a combination of the Croot-Sisask lemma[18] and Chang's inequality) that $\|A * B - 1\|_k$ is similar to $\|V * A * B - 1\|_k$, for some large subspace V. More specifically, if

$$\|A * B - 1\|_k \geq \Omega(1)$$

then

$$\|V * A * B - 1\|_k \geq \Omega(1)$$

for some V of codimension $r \leq 2^{d+O(\log d + \log k)}$. The final point is that such deviation from 1 is impossible if A and B are both spread. Specifically, we apply the claim given below with $D := V * A$ and $D' := V * B$ (noting that $V * V = V$). Overall, the Bloom-Sisask physical space argument shows that if $|\langle A, B * C\rangle - 1| \geq \Omega(1)$, then either $\|A\|_{\perp,r} \geq 1 + \Omega(1)$ or $\|B\|_{\perp,r} \geq 1 + \Omega(1)$, for some

$$r \leq 2^{d+O(\log d)}.$$

Proposition A.2. *Let D and D' be density functions on a finite abelian group G. For $\varepsilon \geq 0$, if*

$$\|D\|_\infty \leq 1 + \varepsilon \text{ and } \|D'\|_\infty \leq 1 + \varepsilon$$

then

$$\|D * D' - 1\|_\infty \leq \varepsilon$$

Proof. The pointwise upper bound $D * D' - 1 \leq \varepsilon$ follows easily by averaging: $\|D * D'\|_\infty \leq \|D\|_\infty \leq 1 + \varepsilon$. Now consider the pointwise lower bound. If $\varepsilon > 1$, such a bound is trivial, so suppose $\varepsilon \leq 1$. We have

$$0 \leq (1 + \varepsilon - D) * (1 + \varepsilon - D') = (1 + \varepsilon)^2 - 2(1 + \varepsilon) + D * D' = -1 + \varepsilon^2 + D * D',$$

and so

$$D * D \geq 1 - \varepsilon^2 \geq 1 - \varepsilon.$$

□

Understanding the power of density increments. Let us say, only in the context of the present section, that a set $A \subseteq \mathbb{F}_q^n$ of size $|A| = 2^{-d}|\mathbb{F}_q^n|$ is simply "spread" if

$$\|A\|_{\perp,r} \leq 1 + O\left(\frac{r(d+1)}{n \log n}\right)$$

for all $r \leq n/2$. One can check that this corresponds to the condition that A has "no strong increments" considered by Bateman and Katz [BK12] within their regime of interest: namely $d = \Theta(\log n)$. We are interested in this specific choice of parameters essentially because it represents the limit of what can be reasonably obtained by a generic density increment argument.[19] Given this definition, we consider the following substantially more concrete variant of Question 2.7.

Question A.3. *Suppose $A, B, C \subseteq \mathbb{F}_q^n$ are sets of size at least $2^{-d}|\mathbb{F}_q^n|$. For what values of d can we deduce that the number of solutions to $a + b + c = 0$ is within a factor 2 of $|A||B||C|/|\mathbb{F}_q^n|$, from an assumption that*

[18] The version of the Croot-Sisask lemma used is notably more efficient in its dependence on the error parameter than, e.g., Lemma C.1. The parameters involved in this alternative version are more subtle and it can be convenient to handle the cases, say, $\|A * B\|_k \leq 10$ and $\|A * B\|_k \geq 10$ separately.

[19] We do not intend this as any kind of formal claim.

(i) one,

(ii) two, or

(iii) three

of the sets are "spread" (in the specific sense described above)?

We note that Gowers asks essentially part (i) of this question in a blog post [Gow11]. More accurately, he asks only about when we can deduce the existence of at least one solution. The Roth-Meshulam argument described above is relevant here: it requires only a spreadness assumption on the single set A. Given that A is spread, it tells us that

$$|\langle A, B * C\rangle - 1| \leq O\left(\frac{(d+1)2^d}{n \log n}\right),$$

and so it provides the following positive answer regarding (i): to ensure that

$$\langle A, B * C\rangle \in \left[\tfrac{1}{2}, 2\right],$$

a spreadness assumption on just A is sufficient, for values of d up to

$$d = \lg n - c$$

where c is some constant. We consider the following example which shows that for this question, the answer obtained from the Roth-Meshulam argument is actually best-possible, despite the fact that it makes use only of a bound on $\|A\|_{\perp,1}$. Let $d = \lg n$, and consider[20]

$$A := \left(\mathbb{F}_2^d \setminus \{0\}\right) \times \mathbb{F}_2^{n-d} \subseteq \mathbb{F}_2^n$$

and

$$B := \{0\} \times \mathbb{F}_2^{n-d} \subseteq \mathbb{F}_2^n.$$

We have $B + B = B$ and $\langle A, B * B\rangle = 0$. One can check that A is indeed spread simply because, for any r,

$$\|A\|_{\perp,r} \leq \|A\|_\infty = \frac{1}{1 - 2^{-d}} \approx 1 + \frac{1}{n}.$$

That is, A has no strong increment onto a large affine subspace in particular because it has *no strong increment onto a set of any kind*. Thus, for

$$d \geq \lg n,$$

it is not possible to control $\langle A, B * C\rangle$ by assuming spreadness for just A; this fully answers the question of Gowers. However, we point out that the situation changes dramatically if we assume spreadness for just A but ask only for *upper bounds*. Then the Roth-Meshulam argument is no longer optimal, and indeed, in Theorem 2.8 we show that k-regularity follows from r-spreadness for $r \approx k^7 d$. In particular we have that $\langle A, B * C\rangle \leq 2$ whenever A is "spread" for d up to roughly

$$d \approx n^{1/9}.$$

Summary for Question A.3. We summarize what we understand regarding Question A.3.

[20] The same example can be constructed over other small fields by considering $\mathbb{F}_q^t \times \mathbb{F}_q^{n-t}$ with $q^t \approx n$.

- For $d \leq \lg n - O(1)$, the Roth-Meshulam argument shows that a spreadness assumption on just A is sufficient to ensure $\langle A, B * C\rangle \in [\frac{1}{2}, 2]$. Moreover, only a bound on $\|A\|_{\perp,1}$ is needed.
- For $d \geq \lg n$, a spreadness assumption on just A does not ensure $\langle A, B * C\rangle > 0$.
- For $d \lessapprox n^{1/9}$, a spreadness assumption on just A ensures $\langle A, B * C\rangle \leq 2$.
- For $d \lessapprox n^{1/9}$, a spreadness assumption on A and B ensures $\langle A * B, C\rangle \in [\frac{1}{2}, 2]$.

