

UC San Diego

UC San Diego Previously Published Works

Title

strong-bounds-for-skew-corner-free-sets

Permalink

<https://escholarship.org/uc/item/0z63x71w>

Journal

Discrete Analysis

ISSN

2397-3129

Authors

Jaber, Michael

Lovett, Shachar

Ostuni, Anthony

Publication Date

2025

DOI

10.19086/da.144040

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Strong Bounds for Skew-Corner-Free Sets

Michael Jaber*

Shachar Lovett†

Anthony Ostuni‡

Received 16 April 2024; Published 30 September 2025

Abstract: Motivated by applications to matrix multiplication algorithms, Pratt asked (ITCS'24) how large a subset of $[n] \times [n]$ could be without containing a *skew corner*: three points (x, y) , $(x, y + h)$, $(x + h, y')$ with $h \neq 0$. We prove any skew-corner-free set has size at most $\exp(-\Omega(\log^{1/12} n)) \cdot n^2$, nearly matching the best known lower bound of $\exp(-O(\sqrt{\log n})) \cdot n^2$ by Beker (arXiv'24). Our techniques generalize those of Kelley and Meka's recent breakthrough on three-term arithmetic progression (FOCS'23), answering a question of Beker (arXiv'24). We note that a similar bound was obtained concurrently and independently by Milićević (arXiv'24).

Key words and phrases: skew corners, corners, arithmetic progressions

1 Introduction

In an attempt to rule out certain groups as candidates for achieving an optimal matrix multiplication algorithm using the framework of [10, 9], Pratt [18] introduced *skew corners*: triples of points (x, y) , $(x, y + h)$, and $(x + h, y')$ for arbitrary x, y, y' , and h . Note that the more traditional form of corners can be recovered by insisting $y = y'$. We call a skew corner trivial if $h = 0$ and nontrivial otherwise, and we call a set skew-corner free if it does not contain any nontrivial skew corners. Pratt's matrix multiplication results are conditional, holding if the largest skew-corner-free subset of $[n] \times [n]$ has size $O(n^{1+\varepsilon})$ for all $\varepsilon > 0$, as well as weaker results for the upper bound $O(n^{4/3-\varepsilon})$ for any fixed $\varepsilon > 0$ [18].

Very recently, these two conjectures were refuted. Pohoata and Zakharov [19] constructed a skew-corner-free set of size $\Omega(n^{5/4})$, which was improved shortly after to $\exp(-O(\sqrt{\log n})) \cdot n^2$ by Beker [3] using a Behrend-type construction (see [2]). In the same paper, Beker also improved the best known upper bounds to $O((\log n)^{-c}) \cdot n^2$ for an absolute constant $c > 0$, and went on to ask if one could use the techniques of the recent breakthrough on three-term arithmetic progressions (3APs) [13] (see also [8, 7]) to obtain a bound of $\exp(-\Omega(\log^c n)) \cdot n^2$ for some constant $c > 0$. We resolve this question affirmatively.

Theorem 1.1. *Any skew-corner-free subset of $[n] \times [n]$ has size at most $\exp\left\{-\Omega\left(\log^{1/12} n\right)\right\} \cdot n^2$.*

*Supported by NSF Grant CCF-2312573 and a Simons Investigator Award (#409864, David Zuckerman).

†Supported by NSF DMS award 1953928, NSF CCF award 2006443, and a Simons investigator award. The research was done while on sabbatical at the Weizmann Institute of Science, partially funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No. 949083.

‡Supported by NSF DMS award 1953928.

In fact, we can prove such a result for any skew-corner-free subset of $G \times G$, where G is an arbitrary finite Abelian group. It is likely that the constant $1/12$ can be somewhat improved using ideas similar to [7]. However, the overall technique's limitations appear to prevent improvements significant enough to approach Beker's lower bound [3], so we instead prioritize a streamlined presentation.

As is somewhat standard, we first exhibit our argument in the simpler finite field setting. Here, we can obtain slightly stronger bounds.

Theorem 1.2. *Let q be a prime power. Any skew-corner-free subset of $\mathbb{F}_q^n \times \mathbb{F}_q^n$ has size at most $q^{2n-\Omega(n^{1/9})}$.*

A version of Theorem 1.2 with the superior bound $O(q^{(2-c)n})$ (for a constant $c > 0$ depending on q) is known via the polynomial method [19]. However, this technique does not appear to generalize to the integer case, whereas our analytic approach is able to do so relatively painlessly.

We largely view this work as a step toward better understanding the traditional corners problem of how large a subset of $[n] \times [n]$ one can have without three points $(x, y), (x, y + h), (x + h, y)$. This problem is much older and better studied than the skew variant, yet the best upper bound of $O(n^2/(\log \log n)^{0.0137\dots})$ [21] remains far from the lower bound of $\exp(-O(\sqrt{\log n})) \cdot n^2$ [2] (see also [14, 11, 12]).

There has been optimism (see, e.g., [15] and [17, Section 1.2]) that the techniques recently used to dramatically improve the best known upper bounds on 3AP-free sets [13] could also be applied in the (traditional) corners setting. Our work provides evidence that this is a potentially fruitful direction. In fact, we are able to either black-box or closely follow the proofs of many lemmas from the excellent exposition of Bloom and Sisask [8] of Kelley and Meka's work, raising the following question about the relationship between these problems.

Question 1.3. *Suppose the largest 3AP-free subset of $[n]$ has size δn . Does the largest skew-corner-free subset of $[n] \times [n]$ have size $\text{poly}(\delta) \cdot n^2$? More generally, is it possible to directly convert bounds from one setting to another?*

Concurrent work. A similar result to Theorem 1.1 was obtained concurrently and independently by Milićević [16]. His techniques are also based on the ideas of [13], although the specifics diverge slightly from our own.

2 Background

Let G be a finite Abelian group throughout (although we will only truly be concerned with vector spaces over a finite field $\mathbb{F}_q^n$ or the integers modulo a positive integer $\mathbb{Z}/N\mathbb{Z}$).

Asymptotics. We use standard asymptotic notation of $O(\cdot)$ and $\Omega(\cdot)$ to suppress fixed constants that do not depend on any parameters. Occasionally, we will use subscripts to hide a dependence on particular parameters (e.g., $O_\varepsilon(\cdot)$).

Distributions. A distribution ν over G is a non-negative function $\nu: G \rightarrow \mathbb{R}_{\geq 0}$ with $\mathbb{E}[\nu] = 1$. For $f: G \rightarrow \mathbb{R}_{\geq 0}$ define $\nu(f) = \mathbb{E}_{x \in G} \nu(x)f(x)$ to be the average of f under ν . We also define ν_f to be f normalized to have expectation 1 under ν ; that is, $\nu_f(x) = \nu(x)f(x)/\nu(f)$. We use μ to denote the uniform distribution on G , so $\mu \equiv 1$. Given a set $A \subseteq G$, we identify A with its indicator function 1_A and define accordingly $\nu(A) = \nu(1_A)$ and $\nu_A = \nu_{1_A}$. In this case, $\mu_A = \frac{|G|}{|A|} \cdot 1_A$, and in particular $\mathbb{E}_{x \in G} \mu_A(x) = 1$. In other words, μ_A is the normalized indicator function of A .

Functions. For functions $f, g: G \rightarrow \mathbb{R}$, we define inner products and norms with the normalized counting measure on G , namely

$$\langle f, g \rangle = \mathbb{E}_{x \in G} f(x)g(x) \quad \text{and} \quad \|f\|_p = \left(\mathbb{E}_{x \in G} |f(x)|^p \right)^{1/p} \quad \text{for } 1 \leq p < \infty,$$

as well as $\|f\|_\infty = \max_{x \in G} |f(x)|$. We will also want to work with other measures on G . For a measure ν on G , we write

$$\langle f, g \rangle_\nu = \mathbb{E}_{x \in G} \nu(x) f(x)g(x) \quad \text{and} \quad \|f\|_{p(\nu)} = \left(\mathbb{E}_{x \in G} \nu(x) |f(x)|^p \right)^{1/p} \quad \text{for } 1 \leq p < \infty.$$

Additionally, we define the convolution and the difference convolution as

$$(f * g)(x) = \mathbb{E}_{y \in G} f(y)g(x-y) \quad \text{and} \quad (f \circ g)(x) = \mathbb{E}_{y \in G} f(y)g(x+y),$$

and the p -fold convolution as $f^{(p)} = f * f^{(p-1)}$, where $f^{(1)} = f$. Note the useful adjoint property $\langle f, g * h \rangle = \langle h \circ f, g \rangle$. Occasionally, the notation $f^g(x) = f(x - g)$ for some $g \in G$ will be useful.

Let $\widehat{G} = \{\text{homomorphisms } \gamma: G \rightarrow \mathbb{C}^\times\}$ be the dual group of G . The Fourier transform of f is $\widehat{f}: \widehat{G} \rightarrow \mathbb{C}$, where

$$\widehat{f}(\gamma) = \mathbb{E}_{x \in G} f(x)\gamma(-x).$$

If $\widehat{f}(\gamma) \geq 0$ for all $\gamma \in \widehat{G}$, we say f is *spectrally non-negative*, denoted $\widehat{f} \geq 0$. Note the basic facts that $\widehat{f * g} = \widehat{f} \cdot \widehat{g}$ and $\widehat{f \circ f} = |\widehat{f}|^2$. In particular, $f \circ f$ is spectrally non-negative. We define the convolution for two functions $f, g: \widehat{G} \rightarrow \mathbb{C}$ using the ordinary counting measure:

$$(f * g)(\gamma) = \sum_{\tau \in \widehat{G}} f(\tau)g(\gamma - \tau).$$

The use of the ordinary counting measure here is so that we have the identity $\widehat{f^p} = \widehat{f}^{(p)}$.

2.1 Skew corners

Sequences of evenly spaced integers, known as *arithmetic progressions*, are central objects of study in additive combinatorics. In an attempt to understand similar structure in higher dimensions, Ajtai and Szemerédi [1] considered the following objects.

Definition 2.1 (Corners). A corner in $G \times G$ is a triple of the form $(x, y), (x, y + h), (x + h, y)$ for arbitrary $x, y, h \in G$. A corner is nontrivial if $h \neq 0$.

By relaxing the structure to allow the third point to lie anywhere on the vertical line $x + h$, one obtains *skew corners*, a generalization recently raised by Pratt [18].

Definition 2.2 (Skew corners). A skew corner in $G \times G$ is a triple of the form $(x, y), (x, y + h), (x + h, y')$ for arbitrary $x, y, y', h \in G$. A skew corner is nontrivial if $h \neq 0$. Additionally, a set is *skew-corner-free* if it does not contain nontrivial skew corners.

One immediate, yet useful, observation is that shifting a set will not change the number of skew corners it contains.

Observation 2.3. The number of skew corners in a set A is invariant to both horizontal shifting ($A \rightarrow \{(x + a, y) : (x, y) \in A\}$) and vertical shifting ($A \rightarrow \{(x, y + b(x)) : (x, y) \in A\}$).

The following lemma allows us to count skew corners in an analytic way.

Lemma 2.4. *Let $A \subseteq G \times G$ and $A_x = \{y : (x, y) \in A\}$. Additionally, let $D : G \rightarrow \mathbb{R}_{\geq 0}$ be given by $D(x) = |A_x|$. Then, the number of skew corners in A is counted by*

$$|G|^2 \sum_{x \in G} \langle 1_{A_x} \circ 1_{A_{x+x}}, D \rangle.$$

Proof. Recall a skew corner in A is a triple of the form

$$(x, y), (x, y+h), (x+h, y')$$

for arbitrary $x, y, y', h \in G$. The number of skew corners is given by

$$\begin{aligned} \sum_{x, y, y', h \in G} 1_A(x, y) 1_A(x, y+h) 1_A(x+h, y') &= \sum_{x, y, y', h} 1_{A_x}(y) 1_{A_x}(y+h) 1_{A_{x+h}}(y') \\ &= \sum_{x, y, h} 1_{A_x}(y) 1_{A_{x+x}}(x+y+h) D(x+h) \\ &= |G| \sum_{x, h} (1_{A_x} \circ 1_{A_{x+x}})(x+h) D(x+h) \\ &= |G|^2 \sum_x \langle 1_{A_x} \circ 1_{A_{x+x}}, D \rangle. \end{aligned} \quad \square$$

Occasionally, the following normalized form will be more useful.

Lemma 2.5. *Let $A \subseteq G \times G$, $A_x = \{y : (x, y) \in A\}$, and $\eta \in \mathbb{R}_{\geq 0}$. Additionally, let $f = \sum_{x \in G} 1_{A_x} \circ 1_{A_{x+x}}$ and $D : G \rightarrow \mathbb{R}_{\geq 0}$ be given by $D(x) = |A_x|$. If $\langle \mu_f, \mu_D \rangle \geq \eta$, then the number of skew corners in A is at least*

$$\eta \cdot \frac{|A|^3}{|G|^2}.$$

Proof. Unraveling the definition of $\langle \mu_f, \mu_D \rangle$, we have

$$\begin{aligned} \eta &\leq \left\langle \frac{1}{\mathbb{E}[f]} \sum_{x \in G} 1_{A_x} \circ 1_{A_{x+x}}, \frac{D}{\mathbb{E}[D]} \right\rangle = \frac{1}{\mathbb{E}[f]} \cdot \frac{1}{\mathbb{E}[D]} \sum_{x \in G} \langle 1_{A_x} \circ 1_{A_{x+x}}, D \rangle \\ &= \frac{|G|^2}{\sum |A_x|^2} \cdot \frac{|G|}{\sum |A_x|} \sum_{x \in G} \langle 1_{A_x} \circ 1_{A_{x+x}}, D \rangle \\ &\leq \frac{|G|^3}{|A|^2} \cdot \frac{|G|}{|A|} \sum_{x \in G} \langle 1_{A_x} \circ 1_{A_{x+x}}, D \rangle \quad (\text{Cauchy-Schwarz}) \\ &= \frac{|G|^2}{|A|^3} \times (\text{number of skew corners in } A). \quad (\text{Lemma 2.4}) \quad \square \end{aligned}$$

2.2 Bohr sets

To prove Theorem 1.2, we will take advantage of the abundance of subspaces in $\mathbb{F}_q^n$. As this no longer holds once we move to more general Abelian groups, we will turn to the machinery of *Bohr sets*, introduced by Bourgain [5], to prove Theorem 1.1. We will only require the very basics, but one may entirely skip over this topic if they are solely interested in a high-level understanding of our results; all the main ideas are present in the simpler finite field case (Section 3).

