Popular differences for corners in Abelian groups

BY AARON BERGER

Department of Mathematics, MIT, 182 *Memorial Drive,*
Cambridge, MA 02142, *U.S.A.*
e-mail: bergera@mit.edu

Abstract

For a compact abelian group G, a *corner* in $G \times G$ is a triple of points (x, y), $(x, y + d)$, $(x + d, y)$. The classical corners theorem of Ajtai and Szemerédi implies that for every $\alpha > 0$, there is some $\delta > 0$ such that every subset $A \subset G \times G$ of density α contains a δ fraction of all corners in $G \times G$, as x, y, d range over G.

Recently, Mandache proved a "popular differences" version of this result in the finite field case $G = \mathbb{F}_p^n$, showing that for any subset $A \subset G \times G$ of density α, one can fix $d \neq 0$ such that A contains a large fraction, now known to be approximately α^4, of all corners with difference d, as x, y vary over G. We generalise Mandache's result to all compact abelian groups G, as well as the case of corners in $\mathbb{Z}^2$.

2020 Mathematics Subject Classification: Primary: 11B30; Secondary: 11B25

————

1. *Introduction*

The following *popular differences* version of Szemerédi's theorem was conjectured by Bergelson, Host, and Kra [**2**] and proved by Green [**5**] for $k = 3$ and Green–Tao [**6**] for $k = 4$: every subset of $[N]$ of size at least αN contains at least $(\alpha^k - o(1))N$ k-term arithmetic progressions, or k-APs, with the same common difference. That is, such a set contains $(\alpha^k - o(1))N$ distinct copies of $\{x, x + d, \ldots, x + (k-1)d\}$ for some fixed $d \neq 0$. These results involve the method of arithmetic regularity developed by Green, and the lower bounds are essentially best possible; a randomised construction gives subsets of density α and only $(\alpha^k + o(1))N$ k-APs with common difference d for each $d \neq 0$. Such polynomial bounds for AP counts are not the norm in additive combinatorics. Indeed, in an appendix to [**2**], Rusza shows that for $k \geq 5$, one can construct sets with density α and fewer than $\alpha^{\omega(1)}$ distinct k-APs[1] with common difference d for each $d \neq 0$. The natural place to look for generalisations is in higher-dimensional configurations. The *corners theorem* of Ajtai and Szemerédi [**1**] is a classical result in this style in two dimensions, implying that any subset of $[N]^2$ with at least αN^2 elements contains $\Omega(N^3)$ corners, which are triples of the form $\{(x, y), (x, y + d), (x + d, y)\}$. As usual, the dependence of the implicit constant in $\Omega(N^3)$

----

[1]The $\omega(1)$ term goes to ∞ as $\alpha \to 0$ and $N \to \infty$.

on the density α is quite poor. One might hope to obtain a better dependence for some fixed d than what one obtains on average by the Ajtai-Szemerédi result. The following result due to Mandache [9] does precisely this, but in the finite field model instead of $[N]$. For a family $\mathcal{F}$ of finite abelian groups, let $M_{\mathcal{F}}(\alpha) \in [0, 1]$ be the minimum constant such that the following statement is true: For every $A \subset G \times G$ with size at least $\alpha|G|^2$, there is some $d \neq 0$ such that A contains $(M_{\mathcal{F}}(\alpha) - o(1))|G|^2$ corners with common difference d, where the $o(1)$ term goes to 0 as $|G| \to \infty$. Mandache shows that for fixed p and $\mathcal{F} = \{\mathbb{F}_p^n\}$, one has

$$m'(\alpha) \leq M_{\mathcal{F}}(\alpha) \leq m(\alpha),$$

where $m'(\alpha)$ and $m(\alpha)$ are polynomially large in terms of α and are given by the solutions to a certain variational problem we describe below. In a somewhat surprising difference from the k-AP case, Mandache shows that the exponents in the growth rates of $m(\alpha), m'(\alpha)$ are strictly greater than 3, whereas random subsets of $G \times G$ have approximately $\alpha^3|G|^2$ corners for each fixed difference $d \neq 0$. The asymptotic growth rates of m and m' were recently determined by Fox, Sah, Sawhney, Stoner, and Zhao [4], who also discuss other possible generalisations and barriers to generalisation for popular differences results. We will include their bounds on m and m' following the discussion of the variational problem itself.

For $\phi : [0, 1]^3 \to [0, 1]$, define

$$T(\phi) := \int_{[0,1]^3} dx\, dy\, dz \int_{[0,1]} \phi(x, y, z')\, dz' \int_{[0,1]} \phi(x, y', z)\, dy' \int_{[0,1]} \phi(x', y, z)\, dx'.$$

We are concerned with the infimum of $T(\phi)$ over ϕ with a fixed expectation:

$$m(\alpha) := \inf_{\substack{\phi:[0,1]^3 \to [0,1] \\ \mathbb{E}[\phi]=\alpha}} T(\phi).$$

This expression may be rewritten by taking independent $X, Y, Z \sim \mathrm{Unif}([0, 1])$, in which case one has

$$T(\phi) = \mathbb{E}\big[\mathbb{E}(\phi|X, Y)\mathbb{E}(\phi|X, Z)\mathbb{E}(\phi|Y, Z)\big]. \tag{1·1}$$

It is clear that the underlying probability space is unimportant here; if X, Y, Z are any independent random variables and ϕ has expectation α, then $T(\phi) \geq m(\alpha)$.

Mandache showed that for any family of finite abelian groups, one has

$$M(\alpha) \leq m(\alpha).$$

Secondly, let $m'(\alpha)$ be the maximal convex function such that $m'(\alpha) \leq m(\alpha)$ pointwise. Mandache proved that for fixed p and $\mathcal{F} = \{F_p^n\}$, one has

$$M_{\mathcal{F}}(\alpha) \geq m'(\alpha).$$

More specifically, for the lower bound Mandache showed that for $A \subseteq G \times G$ with density $\mu(A) = \alpha$, there is a subspace $W \subseteq G$ with codimension bounded in terms of ϵ so that

$$\mathbb{E}_{x,y \in G, d \in W}\big[\mathbb{1}_A(x, y)\mathbb{1}_A(x, y + d)\mathbb{1}_A(x + d, y)\big] \geq m'(\alpha) - \epsilon.$$

Letting $n \to \infty$, by the boundedness of codim(W), the corners with difference $d = 0$ contribute $o(1)$ to this expectation, and so he concludes that there is some $d \neq 0$ with

$$\mathbb{E}_{x,y \in G}\big[\mathbb{1}_A(x, y)\mathbb{1}_A(x, y + d)\mathbb{1}_A(x + d, y)\big] \geq m'(\alpha) - O(\epsilon).$$

Since this inequality holds for every ϵ as $n \to \infty$, we obtain the popular differences result $M_{\mathcal{F}}(\alpha) \geq m'(\alpha)$. Mandache showed that

$$\alpha^4 \leq m'(\alpha) \leq m(\alpha) \leq C\alpha^{3.13}.$$

Fox, Sah, Sawhney, Stoner and Zhao [4] determined more precise asymptotics, showing:

$$\omega(\alpha^4) \leq m'(\alpha) \leq m(\alpha) \leq \alpha^{4-o(1)},$$

where the $o(1)$ term approaches 0 as $\alpha \to 0$, and the $\omega(\alpha^4)$ term is $\alpha^4/o(1)$.

We generalise Mandache's result to all compact abelian groups.

THEOREM 1·1. *For any $\alpha, \epsilon > 0$, there is some absolute $c > 0$ such that the following holds: For any compact abelian group G with Haar probability measure μ and any set $A \subseteq G \times G$ with $\mu(A) = \alpha$, there is a Bohr set $B \subseteq G$ with $\mu(B) \geq c$ such that*

$$\int_{\substack{x,y \in G \\ r \in B}} \mathbb{1}_A(x, y)\mathbb{1}_A(x, y + r)\mathbb{1}_A(x + r, y) \, dx \, dy \, dr \geq m'(\alpha) - \epsilon.$$

From this result and a simple modification we obtain the following two corollaries.

COROLLARY 1·2. *Let G be any finite abelian group and $A \subseteq G \times G$ have size $|A| \geq \alpha|G|^2$. Then there is some $r \neq 0$ such that A contains at least $(m'(\alpha) - o(1))|G|^2$ corners of the form $\{(x, y), (x, y + r), (x + r, y)\}$.*

COROLLARY 1·3. *Let $A \subset [n]^2$ have size $|A| \geq \alpha n^2$. Then there is some $r \neq 0$ such that A contains at least $(m'(\alpha) - o(1))n^2$ corners of the form $\{(x, y), (x, y + r), (x + r, y)\}$.*

1·1. *Notation*

Let $(G, +)$ be a compact abelian group, with Haar probability measure μ, and a (discrete) dual group $\hat{G}$ of characters $\xi : G \to \mathbb{R}/\mathbb{Z}$. We will use function evaluation notation for characters, so $\xi(x)$ denotes the image of $x \in G$ under $\xi \in \hat{G}$. For a measurable function $f : G \to \mathbb{R}$ and a measurable partition P of G, we let $f_P = \mathbb{E}(f|P)$ be the function obtained by averaging f on each part of P. For measurable $X \subseteq G$ with $\mu(X) > 0$, define

$$\mu_X := \frac{\mathbb{1}_X}{\mu(X)},$$

to be the indicator of X, normalised to have integral 1.