The Fourier sum of cubes measure. Recall the estimate

$$|\langle A, B * C\rangle - 1| \leq \sum_{\alpha \neq 0} |\widehat{A}(\alpha)||\widehat{B}(\alpha)||\widehat{C}(\alpha)|$$

which was used in the Roth-Meshulam argument. A sensible approach for part (iii) of Question A.3, and one which is very natural from a Fourier-analytic perspective, is to depart from the earlier argument and treat this quantity more symmetrically by considering

$$\sum_{\alpha \neq 0} |\widehat{A}(\alpha)||\widehat{B}(\alpha)||\widehat{C}(\alpha)| \leq \left(\sum_{\alpha \neq 0} |\widehat{A}(\alpha)|^3\right)^{1/3} \left(\sum_{\alpha \neq 0} |\widehat{B}(\alpha)|^3\right)^{1/3} \left(\sum_{\alpha \neq 0} |\widehat{C}(\alpha)|^3\right)^{1/3}.$$

This shows that a bound of, say,

$$\sum_{\alpha \neq 0} |\widehat{A}(\alpha)|^3 \leq \tfrac{1}{2}$$

on the sum of cubes of nontrivial Fourier coefficients is sufficient to control $\langle A * B, C\rangle$, if we assume such a bound it for all three sets. Indeed, this was the approach taken in the work of Bateman and Katz on the cap-set problem ([BK12]), who show that

$$\sum_{\alpha \neq 0} |\widehat{A}(\alpha)|^3 \geq \Omega(1)$$

is sufficient to infer that A has density increment onto a large affine subspace, for values of d up to $d = (1 + c) \lg n$ for some small constant $c > 0$, notably breaking the technical barrier described above.

It has on occasion been taken as the obvious starting point for any sort of analytic approach to the 3-progression problem to begin with the assumption $\sum_{\alpha \neq 0} |\widehat{A}(\alpha)|^3 \geq \Omega(1)$ and then to see what we can conclude about A. In light of this, it is interesting to note that even in view of our work it seems still unclear whether that approach is viable for obtaining strong bounds. Specifically, we ask the following technical question – we don't know of any specific application related to its resolution, but it seems interesting from a technical perspective for trying to better understand the relation between the Fourier-analytic approach with various "physical space" techniques such as the Croot-Sisask lemma and sifting.

Question A.4. *Let $A \subseteq G$ be a set of size $|A| \geq 2^{-d}|G|$. Suppose that*

$$\sum_{\alpha \neq 0} |\widehat{A}(\alpha)|^3 \geq \Omega(1).$$

Does it follow that

$$\|A \star A - 1\|_k \geq \Omega(1)$$

for some $k \leq \mathrm{poly}(d)$?

We make two remarks regarding this. Firstly, the answer is affirmative if we instead start with, say, the stronger assumption

$$\left|\sum_{\alpha\neq 0}\widehat{A}(\alpha)^3\right| \geq \Omega(1).$$

In this case we can write $\sum_{\alpha\neq 0}\widehat{A}(\alpha)^3 = \langle A * A - 1, B\rangle$, where $B(x) = A(-x)$, and bound this approximately by $\|A\star A - 1\|_d$. Secondly, we note a connection in the opposite direction. For even $k \in \mathbb{N}$ we have the Fourier-analytic interpretation

$$\|A\star A - 1\|_k^k = \sum_{\substack{\alpha_1+\alpha_2+\cdots+\alpha_k=0\\ \alpha_i\neq 0}} |\widehat{A}(\alpha_1)|^2|\widehat{A}(\alpha_2)|^2\cdots|\widehat{A}(\alpha_k)|^2.$$

Applying Young's inequality to this gives

$$\|A\star A - 1\|_k \leq \left(\sum_{\alpha\neq 0}|\widehat{A}(\alpha)|^{2+\frac{2}{k-1}}\right)^{1-\frac{1}{k}}.$$

Finally, we can remark (at least superficially) on a connection between our self-regularity condition and another key quantity considered both in the work of Bateman and Katz as well as the work of Bloom and Sisask ([BK12],[BS20]). For $k \in \mathbb{N}$ we have the alternative Fourier-analytic interpretation

$$\|A\star A\|_{2k}^{2k} = \left\langle (A\star A)^k, (A\star A)^k\right\rangle = \sum_{\beta}\left(\sum_{\alpha_1+\alpha_2+\cdots+\alpha_k=\beta}|\widehat{A}(\alpha_1)|^2|\widehat{A}(\alpha_2)|^2\cdots|\widehat{A}(\alpha_k)|^2\right)^2.$$

By comparison, the authors of the aforementioned works consider various sets of the form

$$\Delta = \left\{\alpha\in G \,:\, \eta_1 \leq |\widehat{A}(\alpha)| \leq \eta_2\right\}$$

as well as the resulting quantity

$$\sum_{\beta}\left(\sum_{\alpha_1+\alpha_2+\cdots+\alpha_k=\beta}\mathbb{1}_\Delta(\alpha_1)\cdot\mathbb{1}_\Delta(\alpha_2)\cdots\mathbb{1}_\Delta(\alpha_k)\right)^2,$$

for various choices of k.

B Detailed account of the proof of Theorem 2.8

We proceed to explain our approach for proving Theorem 2.8. Let us focus only on the second claim, which we recall states essentially the following. Suppose that a large set $A \subseteq \mathbb{F}_q^n$, $|A| \geq 2^{-d}|\mathbb{F}_q^n|$, fails to be self-regular:

$$\|A\star A\|_k \geq 1+\varepsilon.$$

Then we can find an affine subspace $V \subseteq \mathbb{F}_q^n$ of codimension at most $\mathrm{poly}(d,k,1/\varepsilon)$ giving the density increment

$$\langle V, A\rangle \geq 1+\Omega(\varepsilon).$$

For simplicity we also focus only on the case $k = 2$ and $\varepsilon = 1$; this case is already nontrivial and sufficient to illustrate the main obstacles and how they will be overcome.