Definition 2.6 (Bohr sets). For a nonempty $\Gamma \subseteq \widehat{G}$ and $\phi \in [0, 2]$, we define the Bohr set

$$B = \text{Bohr}(\Gamma, \phi) = \{x \in G : |1 - \gamma(x)| \leq \phi \text{ for all } \gamma \in \Gamma\},$$

where Γ is the frequency set, ϕ is the width (or radius), and $|\Gamma|$ is the rank $\text{rk}(B)$.

It follows from the definition that Bohr sets are symmetric, meaning that if $x \in B$, we must also have $-x \in B$. Note also that a Bohr set B does not uniquely define its frequency set or width. Thus, we always implicitly view B as the tuple $(\Gamma, \phi, \text{Bohr}(\Gamma, \phi))$. However, the notation $B' \subseteq B$ only refers to containment of the actual set itself and does not depend on the frequency set or width. Indeed, there are Bohr sets B, B' where $B' \subset B$, but the frequency set of B is not contained in the frequency set of B' , and the width of B' is not smaller than the width of B . To handle these situations, we follow [22] in defining *sub-Bohr* sets.

Definition 2.7. Let $B = \text{Bohr}(\Gamma, \phi)$, $B' = \text{Bohr}(\Gamma', \phi')$ be Bohr sets. We say that B' is a sub-Bohr set of B , denoted $B' \leq B$, if $\Gamma' \supseteq \Gamma$ and $\phi' \leq \phi$.

Observation 2.8. Let $B' \leq B$ for Bohr sets B, B' . Then, $B' \subseteq B$.

A common operation performed on Bohr sets is *dilation*. If $B = \text{Bohr}(\Gamma, \phi)$ is a Bohr set and $\rho > 0$, then the corresponding *dilate* of B is $B_\rho = \text{Bohr}(\Gamma, \rho\phi)$. The following observation about sub-Bohr set dilates will be useful in later arguments.

Observation 2.9. Let $B' \leq B$ for Bohr sets B, B' . Then, $B'_\rho \leq B_\rho$ for all $\rho > 0$. In particular, $B'_\rho \subseteq B_\rho$.

If small dilations do not substantially change the size of a Bohr set B , we call B *regular*. Regular Bohr sets will be particularly useful for our purposes, as informally, one can view a regular Bohr set of rank r analogously to a subspace of co-dimension r .

Definition 2.10 (Regular Bohr sets). A Bohr set B of rank r is regular if for all $0 \leq \kappa \leq 1/100r$, we have

$$|B_{1+\kappa}| \leq (1 + 100\kappa r)|B| \quad \text{and} \quad |B_{1-\kappa}| \geq (1 - 100\kappa r)|B|.$$

We record some useful lemmas, most of which are listed in the appendix of [8]. See [23, Section 4.4] or [4, Chapter 6] for proofs of the first two standard results, as well as a more thorough treatment of Bohr set machinery.

Lemma 2.11. For any Bohr set B , there exists $\rho \in [\frac{1}{2}, 1]$ such that B_ρ is regular.

Lemma 2.12. If $\rho \in (0, 1)$ and B is a Bohr set of rank r , then $|B_\rho| \geq (\rho/4)^r |B|$.

Lemma 2.13 ([6, Lemma 4.6]). There is a constant $c > 0$ such that the following holds. If B is a regular Bohr set of rank r and ν is a probability measure supported on B_ρ , with $\rho \leq c/r$, then

$$\mu_B \leq 2\mu_{B_{1+\rho}} * \nu.$$

Lemma 2.14 ([6, Lemma 4.5]). If B is a regular Bohr set of rank r and ν is a probability measure supported on B_ρ , with $\rho \in (0, 1)$, then

$$\|\mu_B * \nu - \mu_B\|_1 \leq O(\rho r).$$

We will make use of the following corollary.

Corollary 2.15. If B is a regular Bohr set of rank r and ν is a probability measure supported on B_ρ , with $\rho \in (0, 1)$, and $f: G \rightarrow \mathbb{R}$, then

$$|\langle f * \nu, \mu_B \rangle - \langle f, \mu_B \rangle| \leq O(\|f\|_\infty \cdot \rho r).$$

Proof. We have $\langle f * v, \mu_B \rangle = \langle f, \mu_B \circ v \rangle = \langle f, \mu_B \rangle + \langle f, \mu_B \circ v - \mu_B \rangle$. Since Bohr sets are symmetric, we can replace $v(x)$ with $v(-x)$ and maintain that v is supported on B_ρ . It suffices to bound

$$|\langle f, \mu_B * v - \mu_B \rangle| \leq \|f\|_\infty \|\mu_B * v - \mu_B\|_1 \leq O(\|f\|_\infty \cdot \rho r),$$

which follows by Lemma 2.14 and the triangle inequality. $\square$

We briefly remark that in several subsequent arguments, we will again exploit the symmetry of subspaces and Bohr sets V to use the identity $\mu_V(x) = \mu_V(-x)$ for all x .

2.3 Spread sets

A key definition underpinning Kelley and Meka's breakthrough result on 3APs [13] is a pseudorandomness notion they call *spreadness*. A subset of $\mathbb{F}_q^n$ is spread if it has no significant density increment by restricting to a large affine subspace.

Definition 2.16 ((r, λ) -spread). *Let $r \geq 1$ and $\lambda > 1$. A set $A \subseteq \mathbb{F}_q^n$ is (r, λ) -spread if for every subspace $V \subseteq \mathbb{F}_q^n$ of co-dimension at most r and shift $x \in \mathbb{F}_q^n$, we have*

$$\frac{|(A - x) \cap V|}{|V|} \leq \lambda \cdot \frac{|A|}{|\mathbb{F}_q^n|}.$$

In light of Lemma 2.4, we will need the following generalization of spreadness that applies to multiple sets. It should intuitively be clear that the definition we choose needs to be invariant to shifts, since the number of skew corners in a set has this property (Observation 2.3).

Definition 2.17 ((r, λ) -simultaneously spread). *Let $r \geq 1$ and $\lambda > 1$. A collection $\{A_i\}_{i \in S}$ of subsets of $\mathbb{F}_q^n$ is (r, λ) -simultaneously spread if for every subspace $V \subseteq \mathbb{F}_q^n$ of co-dimension at most r and shifts $\{x_i\}_{i \in S}$ in $\mathbb{F}_q^n$, we have*

$$\sum_{i \in S} \left(\frac{|(A_i - x_i) \cap V|}{|V|} \right)^2 \leq \lambda \cdot \sum_{i \in S} \left(\frac{|A_i|}{|\mathbb{F}_q^n|} \right)^2.$$

When our ambient space is a vector space $W \subseteq \mathbb{F}_q^n$, we will say the collection is (r, λ) -simultaneously spread in W if it is (r, λ) -simultaneously spread after being mapped according to an isomorphism from W to $\mathbb{F}_q^{\dim W}$.

Additionally, we require a version with Bohr sets replacing subspaces for general Abelian groups. For technical reasons, it will simplify later analysis to measure density with respect to a particular Bohr set rather than the ambient space (as in Definition 2.17).

Definition 2.18 ((r, δ, λ) -simultaneously spread). *Let $r \geq 1$, $\delta \in (0, 1)$, and $\lambda > 1$. A collection $\{A_i\}_{i \in S}$ of subsets of a regular Bohr set $B \subseteq G$ is (r, δ, λ) -simultaneously spread in B if for every regular Bohr set $B' \leq B$ of rank at most $\text{rk}(B) + r$ and measure $\mu(B') \geq \delta \mu(B)$, and shifts $\{x_i\}_{i \in S}$ in G , we have*

$$\sum_{i \in S} \left(\frac{|(A_i - x_i) \cap B'|}{|B'|} \right)^2 \leq \lambda \cdot \sum_{i \in S} \left(\frac{|A_i|}{|B|} \right)^2.$$

It was useful in [13] to use $\|\mu_A * \mu_V\|_\infty \leq \lambda$ as an equivalent form of Definition 2.16. We will later need to exploit a similar relationship with an alternative definition of simultaneous spreadness. (The proof of this connection explains why the summands in Definition 2.17 and Definition 2.18 are squared.)

Lemma 2.19. *Let $r \geq 1$, $\lambda > 1$, $\{A_i\}_{i \in S}$ be an (r, λ) -simultaneously spread collection of subsets of $\mathbb{F}_q^n$, and $\alpha = \sum_{i \in S} (|A_i|/|\mathbb{F}_q^n|)^2$. Then for every subspace $V \subseteq \mathbb{F}_q^n$ of co-dimension at most r , we have*

$$\left\| \left(\frac{1}{\alpha} \sum_{i \in S} 1_{A_i} \circ 1_{A_i} \right) * \mu_V \right\|_{\infty} \leq \sqrt{\lambda}.$$

Proof. Let $x \in \mathbb{F}_q^n$ be an element maximizing the input to the L^∞ -norm. Then by Cauchy-Schwarz,

$$\left\| \left(\sum_{i \in S} 1_{A_i} \circ 1_{A_i} \right) * \mu_V \right\|_{\infty} = \sum_{i \in S} \langle 1_{A_i} \circ 1_{A_i}, \mu_{V+x} \rangle \leq \sum_{i \in S} \|1_{A_i}\|_1 \|1_{A_i} * \mu_V\|_{\infty} \leq \alpha^{1/2} \left(\sum_{i \in S} \|1_{A_i} * \mu_V\|_{\infty}^2 \right)^{1/2}.$$

Now let x_i be a shift which maximizes $|(A_i - x_i) \cap V|$ for each $i \in S$. By our spreadness assumption, we have

$$\sum_{i \in S} \|1_{A_i} * \mu_V\|_{\infty}^2 = \sum_{i \in S} \left(\frac{|(A_i - x_i) \cap V|}{|V|} \right)^2 \leq \lambda \cdot \sum_{i \in S} \left(\frac{|A_i|}{|\mathbb{F}_q^n|} \right)^2 = \lambda \alpha.$$

Substituting the second equation into the first yields the result. $\square$

We also need an analogous lemma for Bohr sets. While similar, we present the proof here to illustrate why there is an additional $\mu(B)^{-1}$ factor.

Lemma 2.20. *Let $r \geq 1$, $\varepsilon \in (0, 1)$, $\lambda > 1$, $\{A_i\}_{i \in S}$ be an (r, δ, λ) -simultaneously spread collection of subsets of a regular Bohr set $B \subseteq G$, and $\alpha = \sum_{i \in S} (|A_i|/|G|)^2$. Then for every regular Bohr set $B' \leq B$ of rank at most $\text{rk}(B) + r$ and measure $\mu(B') \geq \delta \mu(B)$, we have*

$$\left\| \left(\frac{1}{\alpha} \sum_{i \in S} 1_{A_i} \circ 1_{A_i} \right) * \mu_{B'} \right\|_{\infty} \leq \sqrt{\lambda} \cdot \mu(B)^{-1}.$$

Proof. Let $x \in G$ be an element maximizing the input to the L^∞ -norm. Then by Cauchy-Schwarz,

$$\left\| \left(\sum_{i \in S} 1_{A_i} \circ 1_{A_i} \right) * \mu_{B'} \right\|_{\infty} = \sum_{i \in S} \langle 1_{A_i} \circ 1_{A_i}, \mu_{B'+x} \rangle \leq \sum_{i \in S} \|1_{A_i}\|_1 \|1_{A_i} * \mu_{B'}\|_{\infty} \leq \alpha^{1/2} \left(\sum_{i \in S} \|1_{A_i} * \mu_{B'}\|_{\infty}^2 \right)^{1/2}.$$

Now let x_i be a shift which maximizes $|(A_i - x_i) \cap B'|$. By our spreadness assumption, we have

$$\sum_{i \in S} \|1_{A_i} * \mu_{B'}\|_{\infty}^2 = \sum_{i \in S} \left(\frac{|(A_i - x_i) \cap B'|}{|B'|} \right)^2 \leq \lambda \cdot \sum_{i \in S} \left(\frac{|A_i|}{|B|} \right)^2 = \lambda \alpha \cdot \mu(B)^{-2}.$$

Substituting the second equation into the first yields the result. $\square$

3 The finite field case

In this section, we prove Theorem 1.2, restated below for convenience. Note that Pohoata and Zakharov [19] used the polynomial method to obtain the stronger bound $O(q^{(2-c)n})$ (for a constant $c > 0$ depending on q). However, their techniques do not seem to generalize to the integer case.

Theorem 1.2. *Let q be a prime power. Any skew-corner-free subset of $\mathbb{F}_q^n \times \mathbb{F}_q^n$ has size at most $q^{2n - \Omega(n^{1/9})}$.*

The analogous result for 3APs in [13] is proved by repeatedly restricting the set of interest to a large, denser subspace until no such subspaces exist (i.e., the set is spread). Then, they show spread sets have approximately the same number of 3APs as a random set, and since there cannot be too many trivial 3APs, the original set size cannot be too large.