For asymptotics, we use $x = O(y)$ and $x \lesssim y$ when we would otherwise write $x \leq Cy$ for some absolute constant C. An absolute constant is independent of any variables in the problem. For example, it suffices to prove Theorem 1·1 with $m'(\alpha) - \epsilon$ replaced by $m'(\alpha) - O(\epsilon)$, as the implicit constant is independent of ϵ.

For $f : G \to \mathbb{R}$, we use the L^p norms, normalised as follows.

$$\|f\|_{L^p} = \left(\int_G |f(x)|^p \, dx\right)^{1/p}.$$

For $\hat{f} : \hat{G} \to \mathbb{R}/\mathbb{Z}$, we use ℓ^p norms.

$$\|\hat{f}\|_{\ell^p} = \left(\sum_{\xi \in \hat{G}} (\hat{f}(\xi))^p \right)^{1/p}.$$

Similarly, Fourier transforms are written with an integral over the real domain and a sum over the frequency domain, so $\hat{f}(\xi) = \int_G f(x) e^{-2\pi i \xi(x)} dx$ and $f(x) = \sum_{\xi \in \hat{G}} \hat{f}(\xi) e^{2\pi i \xi(x)}$. Using this notation, Plancherel's theorem states $\|f\|_{L^2} = \|\hat{f}\|_{\ell^2}$. Finally, for $x \in \mathbb{R}$ or $\mathbb{R}/\mathbb{Z}$, we write $\|x\|_{\mathbb{R}/\mathbb{Z}}$ to mean the distance from x to the nearest integer.

2. Bohr set preliminaries

The *Bohr set* given by a finite set of frequencies $S \subset \hat{G}$ and $\rho > 0$ is defined to be

$$B(S, \rho) = \left\{ x \in G : \sup_{\xi \in S} \|\xi(x)\|_{\mathbb{R}/\mathbb{Z}} < \rho \right\}.$$

For $\delta = 1/N$, we also define the Bohr partition $\mathfrak{B}(S, \delta)$ to be the union of parts of the form

$$\left\{ x \in G : \xi_i \cdot x \in \left[\frac{s_i - 1}{N}, \frac{s_i}{N} \right) \ \forall i \in [d] \right\},$$

for each choice of $\{s_i\} \in [N]^d$. The number of parts in a Bohr partition is $|\mathfrak{B}(S, \delta)| = \delta^{-|S|}$. Each Bohr set has size bounded below by a constant depending on ρ and $|S|$:

FACT 2·1. *For any Bohr set $B(S, \rho)$, there exists a constant $C_{|S|,\rho} > 0$ depending only on ρ and $|S|$ such that*:

$$\mu(B(S, \rho)) \geq C_{|S|,\rho}.$$

Proof. Consider the maximal $\delta < \rho$, $\delta = 1/N$. By the triangle inequality, for any $x \in G$, whichever part of $\mathfrak{B}(S, \delta)$ contains x is itself entirely contained in $x + B(S, \rho)$. Choosing one representative x from each nonempty part of $\mathfrak{B}(S, \delta)$, we see that $N^{|S|}$ translates of $B(S, \rho)$ suffice to cover G.

When drawing analogies between the finite field model and the case of general abelian groups, Bohr sets take the role of subspaces. One major problem with the general setting is that Bohr sets, unlike subspaces, are not closed under addition. The common approach to handle this relies on the fact that Bohr sets are approximately closed under *addition by elements of much smaller Bohr sets*. The properties we need are collected in Proposition 2·2 and Corollary 2·4, and may be obtained without relying on the regular neighbourhoods of Bourgain or the smoothed neighbourhoods of Tao (for reference, see [**3**, **5**, **10**]).

In this proposition we look at the interplay between a "coarse" partition $\mathfrak{B}(S, \delta)$, a "fine" partition $\mathfrak{B}(S', \delta')$, and an "intermediate" Bohr set $B(S, \rho)$. As long as ρ is a sufficiently small with respect to δ and a sufficiently large with respect to δ', we have that almost all translates of $B(S, \rho)$ lie inside a single part of $\mathfrak{B}(S, \delta)$, and almost all parts of $\mathfrak{B}(S', \delta')$ that intersect a fixed translate of $B(S, \rho)$ are entirely contained in that translate.

PROPOSITION 2·2. *Let $S \subseteq S' \subseteq \hat{G}$ and fix $\epsilon_0 > 0$. We have*:

(i) *If $\rho \leq \epsilon_0 \delta / |S|$, then for all but an $O(\epsilon_0)$-fraction of $x \in G$, the Bohr set translate $x + B(S, \rho)$ is entirely contained in a single part of the Bohr partition $\mathfrak{B}(S, \delta)$;*

(ii) *If $\delta' \leq \epsilon_0 C_{|S|, \rho} / |S|$,[2] then for all $x \in G$ and all but an $O(\epsilon_0)$-fraction of $y \in B(S, \rho)$, $x + y$ lies in a part of $\mathfrak{B}(S', \delta')$ that is entirely contained in $x + B(S, \rho)$.*

Proof. The strategy is to show that the image of elements of G under a character ξ are either evenly distributed in $\mathbb{R}/\mathbb{Z}$ or do not affect our computation. For those which are evenly distributed, a simple union bound suffices to show that most $x \in G$ are not close to the boundary of a Bohr set or Bohr part in the "direction of" any character.

We begin with the proof of Part 1. To determine which part of a Bohr partition contains $x \in G$, it suffices to determine the values of $\xi(x)$ for each $\xi \in S$. For $\xi \in S$, we consider two possibilities. If there is no $x_0 \in G$ with $0 < \|\xi(x_0)\|_{\mathbb{R}/\mathbb{Z}} < \rho$, then adding any element of $B(S, \rho)$ to any $x \in G$ will not change the value of $\xi(x)$, and so we may ignore such ξ.

Otherwise, there exists $x_0 \in G$ with $0 < \|\xi(x_0)\|_{\mathbb{R}/\mathbb{Z}} < \rho$. In this case, since the map $x \mapsto x + x_0$ is measure-preserving, the sets

$$\{x \in G : \xi(x) \in [(k-1)\xi(x_0), k\xi(x_0)]\}$$

are of equal measure. A union of $\lceil 1/(\xi(x_0)) \rceil = \Theta(1/\xi(x_0))$ of these sets cover G, and so each interval has measure $\Theta(1/(\xi(x_0)))$. By translation, for any interval $I \subset \mathbb{R}/\mathbb{Z}$ with length $|I| \geq \xi(x_0)$, the set $\{x \in G : \xi(x) \in I\}$ has measure $\Theta(|I|)$. Thus, the set

$$S_{\delta, \xi} := \{x \in G : \|\xi(x) - k\delta\|_{\mathbb{R}/\mathbb{Z}} \leq \rho \text{ for some } k \in \mathbb{Z}\}$$

is a union of $O(1/\delta)$ preimages under ξ of intervals of measure $2\rho \geq \xi(x_0)$, and so it has measure

$$\mu(S_{\delta, \xi}) \lesssim \rho/\delta \lesssim \epsilon_0/|S|.$$

For any $x \notin S_{\delta, \xi}$, by triangle inequality, adding any $y \in B$ cannot change the value of the largest multiple of δ less than $\xi(x)$, and summing this up over all $\xi \in S$ gives a subset of measure $O(\epsilon_0)$ which contains all the elements of x that are bad for some ξ, which completes Part 1.

Part 2 proceeds in a similar manner. For any $y \in G$ lying in some part of $p \in \mathfrak{B}'$ we know that $p \subseteq y + B(S', \delta') \subseteq y + B(S, \delta')$. It therefore suffices to show that for all x and all but an ϵ_0-fraction of $y \in x + B(S, \rho)$, we have $y + B(S, \delta') \subset x + B(S, \rho)$. By translation we may assume $x = 0$.

Let $\xi \in S$. If there is no $x_0 \in G$ with $0 < \|\xi(x_0)\|_{\mathbb{R}/\mathbb{Z}} < \rho'$, then adding any element of $B(S', \rho')$ to any $x \in G$ will not change the value of $\xi(x)$, and so we may ignore such ξ.