Our starting point is to instead look for something stronger: a "density increment"

$$\langle V, A \star A \rangle \geq 1 + \Omega(1).$$

This is a key "leap of faith" in the proof, in the sense that we cannot offer a reason to expect (a priori) that this should be possible, even assuming that our original task is possible. However, we note that, firstly, it is certainly sufficient by a simple averaging argument. Secondly, if this approach is indeed workable it would be quite nice from a mechanical perspective, as we start with an assumption on the density function $A \star A$ (i.e., that $\|A \star A\|_2 \geq 2$), and we now seek a conclusion about the *same object*: $\langle V, A \star A \rangle \geq 1 + \Omega(1)$. This is in contrast to the typical structure of an *inverse problem* – considered generally to be fairly difficult – where we make an assumption on $A \star A$ and seek a conclusion regarding A; by taking this approach, we can suppress the fact that we are secretly working on an inverse problem at the outset. To elaborate on the averaging argument, we have

$$\langle V, A \star A \rangle = \langle A * V, A \rangle = \mathop{\mathbb{E}}_{\substack{x \in V \\ a \in A}} A(x + a) = \mathop{\mathbb{E}}_{a \in A} \left[\mathop{\mathbb{E}}_{x \in V + a} A(x) \right],$$

so if $\langle V, A \star A \rangle \geq 1 + \Omega(1)$ then we must have a density increment $\langle V', A \rangle \geq 1 + \Omega(1)$ onto some fixed affine subspace $V' = V + a$.

We pause to consider whether our new goal is at least analytically plausible. By this, we mean that we ask if we can at least obtain

$$\langle F, A \star A \rangle \geq 1 + \Omega(1)$$

for *some* high-entropy density function F – one with $\|F\|_\infty \leq 2^{\mathrm{poly}(d)}$ – without asking that F have any particular additive structure. Ideally, we would also not need to appeal to the additive structure of our density $D = A \star A$, and instead, we see what can be said just from the basic analytic facts $D \geq 0$, $\|D\|_1 = 1$, $\|D\|_2 \geq 2$, and $\|D\|_\infty \leq 2^d$. One can check that the best choice of F satisfying such an entropy constraint on $\|F\|_\infty$ will be (essentially) the uniform density over some super-level set of D. However, there is also a simple choice which is already quite satisfactory, $F = D$:

$$\langle D, D \rangle = \|D\|_2^2 \geq 4.$$

So, what we are asking for is at least analytically plausible – there is a high entropy density function F witnessing $D \gg 1$ – and our task is to argue that we may take F to have a high degree of additive structure.

B.1 Sanders' invariance lemma

To motivate our next step, we consider the following lemma due to Sanders regarding the translation-invariance of the convolution of two large sets $A * B$. Sanders obtains this lemma by combining the Croot-Sisask lemma [CS10], Chang's inequality [Cha02], and some additional Fourier-analytic arguments. The specific formulation given below does not appear explicitly in his work, so we include a proof in the appendix.[21]

Lemma B.1 (Sanders' invariance lemma [San12b])**.** *Suppose $A, B \subseteq \mathbb{F}_q^n$ are sets of sizes $|A| \geq 2^{-d}|\mathbb{F}_q^n|$ and $|B| \geq 2^{-k}|\mathbb{F}_q^n|$. Fix a bounded function $f : \mathbb{F}_q^n \to [-1, +1]$. Then, for any $\varepsilon \geq 2^{-k}$,*

[21] See however [San12b, Appendix: Proof of Theorem 11.1] or [Lov15] for some similar statements.

there exists a linear subspace V (possibly depending on f) with codimension at most $O(dk^3/\varepsilon^2)$ satisfying

$$|\langle V * A * B, f\rangle - \langle A * B, f\rangle| \leq \varepsilon.$$

We suggest the following interpretation of this result. It is "almost" saying that $A * B$ is close in statistical distance to $V * A * B$ for some large subspace V, in light of the dual characterization of the 1-norm distance:

$$\|D - D'\|_1 = \max_{f:\mathbb{F}_q^n \to [-1,+1]} \left\langle D - D', f \right\rangle.$$

However, it is not quite this strong since the subspace V is allowed to depend on the dual witness f. Under this interpretation, the lemma can be productively compared with Chang's inequality via the Fourier-analytic identity

$$(V * A * B)(x) = \sum_{\alpha \in V^\perp} \widehat{A}(\alpha)\widehat{B}(\alpha)e_\alpha(-x).$$

The importance of Sanders' lemma for us can be captured succinctly by the following immediate consequence. We state it in terms of the notation defined by Eq. (1) and Eq. (2) in Section 2, which we repeat here for the reader's convenience.

$$\|f\|_{\perp,r} := \max_{\substack{\text{affine subspace } V \subseteq \mathbb{F}_q^n \\ \mathrm{Codim}(V) \leq r}} \langle V, f\rangle,$$

$$\|f\|_{*,k} := \max_{\substack{B,C \subseteq G \\ \|B\|_\infty, \|C\|_\infty \leq 2^k}} \langle B * C, f\rangle.$$

Corollary B.2. *Let $d \geq 1$ and $\varepsilon \in [2^{-d}, 1]$. For any bounded function $f : \mathbb{F}_q^n \to [0,1]$, we have*

$$\|f\|_{\perp, d^4/\varepsilon^2} \geq \|f\|_{*,d} - O(\varepsilon).$$

That is, Sanders shows us how to bootstrap a high-entropy density function F with mild additive structure witnessing $\langle F, f\rangle \geq \gamma$ into a reasonably high-entropy density function F' with strict additive structure witnessing $\langle F', f\rangle \geq \gamma - O(\varepsilon)$. Sanders applies this fact to the difference-set indicator function $f := \mathbb{1}_{A-A}$ to obtain his solution to the Approximate Bogolyubov Problem (Theorem 1.11): indeed, we clearly have $\|\mathbb{1}_{A-A}\|_{\star,d} = 1$, as witnessed by $A \star A$ itself.