This intuition guides our approach, although the situation is more complicated for skew corners. It is helpful to think about the converse of the above discussion. For a set A of large density, if A contains no 3APs, then A must exhibit a density increment on a reasonably large subspace. In the skew corners case, we start with a large subset A of $\mathbb{F}_q^n \times \mathbb{F}_q^n$, and denote by $\{A_g\}$ its columns. We are asking if there are many solutions of the form $x - y + g \in D$ for $x, y \in A_g$, where D can be thought of as the set of columns of decent density. By Kelley and Meka's result, if each of the sets A_g was spread, then we would have many solutions and thus many skew corners. So, we know that if there are no skew corners, then most of the A_g are *not* spread, and most of the A_g exhibit a density increment on some large subspace V_g . The issue is, however, that a priori these subspaces V_g might be completely uncorrelated.

One might hope that if the sum of $1_{A_g} \circ 1_{A_g+g}$ is far from uniform, then the individual deviations of the $1_{A_g} \circ 1_{A_g+g}$ are cooperating, and so there should be a density increment which is compatible with the A_g on average. Indeed, this ends up being what happens, and we show that there is a common subspace V which “explains” the lack of uniformity for the sum of convolutions. Namely, for most A_g , there is some affine shift $V + x_g$ which admits a density increment on A_g . For technical reasons, we will measure progress in terms of the sum of the squared densities $(|A_g \cap (V + x_g)|/|V|)^2$.

It is tempting to ignore the collection of A_g , and simply proceed with a single A_h of highest density. The Kelley-Meka techniques allow one to restrict A_h to an affine subspace V on which the convolution $1_{A_h} \circ 1_{A_h+h}$ is close to uniform. The issue, however, is that this approach does not ensure that D has large density on V . There may in fact be no nonempty columns indexed by elements from V , and so this naive approach does not work as is. For this reason, our density increment process needs access to many of the A_g at once, so that we can restrict the set of “rows” in which the A_g are contained, as well as the set of “columns” by which the A_g are indexed while maintaining density guarantees.

The main steps then are to argue that (1) a collection of sets $\{A_g\}_{g \in G}$ can be restricted in such a way that it no longer admits a collective density increment on a subspace and (2) if a collection of $\{A_g\}_{g \in G}$ contains enough points but fails to have skew corners, then there must be some subspace on which the A_g collectively obtain a density increment.

3.1 Obtaining simultaneous spreadness

A useful property of spreadness is that it is easy to obtain via an iterative density increment argument. The generalization to simultaneous spreadness preserves this property with only slightly more work.

Lemma 3.1. *Let $\varepsilon \in (0, 1)$, $r \geq 1$, and $\{A_g\}_{g \in \mathbb{F}_q^n}$ be a collection of subsets of $\mathbb{F}_q^n$ with $\sum |A_g|^2 \geq 2^{-d} |\mathbb{F}_q^n|^3$. Then there exists a subspace $V \subseteq \mathbb{F}_q^n$ of co-dimension at most $O(rd/\varepsilon)$ and shifts $\{x_g\}_{g \in \mathbb{F}_q^n}$, $w \in \mathbb{F}_q^n$ such that*

1. *The collection $\{(A_g - x_g) \cap V\}_{g \in V+w}$ is $(r, 1 + \varepsilon)$ -simultaneously spread in V , and*
2. *$\sum_{g \in V+w} |(A_g - x_g) \cap V|^2 \geq 2^{-d} |V|^3$.*

Proof. If $\{A_g\}_{g \in \mathbb{F}_q^n}$ is not $(r, 1 + \varepsilon)$ -simultaneously spread, then there exists a subspace $V \subseteq \mathbb{F}_q^n$ of co-dimension at most r and shifts $\{x_g\}_{g \in \mathbb{F}_q^n}$ in $\mathbb{F}_q^n$ such that

$$\sum_{g \in \mathbb{F}_q^n} \left(\frac{|(A_g - x_g) \cap V|}{|V|} \right)^2 > (1 + \varepsilon) \sum_{g \in \mathbb{F}_q^n} \left(\frac{|A_g|}{|\mathbb{F}_q^n|} \right)^2.$$

By averaging, there must exist an affine subspace $V + w$ with

$$\sum_{g \in V+w} \left(\frac{|(A_g - x_g) \cap V|}{|V|} \right)^2 > (1 + \varepsilon) \frac{|V|}{|\mathbb{F}_q^n|} \sum_{g \in \mathbb{F}_q^n} \left(\frac{|A_g|}{|\mathbb{F}_q^n|} \right)^2. \quad (1)$$

Construct a new collection $A'_g = (A_{g+w} - x_{g+w}) \cap V$ for $g \in V$. If $\{A'_g\}_{g \in V}$ is not $(r, 1 + \varepsilon)$ -simultaneously spread in V , we repeat this process with $\{A'_g\}_{g \in V}$ as the collection and $V \cong \mathbb{F}_q^{\dim(V)}$ as the ambient vector space, noting that Equation (1) and our initial density assumption imply

$$\sum_{g \in V} |A'_g|^2 > (1 + \varepsilon) \cdot 2^{-d} |V|^3.$$

After i iterations, the density will be at least $(1 + \varepsilon)^i \cdot 2^{-d}$. Since it cannot exceed 1, we must obtain the desired collection in $O(d/\varepsilon)$ iterations. $\square$

3.2 Using simultaneous spreadness

Now that we have a simultaneously spread collection, it remains to show the corresponding set must have roughly as many skew corners as a random set of the same density. This section will focus on proving the following theorem.

Theorem 3.2. *Let $\varepsilon \in (0, 1)$. For a collection $\{A_g\}_{g \in \mathbb{F}_q^n}$ of subsets of $\mathbb{F}_q^n$ with $\sum_{g \in \mathbb{F}_q^n} |A_g| \geq 2^{-d} |\mathbb{F}_q^n|^2$, let $F = \sum_{g \in \mathbb{F}_q^n} 1_{A_g} \circ 1_{A_{g+g}}$, $F' = \sum_{g \in \mathbb{F}_q^n} 1_{A_g} \circ 1_{A_g}$, and $D: \mathbb{F}_q^n \rightarrow \mathbb{R}_{\geq 0}$ be given by $D(x) = |A_x|$. Then either*

1. $|\langle \mu_F, \mu_D \rangle - 1| \leq \varepsilon$, or
2. *There exists a subspace V of co-dimension $O_\varepsilon(d^8)$ such that $\|\mu_{F'} * \mu_V\|_\infty \geq 1 + \frac{1}{32} \varepsilon$.*

The proof of Theorem 1.2 follows from Theorem 3.2. Namely, if our collection of sets $\{A_g\}_{g \in G}$ admits no collective density increment, and it is sufficiently dense, then it must contain roughly as many skew corners as a random set.

Proof of Theorem 1.2 from Theorem 3.2. Let $A \subseteq \mathbb{F}_q^n \times \mathbb{F}_q^n$ be skew-corner-free with $|A| = 2^{-d} |\mathbb{F}_q^n|^2$, and define the collection $\{A_g\}_{g \in \mathbb{F}_q^n}$ by $A_g = \{y : (g, y) \in A\}$. By Cauchy-Schwarz, we have $\sum |A_g|^2 \geq 2^{-2d} |\mathbb{F}_q^n|^3$, so applying Lemma 3.1 with $\{A_g\}_{g \in \mathbb{F}_q^n}$, $\varepsilon := 1/128$, $d := 2d$, and $r > 0$ to be determined later yields a collection of sets $\{(A_{g+w} - x_{g+w}) \cap V\}_{g \in V}$, where $V \subseteq \mathbb{F}_q^n$ is a co-dimension $O(dr)$ subspace and $\{x_g\}_{g \in \mathbb{F}_q^n}, w \in \mathbb{F}_q^n$ are shifts. For brevity, define $A'_g := (A_{g+w} - x_{g+w}) \cap V$ and $A' := \{(g, y) : g \in V, y \in A'_g\} \subseteq V \times V$. Note that by Observation 2.3, A has at least as many skew corners as A' , $|A'| \geq 2^{-2d} |V|^2$, and $\{A'_g\}_{g \in V}$ is $(r, 1 + 1/128)$ -simultaneously spread in V .

Apply Theorem 3.2 with $\varepsilon := 1/2$ to the new collection, and set $r = O(d^8)$ to be larger than the co-dimension bound in the second conclusion. If $|\langle \mu_F, \mu_D \rangle - 1| \leq 1/2$, Lemma 2.5 implies A' contains at least

$$\frac{1}{2} \cdot \frac{|A'|^3}{|V|^2} \geq 2^{-6d-1} |V|^4 \geq 2^{-\Omega(d^9)} |\mathbb{F}_q^n|^4$$

skew corners. Since there are at most $|\mathbb{F}_q^n|^3$ trivial skew corners in A' , we are done after rearrangement. Otherwise, there exists a subspace V of co-dimension $\ell = O(d^8)$, with $\ell < r$, such that $\|\mu_{F'} * \mu_V\|_\infty \geq 1 + 1/64$. By Lemma 2.19, this contradicts the fact that $\{A'_g\}_{g \in V}$ is $(r, 1 + 1/128)$ -simultaneously spread. $\square$

The proof of Theorem 3.2 follows from five major steps, which we briefly sketch. Throughout, let $F = \sum_g 1_{A_g} \circ 1_{A_{g+g}}$, $F' = \sum_g 1_{A_g} \circ 1_{A_g}$, and $D(x) = |A_x|$.

1. Non-uniformity (Section 3.2.1) – If there are few skew corners (i.e., $|\langle \mu_F, \mu_D \rangle - 1| > \varepsilon$), then Hölder's inequality implies that $\|\mu_F - 1\|_p \geq \varepsilon/2$ for $p = O(d)$.
2. Unbalancing (Section 3.2.2) – We then show it suffices to consider the spectrally non-negative F' , which we use to prove $\|\mu_{F'}\|_{p'} \geq 1 + \varepsilon/4$ for some $p' = O_\varepsilon(p)$.
3. Dependent random choice (Section 3.2.3) – From here, we rely on a known application of dependent random choice to find two relatively dense sets M_1, M_2 whose difference convolution correlates well with elements of noticeable upward deviation from F' 's mean. That is, $\langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \varepsilon/32$, where $S = \{x : \mu_{F'}(x) > 1 + \varepsilon/8\}$.
4. Almost periodicity (Section 3.2.4) – Next, we use an almost periodicity result of [22] to show there exists a subspace V of co-dimension $O_\varepsilon(d^8)$ such that $\langle \mu_V * \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \varepsilon/16$.
5. Combining steps (Section 3.2.5) – Putting these results together yields the second conclusion of Theorem 3.2.

These are essentially the same steps given in [8], which is an excellent exposition and simplification of [13]. Moreover, we are able to either black-box or closely follow the proofs of many of their results, highlighting the relationship between arithmetic progressions and skew corners.

3.2.1 Non-uniformity

The following lemma will allow us to argue that if there are very few skew corners, then $\|\mu_F - 1\|_p \geq \varepsilon/2$.

Lemma 3.3. *Let $\varepsilon > 0$ and $f : G \rightarrow \mathbb{R}$. If $\{A_g\}_{g \in G}$ is a collection of subsets of G where $\sum_{g \in G} |A_g| \geq 2^{-d} |G|^2$, and $D : G \rightarrow \mathbb{R}_{\geq 0}$ is given by $D(x) = |A_x|$, then either*

1. $|\langle \mu_f, \mu_D \rangle - 1| \leq \varepsilon$ or
2. $\|\mu_f - 1\|_p \geq \varepsilon/2$ for an integer $p = O(d)$.

Proof. If (1) fails, then by Hölder's inequality, we have for any $p \geq 1$ that

$$\varepsilon < |\langle \mu_f - 1, \mu_D \rangle| \leq \|\mu_f - 1\|_p \|\mu_D\|_{1+\frac{1}{p-1}} \leq \|\mu_f - 1\|_p \|\mu_D\|_\infty^{1/p} \|\mu_D\|_1^{1-1/p} = \|\mu_f - 1\|_p \|\mu_D\|_\infty^{1/p}.$$

Note that $\|\mu_D\|_\infty \leq 2^d$, since $D(x) \leq |G|$ and $\mathbb{E}[D] \geq 2^{-d} |G|$. Setting $p = \lceil d \rceil$ yields (2). $\square$

3.2.2 Unbalancing

The next lemma will allow us to consider the function $F' = \sum_g 1_{A_g} \circ 1_{A_g}$ rather than $F = \sum_g 1_{A_g} \circ 1_{A_g+g}$.

Lemma 3.4. *Let $p \geq 1$ be an integer and $\{f_g\}_{g \in S}$ a collection of real-valued functions on G indexed by some set $S \subseteq G$, we have*

$$\left\| \sum_{g \in S} f_g \circ f_g \right\|_p \geq \left\| \sum_{g \in S} f_g \circ f_g^g \right\|_p$$

where f^g denotes the function $f^g(x) = f(x - g)$.

Proof. We have

$$\left\| \sum_{g \in S} f_g \circ f_g^g \right\|_p^p = \sum_{\alpha_1 + \dots + \alpha_p = 0} \prod_{i=1}^p \left(\sum_{g \in S} |\widehat{f_g}(\alpha_i)|^2 \cdot \chi_{\alpha_i}(g) \right) \leq \sum_{\alpha_1 + \dots + \alpha_p = 0} \prod_{i=1}^p \left(\sum_{g \in S} |\widehat{f_g}(\alpha_i)|^2 \right) = \left\| \sum_{g \in S} f_g \circ f_g \right\|_p^p.$$

$\square$

The merit of working with F' is that it is spectrally non-negative, so all of its odd moments are non-negative. Kelley and Meka [13] showed that if a function f admits non-negative odd moments as well as deviations from uniform in the p -norm, then $f + 1$ must also admit strong deviations *upwards* from uniform. The specific version of this lemma we will use comes from [8], which adapts [13, Proposition 5.7] to work when the norm is taken with respect to a spectrally non-negative distribution ν .