Otherwise, there exists $x_0 \in G$ with $0 < \|\xi(x_0)\|_{\mathbb{R}/\mathbb{Z}} < \rho'$. In this case, since the map $x \mapsto x + x_0$ is measure-preserving, the sets

$$\{x \in G : \xi(x) \in [(k-1)\xi(x_0), k\xi(x_0)_o)\}$$

[2]This is the $C_{|S|, \rho}$ from Fact 2·1–we are simply requiring δ' to be smaller than some constant depending on $|S|, \rho, \epsilon_0$.

are of equal measure, and so the exceptional set

$$E_\xi := \{x \in G : \|\xi(x) - \rho\|_{\mathbb{R}/\mathbb{Z}} \le \rho'\} \cup \{x \in G : \|\xi(x) + \rho\|_{\mathbb{R}/\mathbb{Z}} \le \rho'\}$$

has measure bounded by $O(\rho') \lesssim \epsilon/|S|C_{|S|,\rho}$.

By the triangle inequality, we have $x + B(S, \rho') \subseteq B(S, \rho)$ as long as x is not contained in $\bigcup_{\xi \in S} E_\xi$. A simple union bound tells us that this set has size $\lesssim \epsilon_0/|B(S, \rho)|$, as desired.

Reproducing the argument for the second half of this theorem when $B(S, \rho)$ is replaced by an arbitrary translate of an arbitrary part $p \in \mathfrak{B}(S, \rho)$, we can nearly obtain the same conclusion. However, Bohr parts may have wildly varying size. If we replace $C_{|S|,\rho}$ by $\epsilon_0/|\mathfrak{B}(S, \rho)| = \epsilon_0/\rho^{-|S|}$, then the conclusion will hold for all Bohr parts with size at least an ϵ_0-fraction of the average size of a Bohr part, which will be plenty.

This proposition and observation allow us to make the following useful decompositions.

COROLLARY 2·3. *Fix ϵ_0, $|S|$, ρ, and let B be either the Bohr set $B(S, \rho)$ or a part of the Bohr partition $\mathfrak{B}(S, \rho)$ with size $\mu(B) \ge \epsilon_0/|\mathfrak{B}(S, \rho)|$. There exists some C depending only on ϵ_0, $|S|$, ρ such that for any $z_0 \in G$, the set*

$$B_{z_0} := \{(x, y) \in G \times G : x + y + z_0 \in B\}$$

can be expressed as the disjoint union of at most C boxes and a remainder of measure at most $\epsilon_0 \mu(B)$.

Proof. Let us temporarily fix y. We will use a fine partition $\mathfrak{B}' = \mathfrak{B}(S, \rho')$; the boxes of $\mathfrak{B}' \times \mathfrak{B}'$ should mostly cover our set. By Proposition 2·2 and the subsequent comments, as long as ρ' is sufficiently small in terms of ϵ_0, $|S|$, and ρ, for all but an $\epsilon_0/2$-fraction of $x \in y + z_0 + B(S, \rho')$, we have that the part of $\mathfrak{B}'$ containing x lies entirely within $y + z_0 + B$. Varying y, this statement holds for the x-coordinate of all but an ϵ_0-fraction of pairs (x, y) in B_{z_0}. We can repeat the same argument for the y-coordinate. Combining these together, for all but am ϵ_0-fraction of (x, y) in B_{z_0}, we have that the box of $\mathfrak{B}' \times \mathfrak{B}'$ which contains (x, y) is itself fully contained within B_{z_0}. Consequently, B_{z_0} may be partitioned into a union of at most $C := |\mathfrak{B}'|^2$ boxes and an exceptional set of measure at most ϵ_0. Since $|\mathfrak{B}'|$ is bounded in terms of ϵ_0, $|S|$, and ρ, this completes the proof of the corollary.

Integrating the pointwise statements of Proposition 2·2, we can obtain a second useful corollary.

COROLLARY 2·4. *Fix $\epsilon_0, \delta, \delta', \rho > 0$ and $S \subseteq S' \subseteq \hat{G}$. We let $\mathfrak{B} = \mathfrak{B}(S, \delta)$, $\mathfrak{B}' = \mathfrak{B}(S', \delta')$, and $B = B(S, \rho)$. Furthermore, assume*

$$\rho \le \frac{\epsilon_0^2 \delta}{|S|} \quad and \quad \delta' \le \frac{\epsilon_0 C_{|S|,\rho}}{|S|}.$$

Then for any $f : G \to [0, 1]$ we have:

$$\left\| f|_{\mathfrak{B}} - \mu_B * f|_{\mathfrak{B}} \right\|_{L^2} \lesssim \epsilon_0. \tag{2·1}$$

$$\left\| \mu_B * f - \mu_B * f|_{\mathfrak{B}'} \right\|_{L^2} \lesssim \epsilon_0. \tag{2·2}$$

Intuitively, (2·1) says that a function which is constant on a coarse Bohr partition is approximately constant under convolution with a small Bohr set, and (2·2) states that

convolving a function with a Bohr set is approximately the same as first projecting onto a much finer Bohr partition, and then performing the convolution.

Proof. By the first half of Proposition 2·2, the set

$$\{x : f|_{\mathfrak{B}}(x) \neq f|_{\mathfrak{B}} * \mu_{B(S,\rho)}(x)\}$$

has measure bounded by ϵ_0^2. As the difference of two functions with range in $[0, 1]$, we have

$$|f|_{\mathfrak{B}}(x) - f|_{\mathfrak{B}} * \mu_B(x)|^2 \leq 1.$$

Since this function is nonzero on a set of measure at most ϵ_0^2, (2·1) follows immediately.

To show (2·2) we apply the second half of Proposition 2·2. For any $x \in G$, this allows us to partition $x + B(S, \rho)$ into the union of some Bohr parts $b \in \mathfrak{B}$ and an exceptional set E with $\mu(E) \leq \epsilon_0^2 \mu(B)$. Observing that the integral of f equals the integral of $f|_{\mathfrak{B}}$ on such a part b, we obtain:

$$
\begin{aligned}
\mu_B * f - \mu_B * f|_{\mathfrak{B}'}(x) &= \int_{y \in G} (f - f|_{\mathfrak{B}'})(x + y) \, \mu_{B(S,\rho)}(-y) \, dy \\
&= \frac{1}{\mu(B)} \int_{y' \in x + B(S,\rho)} (f - f|_{\mathfrak{B}'})(y') \, dy' \\
&= \frac{1}{\mu(B)} \int_{y' \in E} (f - f|_{\mathfrak{B}'})(y') \, dy'.
\end{aligned}
$$

We take absolute values. The integrand has absolute value bounded by 1, and is supported on a set of measure at most $\epsilon_0 \mu(B)$. We deduce:

$$\left\| \mu_B * f - \mu_B * f|_{\mathfrak{B}'} \right\|_{L^2} \leq \left\| \mu_B * f - \mu_B * f|_{\mathfrak{B}'} \right\|_{L^\infty} \leq \epsilon_0.$$

3. *Regularity lemma*

We will require two types of regularity lemmas. The first is one-dimensional and allows us to decompose a function, or a set of functions, into three parts: one that is constant on a Bohr partition, one that is small in L^1, and one that is Fourier uniform. The second type of regularity is standard two-dimensional strong regularity for graphs or graphons.

LEMMA 3·1. *Fix ϵ, m, and $F : \mathbb{R} \to \mathbb{R}$, a rapidly growing function whose choice may depend on ϵ and m. Then there exist constants D, $R > 0$ such that the following holds. For every set $\mathcal{I}$ of functions $I : G \to [0, 1]$ with cardinality $|\mathcal{I}| = m$, there exists a Bohr partition $\mathfrak{B} = \mathfrak{B}(S, \rho)$ with $|S| < D$, $\rho > R$, and a decomposition*

$$I = I_0 + I_1 + I_2$$

for each $I \in \mathcal{I}$, such that:

$$I_0 = I|_{\mathfrak{B}}, \qquad \|I_1\|_{L^2} \lesssim \frac{1}{F(1)}, \quad \text{and} \quad \left\| \widehat{I_2 \cdot \mathbb{1}_b} \right\|_{\ell^\infty} \lesssim \frac{1}{F(\delta_i^{-1} |\mathfrak{B}|)} \text{ for all } b \in \mathfrak{B}.$$

The proof of this lemma will occupy the remainder of this subsection. For this lemma we use a procedure in which we will be constructing a sequence of Bohr sets $B(S_i, \rho_i)$. Each Bohr set will be accompanied by a Bohr partition $\mathfrak{B}(S_i, \delta_i)$ with δ_i substantially smaller than ρ_i, but by a bounded amount. At each successive refinement, we regularize

an increasingly large family of functions $\mathcal{F}_i$ with respect to the previous Bohr set. The procedure is as follows:

(i)   initialise $S_0 = \emptyset$, $\rho_0 = 1$;

(ii)   set $P_i = \mathfrak{B}(S_i, \delta_i)$, where $1/\delta_i \geq F(1/\rho_i)$ is chosen to be an integer and, for $i \geq 1$, a multiple of $1/\delta_{i-1}$;

(iii)   set $\mathcal{F}_i$ to be the set of pointwise products of functions $I \cdot \mathbb{1}_p$, for all $I \in \mathcal{I}$ and $p \in P_i$;

(iv)   set

$$S_{i+1} = S_i \ \cup \ \{\xi \in \hat{G} : \hat{f}(\xi) \geq 1/F(|\mathcal{F}_i|/\delta_i) \ \text{ for some } f \in \mathcal{F}_i\};$$

(v)   set $\rho_{i+1} = 1/F(|S_{i+1}|/\delta_i)$, and $B_{i+1} = B(S_{i+1}, \rho_{i+1})$;

(vi)   if $\left\| I|_{P_{i+1}} - I|_{P_i} \right\|_{L^2} > 1/F(1)$, then increment i to $i+1$, and return to step 1.