Inspired by this, we might hope (possibly naively) that we could apply such an argument to the function $f = A \star A$, as we have noted already that

$$\langle A \star A, A \star A\rangle \geq 4,$$

and so $\|A \star A\|_{*,d} \geq 4$. Certainly this is not permitted by Sanders' lemma as stated, since $A \star A$ is not a bounded function, but we can for example try to see what can be learned by applying Sanders' lemma to various approximate level-set indicator functions

$$f(x) = \mathbb{1}(\eta_1 \leq (A \star A)(x) \leq \eta_2).$$

For the sake of discussion, let us consider the following (purposefully slightly vague) notion. Say that "F robustly witnesses $D \gg 1 + \eta$" if (firstly)

$$\langle F, D\rangle \geq 1 + \eta,$$

and additionally that F witnesses a deviation of such magnitude typically, rather than just on average. Certainly we would include the following exemplary case: if

$$\langle F, \mathbb{1}_S \rangle \geq 1 - \eta/2,$$

where

$$S = \{x \,:\, D(x) \geq 1 + 2\eta\}$$

and $\eta \leq 1/2$; in this case we indeed have

$$\langle F, D \rangle \geq (1 + 2\eta) \langle F, \mathbb{1}_S \rangle \geq 1 + \eta.$$

In light of Sanders' invariance lemma, we see that to reach our goal it suffices to find some large sets $B, C \subseteq |\mathbb{F}_q^n|$ of sizes $|B|, |C| \geq 2^{-\mathrm{poly}(d)}|\mathbb{F}_q^n|$ such that the convolution $B * C$ robustly witnesses $A \star A \gg 1 + \Omega(1)$. Indeed, if we can show that

$$\|f\|_{*,k} \geq 1 - \eta/2$$

where

$$f(x) = \mathbb{1}((A \star A)(x) \geq 1 + 2\eta)\,,$$

for some $k \leq \mathrm{poly}(d)$ and some $\eta \geq \Omega(1)$, then we can apply Corollary B.2 with $\varepsilon = \Theta(\eta)$ to find a large affine subspace with

$$\langle V, A \star A \rangle \geq (1 + 2\eta) \langle V, f \rangle \geq 1 + \Omega(1).$$

B.2 Finding a robust witness

We proceed to consider the problem of finding a robust witness to $A \star A \gg 1 + \Omega(1)$ of the form $B * C$. It is tempting to check if simply $F = A \star A$ is already a good robust witness for itself. For general density functions D obeying only $\|D\|_2 \geq 2$ and $\|D\|_\infty \leq 2^d$ this is certainly not the case: for example it is not hard to arrange that for uniformly random x, $D(x)$ corresponds to some random variable $X \in [0, 2^d]$ with $\mathbb{E}[X] = 1$ and $\mathbb{E}[X^2] = 4$ and yet

$$\mathbb{E}\left[X \cdot \mathbb{1}(X \geq 1)\right] \leq \frac{O(1)}{2^{2d}}.$$

We can consider whether the additive structure of $D = A \star A$ rescues us from this example. Upon checking some extreme examples, such as when A a subspace of density 2^{-d}, or A is a random set of density 2^{-d}, this looks initially promising; in both cases the function $F = A \star A$ is a reasonably robust witness of $A \star A \gg \|A \star A\|_2$. However, the following third example, which we call the "planted subspace" example, shows this is not the case in general.

Example B.3 (Planted subspace)**.** *Consider a set $A \subseteq \mathbb{F}_2^n$ of the form*

$$A = (W \cup C) \times \mathbb{F}_2^{n-d},$$

where $C \subseteq \mathbb{F}_2^d$ is a random set of size $|C| \leq |\mathbb{F}_2^d|^{1/10}$ and $W \subseteq \mathbb{F}_2^d$ is a subspace of size $|C|^{3/4}$.

It is likely that $|C + C| \approx |C|^2$, $|W + C| \approx |C|^{7/4}$, and (very roughly speaking) that

$$R_{W \cup C} \approx R_W + R_C + R_{W,C} \approx |W| \cdot \mathbb{1}_W + \mathbb{1}_{C+C} + \mathbb{1}_{W+C}.$$

Suppose this is the case, and consider the random variable

$$Z := R_{W\cup C}(y),$$

where $y \sim R_{W\cup C}$.[22] *We have*

$$\mathbb{E}[Z] = \frac{\sum_x R_{W\cup C}(x)^2}{\sum_x R_{W\cup C}(x)} \approx \frac{|C|^{9/4} + |C|^2 + |C|^{7/4}}{|C|^{6/4} + |C|^2 + |C|^{7/4}} \approx \frac{|C|^{9/4}}{|C|^2} = |C|^{1/4},$$

and yet

$$\mathbb{P}(Z \leq O(1)) \approx 1 - \frac{|C|^{6/4}}{|C|^2} = 1 - |C|^{-1/2}.$$

This example shows that it is possible that the majority of the contribution to $\|A \star A\|_2$ is attributable to some additive structure within $A \star A$ which is strong but rare, and we will need to work somewhat harder in order to detect it.

Let us consider once again whether what we ask for is at least plausible analytically: that we can find a high-entropy density function robustly witnessing $D \gg 1 + \Omega(1)$ whenever $\|D\|_2 \geq 2$ and $\|D\|_\infty \leq 2^d$. Again, the best choice of witness F given a constraint on $\|F\|_\infty$ would be some super-level set of D. However, we consider a simple choice which is already quite satisfactory. We denote by $D^{\wedge k}$ the density function which is proportional to D^k; namely

$$D^{\wedge k}(x) := \left(\frac{D(x)}{\|D\|_k}\right)^k. \tag{4}$$

In what follows we refer to this density function as the "degree-k compression" of D. Since $\|D\|_k \geq \|D\|_1 = 1$, it follows that compressions satisfy the entropy-deficit bound

$$\|D^{\wedge k}\|_\infty \leq \|D\|_\infty^k \leq 2^{dk}.$$

Compressions $D^{\wedge k}$ also robustly witness $D \gg (1-\varepsilon)\|D\|_k$.

Proposition B.4 (Compressions escape sub-level sets)**.** *Let $D : \Omega \to \mathbb{R}_{\geq 0}$ be a density function on some arbitrary finite set Ω, and let $k \geq 1$. Consider the sub-level set*

$$S := \{x \in \Omega \,:\, D(x) \leq (1-\varepsilon)\|D\|_k\}.$$

We have

$$\left\langle D^{\wedge k}, \mathbb{1}_S \right\rangle \leq (1-\varepsilon)^k\, \mathbb{E}[\mathbb{1}_S] \leq e^{-\varepsilon k}.$$

Alternatively, consider

$$S' := \left\{x \in \Omega \,:\, D(x) \leq c \cdot \|D\|_k^{1+\frac{1}{k-1}}\right\}.$$

We have

$$\left\langle D^{\wedge k}, \mathbb{1}_{S'} \right\rangle \leq c^{k-1}.$$

Proof. Bound $D^k \cdot \mathbb{1}_S \leq (1-\varepsilon)^k \|D\|_k^k \cdot \mathbb{1}_S$ pointwise and then take the expectation. Alternatively, we may bound $D^k \cdot \mathbb{1}_{S'} \leq c^{k-1} \cdot \|D\|_k^k \cdot D$ pointwise and use $\mathbb{E}[D] = 1$. ☐

[22] That is, $y \in \mathbb{F}_2^d$ is a random variable with pdf proportional to $R_{W\cup C}$.

