BOUNDS IN A POPULAR MULTIDIMENSIONAL NONLINEAR ROTH THEOREM

SARAH PELUSE, SEAN PRENDIVILLE, AND XUANCHENG SHAO

ABSTRACT. A nonlinear version of Roth's theorem states that dense sets of integers contain configurations of the form x, $x+d$, $x+d^2$. We obtain a multidimensional version of this result, which can be regarded as a first step towards effectivising those cases of the multidimensional polynomial Szemerédi theorem involving polynomials with distinct degrees. In addition, we prove an effective "popular" version of this result, showing that every dense set has some non-zero d such that the number of configurations with difference parameter d is almost optimal. Perhaps surprisingly, the quantitative dependence in this result is exponential, compared to the tower-type bounds encountered in the popular linear Roth theorem.

CONTENTS

1. INTRODUCTION

A long-standing programme in additive combinatorics concerns obtaining effective versions of the multidimensional Szemerédi theorem [FK78] and the polynomial Szemerédi theorem [BL96], as well as their common generalisation (also found in [BL96]); see for instance [Gow98, Gow22]. In this paper, we take a first step towards effectivising those cases of the multidimensional polynomial Szemerédi theorem involving polynomials of distinct degrees.

Theorem 1.1 (Density bound)**.** *If $A \subset \{1, 2, \dots, N\} \times \{1, 2, \dots, N\}$ does not contain a triple of the form*

$$(x, y),\ (x + d, y),\ (x, y + d^2) \quad \textit{with} \quad d \neq 0, \tag{1.1}$$

then

$$|A| = O\big(N^2/(\log N)^c\big)\,.$$

Here $c > 0$ is an absolute constant[1].

Simultaneous to our work, Kravitz, Kuca and Leng [KKL] have obtained a result of this type for configurations of the form (x, y), $(x + P(d), y)$, $(x, y + P(d))$, where $P \in \mathbb{Z}[d]$ has an integer root of multiplicity one. Together, these constitute the first effective cases of the

Date: September 13, 2024.
2010 *Mathematics Subject Classification.* 11B30.
[1]We expect that $c = 2^{-300}$ is permissible.

multidimensional polynomial Szemerédi theorem involving both a genuinely multidimensional configuration and a genuinely nonlinear polynomial.

We derive our density bound by taking $\varepsilon = \delta^3/2$ in the following stronger result, which proves the existence of a "popular" difference (with effective bounds).

Theorem 1.2 (Popular difference in two dimensions)**.** *Let* $A \subset \{1, 2, \dots, N\} \times \{1, 2, \dots, N\}$ *with* $|A| \geqslant \delta N^2$ *and let* $0 < \varepsilon \leqslant 1/2$*. Either* $N \leqslant \exp(\varepsilon^{-O(1)})$ *or there exists* $d \neq 0$ *such that*

$$\#\{(x, y) \in A : (x + d, y), (x, y + d^2) \in A\} \geqslant (\delta^3 - \varepsilon) N^2.$$

Theorem 1.2 is a consequence of our more general Theorem 8.1, which we also use (in §8) to derive the existence of a popular difference in the one-dimensional context:

Theorem 1.3 (Popular difference in one dimension)**.** *Let* $A \subset \{1, 2, \dots, N\}$ *with* $|A| \geqslant \delta N$ *and let* $0 < \varepsilon \leqslant 1/2$*. Either* $N \leqslant \exp(\varepsilon^{-O(1)})$ *or there exists* $d \neq 0$ *such that*

$$\#\{x \in A : x + d, x + d^2 \in A\} \geqslant (\delta^3 - \varepsilon) N.$$

We remark that the above exponential bound $N \leqslant \exp(\varepsilon^{-O(1)})$ is in sharp contrast to the tower-type bound necessary for linear three-term progressions, see [FPZ22].

As a corollary to Theorem 1.3, we obtain a new proof that subsets of $\{1, 2, \dots, N\}$ lacking the nonlinear Roth configuration have at most polylogarithmic density, which is the main result of [PP22]. As in [PP22], our argument crucially relies on the advances made in [PP19].

The main input to all of the above results is the following inverse theorem, which characterises those functions that count many configurations of the form (1.1). Importantly, the quantitative dependence in this inverse theorem is polynomial. (In our inverse theorem, it turns out to be more natural to work with the asymmetric grid[2] $[N] \times [N^2]$, rather than $[N] \times [N]$.)

Theorem 1.4 (Inverse theorem)**.** *Let* $f_0, f_1, f_2 : \mathbb{Z}^2 \to \mathbb{C}$ *be 1-bounded functions with support in* $[N] \times [N^2]$*. Suppose that*

$$\left|\sum_{x,y,d} f_0(x, y) f_1(x + d, y) f_2(x, y + d^2)\right| \geqslant \delta N^4. \tag{1.2}$$

Then[3] *either* $N \ll \delta^{-O(1)}$ *or there exist*[4] $\alpha, \beta \in \mathbb{T}$ *and* $q \ll \delta^{-O(1)}$ *with* $\|q\alpha\|_{\mathbb{T}} \ll \delta^{-O(1)}/N$ *and* $\|q\beta\|_{\mathbb{T}} \ll \delta^{-O(1)}/N^2$ *such that*

$$\sum_y \left|\sum_x f_1(x, y) e(\alpha x)\right| \gg \delta^{O(1)} N^3 \quad \text{and} \quad \sum_x \left|\sum_y f_2(x, y) e(\beta y)\right| \gg \delta^{O(1)} N^3. \tag{1.3}$$

This inverse theorem is proved in §6, where it is derived from a combination of Theorem 6.1 and Theorem 6.2. We will bootstrap this result to deduce a variant (Theorem 6.3) concerning the more general configuration $(x, y), (x + qd, y), (x, y + q^2 d^2)$ for small positive integers q, which we will use to run an energy increment argument in §7.