Since each P_{i+1} is a refinement of P_i, we see that $I|_{P_i} - I|_{P_{i-1}}$ is constant on parts of P_i, whereas $I|_{P_{i+1}} - I|_{P_i}$ has integral 0 on such boxes. We obtain the following orthogonality:

$$\left\langle I|_{P_{i+1}} - I|_{P_i}, \, I|_{P_i} - I|_{P_{i-1}} \right\rangle = 0,$$

and so we have the following telescoping sum:

$$\sum_{i=1}^{t} \left\| I|_{P_{i+1}} - I|_{P_i} \right\|_{L^2}^2 = \left\| I|_{P_{t+1}} - I|_{P_1} \right\|_{L^2}^2 \leq 1. \tag{3·1}$$

Consequently there must be some $i \leq mF(1)^2$ (which in turn is bounded in terms of ϵ, m) for which every $I \in \mathcal{I}$ satisfies

$$\left\| I|_{P_{i+1}} - I|_{P_i} \right\|_{L^2} \leq \frac{1}{F(1)}. \tag{3·2}$$

Thus the procedure terminates at such a step i. We now decompose each $I \in \mathcal{I}$:

$$I = I_0 + I_1 + I_2,$$

where

$$I_0 = I|_{P_i},$$

$$I_1 = I * \mu_{B_{i+1}} - I|_{P_i},$$

$$\text{and} \qquad I_2 = I - I * \mu_{B_{i+1}}.$$

We begin by showing that each restriction of f_2 to a part $p \in P_i$ has small Fourier coefficients.

LEMMA 3·2. *We have*

$$\left\| \widehat{I_2 \cdot \mathbb{1}_p} \right\|_{\ell^\infty} \lesssim \frac{1}{F(\rho_i^{-1}|\mathcal{F}_i|)}, \tag{3·3}$$

for every $I \in \mathcal{I}$ and $p \in P_i$.

That is to say, each pointwise product $I_2 \cdot \mathbb{1}_p$ has Fourier coefficients that are arbitrarily small in terms of m, $|P_i|$, ϵ, δ_i.

Proof. We expand

$$\widehat{I_2 \cdot \mathbb{1}_p}(\xi) = \widehat{I \cdot \mathbb{1}_p}(\xi)\left(1 - \widehat{\mu}_{B_{i+1}}(\xi)\right).$$

Noting that $I \cdot \mathbb{1}_p \in \mathcal{F}_i$ for $p \in P_i$, we see that if $\xi \notin B(S_{i+1}, \rho_{i+1})$ we necessarily have

$$\left|\widehat{I \cdot \mathbb{1}_p}(\xi)\right| \le \delta_i = 1/F(|\mathcal{F}_i|/\delta_i). \tag{3.4}$$

We can bound $|\widehat{\mu}_{B_{i+1}}(\xi)|$ by 1 as $\mu_{B_{i+1}}$ is defined to have total mass 1. Consequently, we bound $|1 - \widehat{\mu}_{B_{i+1}}(\xi)|$ by 2, and obtain the claimed inequality in this case.

Otherwise, we have $\xi \in S_{i+1}$. We begin by noting that, trivially, $\mu_{B_{i+1}}$ is supported on $B_{i+1} = B(S_{i+1}, \rho_{i+1})$. For all x in this support, by definition $\xi(x) \le \rho_{i+1} \le 1/F(m|P_i|/\epsilon\delta_i)$. Consequently we have $\exp(2\pi i \xi(x)) = 1 - O(1/F(m|P_i|/\epsilon\delta_i))$. Since the Fourier coefficient $\widehat{\mu}_{B_{i+1}}(\xi)$ is an expectation of such exponentials over $x \in B_{i+1}$, it too must be $1 - O(1/F(|\mathcal{F}_i|/\delta_i))$. Bounding the Fourier coefficient $|\widehat{I \cdot \mathbb{1}_p}(\xi)|$ by 1, the claim follows in this case as well.

LEMMA 3·3. *We have*

$$\|I_1\|_{L^2} \lesssim \frac{1}{F(1)}.$$

Proof. By triangle inequality, we can write

$$
\begin{aligned}
\|I_1\|_L^2 &= \left\|I|_{P_i} - I * \mu_{B_{i+1}}\right\|_{L^2} \\
&\le \left\|I|_{P_i} - (I|_{P_i}) * \mu_{B_{i+1}}\right\|_{L^2} \\
&\quad + \left\|(I|_{P_i}) * \mu_{B_{i+1}} - (I|_{P_{i+1}}) * \mu_{B_{i+1}}\right\|_{L^2} \\
&\quad + \left\|(I|_{P_{i+1}}) * \mu_{B_{i+1}} - I * \mu_{B_{i+1}}\right\|_{L^2}.
\end{aligned}
$$

The first and third terms are bounded by Corollary 2·4; choose $\epsilon_0 = 1/F(1)$ and let F grow quickly enough so that δ_{i+1} and ρ_{i+1} are sufficiently small to satisfy the hypotheses of the Corollary. It remains to bound the second term. Applying Plancherel to (3·2), we see

$$\sum \left|\widehat{I|_{P_{i+1}}} - \widehat{I|_{P_i}}\right|^2 \le \frac{1}{F(1)^2}.$$

Since $\left\|\widehat{\mu}_{B_{i+1}}\right\|_{\ell^\infty} \le 1$, we can multiply this through and obtain

$$\sum \left|\widehat{I|_{P_{i+1}}}\widehat{\mu}_{B_{i+1}} - \widehat{I|_{P_i}}\widehat{\mu}_{B_{i+1}}\right|^2 \le \frac{1}{F(1)^2}.$$

Applying Plancherel again, we obtain

$$\left\|(I|_{P_{i+1}}) * \mu_{B_{i+1}} - (I|_{P_i}) * \mu_{B_{i+1}}\right\|_{L^2} \le \frac{1}{F(1)}.$$

3·1. *Graph regularity*

For this problem we will need to partition a group G with respect to some functions $f : G \times G \to [0, 1]$ in a way that is doubly regular. Specifically, we want a partition Π that is graph-theoretically regular in the sense that our functions f can be replaced to within a good approximation by their averages over boxes of $\Pi \times \Pi$, but we would also like the

parts of Π themselves to be pseudorandom, or Fourier uniform, as subsets of G. For a good reference for the various notions of graph regularity we use, see [8]. We use the box norm, also referred to as the cut norm, which is discussed in [7, section 4]. The relevant property we need is the following:

$$\|F\|_\square = \sup_{g,h:G\to\{0,1\}} \iint F(x,y)g(x)h(y) \lesssim \sup_{g,h:G\to[-1,1]} \iint F(x,y)g(x)h(y). \qquad (3\cdot5)$$

LEMMA 3·4. *Fix $t, \epsilon > 0$ and some quickly growing function F. Then there exists a constant N_0 such that the following holds. Let G be a compact abelian group, and let $\mathcal{F}$ be a family of functions $f : G \to [0,1]$ with cardinality $|\mathcal{F}| \le t$. Then there exist:*

   (i) *three partitions P_s, Π_s, Π of G, where $P_s = \mathfrak{B}(S_s, \rho_s)$, $|\Pi_s| =: m$, $\Pi = P_s \cap \Pi_s$, and $s, |\Pi| \le N_0$;*

 (ii) *for each $f \in \mathcal{F}$, a decomposition into $f = f_0 + f_1 + f_2$, such that $f_0 = f|_{\Pi\times\Pi}$, $\|f_1\|_{L^2} \le 1/F(1/\epsilon)$, and $\|f_2\|_\square \le 1/F(\epsilon/|\Pi|)$;*

(iii) *for each part $\pi \in \Pi_s$, a decomposition of $I = \mathbb{1}_\pi$ into $I_0 + I_1 + I_2$, such that $I_0 = I|_{P_s}$, $\|I_1\|_{L^2} \le F(\epsilon/m)$, and $\left\|\widehat{\mathbb{1}_p \cdot I_2}\right\|_{\ell^\infty} \le F(\epsilon/|\Pi|)$ for every $p \in P_s$.*

Proof. We create these partitions via the following iterative procedure:

   (i) initialise a partition $\Pi_0 = G$, and set $i = 0$;

 (ii) set P_i to be the partition guaranteed by Lemma 3·1, with $\mathcal{I} := \{\mathbb{1}_\pi\}_{\pi\in\Pi_i}$;

(iii) let $\Pi = P_i \cap \Pi_i$ be the common refinement of these partitions. Repeatedly applying weak regularity; create Π_{i+1} a refinement of Π so that $f - f|_{\Pi_i\times\Pi_i}$ has box norm less than $1/F(|\Pi|)$ for each $f \in \mathcal{F}$;[3]

(iv) if $\left\|f|_{\Pi_{i+1}\times\Pi_{i+1}} - f|_{\Pi\times\Pi}\right\|_{L^2} > 1/F(1/\epsilon)$ for any $f \in \mathcal{F}$, increment i to $i+1$ and return to step (2).