Lemma 3.5 ([8, Lemma 7]). *Let $\varepsilon \in (0, 1)$ and $\nu: G \rightarrow \mathbb{R}_{\geq 0}$ be some probability measure such that $\widehat{\nu} \geq 0$. Let $f: G \rightarrow \mathbb{R}$ be such that $\widehat{f} \geq 0$. If $\|f\|_{p(\nu)} \geq \varepsilon$ for some $p \geq 1$, then $\|f + 1\|_{p'(\nu)} \geq 1 + \varepsilon/2$ for some integer $p' = O_\varepsilon(p)$.*

Note that we will only apply Lemma 3.5 when ε is a constant, so the exact dependence on ε is inconsequential. The utility of this lemma comes when applied to the mean zero part of a function f , namely $f_0 := f - \mathbb{E}[f]$. If $\mathbb{E}[f_0^p] \geq 0$ for odd p , then downward deviations from uniform in f_0 imply *upwards* deviations from uniform in f . In particular, we will use the above lemma to exhibit upwards deviations in $F' = \sum_g 1_{A_g} \circ 1_{A_g}$. Later on, we will see how to use this to obtain structural information on the $\{A_g\}_{g \in G}$. Combining the previous two lemmas, we obtain the following corollary.

Corollary 3.6. *Let $\varepsilon \in (0, 1)$, $F = \sum_{g \in G} 1_{A_g} \circ 1_{A_{g+g}}$, $F' = \sum_{g \in G} 1_{A_g} \circ 1_{A_g}$, where $\{A_g\}_{g \in G}$ is a collection of subsets of G . If $\|\mu_F - 1\|_p \geq \varepsilon$ for some integer $p \geq 1$, then $\|\mu_{F'}\|_{p'} \geq 1 + \varepsilon/2$ for some integer $p' = O_\varepsilon(p)$.*

Proof. Let $\alpha_g = |A_g|/|G|$ and $\alpha = \sum_{g \in G} \alpha_g^2$, so that $\frac{1}{\alpha} F = \mu_F$ and $\frac{1}{\alpha} F' = \mu_{F'}$. Setting $f_g = 1_{A_g} - \alpha_g$, we have

$$\sum_{g \in G} f_g \circ f_g = \sum_{g \in G} (1_{A_g} - \alpha_g) \circ (1_{A_g} - \alpha_g) = \sum_{g \in G} 1_{A_g} \circ 1_{A_g} - \alpha_g^2 = F' - \alpha.$$

Similarly, we have

$$\sum_{g \in G} f_g \circ f_g^g = \sum_{g \in G} (1_{A_g} - \alpha_g) \circ (1_{A_g^g} - \alpha_g) = \sum_{g \in G} 1_{A_g} \circ 1_{A_{g+g}} - \alpha_g^2 = F - \alpha.$$

If we apply Lemma 3.4 with $S := G$, this gives $\|\mu_{F'} - 1\|_p \geq \varepsilon$. Since $\mu_{F'} - 1$ is spectrally non-negative, we can apply Lemma 3.5 to obtain $\|\mu_{F'}\|_{p'} \geq 1 + \varepsilon/2$ for some integer $p' = O_\varepsilon(p)$. $\square$

3.2.3 Dependent random choice

At this point, we can assume we are working with a sum of difference convolutions $F' = \sum_{g \in G} 1_{A_g} \circ 1_{A_g}$ which admits upwards deviations from uniform in an appropriate p -norm. We will want to extract information about the collection $\{A_g\}_{g \in G}$ itself from F' . It will be useful for us to find a “robust witness” to upward deviations from the mean. One can view a robust witness as being a convolution which correlates well with the set of points on which F' has a large deviation upwards. We will need the following lemma by Bloom and Sisask [8] using dependent random choice (originally dubbed *sifting* by Kelley and Meka [13]).

Lemma 3.7 ([8, Lemma 8]). *Let $\varepsilon, \delta \in (0, 1)$ and $p \geq 1$ be an integer. Let $B_1, B_2 \subseteq G$, and let $\nu = \mu_{B_1} \circ \mu_{B_2}$. For any finite set $A \subseteq G$, if*

$$S = \{x \in G : (\mu_A \circ \mu_A)(x) > (1 - \varepsilon)\|\mu_A \circ \mu_A\|_{p(\nu)}\},$$

then there are $M_1 \subseteq B_1$ and $M_2 \subseteq B_2$ such that

$$\langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \delta \quad \text{and} \quad \min(\mu_{B_1}(M_1), \mu_{B_2}(M_2)) \geq (\mathbb{E}[1_A] \cdot \|\mu_A \circ \mu_A\|_{p(\nu)})^{2p + O_{\varepsilon, \delta}(1)}.$$

As with the previous lemma (Lemma 3.5), we will only apply Lemma 3.7 for constant ε and δ , so we will not be concerned with the precise dependency on these variables. Observe that since the support of $\mu_{M_1} \circ \mu_{M_2}$ is contained in the support of ν , the conclusion $\langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \delta$ is unchanged by replacing S with $S \cap L$ for any L containing the support of ν . This observation will be used in the proof of Lemma 3.8.

In the 3AP setting, this lemma is enough to obtain a robust witness to $A * A$ being far from uniform. Since we are working with a collection of sets $\{A_g\}_{g \in G}$, we will encode our collection into a large set $A \subseteq G \times G$ while preserving the relevant properties and apply sifting there. Despite the lemma being stated for a single set, we are able to appropriately encode our collection to apply it. First, we prove sifting for an arbitrary Abelian group G , then we apply it specifically to the case of $\mathbb{F}_q^n$. Later, we will apply this result to $G = \mathbb{Z}/N\mathbb{Z}$ as well.

Lemma 3.8. *Let $\varepsilon, \delta \in (0, 1)$ and $p \geq 1$ be an integer. Additionally, let $B_1, B_2 \subseteq G$, $\nu = \mu_{B_1} \circ \mu_{B_2}$, and $f = \sum_{g \in C} 1_{A_g} \circ 1_{A_g}$, where $\{A_g\}_{g \in C}$ is a collection of subsets of G indexed by C with $\sum_{g \in C} |A_g| \geq 2^{-d} |G| |C|$. If*

$$S = \{x \in G : \mu_f(x) > (1 - \varepsilon) \|\mu_f\|_{p(\nu)}\},$$

then there are $M_1 \subseteq B_1$ and $M_2 \subseteq B_2$ such that

$$\langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \delta \quad \text{and} \quad \min(\mu_{B_1}(M_1), \mu_{B_2}(M_2)) \geq \left(2^{-d} \|\mu_f\|_{p(\nu)}\right)^{O_{\varepsilon, \delta}(p)}.$$

Proof. Let $H = G \times G$, $L = \{0\} \times G$, $B'_1 = \{0\} \times B_1$ and $B'_2 = \{0\} \times B_2$. Additionally, let $\pi = \mu_{B'_1} \circ \mu_{B'_2}$, and note that π is supported on L . We embed our collection into a single set $A = \{(g, y) \mid g \in C, y \in A_g\}$ in H , and define

$$S = \{x \in L : (\mu_A \circ \mu_A)(x) > (1 - \varepsilon) \|\mu_A \circ \mu_A\|_{p(\pi)}\}.$$

Recall that it suffices to define S for $x \in L$ (as opposed to H), since L contains the support of π . Applying Lemma 3.7, we obtain $M_1 \subseteq B'_1$ and $M_2 \subseteq B'_2$, such that

$$\langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle_H \geq 1 - \delta \quad \text{and} \quad \min(\mu_{B'_1}(M_1), \mu_{B'_2}(M_2)) \geq (\mathbb{E}[1_A] \cdot \|\mu_A \circ \mu_A\|_{p(\pi)})^{O_{\varepsilon, \delta}(p)}.$$

Observe that for $x = (0, y) \in L$, we have

$$(1_A \circ 1_A)(x) = \mathbb{E}_{z \in H} 1_A(z) 1_A(x+z) = \frac{1}{|G|^2} \sum_{g \in C} \sum_{z \in G} 1_{A_g}(z) 1_{A_g}(y+z) = \frac{1}{|G|} \sum_{g \in C} (1_{A_g} \circ 1_{A_g})(y) = \frac{1}{|G|} f(y).$$

In particular, $(\mu_A \circ \mu_A)(x) = \zeta \cdot \mu_f(y)$, where

$$\zeta = \frac{\mathbb{E}_G[f]}{\mathbb{E}_H[1_A]^2 \cdot |G|} = \frac{|G| \cdot \sum_{g \in C} |A_g|^2}{|\sum_{g \in C} A_g|^2} \geq \frac{|G|}{|C|}, \quad (2)$$

with the final inequality by Cauchy-Schwarz. Similarly, we have

$$\pi(x) = (\mu_{B'_1} \circ \mu_{B'_2})(x) = \frac{|H|^2}{|B'_1| |B'_2|} \cdot (1_{B'_1} \circ 1_{B'_2})(x) = \frac{|G|^4}{|B_1| |B_2|} \cdot \frac{1}{|G|} (1_{B_1} \circ 1_{B_2})(y) = |G| (\mu_{B_1} \circ \mu_{B_2})(y) = |G| \nu(y).$$

Recalling that L contains the support of π , we can rewrite S as $\{(0, y) \in L : \mu_f(y) > (1 - \varepsilon) \|\mu_f\|_{p(\nu)}\}$.

We would like to verify the inner product over G satisfies the conclusion of the theorem. Let $\alpha_1 = |M_1|/|H|$ and $\alpha_2 = |M_2|/|H|$. Unraveling the definition of $\langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle$, we have

$$\mathbb{E}_{x \in H} (\mu_{M_1} \circ \mu_{M_2})(x) 1_S(x) = \frac{1}{|H|} \sum_{x \in H} \left(\frac{1}{|H|} \sum_{y \in H} \mu_{M_1}(y) \mu_{M_2}(x+y) \right) 1_S(x)$$

$$\begin{aligned}
&= \frac{1}{|G|^2} \sum_{x \in H} \left(\frac{1}{|G|^2} \sum_{y \in H} \frac{1_{M_1}(y)}{\alpha_1} \cdot \frac{1_{M_2}(x+y)}{\alpha_2} \right) 1_S(x) \\
&= \frac{1}{|G|^2} \sum_{x \in L} \left(\sum_{y \in L} \frac{1_{M_1}(y)}{\alpha_1 |G|} \cdot \frac{1_{M_2}(x+y)}{\alpha_2 |G|} \right) 1_S(x) \quad (\text{since } M_1, M_2 \subseteq L) \\
&= \frac{1}{|G|} \sum_{x \in L} \left(\frac{1}{|G|} \sum_{y \in L} \frac{1_{M_1}(y)}{\mathbb{E}_L[M_1]} \cdot \frac{1_{M_2}(x+y)}{\mathbb{E}_L[M_2]} \right) 1_S(x).
\end{aligned}$$

Since $L \cong G$, we can interpret M_1, M_2 , and S as subsets of G to obtain $\langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle_G \geq 1 - \delta$. Similarly, we can interpret $M_1 \subseteq B_1$ and $M_2 \subseteq B_2$. We conclude by using Equation (2) to deduce

$$\min(\mu_{B_1}(M_1), \mu_{B_2}(M_2)) \geq (\mathbb{E}_H[1_A] \cdot \zeta \cdot \|\mu_f\|_{p(v)})^{O_{\varepsilon, \delta}(p)} \geq \left(2^{-d} \|\mu_f\|_{p(v)}\right)^{O_{\varepsilon, \delta}(p)}. \quad \square$$

For the case where $G = \mathbb{F}_q^n$, we will not need the same level of generality.

Corollary 3.9. *Let $\varepsilon \in (0, 1)$ and $p \geq 1$ be an integer. Additionally, let $f = \sum_{g \in \mathbb{F}_q^n} 1_{A_g} \circ 1_{A_g}$ where $\{A_g\}_{g \in \mathbb{F}_q^n}$ is a collection of subsets of $\mathbb{F}_q^n$ with $\sum_{g \in \mathbb{F}_q^n} |A_g| \geq 2^{-d} |\mathbb{F}_q^n|^2$. If $\|\mu_f\|_p \geq 1 + \varepsilon$ and $S = \{x : \mu_f(x) > 1 + \varepsilon/2\}$, then there are $M_1, M_2 \subseteq \mathbb{F}_q^n$ both of density at least $2^{-O_{\varepsilon}(dp)}$ such that*

$$\langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \frac{\varepsilon}{8}.$$

Proof. Apply Lemma 3.8 with $C, B_1, B_2 := \mathbb{F}_q^n$, $\varepsilon := \varepsilon/4$, and $\delta := \varepsilon/8$. $\square$

3.2.4 Almost periodicity

At this point, we have obtained a robust witness to the non-uniformity of F' . Namely, we have two dense sets M_1, M_2 whose convolution $M_1 * M_2$ correlates well with the points where F' has upwards deviations. We will want to convert our robust witness into a subspace V on which our collection $\{A_g\}_{g \in \mathbb{F}_q^n}$ admits a collective density increment. Kelley and Meka [13] originally used a result of Sanders [20]; we will follow [8] in using the following theorem to find a large subspace V which correlates well with the upwards deviations of F' via the robust witness we found in the previous step.