The analogous one-dimensional inverse theorem [PP19, PP22] has been used by Krause, Mirek and Tao [KMT22] in their breakthrough work on pointwise convergence of bilinear polynomial ergodic averages. We expect our multidimensional inverse theorem to be useful in the study of ergodic averages of the form

$$\frac{1}{N} \sum_{d=1}^{N} f(T_1^d x) g(T_2^{d^2} x),$$

where T_1 and T_2 commute.

[2] $[N] := \{1, 2, \dots, N\}$.

[3] $f \ll g$ is a shorthand for $f = O(g)$.

[4] $\mathbb{T} := \mathbb{R}/\mathbb{Z}$, $\|\alpha\|_{\mathbb{T}} := \min_{n \in \mathbb{Z}} |\alpha - n|$, $e(\alpha) := e^{2\pi i \alpha}$.

Remark 1.5 (The obstructions in our inverse theorem are necessary)**.** By removing the absolute value in (1.3), at the cost of introducing signs, one can re-phrase our inverse theorem as saying that if the counting operator for our configuration is large, as in (1.2), then the function weighting the $(x+d, y)$ term correlates with a function of the form

$$(x, y) \mapsto e(\alpha x)\phi(y),$$

where α is "major arc at scale N" and $\phi : \mathbb{Z} \to \mathbb{C}$ is 1-bounded. Similarly, the function weighting the $(x, y + d^2)$ term correlates with a function of the form

$$(x, y) \mapsto \psi(x)e(\beta y),$$

where β is major arc at scale N^2 and $\psi : \mathbb{Z} \to \mathbb{C}$ is 1-bounded.

Notice that the above formulation does not posit any structure for the 1-bounded functions ϕ and ψ. There is no hope in being able to extract further structure regarding these functions: their arbitrariness is necessary. This is exhibited in the following example.

Let $\phi : [N^2] \to \{\pm 1\}$ be a random function that takes the value 1 with probability 1/2, independently as the argument ranges over $[N^2]$. For $(x, y) \in [N] \times [N^2]$ set

$$f_0(x, y) := \phi(x)\phi(y)(-1)^{x+y}, \quad f_1(x, y) := (-1)^x\phi(y) \quad \text{and} \quad f_2(x, y) := \phi(x)(-1)^y.$$

Then

$$\sum_{x,y,d} f_0(x, y)f_1(x + d, y)f_2(x, y + d^2) = \sum_{x,y,d} 1_{[N]}(x)1_{[N^2]}(y)1_{[N]}(x + d)1_{[N^2]}(y + d^2) \gg N^4.$$

Yet for any "reasonable" notion of structured function ψ, the randomness of ϕ means that the fibre maps $y \mapsto f_1(x, y)$ and $x \mapsto f_2(x, y)$ almost surely do not correlate with ψ.

1.1. **Previous work.**

1.1.1. *Density bounds for higher-dimensional configurations.* Let us first put our density bound (Theorem 1.1) in context.

Given any two-point polynomial configuration (without congruence obstructions), it is likely that one can show that subsets of $[N]^n$ lacking this configuration have at most polylogarithmic density. One way to prove this is to generalise Sárközy's argument [Sár78] for the configuration $x, x + d^2$; see Lyall and Magyar's [LM09] treatment of $\underline{x}, \underline{x} + (d, d^2, \ldots, d^n)$ for an instance of this approach. It is also plausible that even stronger bounds could be obtained by adapting the argument of Bloom and Maynard [BM22], who have proved the current best known bounds in Sárközy's theorem.

Longer linear configurations in $\mathbb{Z}^n$ can be encoded as *matrix progressions*

$$\underline{x},\ \underline{x} + T_1\underline{d},\ \ldots\ \underline{x} + T_s\underline{d},$$

where T_i are $n \times n$ matrices with integer entries. As was demonstrated in [Pre15], provided that the T_i and their differences $T_i - T_j$ $(i \neq j)$ are all non-singular, one can extract density bounds for sets lacking these configurations by adapting the higher-order Fourier analysis techniques of Gowers [Gow01]. One may be able to synthesise this approach with that of [Pre17] in order to bound the density of sets lacking nonlinear matrix progressions of the form

$$\underline{x},\ \underline{x} + T_1(d_1^k, \ldots, d_n^k),\ \ldots\ \underline{x} + T_s(d_1^k, \ldots, d_n^k),$$

again provided that the T_i and their differences $T_i - T_j$ $(i \neq j)$ are all non-singular.

It is much harder to obtain quantitative bounds for sets lacking *singular* matrix progressions. Preceding this article and that of Kravitz-Kuca-Leng [KKL], there are only two results in this direction (for progressions of length greater than two). The first is due to Shkredov [Shk06] and obtains double-logarithmic bounds for sets lacking *any* three-point matrix progression, regardless of the singularity of the matrices and their differences. This very general result is deduced in [Pre15, Theorem B.2] from Shkredov's much more concrete result which bounds the density of sets lacking two-dimensional corners:

$$(x,y),\ (x+d,y),\ (x,y+d) \quad \text{with} \quad d \neq 0. \tag{1.4}$$

Very recently, the first author [Pel24] obtained the first reasonable bounds for sets lacking a specific four-point singular matrix progression, namely L-shapes:

$$(x,y),\ (x+d,y),\ (x,y+d),\ (x,y+2d) \quad \text{with} \quad d \neq 0. \tag{1.5}$$

Although this argument was carried out over the finite field model $\mathbb{F}_p^n \times \mathbb{F}_p^n$, those conversant with the appropriate machinery will see that adapting the argument to the integer setting is now within reach, albeit technically demanding.