Since Π is a refinement of P_i, we have:

$$\left\|f|_{\Pi_{i+1}\times\Pi_{i+1}} - f|_{\Pi\times\Pi}\right\|_{L^2} \le \left\|f|_{\Pi_{i+1}\times\Pi_{i+1}} - f|_{\Pi_i\times\Pi_i}\right\|_{L^2}.$$

Moreover, as each $f \in F$ has bounded L^2 norm, by the orthogonality of these differences of projections (this is the same statement as $(3\cdot1)$), we may perform Step 4 only a bounded number of times in terms of t, ϵ, F. So the procedure concludes at some step s, at which point for each $f \in \mathcal{F}$, the decomposition

$$f = f_0 + f_1 + f_2,$$

satisfies the conclusions of the theorem, where

$$f_0 = f|_{\Pi\times\Pi}$$
$$f_1 = f|_{\Pi_{s+1}\times\Pi_{s+1}} - f|_{\Pi\times\Pi}$$
$$f_2 = f - f|_{\Pi_{s+1}\times\Pi_{s+1}}.$$

---

[3]For reference, [7, lemmas 3.1 and 3.2] do essentially this. The argument is standard: Initialise $\Pi_{i+1} = \Pi_i$. Then if there is a box $I_1 \times I_2$ on which $\int_{I_1\times I_2}(f - f|_{\Pi_{i+1}^2}) \ge 1/F(|\Pi|)$ for some $f \in \mathcal{F}$, refine Π_{i+1} by intersecting with I_1, I_2. A quick energy increment calculation shows that each f can only force us to refine Π_{i+1} a bounded number of times, after which the construction is complete.

4. *Counting*

We now specialize to the corners problem specifically, in which we are given a subset of $G \times G$ with density α and want to find corners in this set. It will help to use the following symmetric formulation of this problem, in which we embed our set into the hyperplane $P = \{(x, y, z) \in G \times G \times G : x + y + z = 0\}$ by sending $(x, y) \mapsto (x, y, -x - y)$. Under this map, corners are equivalent to triples of points $(x, y, -x - y)$, $(x, -x - z, z)$, $(-y - z, y, z)$, and the difference d equals $-x - y - z$. Let $A \subset P$ have density $\mu(A) = \alpha$, and let $f : G \times G \to \mathbb{R}$ be the indicator function of the projection of A:

$$f(x, y) = \mathbb{1}_A(x, y, -x - y).$$

Define g and h similarly for the projections onto the (x, z) and (y, z)-planes, respectively, and apply Lemma 3·4, regularizing with respect to the set of three functions $\{f, g, h\} =: \mathcal{F}$.

We have now regularised our set with respect to an outer Bohr partition P_s, and an inner uniform partition Π. In the case of $\mathbb{F}_2^n$, Mandache's outer partition that is the analogue of our P_s is given by the cosets of a subspace [**9**]. He then counts the number of corners with common difference lying in that subspace. This is convenient for him as any corner with difference lying in a subspace has all three of its points lying in a single part of P_s^3, and so he may restrict to individual sections of the hyperplane cut out by the boxes of P_s^3. This method relies on the fact that a coset of a subspace is closed under addition by elements of that subspace. Our analogy is the content of Proposition 2·2, in that parts of a Bohr partition are approximately closed under addition by an element of a much smaller Bohr set. Therefore, having regularised with respect to the Bohr partition $P_s = \mathfrak{B}(S_s, \delta_s)$, we now count corners with difference lying in a much smaller set $B(S_s, \rho_s')$. Consequently, the vast majority of all corners we count have all three vertices lying in the same outer box. Here, ρ_s' is an intermediate parameter that should be made sufficiently small with respect to our "large" parameters $\epsilon, m, |P_s|$. Anything assumed to be sufficiently small in terms of these three is also assumed to be sufficiently small in terms of ρ_i'. The set $B(S_s, \rho_s')$ should be thought of as lying between $B(S_s, \rho_s)$ and P_{s+1}, in terms of scale. Define

$$\nu := \mu_{B(S_s, \rho_s')}.$$

Our goal is to count count corners in A with difference weighted by ν. This weighted corner count is given by the integral

$$\int f(x, y) g(x, z) h(y, z) \nu(-x - y - z). \qquad (*)$$

Let $C, D, E \in P_s$ and let $V = C \times D \times E$. We call such V "outer boxes." The partition Π refines each part in P_s into at most $m = |\Pi_s|$ parts; say C is refined into $\{C_1, \ldots, C_m\}$ and similarly for D, E. Then $C \times D \times E$ is refined into m^3 "inner boxes" of the form $C_i \times D_j \times E_k$. We begin by immediately applying regularity to approximate the corner count in A by averages over inner boxes in Π^3.

LEMMA 4·1. *The corner count* $(*)$ *may be approximated as follows*:

$$(*) = O(\epsilon) + \sum_{\substack{C \times D \times E \in P_s^3 \\ i, j, k \in m^3}} f_0(C_i, D_j) g_0(C_i, E_k) h_0(D_j, E_k)$$

$$\times \int \mathbb{1}_{C_i}(x) \mathbb{1}_{D_j}(x) \mathbb{1}_{E_k}(x) \nu(-x - y - z).$$

Proof. We break each occurrence of f, g, h in (∗) into 3 parts by writing $f = f_0 + f_1 + f_2$ (similarly for g and h). This breaks up the integral into 27 terms.

Let's look at contributions of various terms to this integral. A term that contains f_1 can be bounded by taking absolute values and bounding the g, h terms by 1:

$$\int f_1(x, y) g_a(x, z) h_b(y, z) \nu(-x - y - z) \le \int |f_1(x, y)| \nu(-x - y - z).$$

Integrating over z eliminates the ν term and we are left with the L^1-norm of f, which is bounded by ϵ. Thus, such terms contribute $O(\epsilon)$ to the integral.

For terms that contain f_2, we evaluate this integral by first fixing z. We are using the box norm, so it will be convenient to approximate ν by a union of boxes, which is precisely the content of Corollary 2·3.

Choosing $\epsilon_0 = \epsilon$, we obtain an approximation of ν by boxes which differs from the original on a set of measure at most $\epsilon \cdot \mu(B(S_s, \rho_s'))$. Since the value of $|fgh\nu|$ is bounded by $1/\mu(B(S_s, \rho_s'))$, this part of the integral contributes at most ϵ. On the remainder, we have a contribution

$$\int f_2(x, y) g_a(x, z) h_b(y, z) \cdot \frac{1}{\mu(B(S_s, \rho_s'))} \, dx dy \tag{4·1}$$

integrated over a collection of at most $C(\epsilon, |S_i|, \rho_i')$ boxes. For fixed z, we can bound the integral (4·1) over any box by applying (3·5). By assumption the box norm of f_2 is sufficiently small in terms of C and $\mu(B(S_s, \rho_s'))$ so that the sum of these integrals over all boxes in our approximation of ν can be made to be $O(\epsilon)$. Finally, integrating this $O(\epsilon)$ contribution over all z, we conclude that the contribution from the f_2 term is also $O(\epsilon)$.

Consequently, up to an $O(\epsilon)$ error, the number of corners in A is given by the f_0, g_0, h_0 term, which is precisely the expression claimed in the lemma.

It may be worthwhile to provide an outline of the rest of the proof at this point. Having now expressed the corner count in terms of a function on inner boxes, we will group these terms by their outer box. The contributions from each outer box (except a small exceptional set) can be bounded from below by $\mu(V \cap P) \cdot T(\phi_V)$, where T is the functional defining Mandache's variational problem (appearing, for example, in (1·1)), and ϕ_V is some function of three independent random variables that has expectation within $O(\epsilon)$ of $\alpha(V) = \mu(V \cap A)/\mu(V \cap P)$. Consequently, the contribution from each outer box V will be at least $\mu(V \cap P)m(\alpha(V) + O(\epsilon))$, which is at least $m'(\alpha) + O(\epsilon)$ by the fact that m is Lipschitz [9], the pointwise bound $m' \le m$, and the convexity of m'.