Theorem 3.10 ([22, Theorem 3.2]). *Let $\varepsilon \in (0, 1)$. If $M_1, M_2, S \subseteq \mathbb{F}_q^n$ are such that M_1 and M_2 both have density at least 2^{-d} , then there is a subspace V of co-dimension $O_{\varepsilon}(d^4)$ such that*

$$|\langle \mu_V * \mu_{M_1} * \mu_{M_2}, 1_S \rangle - \langle \mu_{M_1} * \mu_{M_2}, 1_S \rangle| \leq \varepsilon.$$

Corollary 3.11. *Let $\varepsilon \in (0, 1)$, and let $M_1, M_2, S \subseteq \mathbb{F}_q^n$ be such that M_1 and M_2 both have density at least 2^{-d} . If*

$$\langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \frac{\varepsilon}{2},$$

then there is a subspace V of co-dimension $O_{\varepsilon}(d^4)$ such that

$$\langle \mu_V * \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \varepsilon.$$

Proof. Apply Theorem 3.10 with $\varepsilon := \varepsilon/2, M_1 := M_2, M_2 := -M_1$, and $S := S$. $\square$

3.2.5 Combining steps

All that remains to prove Theorem 3.2 (restated below) is to combine the prior steps.

Theorem 3.2. *Let $\varepsilon \in (0, 1)$. For a collection $\{A_g\}_{g \in \mathbb{F}_q^n}$ of subsets of $\mathbb{F}_q^n$ with $\sum_{g \in \mathbb{F}_q^n} |A_g| \geq 2^{-d} |\mathbb{F}_q^n|^2$, let $F = \sum_{g \in \mathbb{F}_q^n} 1_{A_g} \circ 1_{A_g+g}$, $F' = \sum_{g \in \mathbb{F}_q^n} 1_{A_g} \circ 1_{A_g}$, and $D: \mathbb{F}_q^n \rightarrow \mathbb{R}_{\geq 0}$ be given by $D(x) = |A_x|$. Then either*

1. $|\langle \mu_F, \mu_D \rangle - 1| \leq \varepsilon$, or
2. *There exists a subspace V of co-dimension $O_\varepsilon(d^8)$ such that $\|\mu_{F'} * \mu_V\|_\infty \geq 1 + \frac{1}{32} \varepsilon$.*

Proof. If (1) does not hold, we apply Lemma 3.3 to arrive at $\|\mu_F - 1\|_p \geq \varepsilon/2$ for integer $p = O(d)$. Corollary 3.6 implies $\|\mu_{F'}\|_{p'} \geq 1 + \varepsilon/4$ for some integer $p' = O_\varepsilon(p)$. Let $S = \{x \in \mathbb{F}_q^n : \mu_{F'}(x) > 1 + \varepsilon/8\}$. We apply Corollary 3.9 to obtain $M_1, M_2 \subseteq \mathbb{F}_q^n$ both of density at least $2^{-O_\varepsilon(d^2)}$ such that

$$\langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \frac{\varepsilon}{32}.$$

We then apply Corollary 3.11 with $\varepsilon := \varepsilon/16$ and $d := O_\varepsilon(d^2)$ to find a subspace V with co-dimension at most $O_\varepsilon(d^8)$ such that

$$\langle \mu_V * \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \frac{\varepsilon}{16}.$$

By the definition of S , we have

$$1 + \frac{\varepsilon}{32} \leq \left(1 + \frac{\varepsilon}{8}\right) \left(1 - \frac{\varepsilon}{16}\right) \leq \langle \mu_V * \mu_{M_1} \circ \mu_{M_2}, \mu_{F'} \rangle \leq \|\mu_{F'} * \mu_V\|_\infty \|\mu_{M_1} \circ \mu_{M_2}\|_1 = \|\mu_{F'} * \mu_V\|_\infty. \quad \square$$

4 The integer case

In this section, we prove the following theorem.

Theorem 4.1. *Let G be a finite Abelian group. Any skew-corner-free subset of $G \times G$ has size at most*

$$\exp\left\{-\Omega\left(\log^{1/12}|G|\right)\right\} \cdot |G|^2.$$

Note that Theorem 4.1 immediately implies Theorem 1.1 (restated below), since embedding $A \subseteq [n] \times [n]$ in $\mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ for $N = 2n$ does not change the number of skew corners or (asymptotic) density,

Theorem 1.1. *Any skew-corner-free subset of $[n] \times [n]$ has size at most $\exp\left\{-\Omega\left(\log^{1/12}n\right)\right\} \cdot n^2$.*

All the main ideas of Theorem 4.1's proof are the same as those in Section 3, although some additional technicalities arise from replacing subspaces with regular Bohr sets. (It may be helpful to revisit Section 2.2.) For the remainder of the discussion, we will assume we are working with $G = \mathbb{Z}/N\mathbb{Z}$, but everything we are doing works over any finite Abelian group.

We will aim to replicate what we did in the finite field case when working with $\mathbb{Z}/N\mathbb{Z}$. The main issue is that there are potentially no nontrivial subgroups to work with over $\mathbb{Z}/N\mathbb{Z}$. This ends up not being so much of an issue, since we can replace subgroups with regular Bohr sets, which approximate subgroups in many important ways. And, not unlike the case over $\mathbb{F}_q^n$, we will be able to argue that if a collection $\{A_g\}_{g \in G}$ contains very few skew corners, then there must be some collective density increment on a Bohr set B .

In the finite field case, when we found a subspace V with a collective density increment, we could restrict our attention to V since $V \cong \mathbb{F}_q^{\dim(V)}$. Here, we will not have this luxury, since Bohr sets are not necessarily subgroups! We will have to maintain our ambient group as $\mathbb{Z}/N\mathbb{Z}$, all while restricting our sets $\{A_g\}_{g \in G}$ to

smaller and smaller Bohr sets. Thus, we will work with functions defined over $\mathbb{Z}/N\mathbb{Z}$. The supports of these functions, however, will only be contained in some Bohr set R . So, when taking norms and inner products, we will want to zoom in on the action by taking everything with respect to a certain measure ν . For the most part, ν will be some convolution of dilates of R , which we will be able to control since Bohr sets behave well under convolutions. Dealing with these technicalities surrounding Bohr sets and non-uniform measures forms the major differences between working over $\mathbb{F}_q^n$ and $\mathbb{Z}/N\mathbb{Z}$, but the overall proof structures are conceptually the same.

4.1 Obtaining simultaneous spreadness

As in the finite field case, we will want to perform a density increment argument on the rows and columns of $A \subseteq G \times G$ until we obtain a collection $\{A_g\}_{g \in C}$ of subsets of R for regular Bohr sets $R, C \subseteq G$ which is simultaneously spread. For technical reasons, we will want C to be a translate of a dilate of R . This will end up being necessary when we prove Lemma 4.4 and Lemma 4.5.

Another difference between the Abelian group case and the finite field case comes from our definition of simultaneous spreadness. In particular, if our sets $\{A_g\}_{g \in C}$ are each contained in R , then the Bohr set B on which we find a density increment must be a sub-Bohr set of R , which to recall we notate by $B \leq R$. The reader will notice in the proof of Lemma 4.2 that this becomes necessary when trying to obtain $B_\rho \subseteq R_\rho$ from $B \leq R$ for some $\rho \in (0, 1)$. If we try to start with an assumption of the form $B \subseteq R$, it is not true in general that $B_\rho \subseteq R_\rho$.

Overall, the proof idea is conceptually the same as Lemma 3.1 in the finite field case. The main differences come from ensuring that the index C is a translate of an appropriate dilate of R .

Lemma 4.2. *Let $\varepsilon, \delta \in (0, 1), r \geq 1, d \geq 0$, and $\{A_g\}_{g \in G}$ be a collection of subsets of G with $\sum_{g \in G} |A_g|^2 \geq 2^{-d} |G|^3$. Then there exists a regular Bohr set $R \subseteq G$ of rank $\text{rk}(R) \leq O(dr/\varepsilon)$ and density $\mu(R) \geq \delta^{O(d/\varepsilon)}$, a shift of a regular Bohr set $C = R_\rho + w$ where $\rho = (\varepsilon^2 / rd 2^d)^{O(d/\varepsilon)}$ and $w \in G$, and shifts $\{x_g\}_{g \in C}$ with $x_g \in G$ such that*

1. *The collection $\{(A_g - x_g) \cap R\}_{g \in C}$ is $(r, \delta, 1 + \varepsilon)$ -simultaneously spread in R , and*
2. *$\sum_{g \in C} |(A_g - x_g) \cap R|^2 \geq 2^{-d} |R|^2 |C|$.*

Proof. We proceed by a similar density increment argument as in the finite field case (Lemma 3.1), where regular Bohr sets replace the role of vector spaces.

Let B be a regular Bohr set and $\{A_g\}_{g \in B_\sigma + w}$ be a collection of subsets of B indexed by $B_\sigma + w$, where initially $B = G$ and $\sigma = 1$. We will assume $w = 0$ for simplicity; the proof is essentially the same for nonzero shifts. If $\{A_g\}_{g \in B_\sigma}$ is not $(r, \delta, 1 + \varepsilon)$ -simultaneously spread in B , then by definition, there exists a regular Bohr set $B' \leq B$ with rank $\text{rk}(B') \leq \text{rk}(B) + r$ and density $\mu(B') \geq \delta \mu(B)$, and shifts $\{x_g\}_{g \in B_\sigma}$ in G , such that for $A'_g := (A_g - x_g) \cap B'$ we have

$$\sum_{g \in B_\sigma} \left(\frac{|A'_g|}{|B'|} \right)^2 \geq (1 + \varepsilon) \sum_{g \in B_\sigma} \left(\frac{|A_g|}{|B|} \right)^2. \quad (3)$$

From here, the proof proceeds by filtering the index set of our collection to be some translate of a dilate of B' . In other words, we will restrict our index set to some translate of $B'_{\sigma'}$, where σ' will be chosen later. Consider the function $f(g) := (|A'_g|/|B'|)^2$. Then,

$$\frac{1}{|B_\sigma|} \sum_{g \in B_\sigma} \left(\frac{|A'_g|}{|B'|} \right)^2 = \langle \mu_{B_\sigma}, f \rangle \quad \text{and} \quad \frac{1}{|B_\sigma| |B'_{\sigma'}|} \sum_{g \in B_\sigma} \sum_{h \in B'_{\sigma'}} \left(\frac{|A'_{g+h}|}{|B'|} \right)^2 = \langle \mu_{B_\sigma} * \mu_{B'_{\sigma'}}, f \rangle.$$

Note that $\langle \mu_{B_\sigma}, f \rangle \geq 2^{-d}$, since the density of our collection has only increased from the initial assumption that $\sum_{g \in G} |A_g|^2 \geq 2^{-d} |G|^3$.

We want to apply Corollary 2.15 to B_σ and $B'_{\sigma'}$. Set $\sigma' = c\sigma\tau$ where we will choose τ later, and $c \in [1/2, 1]$ according to Lemma 2.11 so that $B'_{\sigma'}$ is regular. Since $B' \leq B$, Observation 2.9 implies the inclusion $B'_{\sigma'} \subseteq B_{\sigma\tau} = (B_\sigma)_\tau$. Applying Corollary 2.15 gives

$$\begin{aligned} \langle \mu_{B_\sigma} * \mu_{B'_{\sigma'}}, f \rangle &\geq \langle \mu_{B_\sigma}, f \rangle - |\langle \mu_{B_\sigma} * \mu_{B'_{\sigma'}} - \mu_{B_\sigma}, f \rangle| \\ &\geq \langle \mu_{B_\sigma}, f \rangle - O(\tau \cdot \text{rk}(B_\sigma)) \quad (\text{since } \|f\|_\infty \leq 1) \\ &\geq (1 - \varepsilon/4) \langle \mu_{B_\sigma}, f \rangle \end{aligned}$$

for $\tau \leq c'\varepsilon/2^d \text{rk}(B_\sigma)$ with constant $c' > 0$ chosen sufficiently small. By our setting of τ , we have

$$\frac{1}{|B_\sigma| |B'_{\sigma'}|} \sum_{g \in B_\sigma} \sum_{h \in B'_{\sigma'}} \left(\frac{|A'_{g+h}|}{|B'|} \right)^2 \geq (1 - \varepsilon/4) \frac{1}{|B_\sigma|} \sum_{g \in B_\sigma} \left(\frac{|A'_g|}{|B'|} \right)^2. \quad (4)$$

By combining Equation (3) with Equation (4) and averaging, there is some choice of $g \in B_\sigma$ so that

$$\frac{1}{|B'_{\sigma'}|} \sum_{h \in B'_{\sigma'}} \left(\frac{|A'_{g+h}|}{|B'|} \right)^2 \geq (1 - \varepsilon/4)(1 + \varepsilon) \frac{1}{|B_\sigma|} \sum_{g \in B_\sigma} \left(\frac{|A_g|}{|B|} \right)^2 \geq (1 + \varepsilon/2) \frac{1}{|B_\sigma|} \sum_{g \in B_\sigma} \left(\frac{|A_g|}{|B|} \right)^2.$$

If our new collection $\{A'_h\}_{h \in B'_{\sigma'}+g}$ is not $(r, \delta, 1 + \varepsilon)$ -simultaneously spread in B' , we can iterate this procedure. After i iterations, we obtain Bohr sets $B^{(i)}, C^{(i)}$ and a collection $\{A_g^{(i)}\}_{g \in C^{(i)}}$ such that:

1. $B^{(i)}$ is a regular sub-Bohr set of $B^{(i-1)}$ of rank at most $i \cdot r$ and density at least δ^i .
2. Each $C^{(i)}$ is a translate of $B_{\sigma_i}^{(i)}$ for $\sigma_i \geq (\tau/2)^i$.
3. Each $A_g^{(i)}$ equals $(A_g^{(i-1)} - x_g^{(i)}) \cap B^{(i)}$ for some shift $x_g^{(i)}$.
4. We have

$$\frac{1}{|B_{\sigma_i}^{(i)}|} \sum_{g \in C^{(i)}} \left(\frac{|A_g^{(i)}|}{|B^{(i)}|} \right)^2 \geq (1 + \varepsilon/2)^i \frac{1}{|G|} \sum_{g \in G} \left(\frac{|A_g|}{|G|} \right)^2.$$

The process must end within $O(d/\varepsilon)$ iterations, since the density on the right-hand side of (4) cannot exceed 1. This implies that $\text{rk}(B^{(i)}) \leq O(rd/\varepsilon)$ for all i , so we can obtain the claimed bound on the final σ_i by ultimately choosing $\tau = c''\varepsilon^2/(rd2^d)$ for some constant $c'' > 0$. $\square$

4.2 Using simultaneous spreadness

Assuming we have a simultaneously spread collection, it remains to show the corresponding set must have roughly as many skew corners as a random set of the same density. This can be viewed as a “structure vs. pseudo-randomness” result, where we say that either a set $A \times A \subseteq G$ has roughly as many skew-corners as a random set of the same density, or A admits a type of density increment on a structured set.