We believe the arguments of this paper open up the possibility of obtaining bounds for sets lacking polynomial progressions of the form

$$\underline{x},\ \underline{x}+P_1(d)\underline{v}_1,\ \dots\ ,\ \underline{x}+P_s(d)\underline{v}_s \quad \text{with} \quad d \neq 0, \tag{1.6}$$

where $\underline{v}_1, \dots, \underline{v}_s$ are fixed integer vectors and the polynomials P_i have integer coefficients, zero constant term and (importantly) *distinct* degrees. Our configuration (1.1) corresponds to taking $P_1 = d$, $P_2 = d^2$, $\underline{v}_1 = (1,0)$ and $\underline{v}_2 = (0,1)$. For the analogous problem over the finite field $\mathbb{F}_p$, our paper corresponds to work of Han-Lacey-Yang [HLY21], and we posit that it is possible to obtain the analogue of Kuca's more general result [Kuc24] in the integers. In this paper we make heavy use of the one-dimensional machinery developed in [PP19] for $x, x+d, x+d^2$, and for (1.6) we expect that the appropriate one-dimensional machinery is available in [Pel20].

1.1.2. *Existence of a popular difference.* We now turn to the context surrounding our popular difference results (Theorem 1.2 and Theorem 1.3). To our knowledge, the first results in this direction concern three-term and four-term arithmetic progressions [BHK05]. These results were obtained via ergodic methods and as a result concern the weaker notion of upper Banach density, rather than the lower density we deal with in Theorem 1.2 and Theorem 1.3. There have since been a number of ergodic works dealing with other configurations, see for instance [Fra08, Chu11, DLMS21]. In particular, Chu, Frantzikinakis and Host [CFH11] have proved the ergodic (and hence qualitative) analogue of Theorem 1.2.

Green [Gre05b] pioneered the use of arithmetic regularity to study these problems, proving the existence of a popular difference for three-term arithmetic progressions and thereby strengthening [BHK05] from upper Banach density to lower density. Green's theorem states that for any $\varepsilon > 0$ and any $A \subset [N]$ with $|A| \geqslant \delta N$, either $N \leqslant C(\varepsilon)$ or there exists $d \neq 0$ such that

$$\#\{x \in A : x+d,\ x+2d \in A\} \geqslant (\delta^3 - \varepsilon)N. \tag{1.7}$$

Green and Tao [GT10] used higher-order Fourier analysis to obtain the analogous result for four-term arithmetic progressions. The result is false for longer progressions; see Ruzsa's appendix to [BHK05].

In Green's argument for (1.7), the constant $C(\varepsilon)$ is bounded in terms of a tower of twos of height $\varepsilon^{-O(1)}$. This tower height was reduced to $O(\log(1/\varepsilon))$ by Fox, Pham and Zhao [FP21, FP19, FPZ22], who went on to demonstrate (remarkably) that this bound is best possible.

In higher dimensions, existence of a popular common difference for certain non-singular matrix progressions of length 3 and 4 was established in [BSST22]. For corners (1.4), the ergodic version of a popular difference was studied by Chu [Chu11], with combinatorial analogues obtained by Mandache [Man21]. Both of these authors establish that the naive conjecture is not correct for the popular difference version of the corners theorem. The "correct" conjecture was emphatically addressed in [FSS+20]. In [SSZ21], linear configurations were classified according to when the naive conjecture is correct.

All of the above results concerning popular differences have either been qualitative or the bounds obtained have been tower-type. The only result we are aware of with reasonable bounds is due to Lyall and Magyar [LM13], and concerns two-point polynomial progressions in one dimension. Lyall and Magyar show that if P is a polynomial with integer coefficients and zero constant term, then for any $\varepsilon > 0$ and $A \subset [N]$ with $|A| \geqslant \delta N$, either there exists $d \in \mathbb{Z} \setminus \{0\}$ such that

$$\#\{x \in A : x + P(d) \in A\} \geqslant (\delta^2 - \varepsilon)N,$$

or $N \leqslant \exp\exp(O(\varepsilon^{-1}\log(\varepsilon^{-1})))$. We believe that by suitably adapting the methods of this paper, one should be able to replace this double exponential bound with $N \leqslant \exp\big(\varepsilon^{-O(1)}\big)$. Determining the correct bound on N seems an interesting problem, even in the simplest situation of x, $x + d^2$. In forthcoming work by the second author, the third author and Mengdi Wang, we generalize Lyall and Magyar's result to certain multi-point polynomial progressions, proving that if $P_1, \cdots, P_m$ are polynomials with integer coefficients and zero constant terms with distinct degrees, then for any $\varepsilon > 0$ and $A \subset [N]$ with $|A| \geqslant \delta N$, either there exists $d \in \mathbb{Z} \setminus \{0\}$ such that

$$\#\{x \in A : x + P_1(d), \cdots, x + P_m(d) \in A\} \geqslant (\delta^{m+1} - \varepsilon)N,$$

or $N \leqslant \exp\exp(O(\varepsilon^{-1}))$.

Acknowledgments. We thank Zach Hunter and the anonymous referee for helpful comments and suggestions. The first and third authors were supported by NSF grants DMS-2401117 and DMS-2200565, respectively.

2. Notation

2.1. **Standard conventions.** We use $\mathbb{N}$ to denote the positive integers. For real $X \geqslant 1$, write $[X] = \{1, 2, \ldots, \lfloor X \rfloor\}$. A complex-valued function is *1-bounded* if the modulus of the function does not exceed 1.