We will begin the next section by defining the function ϕ for each V and evaluating $\mathbb{E}[\phi]$ and $T(\phi)$, and conclude by showing that $T(\phi)$ is indeed a lower bound for the corner count derived in Lemma 4·1.

5. *Reduction to variational problem*

We perform the reduction described in the previous section. This follows generally the strategy in [9, section 3.3], although some counts which are very easy to compute in the finite field case become more involved in the general setting (notably, Lemmas 5·4 and 5·5 may each be replaced by a single line of computation or less, in the finite field setting).

Fix $V = C \times D \times E \in P_s^3$, and let Π refine V into m^3 inner boxes of the form $C_i \times D_j \times E_k$. Let X be a random $C_i \subset C$, with probabilities given by

$$\mathbb{P}(X = C_i) = \frac{\mu(C_i)}{\mu(C)} =: \delta_{C_i}.$$

Similarly define Y, Z to be random D_j and E_k.

We will define functions ϕ'_V and ϕ_V so that $T(\phi_V)$ will be a lower bound for the corner count within box V. When the choice of V is clear, we drop the subscript and use ϕ', ϕ.

Having fixed C, D, and E, let $\phi' : \{C_i\} \times \{D_j\} \times \{E_k\} \to \mathbb{R}$ be defined as follows:

$$\phi'(C_i, D_j, E_k) = \frac{1}{\delta_{C_i}\delta_{D_j}\delta_{E_k}} \cdot \frac{\mu(A \cap C_i \times D_j \times E_k)}{\mu(P \cap C \times D \times E)}.$$

For now, we note:

$$\mathbb{E}[\phi'] = \frac{\mu(V \cap A)}{\mu(V \cap P)}.$$

The average of these values of $\mathbb{E}[\phi']$ over all boxes $V \in P_s^3$, weighted by $\mu(V \cap P)$, equals $\mu(A) = \alpha$. Indeed, as we will only ever consider the set of $V \in P_s^3$ as weighted by $\mu(V \cap P)$, we will sometimes make this implicit when referring to small fractions of the set: when we say a collection of outer boxes $X \subset P_s^3$ is at most an ϵ-fraction of all outer boxes, we mean

$$\sum_{V \in X} \mu(V \cap P) \leq \epsilon.$$

This is often quite different than the measure of X as a subset of G^3. Similarly, when taking the expectation of some function over all outer boxes V, we will always do so with respect to this measure induced by the hyperplane.

The desired minimisation problem requires that ϕ has range in $[0, 1]$, whereas our ϕ' might not; we will fix this, along with some similar normalisation problems with ϕ', as follows. Define:

$$\phi(C_i, D_j, E_k) = \begin{cases} 0 & \text{if } \min(\delta_{C_i}, \delta_{D_j}, \delta_{E_k}) < \epsilon^2/m, \\ \min(\phi', 1) & \text{otherwise.} \end{cases}$$

We show this does not affect our expectation by much. To that end, we begin with a lemma:

LEMMA 5·1. *For all but an $O(\epsilon)$-fraction of boxes $V \in P_s^3$, we have $\mathbb{E}(\phi' - \phi) = O(\epsilon)$.*

Proof. We have

$$\phi'(C_i, D_j, E_k) = \frac{1}{\delta_{C_i}\delta_{D_j}\delta_{E_k}} \cdot \frac{\mu(A \cap C_i \times D_j \times E_k)}{\mu(P \cap C \times D \times E)}$$

$$\leq \frac{1}{\delta_{C_i}\delta_{D_j}\delta_{E_k}} \cdot \frac{\mu(P \cap C_i \times D_j \times E_k)}{\mu(P \cap C \times D \times E)}.$$

Let's evaluate $\mu(P \cap C_i \times D_j \times E_k)$. The set C_i is the intersection of the parts $C \in P_s$ and $p_i \in \Pi_s$. For consistency of notation, write $I = \mathbb{1}_{p_i}$, $J = \mathbb{1}_{p_j}$, $K = \mathbb{1}_{p_k}$. Consequently we can

write $\mathbb{1}_{C_i} = I \cdot \mathbb{1}_C$, and similarly for D_j and E_k. Thus we want to evaluate

$$\frac{1}{\delta_{C_i}\delta_{D_j}\delta_{E_k}\mu(V \cap P)} \int_{x,y} I\mathbb{1}_C(x) J\mathbb{1}_E(y) K\mathbb{1}_E(-x-y).$$

Ideally, we would show that this quantity cannot be much larger than 1.

To begin, break up I, J, K as described in the regularity section. We can write $I = I_0 + I_1 + I_2$, where these functions satisfy the conclusion of Lemma 3·1. This breaks the integral into 27 terms.

We first bound terms that contain I_1, J_1, or K_1; without loss of generality, assume the term contains I_1. Bounding $|J_a|, |K_b|$ by 1, this term contributes

$$\frac{1}{\delta_{C_i}\delta_{D_j}\delta_{E_k}\mu(V \cap P)} \int_{V\cap P} |I_1|.$$

Now by the L^2 bound (which also bounds L^1), we have

$$\mathbb{E}_{V\in P_s^3}\left[\frac{1}{\mu(V \cap P)} \int_{V\cap P} |I_1|\right] = \|I_1\|_{L^1} \leq F(\epsilon/m) \leq (\epsilon/m)^{100}.$$

As a consequence, in all but an ϵ/m fraction of outer boxes V, we have

$$\frac{1}{\mu(V \cap P)} \int_{V\cap P} |I_1| \leq (\epsilon/m)^{99}.$$

We will call such an outer box *exceptional* with respect to p_i. There are $3m$ choices of I, J, K, for a total of $O(\epsilon)$ exceptional outer boxes. In the rest, the I_1, J_1, K_1 terms always contribute less than

$$\frac{(\epsilon/m)^{99}}{\delta_{C_i}\delta_{D_j}\delta_{E_k}}. \tag{5·1}$$

For terms that contain I_2 (or equivalently J_2 or K_2), we express our integral in terms of Fourier coefficients:

$$\int_{x,y} I_2\mathbb{1}_C(x) J_a\mathbb{1}_D(y) P_b\mathbb{1}_E(-x-y) = \sum_\xi \hat{I}_2 * \hat{\mathbb{1}}_C(\xi)\hat{J}_a * \hat{\mathbb{1}}_D(\xi)\hat{K}_b * \hat{\mathbb{1}}_E(\xi).$$

By our regularity assumptions we may assume the leftmost term is bounded in magnitude by some small ϵ_2, so this sum is bounded by

$$\epsilon_2 \sum_\xi |\hat{J}_a * \hat{\mathbb{1}}_D(\xi)||\hat{K}_b * \hat{\mathbb{1}}_E(\xi)|.$$

By Cauchy–Schwarz, this in turn is at most

$$\epsilon_2 \left\|\hat{J}_a * \hat{\mathbb{1}}_D\right\|_{\ell^2} \left\|\hat{K}_b * \hat{\mathbb{1}}_E\right\|_{\ell^2} = \epsilon_2 \|J_a\mathbb{1}_D\|_{L^2} \|K_b\mathbb{1}_E\|_{L^2} \leq \epsilon_2.$$

So terms of this form contribute an error on the order of $\epsilon_2/\mu(V \cap P)$, so we need to make sure $\mu(V \cap P)$ is not too small. This can be achieved easily; consider the boxes in $V \in P_s^3$ such that $\mu(V \cap P) \leq \epsilon/|P_s|^3$. Summing over all such boxes, the total fraction of P contained in any of these small outer boxes is at most ϵ. Therefore all but an ϵ-fraction of V

have $\mu(V \cap P) \geq \epsilon/|P_s|^3$. Returning to our computation, we can take

$$\left\| \widehat{I_2 \cdot \mathbb{1}_C} \right\|_{\ell^\infty} \leq \epsilon_2 \leq \frac{(\epsilon/m)^{100}}{|P_s|^3}.$$

For all of the outer boxes that are not too small, we then get a contribution from I_2 terms of

$$O\left(\frac{(\epsilon/m)^{99}}{\delta_{C_i}\delta_{D_j}\delta_{E_k}} \right).$$

The only term left contains I_0, J_0, K_0 and is simply equal to 1, as (for example) I_0 is defined to be the expectation of $\mathbb{1}_{P_i}$ on C, which is precisely $\mu(C_i)/\mu(C) = \delta_{C_i}$. Putting everything together, we have

$$\phi'(C_i, D_j, E_k) \leq \frac{1}{\delta_{C_i}\delta_{D_j}\delta_{E_k}} \cdot \frac{\mu(P \cap C_i \times D_j \times E_k)}{\mu(P \cap C \times D \times E)} \leq 1 + O\left(\frac{(\epsilon/m)^{99}}{\delta_{C_i}\delta_{D_j}\delta_{E_k}} \right). \tag{5.2}$$