We apply this to our current situation for some choice of $k \geq 2$, noting that $\|D\|_k \geq \|D\|_2 \geq 2$. We find that it suffices to choose some $k \leq O(1)$ to get a (high-entropy) robust witness to $D \gg 1+\Omega(1)$, as desired.

Importantly for us, for integral values of k, degree-k compressions of a convolution $D = A \star A$ also retain a certain amount of additive structure. More specifically, we will soon see that the fact that

$$\left\langle (A \star A)^{\wedge k}, \mathbb{1}_S \right\rangle \leq \eta,$$

entails the following consequence: that there is some subset $A' \subseteq A$, of size $|A'| \geq 2^{-O(dk)}|A|$, such that

$$\left\langle A' \star A', \mathbb{1}_S \right\rangle \leq O(\eta).$$

More specifically, A' is of the form

$$A' = A \cap (A + s_1) \cap (A + s_2) \cap \cdots \cap (A + s_{k-1})$$

for some choice of additive shifts s_i. Thus we can obtain a robust witness to $A \star A \gg 1 + \Omega(1)$, which has the form $A' \star A'$, as desired.

We offer the following interpretation of this. Taking multiple additive perturbations of the set A and considering their common intersection is sufficient to "uncover" or "reveal" the additive structure hidden within A. This technique has been used in some form by various works in the existing literature; we call it "sifting", and in this work we develop some refinements to it.

B.3 The Pre-BSG Lemma and the sifting lemma

Let us consider the following lemma due to Schoen [Sch15], which we state in the density formulation.

Lemma B.5 (Schoen's Pre-BSG Lemma)**.** *Let A be subset of a finite abelian group G of size $|A| \geq \delta|G|$. Consider the sub-level set*

$$S := \left\{ x \in G \,:\, (A \star A)(x) \leq c \cdot \|A \star A\|_2^2 \right\}.$$

There is a subset $A' \subseteq A$ with

$$\left\langle A' \star A', \mathbb{1}_S \right\rangle \leq 16c$$

and

$$\frac{|A'|}{|G|} \geq \tfrac{1}{3}\delta^2 \cdot \|A \star A\|_2^2 \geq \tfrac{1}{3}\delta^2.$$

Specifically, A' is of the form

$$A' = A \cap (A + s)$$

for some $s \in G$.

We call this the "Pre-BSG Lemma" as it can be considered as a sort of soft, analytic precursor to the Balog-Szemerédi-Gowers lemma – indeed, Schoen shows that it readily implies the following.

Lemma B.6 (Balog-Szemerédi-Gowers [Sch15])**.** *Suppose that $A \subseteq G$ has*

$$E(A) = \sum_{x \in G} R_A(x)^2 = \kappa |A|^3.$$

Then there is a subset $B \subseteq A$ with

$$|B - B| \leq O(\kappa^{-4}|B|)$$

and

$$|B| \geq \Omega(\kappa|A|).$$

The name "Pre-BSG Lemma" is certainly ahistorical, however in a sense it is fairly appropriate in spirit: the technique used to prove it (sifting) is also the technique Gowers uses to prove his own combinatorial precursor-lemma ([Gow01, Lemma 7.3]) from which he deduces (a form of) the BSG Lemma.

We remark on a peculiar feature of the Pre-BSG Lemma. It is a tool that is useful for studying both sparse and dense subsets[23] of a finite abelian group G, which we roughly delineate as the regimes $|A| \leq |G|^{1/2}$ and $|A| \geq |G|^{1/2}$. This is in contrast to e.g. other tools such as Chang's inequality, which says nothing interesting regarding sparse sets, or *the BSG Lemma itself* which says nothing interesting regarding dense sets except in some very extreme cases.[24] While the density-formulation of the Pre-BSG Lemma stated above is technically equivalent[25] to the counting-measure formulation stated in [Sch15], translating between the two settings is a nontrivial exercise in bookkeeping and for studying sparse sets $A \subseteq G$ one would much prefer to work from Schoen's original formulation. It is possibly better for practical purposes even to consider the tool as "conceptually different" as it applies to the two settings.

One can check that Schoen's Pre-BSG Lemma, together with the plan laid out above, is sufficient to obtain a density increment

$$\langle V, A \rangle \geq 1 + \Omega(1)$$

from an assumption such as

$$\|A \star A\|_2^2 \geq 2^7.$$

To obtain a density increment from milder assumptions of the form $\|A \star A\|_k \geq 1 + \varepsilon$, we turn to the following.

Lemma B.7 (Sifting lemma – simplified version)**.** *Suppose $A \subseteq G$ has size $|A| = \delta|G|$. Fix a nonnegative function $f : G \to \mathbb{R}_{\geq 0}$ and an integer $k \geq 2$. Suppose that*

$$\left\langle (A \star A)^{\wedge k}, f \right\rangle = \eta.$$

Then there is a subset $A' \subseteq A$ with

$$\left\langle A' \star A', f \right\rangle \leq 2\eta$$

and

$$\frac{|A'|}{|G|} \geq \tfrac{1}{2}\delta^k.$$

Specifically, A' is of the form

$$A' = A \cap (A + s_1) \cap (A + s_2) \cap \cdots \cap (A + s_{k-1})$$

for some shifts $s_i \in G$.

[23] Certainly it seems to have seen more applications to the study of sparse sets (via the BSG Lemma). However, see e.g. Sanders' work [San10] where he applies a (chronologically earlier) variant to the study of dense sets.

[24] Indeed, for a set of size $|A| = \delta|G|$, consider the ratio $\kappa = E(A)/|A|^3$. We have $\kappa = \delta\|A \star A\|_2^2 \in [\delta, 1]$. In our present context, we consider A to have noticeable additive structure already when $\|A \star A\|_2 \geq 1 + \Omega(1)$; in contrast the BSG Lemma above is not better than the trivial bound $|A - A| \leq |G| \leq \delta^{-1}|A|$ unless $\|A \star A\|_2 \geq \delta^{-3/8}$.

[25] That is, equivalent in the case of finite groups G.

Combining this with the plan laid out above suffices to prove Theorem 2.8; it remains only to prove the sifting lemma and to work out the quantitative details. We note that as stated this claim is incomparable with Schoen's Pre-BSG Lemma – indeed, it says nothing interesting regarding sparse sets $A \subseteq G$ because of the size guarantee on A' – but it does suffice for the applications considered in this work. With some extra effort one can derive a version of the sifting lemma which leads to the following strict improvement which may be useful elsewhere; we state it below in the counting-measure formulation.