We use counting measure on $\mathbb{Z}^d$, so that for $f, g : \mathbb{Z}^d \to \mathbb{C}$ we have

$$\langle f, g\rangle := \sum_x f(x)\overline{g(x)} \qquad \text{and} \qquad \|f\|_{L^p} := \Big(\sum_x |f(x)|^p\Big)^{\frac{1}{p}}. \tag{2.1}$$

Any sum of the form $\sum_x$ is to be interpreted as a sum over $\mathbb{Z}^d$. We use Haar probability measure on $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$, so that for measurable $F : \mathbb{T}^d \to \mathbb{C}$ we have

$$\|F\|_{L^p} := \Big(\int_{\mathbb{T}^d} |F(\alpha)|^p \mathrm{d}\alpha\Big)^{\frac{1}{p}} = \Big(\int_{[0,1)^d} |F(\alpha)|^p \mathrm{d}\alpha\Big)^{\frac{1}{p}}$$

For $\alpha \in \mathbb{T}$ we write $\|\alpha\|_{\mathbb{T}}$ for the distance to the nearest integer.

For a finite set S and function $f : S \to \mathbb{C}$, denote the average of f over S by

$$\mathbb{E}_{s\in S} f(s) := \frac{1}{|S|}\sum_{s\in S} f(s).$$

Given functions $f, g : G \to \mathbb{C}$ on an additive group with measure μ_G we define their convolution by

$$f * g(x) := \int_G f(x - y)g(y)\mathrm{d}\mu_G, \tag{2.2}$$

when this makes sense. On the integers $G = \mathbb{Z}$, we take μ_G to be counting measure.

We define the Fourier transform of $f : \mathbb{Z}^d \to \mathbb{C}$ by

$$\hat{f}(\alpha) := \sum_x f(x)e(\alpha \cdot x) \qquad (\alpha \in \mathbb{T}^d), \tag{2.3}$$

again, when this makes sense. Here $e(\beta)$ stands for $e^{2\pi i\beta}$.

The *difference function* of $f : \mathbb{Z}^d \to \mathbb{C}$ with respect to $h \in \mathbb{Z}^d$ is the function $\Delta_h f : \mathbb{Z}^d \to \mathbb{C}$ given by

$$\Delta_h f(x) = f(x)\overline{f(x+h)}. \tag{2.4}$$

Iterating, we set

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

This allows us to define the *Gowers U^s-norm*

$$\|f\|_{U^s(\mathbb{Z}^d)} := \left(\sum_{x,h_1,\dots,h_s \in \mathbb{Z}^d} \Delta_{h_1,\dots,h_s} f(x) \right)^{1/2^s}. \tag{2.5}$$

When $S \subset \mathbb{Z}^d$ we define the *localised* Gowers U^s-norm

$$\|f\|_{U^s(S)} := \|f 1_S\|_{U^s}. \tag{2.6}$$

For a function f and positive-valued function g, write $f \ll g$ or $f = O(g)$ if there exists a constant C such that $|f(x)| \le Cg(x)$ for all x. We write $f = \Omega(g)$ if $f \gg g$. We sometimes opt for a more explicit approach, using C to denote a large absolute constant, and c to denote a small positive absolute constant. The values of C and c may change from line to line.

2.2. **Local conventions.** Up to normalisation, all of the above are well-used in the literature. Next we list notation specific to our paper. We have tried to minimise this in order to aid the casual reader.

For a real parameter $H \geqslant 1$, we use $\mu_H : \mathbb{Z} \to [0,1]$ to represent the following normalised Fejér kernel

$$\mu_H(h) := \frac{1}{\lfloor H \rfloor}\left(1 - \frac{|h|}{\lfloor H \rfloor}\right)_+ = \frac{(1_{[H]} * 1_{-[H]})(h)}{\lfloor H \rfloor^2}, \tag{2.7}$$

where $x_+ := \max\{x, 0\}$ is the positive part of a real number x.

Strictly speaking, "Fejér kernel" usually refers to the Fourier transform of this function, which is a function on the torus $\mathbb{T}$. We adhere to our unconventional nomenclature since most of our arguments take place in $\mathbb{Z}$. For a multidimensional vector $h \in \mathbb{Z}^d$ we write

$$\mu_H(h) = \mu_H(h_1, \dots, h_d) := \mu_H(h_1) \cdots \mu_H(h_d). \tag{2.8}$$

We observe that this is a probability measure on $\mathbb{Z}^d$ with support in the box $(-H, H)^d$.

Define a *counting operator* on the functions $f_0, f_1, f_2 : \mathbb{Z}^2 \to \mathbb{C}$ by

$$\Lambda_N(f_0, f_1, f_2) := \mathbb{E}_{x\in[N]}\mathbb{E}_{y\in[N^2]} \sum_d \mu_N(d) f_0(x,y) f_1(x+d, y) f_2(x, y+d^2). \tag{2.9}$$

When the f_i are all equal to f we simply write $\Lambda_N(f)$.

3. An outline of our argument

All statements in this section are of a heuristic and informal nature, in particular much of what we write is, strictly speaking, false. To avoid obfuscating our sketch by tracking ranges of summation, we simply write $\mathbb{E}_x f(x)$ to indicate a normalised sum, with the range of summation to be inferred from the context.

Given a set $A \subset [N]^2$ of density $\delta := |A|/N^2$, our starting point is to study the count

$$\mathbb{E}_{x,y,d} 1_A(x,y) 1_A(x+d, y) 1_A(x, y+d^2). \tag{3.1}$$

Traditionally in additive combinatorics, one compares this with the count of configurations lying in a random set of density δ, which (after normalisation) is of order δ^3. If these counts are close, then one has a profusion of configurations in the set A. If not, then we hope to show that A has some exploitable structure. In a density increment argument, we typically exploit this by passing to a substructure where the density is larger. Iterating this argument eventually yields a substructure where the count of configurations *is* close to the random count, finishing our proof.