Summing this up, we see

$$\mathbb{E}[\phi' - \min(\phi', 1)] \leq \sum_{i,j,k} \delta_{C_i}\delta_{D_j}\delta_{E_k} \cdot O\left(\frac{(\epsilon/m)^{99}}{\delta_{C_i}\delta_{D_j}\delta_{E_k}} \right) = O(\epsilon),$$

which nearly finishes the proof. We still need to show that ignoring points with δ_{C_i}, δ_{D_j}, or δ_{E_k} much smaller than average does not affect our computation by much. Let X_i be the exceptional set of $p = (x, y, z) \in P$ with $\delta_{C_i(p)} \leq \epsilon^2/m$, where $C_i(p)$ is the C_i containing p. Summing this up over all i, we see the union of all X_i has measure at most ϵ^2, and performing the same process for the y and z coordinates gives a set of exceptional points X of size $O(\epsilon^2)$. Then for all but an ϵ-fraction of V, we have $\mu(V \cap X)/\mu(V \cap P) \leq \epsilon$. In such cases, removing all points in X reduces $\mathbb{E}[\phi']$ by an $O(\epsilon)$-fraction. As a consequence we have

$$\mathbb{E}(\phi) \geq \mathbb{E}(\phi')(1 - O(\epsilon)) - O(\epsilon).$$

Noting that $\mathbb{E}(\phi) \leq 1$, we see $\mathbb{E}(\phi') \leq 1 + O(\epsilon)$. Concluding,

$$\mathbb{E}(\phi' - \phi) = -O(\epsilon)\mathbb{E}(\phi') - O(\epsilon) = O(\epsilon).$$

COROLLARY 5·2. *On all but an ϵ fraction of $V \in P_s^3$ we have:*

$$T(\phi') = T(\phi) + O(\epsilon).$$

Proof. We expand by linearity of expectation, bounding terms like $\mathbb{E}[\phi'|X, Y]$ and $\mathbb{E}[\phi|X, Y]$ by $1 + O(\epsilon)$ (this estimate follows, e.g., from (5·3) below, which does not rely on this corollary).

$$T(\phi') = \mathbb{E}\big[\mathbb{E}(\phi'|X, Y)\mathbb{E}(\phi'|X, Z)\mathbb{E}(\phi'|Y, Z)\big]$$
$$= \mathbb{E}\big[\mathbb{E}(\phi|X, Y)\mathbb{E}(\phi|X, Z)\mathbb{E}(\phi|Y, Z)\big]$$
$$+ \mathbb{E}\big[\mathbb{E}(\phi' - \phi|X, Y)\mathbb{E}(\phi|X, Z)\mathbb{E}(\phi|Y, Z)\big]$$
$$+ \mathbb{E}\big[\mathbb{E}(\phi'|X, Y)\mathbb{E}(\phi' - \phi|X, Z)\mathbb{E}(\phi|Y, Z)\big]$$
$$+ \mathbb{E}\big[\mathbb{E}(\phi'|X, Y)\mathbb{E}(\phi'|X, Z)\mathbb{E}(\phi' - \phi|Y, Z)\big]$$

$$\leq T(\phi) + (1 + O(\epsilon))(\mathbb{E}[\mathbb{E}(\phi' - \phi | X, Y)]$$
$$+ \mathbb{E}[\mathbb{E}(\phi' - \phi | X, Z)] + \mathbb{E}[\mathbb{E}(\phi' - \phi | Y, Z)])$$
$$\leq T(\phi) + O(\epsilon).$$

As a consequence of this, we have

$$\mathbb{E}_{V \in P_s^3}\big[\mathbb{E}[T(\phi_V)]\big] = \alpha + O(\epsilon).$$

5·1. *Computing $T(\phi)$*

We define an auxiliary function $T(V)$ as follows:

$$T(V) = \sum_{i,j,k} \begin{cases} 0 & \min(\delta_{B_i}, \delta_{C_j}, \delta_{D_k}) < \epsilon^2/m, \\ \delta_{C_i}\delta_{D_j}\delta_{E_k}\mathbb{E}(\phi'|B_i, C_j)\mathbb{E}(\phi'|C_i, E_k)\mathbb{E}(\phi'|D_j, E_k) & \text{else.} \end{cases}$$

Since $T(\phi) \leq T(V)$, it suffices to show $T(V)$ gives us a lower bound on corner counts up to an additive error of $O(\epsilon)$.

LEMMA 5·3. *On all boxes $C_i \times D_j \times E_k$ that contribute a nonzero amount to $T(V)$, we have*

$$\frac{\mu(P \cap C_i \times D_j \times E_k)}{\delta_{C_i}\delta_{D_j}\delta_{E_k}\mu(P \cap C \times D \times E)} = 1 + O(\epsilon).$$

Proof. For contributing boxes, we have $\delta_{B_i}, \delta_{C_j}, \delta_{D_k} \geq \epsilon^2/m$. Plug these bounds into (5·2).

To evaluate $T(V)$ we need to evaluate expressions of the form $\mathbb{E}[\phi' \mid X = C_i, Y = D_j]$. Readers familiar with Mandache's proof may recall that this was a simple computation in $\mathbb{F}_2^n$; that is unfortunately not the case here. We perform these calculations now.

LEMMA 5·4. *For all but an ϵ-fraction of $V \in P_s^3$, we have*

$$T(V) = O(\epsilon)$$
$$+ \sum_{i,j,k} \begin{cases} 0 & \min(\delta_{C_i}, \delta_{D_j}, \delta_{E_k}) < \epsilon^2/m, \\ \delta_{C_i}\delta_{D_j}\delta_{E_k} f_0(C_i, D_j)g_0(C_i, E_k)h_0(D_j, E_k) & \text{else.} \end{cases}$$

Proof. We begin by computing

$$\mathbb{E}[\phi' \mid X = C_i, Y = D_j] = \frac{1}{\delta_{C_i}\delta_{D_j}} \cdot \frac{\mu(A \cap C_i \times D_j \times E)}{\mu(V \cap P)}$$
$$= \frac{1}{\delta_{C_i}\delta_{D_j}\mu(V \cap P)} \int I\mathbb{1}_C(x)J\mathbb{1}_D(y)f(x, y)\mathbb{1}_E(-x - y).$$

We also note that for terms contributing a nonzero amount to $T(V)$, we have

$$\frac{1}{\delta_{C_i}\delta_{D_j}} \leq \frac{m^2}{\epsilon^4},$$

and that for all but an ϵ-fraction of outer boxes V, we have

$$\frac{1}{\mu(V \cap P)} \leq \frac{|P_s|^3}{\epsilon}.$$

We break up the contribution to our integral into various pieces.

First we write f as $f_0 + f_1 + f_2$. For terms that contain f_2, we want to use the box norm bound. The function $\mathbb{1}_D$ is the indicator function of a Bohr part; as such it can be broken up into boxes on which it is constant by Corollary 2·3. This Corollary does not hold for the ϵ-fraction of boxes with $\mu(D)$ too small, so we discard those exceptional boxes. On the rest, we can write $\mathbb{1}_D$ as the union of a set with measure $\leq \epsilon_0$ and a collection of C boxes, where C is bounded in terms of $\epsilon_0, |P_s|$. We choose ϵ_0 sufficiently small in terms of $\epsilon, m, |P_s|$ so that this leftover set has measure less than $\epsilon/(\delta_{C_i}\delta_{D_j}\mu(V \cap P))$, so this part contributes at most ϵ to the integral. Since f_2 has sufficiently small box norm in terms of $\epsilon, m, |P_s|$, the contributions from the boxes sum to $O(\epsilon)$ as well, which finishes the bounds on the f_2 term.

Next we consider the f_1 term. Since we have a global bound on $\|f_1\|_{L^2}$, we want to handle this term globally as well. The contribution to $\mathbb{E}[T(V)]$ from f_1 terms is bounded by:

$$\leq \mathbb{E}_{V \in P_s^3} \sum_{i,j,k} \delta_{C_i}\delta_{D_j}\delta_{E_k} \cdot \frac{1}{\delta_{C_i}\delta_{D_j}\mu(V \cap P)} \int_{P \cap C_i \times D_j \times E} |f_1|(1 + O(\epsilon))$$

$$= \sum_{V \in P_s^3} \sum_{i,j} \int_{P \cap C_i \times D_j \times E} |f_1|(1 + O(\epsilon))$$

$$= \|f_1\|_{L^1}(1 + O(\epsilon))$$

$$= O(\epsilon^2).$$

By Markov then, on all but an ϵ-fraction of outer boxes the f_1 terms contribute $O(\epsilon)$ to $T(V)$. The only remaining terms contain all of f_0, g_0, h_0. Such a term evaluates to

$$\frac{\mu(P \cap C_i \times D_j \times E)}{\delta_{C_i}\delta_{D_j}\mu(V \cap P)} f_0(C_i, D_j) = (1 + O(\epsilon))f_0(C_i, D_j),$$

by applying Lemma 5·3, and so we have:

$$\mathbb{E}[\phi' \mid X = C_i, Y = D_j] = f_0(C_i, D_j) + O(\epsilon). \tag{5·3}$$

Combining these terms gives the desired expression.