Theorem 4.3. *There exists an absolute constant $c > 0$ such that the following holds. Let $\varepsilon \in (0, 1)$ and $R \subseteq G$ be a regular Bohr set of rank r . Furthermore, let $\{A_g\}_{g \in C}$ be a collection of subsets of R indexed by a regular Bohr set $C = R_\sigma$ with $\sigma \leq c\varepsilon/r2^d$, where $\sum_{g \in C} |A_g| \geq 2^{-d} |R| |C|$. Moreover, let $F = \sum_{g \in C} 1_{A_g} \circ 1_{A_g+g}$ and $F' = \sum_{g \in C} 1_{A_g} \circ 1_{A_g}$, and let $D(x) = |A_x|$ when $x \in C$ and 0 otherwise. Then either*

1. $|\langle \mu_F, \mu_D \rangle - \mu(R)^{-1}| \leq \varepsilon \mu(R)^{-1}$, or

2. There exists a sub-Bohr set $B \leq R$ with $\text{rk}(B) \leq \text{rk}(R) + O_\varepsilon(d^8)$ and $\mu(B) \geq \delta \mu(R)$ such that

$$\|\mu_{F'} * \mu_B\|_\infty \geq (1 + \varepsilon/80) \mu(R)^{-1},$$

where $\delta = \exp(-O_\varepsilon(r \log(1/\sigma) + rd^2 + rd \log r + d^{10}))$.

A quick aside on normalization: say for a moment that the $\{A_g\}_{g \in C}$ are random subsets of R , and imagine we were in the perfectly structured situation that $R + R = R$. Then, we would expect μ_F to be a very uniform function supported on $R + C$. Since C is a sufficiently narrow dilate of R , by regularity we can think of the distribution $R + C$ as being “close” to the uniform distribution on R . Thus, we are left with μ_F which is roughly uniform on R , and zero everywhere else. Since μ_F has mean 1 on all of G , we can really think of μ_F as being a uniform function taking values close to $\mu(R)^{-1}$ on R , and zero everywhere else. Making these structured statements into approximate ones will follow from Bohr set machinery.

For our application (Theorem 4.1), we will instantiate σ to be the value given by Lemma 4.2, which is smaller than the bound in the hypothesis of Theorem 4.3. This does not affect our final bound.

Proof of Theorem 4.1 from Theorem 4.3. Let $A \subseteq G \times G$ be skew-corner-free with $|A| = 2^{-d}|G|^2$, and define the collection $\{A_g\}_{g \in G}$ by $A_x = \{y : (x, y) \in A\}$. By Cauchy-Schwarz, we have $\sum |A_g|^2 \geq 2^{-2d}|G|^3$, so applying Lemma 4.2 with $\{A_g\}_{g \in G}$, $\varepsilon := 1/80$, $\delta := \exp(-cd^{11})$, $d := 2d$, and $r := c'd^8$ for constants $c, c' > 0$ to be determined later yields a collection of sets $\{(A_g - x_g) \cap R\}_{g \in C+w}$, where $R \subseteq G$ is a regular Bohr set of rank at most $O(d^9)$ and density $\mu(R) \geq 2^{-O(d^{12})}$, $C = R_\rho$ for $\rho = 2^{-O(d^2)}$ is a regular Bohr set, and $\{x_g\}_{g \in C+w}$, $w \in G$ are shifts. Observe that $\mu(C) \geq 2^{-O(d^{12})}$ by Lemma 2.12.

For brevity, define $A'_g := (A_{g+w} - x_{g+w}) \cap R$ for $g \in C$ and $A' := \{(g, y) : g \in C, y \in A'_g\} \subseteq C \times R$. Note that by Observation 2.3, A has at least as many skew corners as A' , $|A'| \geq 2^{-2d}|R||C|$, and $\{A'_g\}_{g \in C}$ is $(r, \delta, 1 + \varepsilon)$ -simultaneously spread in R .

Apply Theorem 4.3 with $\varepsilon := 1/2$ to the new collection, and set c' large enough for r to exceed the rank bound in the second conclusion. If $|\langle \mu_F, \mu_D \rangle - \mu(R)^{-1}| \leq \frac{1}{2} \mu(R)^{-1}$, Lemma 2.5 implies A' contains at least

$$\frac{\mu(R)^{-1}}{2} \cdot \frac{|A'|^3}{|G|^2} \geq 2^{-6d-1} \mu(R)^2 \mu(C)^3 |G|^4 \geq 2^{-O(d^{12})} |G|^4$$

skew corners. Since there are at most $|G|^3$ trivial skew corners in A' , we are done after rearrangement. Otherwise, there exists a sub-Bohr set $B \leq R$ with $\text{rk}(B) \leq \text{rk}(R) + r$ and $\mu(B) \geq \delta' \mu(R)$ such that

$$\|\mu_{F'} * \mu_B\|_\infty \geq \left(1 + \frac{1}{160}\right) \mu(R)^{-1},$$

where $\delta' = \exp(-O(d^{11}))$. By Lemma 2.20, this contradicts the fact that $\{A'_g\}_{g \in C}$ is $(r, \delta, 1 + 1/80)$ -simultaneously spread in R , assuming c is large enough to ensure $\delta < \delta'$. $\square$

4.2.1 Non-uniformity

The following lemma shows that if no corners are present, then we obtain some non-uniformity on our collection of sets in an appropriate p -norm. Here, it will be critical that C is contained in a sufficiently narrow dilate of R so that we can apply Lemma 2.14. This combined with an application of Hölder’s inequality constitutes most of the proof.

Lemma 4.4. *There exists an absolute constant $c > 0$ such that the following holds. Let $\varepsilon > 0$ and $\{A_g\}_{g \in C}$ be a collection of subsets of a regular Bohr set R of rank r , indexed by a regular Bohr set $C = R_\sigma$ with $\sigma \leq c\varepsilon/r2^d$, where $\sum_{g \in C} |A_g| \geq 2^{-d}|R||C|$. Define $\alpha_g = |A_g|/|G|$ and $\alpha = \sum_{g \in C} \alpha_g^2$. Additionally, let $D(x) = |A_x|$ when $x \in C$ and 0 otherwise, and let $f = \sum_{g \in C} 1_{A_g} \circ 1_{A_g+g}$. Then either*

1. $|\langle \mu_f, \mu_D \rangle - \mu(R)^{-1}| \leq \varepsilon \mu(R)^{-1}$ or
2. *There is some even integer $p = O(d)$ such that*

$$\left\| \frac{1}{\alpha} \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_g+g} - \alpha_g \mu_{R+g}) \right\|_{p(\mu_C)} \geq \frac{\varepsilon}{2} \mu(R)^{-1}.$$

Proof. Our proof follows closely [8, Proposition 20]. Assume the first conclusion does not hold, and note that

$$\left\langle \frac{1}{\alpha} \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_g+g} - \alpha_g \mu_{R+g}), \mu_D \right\rangle = \left(\frac{1}{\alpha} \sum_{g \in C} \langle 1_{A_g} \circ 1_{A_g+g}, \mu_D \rangle - \mu(R)^{-1} \right) \quad (5)$$

$$- \left(\frac{1}{\alpha} \sum_{g \in C} \alpha_g \langle 1_{A_g} \circ \mu_{R+g}, \mu_D \rangle - \mu(R)^{-1} \right) \quad (6)$$

$$- \left(\frac{1}{\alpha} \sum_{g \in C} \alpha_g \langle \mu_R \circ 1_{A_g+g}, \mu_D \rangle - \mu(R)^{-1} \right) \quad (7)$$

$$+ \left(\frac{1}{\alpha} \sum_{g \in C} \alpha_g^2 \langle \mu_R \circ \mu_{R+g}, \mu_D \rangle - \mu(R)^{-1} \right). \quad (8)$$

We will bound the last three terms in magnitude, then apply Hölder's inequality to the first term (5) to obtain the second conclusion. Consider the second term (6); we have

$$\begin{aligned} \frac{1}{\alpha} \sum_{g \in C} \alpha_g \langle 1_{A_g} \circ \mu_{R+g}, \mu_D \rangle - \mu(R)^{-1} &= \frac{1}{\alpha} \sum_{g \in C} \alpha_g \langle 1_{A_g}, \mu_{R+g} * \mu_{-D} \rangle - \mu(R)^{-1} \\ &= \frac{1}{\alpha} \sum_{g \in C} \alpha_g \langle 1_{A_g}, \mu_R * \mu_{-D}^g \rangle - \mu(R)^{-1} \\ &= \left(\frac{1}{\alpha} \sum_{g \in C} \alpha_g \langle 1_{A_g}, \mu_R \rangle \right) + \left(\frac{1}{\alpha} \sum_{g \in C} \alpha_g \langle 1_{A_g}, \mu_R * \mu_{-D}^g - \mu_R \rangle \right) - \mu(R)^{-1}. \end{aligned} \quad (9)$$

First, since $A_g \subseteq R$, we know that $\langle 1_{A_g}, \mu_R \rangle = \alpha_g \mu(R)^{-1}$. Summing over $g \in C$ gives

$$\frac{1}{\alpha} \sum_{g \in C} \alpha_g \langle 1_{A_g}, \mu_R \rangle = \frac{1}{\alpha} \sum_{g \in C} \alpha_g^2 \cdot \mu(R)^{-1} = \mu(R)^{-1},$$

so the first and third terms of (9) cancel out. We next bound the second term of (9). Since μ_D is supported on $C = R_\sigma$ and C is symmetric, we have μ_{-D} is supported on C as well. As $g \in C$, we can infer that μ_{-D}^g is supported on $C + C \subseteq R_{2\sigma}$. The last point to note is that $\frac{1}{\alpha} (\sum_{g \in C} \alpha_g) \leq 2^d \mu(R)^{-1}$ by Cauchy-Schwarz. We will now use the assumption that $\sigma \leq c\varepsilon/2^d \text{rk}(R)$ for a small enough constant $c > 0$. By Corollary 2.15, we know

$$\left| \frac{1}{\alpha} \sum_{g \in C} \alpha_g \langle 1_{A_g}, \mu_R * \mu_{-D}^g - \mu_R \rangle \right| \leq \frac{1}{16\alpha} \sum_{g \in C} \alpha_g \varepsilon 2^{-d} \leq \frac{\varepsilon}{16} \mu(R)^{-1}.$$

Similar reasoning bounds the third (7) and fourth (8) terms by the same quantity. Finally, let $f_0 = \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_{g+g}} - \alpha_g \mu_{R+g})$. Combining the above bounds gives

$$|\langle \mu_f, \mu_D \rangle - \mu(R)^{-1} - \langle \mu_{f_0}, \mu_D \rangle| \leq \frac{\varepsilon}{4} \mu(R)^{-1}.$$

For simplicity, define $D'(x) = D(x)/|R|$, so that $\|D'\|_\infty \leq 1$. Note that $\mu_D = \mu_{D'}$. If the first conclusion fails, and using the fact that D' is supported on C , we have

$$\frac{3}{4} \varepsilon \mu(R)^{-1} \leq |\langle \mu_{f_0}, \mu_D \rangle| = |\langle \mu_{f_0}, \mu_{D'} \rangle| = \mathbb{E}_C[D']^{-1} \cdot |\langle \mu_{f_0}, D' \rangle_{\mu_C}|.$$

An application of Hölder's inequality, for any $p \geq 1$, gives

$$|\langle \mu_{f_0}, D' \rangle_{\mu_C}| \leq \|\mu_{f_0}\|_{p(\mu_C)} \cdot \|D'\|_{(\frac{p}{p-1})(\mu_C)} \leq \|\mu_{f_0}\|_{p(\mu_C)} \cdot \|D'\|_\infty \cdot \|D'\|_{1(\mu_C)}^{1-1/p}.$$

We have $\|D'\|_{1(\mu_C)} = \mathbb{E}_C[D']^{-1} \geq 2^{-d}$. Combining the above inequalities gives

$$2^{d/p} \|\mu_{f_0}\|_{p(\mu_C)} \geq |\langle \mu_{f_0}, \mu_{D'} \rangle| \geq \frac{3}{4} \varepsilon \mu(R)^{-1}.$$

Choosing p an even integer so that $2^{d/p} \geq 3/2$ gives the result. $\square$

4.2.2 Unbalancing

At this point, we have a non-uniformity assumption on the function

$$F_0 = \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_{g+g}} - \alpha_g \mu_{R+g}),$$

and we would like to switch to working with the function $F' = \sum_{g \in C} 1_{A_g} \circ 1_{A_g}$. The reason for this is mostly technical; F' has non-negative Fourier coefficients, while F_0 in general may not. There are minor complications that arise when working with non-uniform measures, which accounts for the main differences between the following lemma and the corresponding one in the finite field case.