For many multidimensional configurations, we contend that the random count δ^3 is not a useful comparison to make. One reason for this is that there are very unstructured sets that do not have the random number of configurations: taking an example from [Gre05a, §5], consider a product of random sets $B_1 \times B_2$ where each $B_i \subset [N]$ has density $\sqrt{\delta}$. In the corners problem (1.4), one is forced to deal with these very loosely structured objects - they only possess the structure of a Cartesian product. However, the inverse theorem for our configuration (Theorem 1.4) says that the sets we need to deal with have more structure: they take the form $B \times P$ where $P \subset [N]$ is a *dense* arithmetic progression (i.e., of length proportional to N and with small common difference) and $B \subset [N]$ is arbitrary. After moving to a grid with slightly larger step size, this purported "bad" set can be thought of as taking the form $B' \times [N']$, where B' has no exploitable structure.

With the above discussion in mind, and following [HLY21, Kuc24], we contend that it is more sensible to compare (3.1) with the following count, where we have essentially replaced the third indicator function $1_A(x, y + d^2)$ with its average $\mathbb{E}_{y'} 1_A(x, y')$ on vertical fibres:

$$\begin{aligned}\mathbb{E}_{x,y,d} 1_A(x,y) 1_A(x+d,y) \mathbb{E}_{d'} 1_A(x, y+d'+d^2) \\ \approx \mathbb{E}_{x,y,d,d'} 1_A(x,y) 1_A(x+d,y) 1_A(x,y+d') \\ \approx \mathbb{E}_{x,y,x',y'} 1_A(x,y) 1_A(x',y) 1_A(x,y'). \quad (3.2)\end{aligned}$$

A (non-obvious[5]) application of Hölder's inequality gives that

$$\mathbb{E}_{x,y,x',y'} 1_A(x,y) 1_A(x',y) 1_A(x,y') \geqslant \left(\mathbb{E}_{x,y} 1_A(x,y)\right)^3. \quad (3.3)$$

Hence, we have a profusion of configurations, and, in particular, the existence of a popular difference, if (3.1) is ε-close to (3.2).

If (3.1) is not ε-close to (3.2), then, on setting

$$f_2(x,y) := 1_A(x,y) - \mathbb{E}_{d'} 1_A(x, y+d'), \quad (3.4)$$

we have

$$\left|\mathbb{E}_{x,y,d} 1_A(x,y) 1_A(x+d,y) f_2(x, y+d^2)\right| \geqslant \varepsilon.$$

At this point, we apply our inverse theorem (Theorem 1.4) to deduce the existence of $\alpha \approx a/q$ with $q \leqslant \varepsilon^{-O(1)}$ such that

$$\mathbb{E}_x |\mathbb{E}_y f_2(x,y) e(\alpha y)| \geqslant \varepsilon^{O(1)}.$$

The function $y \mapsto e(\alpha y)$ is approximately invariant under small shifts of the form $y \mapsto y + q^2 d'$, so by a little averaging, there exists $N' \geqslant \varepsilon^{O(1)} N$ such that

$$\mathbb{E}_{x,y} \left|\mathbb{E}_{|d'|<N'} f_2(x, y + q^2 d')\right| \geqslant \varepsilon^{O(1)}. \quad (3.5)$$

Equation (3.5) is saying that we can convolve f_2 in the second coordinate with a long arithmetic progression of small common difference, in order to yield a function with large L^1-norm. Recall from (3.4) that f_2 is, essentially, 1_A minus the convolution of 1_A in the second coordinate with the interval $(-N, N)$. We can, therefore, use (3.5), together with approximate orthogonality properties of convolution, in order to obtain an *energy increment*:

$$\mathbb{E}_{x,y} \left|\mathbb{E}_{|d'|<N'} 1_A(x, y+q^2 d')\right|^2 \geqslant \mathbb{E}_{x,y} \left|\mathbb{E}_{|d|<N} 1_A(x, y+d)\right|^2 + \varepsilon^{O(1)}.$$

We now repeat the above argument, this time focusing on the *restricted* counting operator

$$\mathbb{E}_{x,y} \mathbb{E}_{|d|<N'} 1_A(x,y) 1_A(x+qd, y) 1_A(x, y+q^2 d^2),$$

[5]See https://mathoverflow.net/questions/189222.

which counts configurations where the difference parameter is divisible by q and ranges over a shorter interval. We compare this restricted counting operator with the restricted analogue of (3.2), namely

$$\mathbb{E}_{x,y}\mathbb{E}_{|d|,|d'|<N'}1_A(x,y)1_A(x+qd,y)1_A(x,y+q^2d').$$

Again, this comparison either yields a profusion of configurations in our set A, or alternatively we have an energy increment of the form

$$\mathbb{E}_{x,y}\left|\mathbb{E}_{|\tilde{d}|<\tilde{N}}1_A(x,y+q^2\tilde{q}^2\tilde{d})\right|^2 \geqslant \mathbb{E}_{x,y}\left|\mathbb{E}_{|d'|<N'}1_A(x,y+q^2d')\right|^2+\varepsilon^{O(1)},$$

for some $\tilde{q} \ll \varepsilon^{-O(1)}$ and some $\tilde{N} \gg \varepsilon^{O(1)}N'$.

Iterating the energy increment n times, we obtain a convolution whose energy is at least $n\varepsilon^{O(1)}$. Since the energy is bounded above by 1, the energy increment must terminate after at most $n \leqslant \varepsilon^{-O(1)}$ steps. Setting $q_1 := q$, $q_2 := q\tilde{q}$ and so on, the iteration yields some common difference q_n of size

$$q_n \leqslant \underbrace{\varepsilon^{-O(1)}\cdots\varepsilon^{-O(1)}}_{\varepsilon^{-O(1)}\text{ times}} = \exp\bigl(\varepsilon^{-O(1)}\bigr)$$

and an interval length N_n satisfying

$$N_n \geqslant \underbrace{\varepsilon^{O(1)}\cdots\varepsilon^{O(1)}}_{\varepsilon^{-O(1)}\text{ times}}N = N/\exp\bigl(\varepsilon^{-O(1)}\bigr),$$

such that we have the comparison

$$\begin{aligned}|\mathbb{E}_{x,y}\mathbb{E}_{|d|<N_n}&1_A(x,y)1_A(x+q_nd,y)1_A(x,y+q_n^2d^2)\\ &-\mathbb{E}_{x,y}\mathbb{E}_{|d|,|d'|<N_n}1_A(x,y)1_A(x+q_nd,y)1_A(x,y+q_n^2d')| \leqslant \varepsilon.\end{aligned}$$