Lemma B.8 (Extended Pre-BSG Lemma)**.** *Suppose A is a finite subset of an abelian group G with*

$$E_k(A) = \sum_{x \in G} R_A^-(x)^k = \kappa |A|^{k+1}$$

for some integer $k \geq 2$. Consider the sub-level set

$$S := \left\{ x \in G \,:\, R_A^-(x) \leq c \cdot \kappa^{\frac{1}{k-1}} \cdot |A| \right\}.$$

There is a subset $A' \subseteq A$ with

$$\frac{1}{|A'|^2} \sum_{a,b \in A'} \mathbb{1}_S(a-b) \leq 2c^{k-1}$$

and

$$|A'| \geq \tfrac{1}{2}\kappa |A|.$$

C Sanders' invariance lemma

C.1 The Croot-Sisask lemma

For a point $p \in G$, define the shift operator T_p by

$$(T_p f)(x) := f(x-p).$$

This operation can be equivalently described as convolution of f with the point-mass density at p (and hence it commutes with other convolutions). For a density B, we define the linear operator T_B analogously: $(T_B f)(x) := (B * f)(x)$.

Lemma C.1 (Croot-Sisask lemma)**.** *Let G be a finite abelian group, and suppose $A, A' \subseteq G$ are such that*

$$|A| \geq 2^{-d}|A + A'|,$$

which is in particular the case if simply $A' = G$ and $|A| \geq 2^{-d}|G|$. Fix an even integer $k \geq 2$. and a function $f : G \to \mathbb{R}$ such that $\|f\|_k \leq 1$ (which is in particular the case if f maps into $[-1, +1]$). For any $\varepsilon \in [0,1]$, there exists a set $S \subseteq A'$ of size at least

$$|S| \geq 2^{-O(kd/\varepsilon^2)}|A'|$$

such that for any points $p, p' \in S$,

$$\left\| T_{p-p'} A * f - A * f \right\|_k \leq \varepsilon.$$

Consequently, for any integer t and any point $p \in tS - tS$,

$$\left\| T_p A * f - A * f \right\|_k \leq t\varepsilon.$$

Proof. Fix an integer $\ell \in \mathbb{N}$ which is somewhat larger than k. We consider the idea of approximating $A * f$ by a random "sketch"

$$s(a_1, a_2, \ldots, a_\ell) := \frac{1}{\ell} \sum_{i=1}^{\ell} T_{a_i} f$$

If each a_i is drawn independently from A, then $\mathbb{E}[T_{a_i} f] = A * f$, and the vector-valued Khintchine inequality (proved below for even k) gives

$$\mathop{\mathbb{E}}_{a} \|s(a) - A * f\|_k^k = \mathop{\mathbb{E}}_{a} \left\| \frac{1}{\ell} \sum_{i=1}^{\ell} (T_{a_i} - T_A) f \right\|_k^k \leq \left(\frac{k}{\ell}\right)^{k/2} \cdot \|f\|_k^k,$$

So by a Markov inequality, the chance that $\|s(a) - A * f\|_k \geq 2\sqrt{\frac{k}{\ell}}$ is at most 2^{-k}.

Now, we seek a simultaneous approximation of $T_p A * f = T_{A+p} f$ for all points $p \in A'$ by some sketches

$$s(y) = s(y_1, y_2, \ldots y_\ell) := \frac{1}{\ell} \sum_{i=1}^{\ell} T_{y_i} f,$$

this time allowing $y \in (A + A')^\ell$ instead of just $a \in A^\ell$. We say that y is a *plausible* sketch for $T_{A+p} f$ if $y_i \in (A + p)$ for all i. We say that y is a *good* sketch for $T_{A+p} f$ if

$$\|s(y) - T_{A+p} f\|_k \leq 2\sqrt{\frac{k}{\ell}}.$$

By the same argument as above, for every p, at least a fraction $(1 - 2^{-k})$ of the sketches which are plausible for $T_{A+p} f$ are in fact good for $T_{A+p} f$.

Now we form a bipartite graph on the vertex-set $A' \times (A \times A')^\ell$, including the edge (p, y) whenever $s(y)$ is a good sketch for $T_{A+p} f$. We wish to find a sketch $s(y)$ on the right side which is simultaneously good for as many points p as possible, and we will do so by a pigeonhole argument:

- Vertices on the left side have degree at least $\frac{1}{2}|A|^\ell$.
- We have the following relationship between the average degrees of both sides (D_L, D_R), the sizes of the vertex-sets of both sides (N_L, N_R), and the total number of edges E:

$$D_L \cdot N_L = E = D_R \cdot N_R.$$

- Thus, $D_R \geq \frac{D_L}{N_R} \cdot N_L \geq \frac{1}{2} \left(\frac{|A|}{|A+A'|}\right)^\ell |A'|$.

So, we can find a fixed sketch $s(y)$ on the right adjacent to a large set of points $S \subseteq A'$ on the left, meaning that $s(y)$ is a good sketch for every $p \in S$. Since $T_{A+p} f$ are all close to $s(y)$, they are all close to each other:

$$\left\|T_p A * f - T_{p'} A * f\right\|_k = \left\|T_{A+p} f - T_{A+p'} f\right\|_k \leq \|T_{A+p} f - s(y)\|_k + \left\|s(y) - T_{A+p'} f\right\|_k \leq 4\sqrt{\frac{k}{\ell}}$$

for all $p, p' \in S$. To conclude, note that since $\|T_x g\|_k = \|g\|_k$ for any function g and any point x, we have

$$\left\|T_p A * f - T_{p'} A * f\right\|_k = \left\|T_{-p'}(T_p A * f - T_{p'} A * f)\right\|_k = \left\|T_{p-p'} A * f - A * f\right\|_k.$$