The goal of this section is to prove the following:

Lemma 4.5. *There exists an absolute constant $c > 0$ such that the following holds. Let $\varepsilon \in (0, 1)$ and $p \geq 2$ an even integer. Furthermore, let $\{A_g\}_{g \in C}$ be a collection of subsets of a regular Bohr set R of rank r indexed by a regular Bohr set $C = R_\sigma$ for $\sigma \leq c\varepsilon/r2^d$. Moreover, let $C', C'' \subseteq C_\tau$ both be regular Bohr sets for any $\tau \leq cr^{-1}$, and define $\alpha_g = |A_g|/|G|$ and $\alpha = \sum_{g \in C} \alpha_g^2$. Then*

$$\left\| \frac{1}{\alpha} \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_{g+g}} - \alpha_g \mu_{R+g}) \right\|_{p(\mu_C)} \geq \varepsilon \mu(R)^{-1}$$

implies that

$$\left\| \frac{1}{\alpha} \sum_{g \in C} 1_{A_g} \circ 1_{A_g} \right\|_{p'(\mu_{C'} * \mu_{C''-x})} \geq \left(1 + \frac{\varepsilon}{8}\right) \mu(R)^{-1}$$

for some shift $x \in G$ and integer $p' = O_\varepsilon(p)$.

To prove Lemma 4.5, we would like to remove the “shifts” in F_0 while maintaining the non-uniformity assumption. As a first step, we would like to switch to working with the function

$$F'_0 = \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_g} - \alpha_g \mu_R),$$

which enjoys the property of being spectrally non-negative. For technical reasons, we will want to start working with a non-uniform measure which is also spectrally non-negative, so that we can eventually apply Lemma 3.5. The following lemma accomplishes both of these goals with little loss in parameters.

Lemma 4.6. *There exists an absolute constant $c > 0$ such that the following holds. Let $p \geq 2$ be an even integer and $\{f_g\}_{g \in S}$ be a collection of real-valued functions on G indexed by some set $S \subseteq G$. Additionally, let $B \subseteq G$ and $B', B'' \subseteq B_\tau$ all be regular Bohr sets for $\tau \leq c \cdot \text{rk}(B)^{-1}$. Then*

$$\left\| \sum_{g \in S} f_g \circ f_g \right\|_{p(\mu_{B'} \circ \mu_{B'} * \mu_{B''} \circ \mu_{B''})} \geq \frac{1}{2} \left\| \sum_{g \in S} f_g \circ f_g^g \right\|_{p(\mu_B)}$$

where f^g denotes the function $f^g(x) = f(x - g)$.

Proof. For simplicity, we will define

$$F_1 := \sum_{g \in S} f_g \circ f_g^g \quad \text{and} \quad F_2 := \sum_{g \in S} f_g \circ f_g.$$

It follows by linearity that for $\gamma \in \widehat{G}$, we have

$$\widehat{F}_1(\gamma) = \sum_{g \in S} |\widehat{f}_g(\gamma)|^2 \gamma(g) \quad \text{and} \quad \widehat{F}_2(\gamma) = \sum_{g \in S} |\widehat{f}_g(\gamma)|^2,$$

and by the triangle inequality, we have $|\widehat{F}_1(\gamma)| \leq |\widehat{F}_2(\gamma)|$.

From here, the proof is almost identical to [8, Proposition 19]. Let $\nu = \mu_{B'} \circ \mu_{B'} * \mu_{B''} \circ \mu_{B''}$, and observe that ν is supported on B_ρ for $\rho = 4\tau$. Additionally, note that ν is spectrally non-negative, since $\widehat{\nu}(\gamma) = \widehat{\mu_{B'}}(\gamma)^2 \widehat{\mu_{B''}}(\gamma)^2$. Apply Lemma 2.13 to obtain

$$\mu_B \leq 2\mu_{B_{1+\rho}} * \nu.$$

Expanding, we have

$$\begin{aligned} \|F_1\|_{p(\mu_B)}^p &= \mathbb{E}_{x \in G} \mu_B(x) F_1(x)^p \\ &\leq 2 \mathbb{E}_{x \in G} (\mu_{B_{1+\rho}} * \nu)(x) F_1(x)^p \\ &= 2 \mathbb{E}_{t \in B_{1+\rho}} \mathbb{E}_{x \in G} \nu(x - t) F_1(x)^p. \end{aligned}$$

Recall that $\widehat{f^p} = \widehat{f}^{(p)}$, where $\widehat{f}^{(p)}$ is the p -fold convolution of $\widehat{f}$. By averaging, there exists some $t \in B_{1+\tau}$ such that

$$\begin{aligned} \|F_1\|_{p(\mu_B)}^p &\leq 2 \mathbb{E}_{x \in G} \nu(x - t) F_1(x)^p \\ &= 2 \sum_{\gamma \in \widehat{G}} \widehat{\nu}(\gamma) \gamma(-t) \left(\widehat{F}_1 \right)^{(p)}(\gamma) \\ &\leq 2 \sum_{\gamma \in \widehat{G}} \widehat{\nu}(\gamma) \left(|\widehat{F}_1| \right)^{(p)}(\gamma) \\ &\leq 2 \sum_{\gamma \in \widehat{G}} \widehat{\nu}(\gamma) \left(|\widehat{F}_2| \right)^{(p)}(\gamma) \\ &= 2 \|F_2\|_{p(\nu)}^p. \end{aligned}$$

□

From here we can instantiate the previous lemma to obtain the following corollary.

Corollary 4.7. *There exists an absolute constant $c > 0$ such that the following holds. Let $\varepsilon \in (0, 1)$ and $p \geq 2$ an even integer. Furthermore, let $\{A_g\}_{g \in C}$ be a collection of subsets of $R \subseteq G$ indexed by a regular Bohr set C . Define $\alpha_g = |A_g|/|G|$ and $\alpha = \sum_{g \in C} \alpha_g^2$. Moreover, let $C', C'' \subseteq C_\tau$ be regular Bohr sets for some $\tau \leq c \cdot \text{rk}(C)^{-1}$, and let $\nu = \mu_{C'} \circ \mu_{C'} * \mu_{C''} \circ \mu_{C''}$. Then*

$$\left\| \frac{1}{\alpha} \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_{g+g}} - \alpha_g \mu_{R+g}) \right\|_{p(\mu_C)} \geq \varepsilon \mu(R)^{-1}$$

implies that

$$\left\| \frac{1}{\alpha} \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_g} - \alpha_g \mu_R) \right\|_{p(\nu)} \geq \frac{1}{2} \varepsilon \mu(R)^{-1}.$$

Proof. Apply Lemma 4.6 with $B := C$, $S := C$, $B' := C'$, $B'' := C''$, and $f_g := 1_{A_g} - \alpha_g \mu_R$. We have

$$\sum_{g \in C} f_g \circ f_g = \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_g} - \alpha_g \mu_R)$$

and

$$\sum_{g \in C} f_g \circ f_g^g = \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_g^g} - \alpha_g \mu_R^g) = \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_{g+g}} - \alpha_g \mu_{R+g}). \quad \square$$

The following lemma lets us move from a non-uniformity assumption on a function F'_0 whose mean is zero to an assumption on F' . This will be critical when we want to apply “sifting” later (see Section 4.2.3). The lemma applies Lemma 3.5 after making some estimates using Bohr sets and Lemma 2.14.

Lemma 4.8. *There exists an absolute constant $c > 0$ such that the following holds. Let $\varepsilon \in (0, 1)$ and $p \geq 2$ be an integer. Additionally, let $\{A_g\}_{g \in C}$ be a collection of subsets of a regular Bohr set R of rank r indexed by a regular Bohr set C with $\sum_{g \in C} |A_g| \geq 2^{-d} |R| |C|$. Define $\alpha_g = |A_g|/|G|$ and $\alpha = \sum_{g \in C} \alpha_g^2$, and let $\nu: G \rightarrow \mathbb{R}_{\geq 0}$ be a probability measure supported on R_σ for $\sigma \leq c\varepsilon/r2^d$ such that $\widehat{\nu} \geq 0$. If*

$$\left\| \frac{1}{\alpha} \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_g} - \alpha_g \mu_R) \right\|_{p(\nu)} \geq \varepsilon \mu(R)^{-1},$$

then there is an integer $p' = O_\varepsilon(p)$ such that

$$\left\| \frac{1}{\alpha} \sum_{g \in C} 1_{A_g} \circ 1_{A_g} \right\|_{p'(\nu)} \geq (1 + \varepsilon/4) \mu(R)^{-1}.$$

Proof. The proof follows [8, Proposition 18]. We write

$$f = \frac{1}{\alpha} \sum_{g \in C} \alpha_g (1_{A_g} \circ \mu_R) + \alpha_g (\mu_R \circ 1_{A_g}) - \alpha_g^2 (\mu_R \circ \mu_R).$$

For any $p' \geq 1$,

$$\left\| \frac{1}{\alpha} \sum_{g \in C} 1_{A_g} \circ 1_{A_g} \right\|_{p'(\nu)} = \left\| \frac{1}{\alpha} \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_g} - \alpha_g \mu_R) + \mu(R)^{-1} + f - \mu(R)^{-1} \right\|_{p'(\nu)}$$

$$\geq \left\| \frac{1}{\alpha} \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_g} - \alpha_g \mu_R) + \mu(R)^{-1} \right\|_{p'(\nu)} - \|f - \mu(R)^{-1}\|_{p'(\nu)}.$$

First, we use spectral positivity and Lemma 3.5 to obtain that for $p' = O_\varepsilon(p)$ it holds

$$\left\| \frac{1}{\alpha} \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_g} - \alpha_g \mu_R) + \mu(R)^{-1} \right\|_{p'(\nu)} \geq \left(1 + \frac{\varepsilon}{2}\right) \mu(R)^{-1}.$$

Next, we use Corollary 2.15 to bound the error term. For $x \in \text{supp}(\nu)$, we have

$$\begin{aligned} \frac{1}{\alpha} \sum_{g \in C} \alpha_g |(1_{A_g} \circ \mu_R)(x) - (1_{A_g} \circ \mu_R)(0)| &= \frac{1}{\alpha} \sum_{g \in C} \alpha_g |\langle 1_{A_g}, \mu_{R-x} - \mu_R \rangle| \\ &= \frac{1}{\alpha} \sum_{g \in C} \alpha_g |\langle 1_{A_g}, \mu_R * 1_{-x} - \mu_R \rangle| \\ &\leq \frac{1}{\alpha} \sum_{g \in C} \alpha_g \cdot \sigma \text{rk}(R) \\ &= \frac{\varepsilon}{4} \mu(R)^{-1}, \end{aligned}$$

where the implicit constant in σ is set sufficiently small. Similar reasoning gives the same bounds for the other terms in f . We have

$$\|f - \mu(R)^{-1}\|_{p'(\nu)} \leq \|f - \mu(R)^{-1}\|_{L^\infty(\text{supp}(\nu))} \leq \frac{1}{4} \mu(R)^{-1}.$$

This concludes the proof. $\square$

Putting everything together, we can now prove the main result of this section (restated below). The choice of measure $\mu_{C'} * \mu_{C''-x}$ in the conclusion may seem a bit mysterious, but it will be necessary when we apply (in Corollary 4.9) that the measure we are working with is a convolution of two sets.

Lemma 4.5. *There exists an absolute constant $c > 0$ such that the following holds. Let $\varepsilon \in (0, 1)$ and $p \geq 2$ an even integer. Furthermore, let $\{A_g\}_{g \in C}$ be a collection of subsets of a regular Bohr set R of rank r indexed by a regular Bohr set $C = R_\sigma$ for $\sigma \leq c\varepsilon/r2^d$. Moreover, let $C', C'' \subseteq C_\tau$ both be regular Bohr sets for any $\tau \leq cr^{-1}$, and define $\alpha_g = |A_g|/|G|$ and $\alpha = \sum_{g \in C} \alpha_g^2$. Then*

$$\left\| \frac{1}{\alpha} \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_{g+g}} - \alpha_g \mu_{R+g}) \right\|_{p(\mu_C)} \geq \varepsilon \mu(R)^{-1}$$

implies that

$$\left\| \frac{1}{\alpha} \sum_{g \in C} 1_{A_g} \circ 1_{A_g} \right\|_{p'(\mu_{C'} * \mu_{C''-x})} \geq \left(1 + \frac{\varepsilon}{8}\right) \mu(R)^{-1}$$

for some shift $x \in G$ and integer $p' = O_\varepsilon(p)$.

Proof. We first apply Corollary 4.7 to infer that

$$\left\| \frac{1}{\alpha} \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_g} - \alpha_g \mu_R) \right\|_{p(\nu)} \geq \frac{\varepsilon}{2} \mu(R)^{-1}.$$

Then Lemma 4.8 implies that

$$\left\| \frac{1}{\alpha} \sum_{g \in C} 1_{A_g} \circ 1_{A_g} \right\|_{p'(v)} \geq (1 + \varepsilon/8) \mu(R)^{-1}$$

for some integer $p' = O_\varepsilon(p)$ and $v = \mu_{C'} \circ \mu_{C'} * \mu_{C''} \circ \mu_{C''}$. By averaging, for some choice of $x \in C' + C''$, we must have

$$\left\| \frac{1}{\alpha} \sum_{g \in C} 1_{A_g} \circ 1_{A_g} \right\|_{p'(\mu_{C'} * \mu_{C''-x})} \geq (1 + \varepsilon/8) \mu(R)^{-1}. \quad \square$$

4.2.3 Dependent random choice

Here, we perform the analogous step to what Kelley and Meka call “sifting,” relying on Lemma 3.8.