This delivers a popular difference, since an argument using Hölder's inequality again gives something of the form

$$\mathbb{E}_{x,y}\mathbb{E}_{|d|,|d'|<N_n}1_A(x,y)1_A(x+q_nd,y)1_A(x,y+q_n^2d') \gtrapprox \delta^3.$$

3.1. **Paper organisation.** Proving our inverse theorem (Theorem 1.4) occupies §§4-6. The first part of this argument (§4) shows that our configuration is controlled by a Gowers U^5-norm in the vertical direction. This uses the PET induction scheme of Bergelson-Leibman [BL96] together with the "quantitative concatenation" machinery developed in [PP19]. In §5, we use the degree lowering technique as developed in [PP19] to reduce the U^5-norm to a U^1-norm (after passing to a long progression with small common difference). In §6, we combine the results of the previous sections in order to derive our inverse theorem. Our energy increment argument is carried out in §7 and we deduce the existence of a popular common difference in §8.

4. PET INDUCTION AND U^5-CONTROL IN THE VERTICAL DIRECTION

We begin by showing that our counting operator is controlled by a Gowers U^5-norm in the vertical direction by combining a PET induction argument with the main concatenation result from [PP19]. Those familiar with PET induction should not be surprised that the Gowers norm control we obtain is solely in the vertical direction, as all traces of the linear polynomial d are erased when applying the Cauchy–Schwarz inequality enough times to linearise d^2.

In order to establish the existence of a popular common difference (Theorem 1.2), it is convenient to work with a counting operator that incorporates a nicer weight than the sharp cut-off $d \in [-N, N]$ used in the inverse theorem stated in our introduction (Theorem 1.4).

Theorem 4.1 (U^5-control in the vertical direction)**.** *Let $f_0, f_1, g : \mathbb{Z}^2 \to \mathbb{C}$ be 1-bounded functions and let μ_N denote the probability measure*[6]

$$\mu_N := \tfrac{1}{N^2} 1_{[N]} * 1_{-[N]} = \tfrac{1}{N}\Big(1 - \tfrac{|\cdot|}{N}\Big)_+ .$$

Write

$$\Lambda_N(f_0, f_1, g) := \mathbb{E}_{x\in[N]}\mathbb{E}_{y\in[N^2]} f_0(x,y) \sum_d \mu_N(d) f_1(x+d, y) g(x, y+d^2).$$

If $|\Lambda_N(f_0, f_1, g)| \geqslant \delta$, then

$$\mathbb{E}_{x\in[N]} \big\| g_x 1_{[2N^2]} \big\|_{U^5(\mathbb{Z})}^{2^5} \gg \delta^{O(1)} \big\| 1_{[2N^2]} \big\|_{U^5(\mathbb{Z})}^{2^5},$$

where $g_x(y) := g(x,y)$ and the U^5-norm is defined via (2.5)*.*

Proof. Rather than summing x and y over intervals, it is convenient to be able to sum over $\mathbb{Z}$, as this allows us to change variables in a clean manner, without having to keep track of how this affects the range of summation. We may therefore assume that

$$\operatorname{supp}(f_0) \subset [N]\times[N^2], \quad \operatorname{supp}(f_1) \subset [-N, 2N]\times[N^2], \quad \operatorname{supp}(g) \subset [N]\times[2N^2].$$

Notice that this affects neither the hypothesis nor the conclusion of the theorem.

Following the procedure of [PP19, Lemma 3.3], we repeatedly apply the Cauchy–Schwarz inequality and a change of variables. It is convenient to renormalise by writing $\tilde{\mu}_N := N\mu_N$, to give a 1-bounded function with support contained in $[-N, N]$.

Averaging the d parameter by shifts $d \mapsto d + a$ with $a \in [N]$, and taking into account the support of $\tilde{\mu}_N$, we have that

$$\Lambda_N(f_0, f_1, g) = \frac{1}{N^4} \sum_{x,y} f_0(x,y) \sum_{|d|\leqslant N} \mathbb{E}_{a\in[N]} \tilde{\mu}_N(d+a) f_1(x+d+a, y) g(x, y+(d+a)^2).$$

We apply the Cauchy–Schwarz inequality to double the a variable in the above sum. After a further change of variables, recalling that $\Delta_{(a,0)} f_1(x,y) = f_1(x,y)\overline{f_1(x+a,y)}$ as in (2.4) and $\mu_N(a) := N^{-2}\sum_{a_1-a_2=a} 1_{[N]}(a_1)1_{[N]}(a_2)$ as in (2.7), this gives that

$$|\Lambda_N(f_0, f_1, g)| \leqslant \left(\frac{1}{N^4} \sum_{x,y} |f_0(x,y)|^2 \sum_{|d|\leqslant N} 1 \right)^{1/2}$$
$$\left(\frac{1}{N^4} \sum_{x,y,d,a} \mu_N(a) \Delta_a \tilde{\mu}_N(d) \Delta_{(a,0)} f_1(x,y) g(x-d, y+d^2) \overline{g(x-d, y+(d+a)^2)} \right)^{1/2}.$$

We now repeat the above process, averaging d by shifts $d \mapsto d+b$ with $b \in [N]$, then applying Cauchy–Schwarz to double the b variable. After further change of variables, this gives

$$|\Lambda_N(f_0, f_1, g)|^4 \ll \frac{1}{N^4} \sum_{a,b,x,y,d} \mu_N(a)\mu_N(b) \Delta_{a,b}\tilde{\mu}_N(d) g(x,y) \overline{g\big(x, y+2ad+a^2\big)}$$
$$\overline{g\big(x-b, y+2bd+b^2\big)} g\big(x-b, y+2(a+b)d+(a+b)^2\big).$$

[6]We normalise convolution on $\mathbb{Z}$ with counting measure, see (2.2). We write $x_+ := \max\{x, 0\}$ for the positive part of a real number.