Setting $\ell = 16 \cdot k/\varepsilon^2$ gives the desired parameters. □

Proposition C.2 (Vector-valued Khintchine inequality)**.** *Let $k \geq 2$ be an even integer. Let $v_1, v_2, \ldots, v_\ell \in \mathbb{R}^m$ be some independent vector-valued random variables with means $\mu_i := \mathbb{E}[v_i]$. Suppose that*

$$\mathbb{E}\,\|v_i\|_k^k := \mathbb{E}\left[\frac{1}{m}\sum_{j\in[m]} v_i(j)^k\right] \leq M$$

for all i. Consider the average

$$v := \frac{1}{\ell}\sum_{i=1}^{\ell}(v_i - \mu_i)$$

of the (centered versions of) the v_i's. We have the following bound on the average value of $\|v\|_k^k$:

$$\mathbb{E}\left[\|v\|_k^k\right] \leq \left(\frac{k}{\ell}\right)^{k/2} M.$$

Proof. By re-scaling we may assume $M = 1$. We may also assume $\mu_i = 0$ for all i, since $\mathbb{E}\,\|v_i - \mu_i\|_k^k \leq \mathbb{E}\,\|v_i\|_k^k$ for $k \geq 2$. Define the k-wise dot product of some k vectors $u_i \in \mathbb{R}^m$ by

$$u_1 \cdot u_2 \cdots u_k := \mathop{\mathbb{E}}_{j\in[m]} u_1(j)u_2(j)\cdots u_k(j).$$

Now, we compute

$$\|v\|_k^k = \mathop{\mathbb{E}}_{(i_1,i_2,\ldots,i_k)\in[\ell]^k} v_{i_1}\cdot v_{i_2}\cdots v_{i_k}$$

Upon taking the expectation of this quantity, any term which contains a certain random variable v_i only once in the dot product becomes zero. For the remaining terms, we can bound

$$\mathbb{E}[v_{i_1}\cdot v_{i_2}\cdots v_{i_k}] \leq \prod_{j=1}^{k}\left(\mathbb{E}\,\|v_{i_j}\|_k^k\right)^{1/k} \leq 1.$$

So we reduce to the combinatorial problem of counting the number of tuples $\mathbf{i} \in [\ell]^k$ which have no unique entries. Let $T(k,\ell)$ denote this quantity. We can bound this number recursively as follows: Consider the first entry i_1 of such a tuple. There must be some i_j with $i_1 = i_j$; we consider all the possibilities for this, then remove the first and j-th entry from the tuple, and then count the number of tuples in $[\ell]^{k-2}$ which have no unique entries. This argument gives the bound

$$T(k,\ell) \leq k\cdot\ell\cdot T(k-2,\ell) \leq \cdots \leq (k\ell)^{k/2}.$$

So overall we have

$$\mathbb{E}\left[\|v\|_k^k\right] \leq \frac{(k\ell)^{k/2}}{\ell^k} = \left(\frac{k}{\ell}\right)^{k/2}. \qquad \square$$

C.2 Sanders' invariance lemma

Lemma C.3 (Sanders' invariance lemma, restated)**.** *Suppose $A, B \subseteq \mathbb{F}_q^n$ are sets of sizes $|A| \geq 2^{-d}|\mathbb{F}_q^n|$ and $|B| \geq 2^{-k}|\mathbb{F}_q^n|$. Fix a bounded function $f : \mathbb{F}_q^n \to [-1,+1]$. Then, for any $\varepsilon \geq 2^{-d}$, there exists a linear subspace V (possibly depending on f) with codimension at most $O(kd^3/\varepsilon^2)$ satisfying*

$$|\langle V * A * B, f\rangle - \langle A * B, f\rangle| \leq \varepsilon.$$

More specifically, we have the pointwise bound

$$\left| \mathop{\mathbb{E}}_{\substack{a\in A\\ b\in B}} f(x+v+a+b) - \mathop{\mathbb{E}}_{\substack{a\in A\\ b\in B}} f(x+a+b) \right| \le \varepsilon$$

for all $v \in V$ *and* $x \in \mathbb{F}_q^n$.

Proof. We seek a large linear subspace V such that

$$\|T_v A \star f - A \star f\|_k \le \frac{\varepsilon}{2}$$

for all $v \in V$. We note that this suffices, since

$$|\langle T_x B, T_v A \star f - A \star f\rangle| \le \|T_x B\|_1 \, \|T_x B\|_\infty^{1/k} \, \|T_v A \star f - A \star f\|_k \le 2\|T_v A \star f - A \star f\|_k$$

for any $x \in \mathbb{F}_q^n$. First we apply the Croot-Sisask lemma to obtain a large set S of shifts T_p which leave $A \star f$ approximately unchanged: $\|T_{p-p'} A \star f - A \star f\|_k \le \eta$ for all $p, p' \in S$ where η is some parameter we will pick later. We can obtain

$$|S| \ge 2^{-O(kd/\eta^2)} |\mathbb{F}_q^n|.$$

Continuing on, let's write $g := A \star f$ for brevity. We now describe how to pick the linear subspace V. First, consider the large spectrum of S:

$$\mathrm{Spec}_{1/2}(S) := \left\{ \alpha \in \mathbb{F}_q^n \ : \ |\widehat{S}(\alpha)| \ge \tfrac{1}{2} \right\}.$$

We let $W := \mathrm{Span}(\mathrm{Spec}_{1/2}(S))$. By Chang's inequality for vector spaces over finite fields (proved below), the dimension of W is at most $O(kd/\eta^2)$. Finally, we let $V := W^\perp$.

Consider the density function $D = S \star S$, and let X denote the of the t-iterated convolution $D * D * \cdots * D$. We have

$$X(x) = \sum_\alpha |\widehat{S}(\alpha)|^{2t} e_\alpha(-x),$$

and in particular for any $\alpha \notin W$, we have $|\widehat{X}(\alpha)| \le 2^{-2t}$. We note that

$$(V * X * g - X * g)(x) = -\sum_{\alpha \notin W} \widehat{X}(\alpha) \widehat{A}(\alpha) \widehat{f}(\alpha) e_\alpha(-x).$$

Thus

$$\|V * X * g - X * g\|_\infty \le 2^{-2t} \sqrt{\sum_\alpha |\widehat{A}(\alpha)|^2} \sqrt{\sum_\alpha |\widehat{f}(\alpha)|^2} \le 2^{-2t} \cdot 2^d.$$

Finally, we estimate

$$\begin{aligned}
\|g - V * g\|_k &\le \|g - X * g\|_k + \|X * g - V * X * g\|_k + \|V * X * g - V * g\|_k \\
&= \|g - X * g\|_k + \|X * g - V * X * g\|_k + \|V * (X * g - g)\|_k \\
&\le 2\,\|g - X * g\|_k + \|X * g - V * X * g\|_\infty \\
&\le 2t\eta + 2^{d-2t},
\end{aligned}$$

and

$$\|T_v g - V * g\|_k = \|g - T_{-v} V * g\|_k = \|g - V * g\|_k \le 2t\eta + 2^{d-2t},$$

which together give a bound on $\|T_v g - g\|_k$. Choosing $t := O(d)$ and $1/\eta := O(t/\varepsilon)$, we obtain the desired error bound. The resulting bound on the codimension of V is $O(kdt^2/\varepsilon^2) = O(kd^3/\varepsilon^2)$. $\square$