We now know that $T(V)$, and consequently $\mathbb{E}[T(V)]$, can be approximated by a nice sum of terms involving only the averages $f_0, g_0, h_0, \delta_{C_i}, \delta_{D_j}, \delta_{E_k}$, and moreover $\mathbb{E}[T(V)]$ is within $O(\epsilon)$ of $\mathbb{E}[T(\phi)]$ and therefore lower bounded by the solution to Mandache's variational problem. It remains to show that this expression is a lower bound for the corner count derived in Lemma (4·1).

LEMMA 5·5. *For all but an ϵ-fraction of $V \in P_s^3$ and all $C_i \times D_j \times E_k$ contributing to $T(V)$, we have*

$$\int \mathbb{1}_{C_i}(x)\mathbb{1}_{D_j}(x)\mathbb{1}_{E_k}(x)\nu(-x-y-z) \geq (1 + O(\epsilon))\delta_{C_i}\delta_{D_j}\delta_{E_k}\mu(P \cap C \times D \times E).$$

Proof. Expanding products of indicator functions, the left-hand side above becomes:

$$\int I \mathbb{1}_C(x) J \mathbb{1}_C(y) K \mathbb{1}_D(z) \nu(-x - y - z).$$

Break up the $I = I_0 + I_1 + I_2$, and similarly for J, K. Recall that on contributing inner boxes in non-exceptional outer boxes, we have

$$\delta_{C_i} \delta_{D_j} \delta_{E_k} \mu(P \cap C \times D \times E) \gg_{\epsilon, m, |P_s|} 1.$$

Assume we are dealing with a term containing I_2, J_2 or K_2. Then since $\left\| \widehat{I_2 \mathbb{1}_C} \right\|_{\ell^\infty}$ may be assumed to be sufficiently small in terms of ϵ, m, $|P_s|$, this contribution may immediately be bounded by $\epsilon \delta_{C_i} \delta_{D_j} \delta_{E_k} \mu(P \cap C \times D \times E)$ via Plancherel and Cauchy–Schwarz. If we are dealing with an I_1 term, take absolute values and bound $|J_a|$, $|K_b|$ by 1, obtaining a contribution of

$$\int |I_1| \mathbb{1}_C(x) \mathbb{1}_C(y) \mathbb{1}_D(z) \nu(-x - y - z).$$

We integrate over z first: applying Corollary 2·4, we have

$$\int_z \mathbb{1}_D(z) \nu(-x - y - z) = \mathbb{1}_D * \nu(-x - y) \approx \mathbb{1}_D(-x - y).$$

In particular we may replace one for the other and incur an arbitrarily small L^2 penalty (in terms of, say, ϵ, m, $|P_s|$). Making this substitution, we now want to compute

$$\int |I_1| \mathbb{1}_C(x) \mathbb{1}_D(y) \mathbb{1}_E(-x - y),$$

which is $O(\epsilon/m)^{99} \mu(P \cap C \times D \times E)$ by (5·1). Applying our lower bounds on δ_{c_i}, δ_{D_j}, δ_{E_k}, this error is indeed $O(\epsilon) \cdot \delta_{C_i} \delta_{D_j} \delta_{E_k} \mu(P \cap C \times D \times E)$. Finally for terms that are constant on $\mathbb{1}_C$, that is just

$$\delta_i \delta_j \delta_k \int \mathbb{1}_C \mathbb{1}_D \mathbb{1}_E \nu(-x - y - z).$$

Applying Corollary 2·4 again, this integral is within $1 + O(\epsilon)$ of $\int \mathbb{1}_C(x) \mathbb{1}_D(y) \mathbb{1}_E(-x - y)$, which completes the proof.

We are now in a position to prove our main theorem.

Proof of Theorem 1·1. Putting everything together, we have

$$\int f(x, y) g(x, z) h(y, z) \nu(-x - y - z)$$

$$= O(\epsilon) + \sum_{\substack{C \times D \times E \in P_s^3 \\ i, j, k \in [m]^3}} f_0(C_i, D_j) g_0(C_i, E_k) h_0(D_j, E_k) \int \mathbb{1}_{C_i}(x) \mathbb{1}_{D_j}(x) \mathbb{1}_{E_k}(x) \nu(-x - y - z).$$

$$\geq O(\epsilon) + \sum_{\substack{C \times D \times E \in P_s^3 \\ i, j, k \in [m]^3}} f_0(C_i, D_j) g_0(C_i, E_k) h_0(D_j, E_k) \delta_i \delta_j \delta_k \, \mu(C \times D \times E \cap P)$$

$$\geq O(\epsilon) + \sum_{C \times D \times E \in P_s^3} \mu(C \times D \times E \cap P) T(V)$$

$$= O(\epsilon) + \sum_{C \times D \times E \in P_s^3} \mu(C \times D \times E \cap P) T(\phi_{C \times D \times E})$$
$$\geq O(\epsilon) + \mathbb{E}_V \big[m(\alpha(V) + O(\epsilon)) \big]$$
$$= O(\epsilon) + m'(\alpha),$$

which completes the proof.

6. *Concluding remarks*

We conclude with a proof of Corollary 1·3. To do this we simply need to include an extra character in our Bohr sets when performing the proof in $\mathbb{Z}/n\mathbb{Z}$; this strategy appears in Green's work [**5**].

Proof of Corollary 1·3. Embed $A \subseteq [n]^2$ in the natural way into $(\mathbb{Z}/n\mathbb{Z})^2$ and perform the proof to count corners in $(\mathbb{Z}/n\mathbb{Z})^2$. However, when choosing each Bohr set S_i, include (if it is not already present) the character $x \mapsto \exp(2\pi i x/n)$ as one of the frequencies. The rest of the proof proceeds unmodified, and one obtains the correct corner count in $(\mathbb{Z}/n\mathbb{Z})^2$, but some corners in $(\mathbb{Z}/n\mathbb{Z})^2$ do not pull back to corners in $[n]^2$, e.g., triples that look like $(x, y), (x + d, y), (x, y + d - n)$. Here is where the modification helps. Since we are only allowing differences in a Bohr set B which contains $x \mapsto \exp(2\pi i x/n)$ of some radius ρ, every $d \in B$ lies in the interval $[-\rho n, \rho n]$. Consequently the only corners in $(\mathbb{Z}/n\mathbb{Z})^2$ with difference in B which do not pull back to corners in $[n]^2$ must have either x or y lying in $[0, \rho n] \cup [n - \rho n, n]$. Ensuring $\rho = O(\epsilon)$, the number of such corners is $O(\epsilon n^2 |B|)$, and so deleting these bad corners we are still left with $(m'(\alpha) - O(\epsilon)) n^2 |B|$ corners in A, which is sufficient for the claim.

REFERENCES

[1] M. AJTAI and E. SZEMERÉDI. Sets of lattice points that form no squares. *Stud. Sci. Math. Hungar* **9**, 1975 (1974), 9–11.
[2] V. BERGELSON, B. HOST and B. KRA. Multiple recurrence and nilsequences. *Invent. Math.* **160**, 2 (2005), 261–303. With an appendix by Imre Rusza.
[3] J. BOURGAIN. On triples in arithmetic progression. *Geom. Funct. Anal.* **9** (1999), 968–984.
[4] J. FOX, A. SAH, M. SAWHNEY, D. STONER and Y. ZHAO. Triforce and corners. *Math. Proc. Cambridge Philos. Soc.*, 1–15.
[5] B. GREEN. A Szemerédi-type regularity lemma in abelian groups, with applications. *Geom. Funct. Anal.* **15**, 2 (2005), 340–376.
[6] B. GREEN and T. TAO. An arithmetic regularity lemma, an associated counting lemma, and applications. In *An Irregular Mind (Szemerédi is 70)* ed. I. Bárány, J. Solymosi, G. Sági. (Springer, 2010), pp. 261–334.
[7] L. LOVÁSZ and B. SZEGEDY. Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96**, 6 (2006), 933–957.
[8] L. LOVÁSZ and B. SZEGEDY. Szemerédi's lemma for the analyst. *Geom. Funct. Anal.* **17**, 1 (2007), 252–270.
[9] M. MANDACHE. A variant of the corners theorem. *arXiv preprint* arXiv:1804.03972 (2018).
[10] T. TAO. A proof of Roth's theorem. https://terrytao.wordpress.com/2014/04/24/a-proof-of-roths-theorem/ (2014). Accessed: 2019-8-21.