Corollary 4.9. *Let $\varepsilon \in (0, 1)$ and $p \geq 1$ be an integer. Furthermore, let $\{A_g\}_{g \in C}$ be a collection of subsets of a set $R \subseteq G$ indexed by a set $C \subseteq G$. Assume $\sum_{g \in C} |A_g| = 2^{-d} |R| |C|$. Moreover, let $B, B' \subseteq G$ and $f = \sum_{g \in C} 1_{A_g} \circ 1_{A_g}$. If*

$$\|\mu_f\|_{p(\mu_B * \mu_{B'})} \geq (1 + \varepsilon) \mu(R)^{-1}$$

and

$$S = \{x \in G : \mu_f(x) > (1 - \varepsilon/2) \|\mu_f\|_{p(\mu_B * \mu_{B'})}\},$$

then there exists $M_1 \subseteq B$, $M_2 \subseteq B'$ such that

$$\langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \frac{\varepsilon}{4} \quad \text{and} \quad \min(\mu_B(M_1), \mu_{B'}(M_2)) \geq 2^{-O_\varepsilon(dp)}.$$

Proof. Apply Lemma 3.8 to R and C with $B_1 := B$, $B_2 := -B'$, $d := d - \log(\mu(R))$, $\varepsilon := \varepsilon/2$, and $\delta := \varepsilon/4$. $\square$

4.2.4 Almost periodicity

Corollary 4.9 allowed us to turn a non-uniformity assumption on F' into dense sets M_1, M_2 which are robust witnesses to upwards deviations in F' . The following result allows us to obtain a density increment on some translate of a sufficiently large Bohr set.

Theorem 4.10 ([22, Theorem 5.4], see also [8, Theorem 17]). *Let $\varepsilon > 0$ and $B, B' \subseteq G$ be regular Bohr sets of rank r . Suppose that $M_1 \subseteq B$ with $\mu_B(M_1) = 2^{-d_1}$ and $M_2 \subseteq B'$ with $\mu_{B'}(M_2) = 2^{-d_2}$. Let S be any set with $|S| \leq 2|B|$. Then there is a regular Bohr set $B'' \leq B'$ of rank at most*

$$\text{rk}(B'') \leq r + O_\varepsilon(d_1^3 d_2)$$

and density

$$\mu(B'') \geq \exp(-O_\varepsilon((d_1^3 d_2 + r)(d_1 + d_2 + \log r))) \mu(B')$$

such that

$$|\langle \mu_{B''} * \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle - \langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle| \leq \varepsilon.$$

Note that the Bohr set B'' is a sub-Bohr set of B' in the conclusion of the above theorem. We will apply Theorem 4.10 to obtain the corollary below. The choice of dilates for C', C'' in the hypothesis are simply to ensure that $|C' + C''| \leq 2|C'|$ via regularity.

Corollary 4.11. *Let $\varepsilon \in (0, 1)$ and $p \geq 1$ an integer. Furthermore, let $\{A_g\}_{g \in C}$ be a collection of subsets of a Bohr set R of rank r indexed by a regular Bohr set C , and assume $\sum_{g \in C} |A_g| = 2^{-d} |R| |C|$. Let C' be a rank r regular Bohr set, and let $C'' = C'_\sigma$ where $\sigma = O(r^{-1})$ is chosen to ensure C'' is regular. Moreover, let $M_1 \subseteq C'$ and $M_2 \subseteq C'' - x$ for some shift $x \in G$ with density guarantees $\mu_{C'}(M_1), \mu_{C''-x}(M_2) \geq 2^{-O_\varepsilon(dp)}$. Let $S \subseteq G$ be any set. If*

$$\langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \varepsilon/2,$$

then there exists a regular Bohr set $B \leq C''$ such that

$$\langle \mu_B * \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \varepsilon$$

where

$$\text{rk}(B) \leq r + O_\varepsilon(d^4 p^4) \quad \text{and} \quad \mu(B) \geq \exp(-O_\varepsilon(rp(d + \log r) + d^5 p^5)) \mu(C'').$$

Proof. Since $\mu_{M_1} \circ \mu_{M_2}$ is supported on $C' + C'' - x$, we can assume S is supported on the same set. We can also shift M_2, S by x so that $M_2 \subseteq C''$ without changing any assumptions or conclusions. Thus, in order to apply Theorem 4.10, it suffices to check that $|C' + C''| \leq 2|C'|$. By regularity and our choice of σ , we have

$$|C' + C''| = |C' + C'_\sigma| \leq (1 + O(\sigma \cdot r)) |C'| \leq 2|C'|.$$

The proof concludes by applying Theorem 4.10 with $B := C', B' := C''$ and M_1, M_2, S as given. $\square$

4.2.5 Combining steps

All that remains to prove Theorem 4.3 (restated below) is to combine the prior steps.

Theorem 4.3. *There exists an absolute constant $c > 0$ such that the following holds. Let $\varepsilon \in (0, 1)$ and $R \subseteq G$ be a regular Bohr set of rank r . Furthermore, let $\{A_g\}_{g \in C}$ be a collection of subsets of R indexed by a regular Bohr set $C = R_\sigma$ with $\sigma \leq c\varepsilon/r2^d$, where $\sum_{g \in C} |A_g| \geq 2^{-d} |R| |C|$. Moreover, let $F = \sum_{g \in C} 1_{A_g} \circ 1_{A_{g+g}}$ and $F' = \sum_{g \in C} 1_{A_g} \circ 1_{A_g}$, and let $D(x) = |A_x|$ when $x \in C$ and 0 otherwise. Then either*

1. $|\langle \mu_F, \mu_D \rangle - \mu(R)^{-1}| \leq \varepsilon \mu(R)^{-1}$, or
2. *There exists a sub-Bohr set $B \leq R$ with $\text{rk}(B) \leq \text{rk}(R) + O_\varepsilon(d^8)$ and $\mu(B) \geq \delta \mu(R)$ such that*

$$\|\mu_{F'} * \mu_B\|_\infty \geq (1 + \varepsilon/80) \mu(R)^{-1},$$

where $\delta = \exp(-O_\varepsilon(r \log(1/\sigma) + rd^2 + rd \log r + d^{10}))$.

Proof. If we are not in the first condition, then Lemma 4.4 implies there is some integer $p = O(d)$ such that

$$\left\| \frac{1}{\alpha} \sum_{g \in C} (1_{A_g} - \alpha_g \mu_R) \circ (1_{A_{g+g}} - \alpha_g \mu_{R+g}) \right\|_{p(\mu_C)} \geq \frac{1}{2} \varepsilon \mu(R)^{-1}.$$

We then apply Lemma 4.5 to infer that

$$\left\| \frac{1}{\alpha} \sum_{g \in C} 1_{A_g} \circ 1_{A_g} \right\|_{p'(\mu_{C'} * \mu_{C''-x})} \geq \left(1 + \frac{\varepsilon}{16}\right) \mu(R)^{-1},$$

for some shift $x \in G$, where $C' \subseteq C_{\tau_1}$ and $C'' \subseteq C'_{\tau_2}$ are regular Bohr sets for some $\tau_1, \tau_2 = O(r^{-1})$ chosen so that C', C'' are regular. In particular, the choice of dilates is so that we can later use Corollary 4.11. Let

$$S = \left\{ x \in G : \sum_{g \in C} 1_{A_g} \circ 1_{A_g}(x) > (1 - \varepsilon/32) \left\| \sum_{g \in C} 1_{A_g} \circ 1_{A_g} \right\|_{p'(\mu_{C'} * \mu_{C''-x})} \right\}.$$

Via Corollary 4.9, we obtain $M_1 \subseteq C'$, $M_2 \subseteq C''$ such that

$$\langle \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \varepsilon/128,$$

where M_1 and M_2 satisfy the density constraints

$$\min(\mu_B(M_1), \mu_{B'}(M_2)) \geq 2^{-O_\varepsilon(dp)}.$$

We then apply Corollary 4.11 to produce a regular Bohr set $B \subseteq C''$ such that

$$\langle \mu_B * \mu_{M_1} \circ \mu_{M_2}, 1_S \rangle \geq 1 - \varepsilon/64,$$

where

$$\text{rk}(B) = r + O_\varepsilon(d^8) \quad \text{and} \quad \mu(B) \geq \exp(-O_\varepsilon(rd^2 + rd \log r + d^{10})) \mu(C'').$$

Since $C = R_\sigma$, $C' = R_{\sigma\tau_1}$, and $C'' \subseteq R_{\sigma\tau_1\tau_2}$, Lemma 2.12 implies

$$\mu(C'') \geq \left(\frac{\sigma\tau_1\tau_2}{4}\right)^r \mu(R) \geq \exp(-O(r \log(1/\sigma) + r \log r)) \mu(R).$$

Combining the above inequalities gives

$$\mu(B) \geq \exp(-O_\varepsilon(r \log(1/\sigma) + rd^2 + rd \log r + d^{10})).$$

The last step is to note that

$$\begin{aligned} (1 + \varepsilon/80)\mu(R)^{-1} &\leq (1 - \varepsilon/64)(1 - \varepsilon/32)(1 + \varepsilon/16)\mu(R)^{-1} \\ &\leq \langle \mu_B * \mu_{M_1} \circ \mu_{M_2}, \mu_{F'} \rangle \\ &\leq \|\mu_{F'} * \mu_B\|_\infty \|\mu_{M_1} \circ \mu_{M_2}\|_1 \\ &= \|\mu_{F'} * \mu_B\|_\infty. \end{aligned}$$

□

Acknowledgments

We thank an anonymous reviewer for a number of helpful comments. MJ and AO thank Lior Lovett for his constant encouragement.

References

- [1] Miklós Ajtai and Endre Szemerédi, *Sets of lattice points that form no squares*, Stud. Sci. Math. Hungar **9** (1974), no. 1975, 9–11. 3
- [2] Felix A Behrend, *On sets of integers which contain no three terms in arithmetical progression*, Proceedings of the National Academy of Sciences **32** (1946), no. 12, 331–332. 1, 2
- [3] Adrian Beker, *Improved bounds for skew corner-free sets*, arXiv preprint arXiv:2402.19169 (2024). 1, 2
- [4] Thomas F Bloom, *Additive combinatorics*, 2021, Course notes available at <http://thomasbloom.org/teaching/AC2021.pdf>. 5
- [5] Jean Bourgain, *On triples in arithmetic progression*, Geometric And Functional Analysis **9** (1999), no. 5, 968–984. 4

- [6] Thomas F Bloom and Olof Sisask, *Breaking the logarithmic barrier in Roth’s theorem on arithmetic progressions*, arXiv preprint arXiv:2007.03528 (2020). 5
- [7] Thomas F Bloom and Olof Sisask, *An improvement to the Kelley–Meka bounds on three-term arithmetic progressions*, arXiv preprint arXiv:2309.02353 (2023). 1, 2
- [8] Thomas F Bloom and Olof Sisask, *The Kelley–Meka bounds for sets free of three-term arithmetic progressions*, Essential Number Theory **2** (2023), no. 1, 15–44. 1, 2, 5, 10, 11, 13, 18, 20, 21, 23
- [9] Henry Cohn, Robert Kleinberg, Balázs Szegedy, and Christopher Umans, *Group-theoretic algorithms for matrix multiplication*, 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS’05), IEEE, 2005, pp. 379–388. 1
- [10] Henry Cohn and Christopher Umans, *A group-theoretic approach to fast matrix multiplication*, 44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings., IEEE, 2003, pp. 438–449. 1
- [11] Ben Green, *Lower bounds for corner-free sets*, New Zealand J. Math **51** (2021), 1–2. 2
- [12] Zach Hunter, *Corner-free sets via the torus*, arXiv preprint arXiv:2209.10012 (2022). 2
- [13] Zander Kelley and Raghu Meka, *Strong bounds for 3-progressions*, 2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS), 2023, pp. 933–973. 1, 2, 6, 8, 10, 11, 13
- [14] Nati Linial and Adi Shraibman, *Larger corner-free sets from better NOF Exactly- n protocols*, arXiv preprint arXiv:2102.00421 (2021). 2
- [15] Raghu Meka, *Strong bounds for 3-progressions*, 2023, Talk as part the Breakthroughs lecture series at the Simons Institute for the Theory of Computing. [Accessed: 2024 03 16] <https://www.youtube.com/watch?v=WN7rJPWy6z8>. 2
- [16] Luka Milićević, *Good bounds for sets lacking skew corners*, arXiv preprint arXiv:2404.07180 (2024). 2
- [17] Sarah Peluse, *Finite field models in arithmetic combinatorics—twenty years on*, arXiv preprint arXiv:2312.08100 (2023). 2
- [18] Kevin Pratt, *On Generalized Corners and Matrix Multiplication*, 15th Innovations in Theoretical Computer Science Conference (ITCS 2024) (Dagstuhl, Germany) (Venkatesan Guruswami, ed.), Leibniz International Proceedings in Informatics (LIPIcs), vol. 287, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024, pp. 89:1–89:17. 1, 3
- [19] Cosmin Pohoata and Dmitrii Zakharov, *On skew corner-free sets*, arXiv preprint arXiv:2401.17507 (2024). 1, 2, 7
- [20] Tom Sanders, *On the Bogolyubov–Ruzsa lemma*, Analysis & PDE **5** (2012), no. 3, 627–655. 13
- [21] Ilya D Shkredov, *On a generalization of Szemerédi’s theorem*, Proceedings of the London Mathematical Society **93** (2006), no. 3, 723–760. 2
- [22] Tomasz Schoen and Olof Sisask, *Roth’s theorem for four variables and additive structures in sums of sparse sets*, Forum of Mathematics, Sigma, vol. 4, Cambridge University Press, 2016, p. e5. 5, 10, 13, 23
- [23] Terence Tao and Van H Vu, *Additive combinatorics*, vol. 105, Cambridge University Press, 2006. 5

AUTHORS

Michael Jaber
Department of Computer Science, UT Austin
USA
mjjaber@cs.utexas.edu
<https://michaeljaber.github.io/>

Shachar Lovett
Department of Computer Science and Engineering, UC San Diego
USA
slovett@ucsd.edu
<https://cseweb.ucsd.edu/~slovett/>

Anthony Ostuni
Department of Computer Science and Engineering, UC San Diego
USA
aostuni@ucsd.edu
<https://aostuni.github.io/>