We continue to iterate this procedure, this time averaging d by shifts $d \mapsto d + h_1$ with $h_1 \in [H]$, where $1 \leqslant H \leqslant N$. Doing so gives

$$|\Lambda_N(f_0, f_1, g)|^8 \ll \frac{1}{N^4} \sum_{a,b,h_1,x,y,d} \mu_N(a,b)\mu_H(h_1)\Delta_{a,b,h_1}\tilde{\mu}_N(d)\overline{\Delta_{(0,2ah_1)}g\big(x, y + a^2\big)}$$
$$\overline{\Delta_{(0,2bh_1)}g\big(x - b, y + 2(b-a)d + b^2\big)}\Delta_{(0,2(a+b)h_1)}g\big(x - b, y + 2bd + (a+b)^2\big),$$

using the notation in (2.8). Writing $g_x(y) := g(x,y)$, another iteration yields

$$|\Lambda_N(f_0, f_1, g)|^{16} \ll \frac{1}{N^4} \sum_{a,b,h_1,h_2,x,y,d} \mu_N(a,b)\mu_H(h_1,h_2)\Delta_{a,b,h_1,h_2}\tilde{\mu}_N(d)$$
$$\overline{\Delta_{2bh_1,2(b-a)h_2}g_x\big(y + b^2\big)}\Delta_{2(a+b)h_1,2bh_2}g_x\big(y + 2ad + (a+b)^2\big).$$

Introducing an additional average over $h_3 \in [N]$ and changing variables, we have

$$|\Lambda_N(f_0, f_1, g)|^{16} \ll \frac{1}{N^4} \sum_{a,b,\underline{h},x,y,d} \mu_N(a,b)\mu_H(\underline{h})\mathbb{E}_{h_3\in[H]}\Delta_{a,b,\underline{h}}\tilde{\mu}_N(d + h_3)$$
$$\overline{\Delta_{2bh_1,2(b-a)h_2}g_x\big(y - 2ad + b^2\big)}\Delta_{2(a+b)h_1,2bh_2}g_x\big(y + 2ah_3 + (a+b)^2\big)$$

We next observe that, for $1 \leqslant h_3 \leqslant H \leqslant N$, we have

$$\Delta_{a,b,\underline{h}}\tilde{\mu}_N(d + h_3) = \Delta_{a,b,\underline{h}}\tilde{\mu}_N(d) + O\big(1_{[-2N,2N]}(d)H/N\big).$$

Hence,

$$|\Lambda_N(f_0, f_1, g)|^{16} \ll \frac{H}{N} + \frac{1}{N^4}\Bigg| \sum_{a,b,\underline{h},x,y} \mu_N(a,b)\mu_H(\underline{h}) \sum_d \Delta_{a,b,\underline{h}}\tilde{\mu}_N(d)$$
$$\overline{\Delta_{2bh_1,2(b-a)h_2}g_x\big(y - 2ad + b^2\big)}\mathbb{E}_{h_3\in[H]}\Delta_{2(a+b)h_1,2bh_2}g_x\big(y + 2ah_3 + (a+b)^2\big)\Bigg|.$$

So, a final application of the Cauchy–Schwarz inequality gives

$$|\Lambda_N(f_0, f_1, g)|^{32} \ll \frac{H^2}{N^2} + \left| \frac{1}{N^3} \sum_{a,b,\underline{h},x,y} \mu_N(a,b)\mu_H(\underline{h})\Delta_{2(a+b)h_1,2bh_2,2ah_3}g_x\big(y\big) \right|.$$

It follows that if $|\Lambda_N(f_0, f_1, g)| \geqslant \delta$, then either $H \gg \delta^{16}N$, or there exists a set $S \subset [N]$ of size $|S| \gg \delta^{32}N$ such that, for each $x \in S$, we have

$$\sum_{a,b\in(-2N,2N)} \left| \sum_h \mu_H(h) \sum_y \Delta_{(a+b)h_1,\, bh_2,\, ah_3}g_x(y) \right| \gg \delta^{32}N^4.$$

For each such $x \in S$, the above gives the same conclusion as the first displayed equation in the proof of [PP19, Theorem 5.6]. We note that, in the notation of [PP19, Theorem 5.6], we are replacing N with $2N^2$, and taking $M := N$, $q := 1$ and $f := g_x$ (which is a 1-bounded function with support contained in $[2N^2]$). Applying the proof of [PP19, Theorem 5.6], for each $x \in S$ we deduce that either $N \ll 1$ or

$$\|g_x\|_{U^5}^{32} \gg \delta^{O(1)}\big\|1_{[2N^2]}\big\|_{U^5}^{32}.$$

The conclusion of our theorem being trivial in the case that $N \ll 1$, the result follows on summing over $x \in S$. □

As is exploited repeatedly in [PP19, PP22, Pel20], it is often more convenient to replace arbitrary bounded functions with dual functions, as these functions possess more structure. The next result shows that this is possible in Theorem 4.1. We eventually use this in §6.