Lemma C.4 (Chang's inequality for vector spaces)**.** *Suppose $A \subseteq \mathbb{F}_q^n$ has size $|A| \geq 2^{-d}|\mathbb{F}_q^n|$, where $d \geq 1$. Then any subset of linearly independent vectors $L \subseteq Spec_\varepsilon(S)$ has size at most*

$$|L| \leq 4d/\varepsilon^2.$$

Proof. Suppose $\alpha_1, \ldots, \alpha_\ell \subseteq \mathrm{Spec}_\varepsilon(S)$ are linearly independent. Let $c_i := \widehat{A}(\alpha_i)$, and consider the auxiliary function

$$f(x) := \sum_{i=1}^{\ell} c_i \cdot e_{\alpha_i}(-x).$$

On one hand we have

$$\langle A, f \rangle = \sum_{i=1}^{\ell} c_i \mathop{\mathbb{E}}_{x \in A} e_{\alpha_i}(-x) = \sum_{i=1}^{\ell} |\widehat{A}(\alpha_i)|^2.$$

On the other hand (choosing $k = d$) we can bound

$$\langle A, f \rangle \leq \|A\|_1 \|A\|_\infty^{1/k} \|f\|_k = 2^{d/k} \cdot \|f\|_k \leq 2 \cdot \|f\|_k.$$

Now, since the vectors $\alpha_i \in \mathbb{F}_q^n$ are linearly independent, the complex random variables defined by $X_i := e_{\alpha_i}(x)$ (where x is uniformly random in $\mathbb{F}_q^n$) are in fact mutually independent. So by a Khintchine inequality, $\|f\|_k \leq \sqrt{k} \cdot \|f\|_2 = \sqrt{k} \cdot \sqrt{\sum_i |\widehat{A}(\alpha_i)|^2}$. We conclude that

$$\varepsilon^2 \cdot \ell \leq \sum_{i=1}^{\ell} |\widehat{A}(\alpha_i)|^2 \leq 4k = 4d,$$

and so $\ell \leq 4d/\varepsilon^2$. □

D Odd moments

Proof of Proposition 5.7. We wish to show that $\mathbb{E}\, X^{k'} \geq \frac{1}{4}(1+\varepsilon)^{k'}$. We express

$$\begin{aligned}
\mathbb{E}\, X^{k'} &= \mathbb{E}(1 + (X-1))^{k'} \\
&= \sum_{k=0}^{k'} \binom{k'}{k} \cdot \mathbb{E}(X-1)^k \\
&\geq \sum_{\substack{k \text{ even} \\ k_0 \leq k \leq k'}} \binom{k'}{k} \cdot \varepsilon^k.
\end{aligned}$$

Towards understanding this quantity we apply the following trick. We introduce a uniform random variable $\theta \in \{+1, -1\}$ and consider the value $\mathbb{E}[(1+\theta\varepsilon)^{k'}]$. On one hand this value is of course

$$\mathop{\mathbb{E}}_{\theta}(1+\theta\varepsilon)^{k'} = \tfrac{1}{2}(1+\varepsilon)^{k'} + \tfrac{1}{2}(1-\varepsilon)^{k'}.$$

It can be expressed also as

$$\mathop{\mathbb{E}}_{\theta}(1+\theta\varepsilon)^{k'} = \sum_{k=0}^{k'} \binom{k'}{k} \mathbb{E}[\theta^k]\varepsilon^k = \sum_{\substack{k \text{ even} \\ 0 \leq k \leq k'}} \binom{k'}{k} \varepsilon^k;$$

we conclude that

$$\sum_{\substack{k \text{ even} \\ 0 \le k \le k'}} \binom{k'}{k} \varepsilon^k \ge \tfrac{1}{2}(1+\varepsilon)^{k'}.$$

At this point, let us normalize by $(1+\varepsilon)^{k'}$ so that we can proceed via a probabilistic interpretation: we are interested in the probability of a certain event concerning a random variable $k \in \mathbb{N}$ drawn according to a binomial distribution with k' trials and success-probability $p = \frac{\varepsilon}{1-\varepsilon}$:

$$k \sim \mathrm{Bin}\left(k', \frac{\varepsilon}{1-\varepsilon}\right).$$

In particular, we are interested in the chance that both

1. k is even, and
2. $k \ge k_0$.

So far, we have calculated that the chance that k is even is at least a half. In addition, it is known that a median of a binomial distribution $\mathrm{Bin}(k', p)$ is at least $\lfloor pk' \rfloor$ [KB80]. Since $k_0 \le \varepsilon k'/2 \le \varepsilon k'/(1-\varepsilon)$, we also have that the chance that $k \ge k_0$ is at least a half. One might then suspect that the probability that both events occur simultaneously should be roughly at least 1/4; the following calculation confirms this. Suppose $t \le \varepsilon k'$ is an odd natural number. Then

$$\begin{aligned}
\mathbb{P}(k = t) &= \frac{k'-t}{t}\left(\frac{\varepsilon}{1-\varepsilon}\right) \cdot \mathbb{P}(k = t-1) \\
&\ge \left(\frac{1}{\varepsilon} - 1\right)\left(\frac{\varepsilon}{1-\varepsilon}\right) \cdot \mathbb{P}(k = t-1) \\
&= \mathbb{P}(k = t-1)
\end{aligned}$$

and so

$$\mathbb{P}(k \text{ is even and } k < k_0) \le \tfrac{1}{2} \cdot \mathbb{P}(k < k_0) \le \tfrac{1}{4}.$$

We infer that

$$\mathbb{P}(k \text{ is even and } k \ge k_0) \ge \tfrac{1}{2} - \tfrac{1}{4} = \tfrac{1}{4}$$

and so

$$\mathbb{E}\, X^{k'} \ge \tfrac{1}{4} \cdot (1+\varepsilon)^{k'}.$$

Equivalently,

$$\left(\mathbb{E}\, X^{k'}\right)^{1/k'} \ge 4^{-1/k'} \cdot (1+\varepsilon).$$

To conclude, one can check that

$$4^{-\varepsilon/4} \cdot (1+\varepsilon) \ge 1 + \frac{\varepsilon}{2}$$

for $\varepsilon \in [0, \frac{1}{2}]$. □