Corollary 4.2 (Vertical U^5-control of the dual)**.** *Let $f_0, f_1, g : \mathbb{Z}^2 \to \mathbb{C}$ be 1-bounded functions and let μ_N denote the probability measure*

$$\mu_N := \tfrac{1}{N^2} 1_{[N]} * 1_{-[N]} = \tfrac{1}{N}\Big(1 - \tfrac{|\cdot|}{N}\Big)_+ .$$

Write

$$\Lambda_N(f_0, f_1, g) := \mathbb{E}_{x\in[N]}\mathbb{E}_{y\in[N^2]} \sum_d \mu_N(d) f_0(x,y) f_1(x+d,y) g(x, y+d^2).$$

If $|\Lambda_N(f_0, f_1, g)| \geqslant \delta$ then, on defining the dual

$$F_x(y) = F(x,y) := \sum_d \mu_N(d) f_0(x, y-d^2) f_1(x+d, y-d^2), \tag{4.1}$$

we have

$$\mathbb{E}_{x\in[N]} \big\| F_x 1_{[2N^2]} \big\|_{U^5(\mathbb{Z})}^{2^5} \gg \delta^{O(1)} \big\| 1_{[2N^2]} \big\|_{U^5(\mathbb{Z})}^{2^5},$$

where the U^5-norm is defined via (2.5).

Proof. We may assume that

$$\operatorname{supp}(f_0) \subset [N] \times [N^2], \quad \operatorname{supp}(f_1) \subset [-2N, 2N] \times [N^2], \quad \operatorname{supp}(g) \subset [N] \times [2N^2],$$

so that $\operatorname{supp}(F) \subset [N] \times [2N^2]$. With these assumptions on the support, we may remove the restriction that $(x,y) \in [N] \times [N^2]$ in our counting operator, instead summing over $\mathbb{Z}^2$. Thus

$$\delta \leqslant |\Lambda_N(f_0, f_1, g)| = \frac{1}{N^3}\Bigg| \sum_{x,y} F(x,y) g(x,y) \Bigg| .$$

The Cauchy–Schwarz inequality then gives

$$\delta^2 \ll \frac{1}{N^3} \sum_{x,y} \overline{F(x,y)} F(x,y) = \Lambda_N\big(\overline{f}_0, \overline{f}_1, F\big) .$$

By Theorem 4.1, we deduce that

$$\mathbb{E}_{x\in[N]} \big\| F_x 1_{[2N^2]} \big\|_{U^5(\mathbb{Z})}^{2^5} \gg \delta^{O(1)} \big\| 1_{[2N^2]} \big\|_{U^5(\mathbb{Z})}^{2^5} .$$

□

5. Degree lowering

The key result of this section is Lemma 5.5, which shows that we can replace the U^5-norm in Corollary 4.2 with the U^1-norm, after passing to a long subprogression of small common difference. Before proving this, we begin with a number of technical lemmas that underlie the argument.

Lemma 5.1 (Weyl bound)**.** *Let μ_N denote the probability measure*

$$\mu_N := \tfrac{1}{N^2} 1_{[N]} * 1_{-[N]} = \tfrac{1}{N}\Big(1 - \tfrac{|\cdot|}{N}\Big)_+ .$$

For any $\alpha, \beta \in \mathbb{T}$ we have the implication

$$\Bigg| \sum_d \mu_N(d) e(\alpha d^2 + \beta d) \Bigg| \geqslant \delta \implies \big(\exists q \ll \delta^{-O(1)}\big) \left[\|q\alpha\|_{\mathbb{T}} \ll \delta^{-O(1)}/N^2 \text{ and } \|q\beta\|_{\mathbb{T}} \ll \delta^{-O(1)}/N \right] . \tag{5.1}$$

Furthermore, for any integers $a, b \in \mathbb{Z}$ we have

$$\left|\sum_d \mu_N(d+a)\mu_N(d+b)e(\alpha d^2+\beta d)\right| \geqslant \delta/N \implies \left(\exists q \ll \delta^{-O(1)}\right)\left[\|q\alpha\|_{\mathbb{T}} \ll \delta^{-O(1)}/N^2\right]. \tag{5.2}$$

Proof. Let us prove (5.2), the remaining assertions being similar and simpler.

Expanding the definition of the Fejér kernel (2.7), we have

$$\sum_d \mu_N(d+a)\mu_N(d+b)e(\alpha d^2+\beta d)$$
$$= N^{-4}\sum_{d_1,d_2,d_3} 1_{[N]}(d_1)1_{[N]}(d_2)1_{[N]}(d_3)1_{[N]}(a-b-d_1+d_2+d_3)$$
$$e\big(\alpha(d_1-d_2-a)^2+\beta(d_1-d_2-a)\big)$$
$$\leqslant N^{-2}\max_{c_1,c_2}\left|\sum_{d_1} 1_{[N]}(d_1)1_{[N]}(c_1-d_1)e\big(\alpha d_1^2+(\beta-c_2\alpha)d_1\big)\right|.$$

Since the final summation in d_1 is over a subinterval of $[N]$, the claimed conclusion follows from Weyl's inequality, as formulated in [Tao12, Lemma 1.1.16] or [GT12nil, Proposition 4.3]. □

Lemma 5.2 (Shifted quadratic correlations are major arc)**.** *Let $f : \mathbb{Z} \to \mathbb{C}$ be a 1-bounded function and let μ_N denote the probability measure*

$$\mu_N := \tfrac{1}{N^2}1_{[N]} * 1_{-[N]} = \tfrac{1}{N}\left(1-\tfrac{|\cdot|}{N}\right)_+.$$

Suppose that for some $\alpha \in \mathbb{T}$ we have

$$\mathbb{E}_{x\in[N]}\left|\sum_d \mu_N(d)f(x+d)e\big(\alpha d^2\big)\right| \geqslant \delta. \tag{5.3}$$

Then there exists $q \ll \delta^{-O(1)}$ such that $\|q\alpha\|_{\mathbb{T}} \ll \delta^{-O(1)}/N^2$. Furthermore there exists $\beta \in \mathbb{T}$ such that $\|q\beta\|_{\mathbb{T}} \ll \delta^{-O(1)}/N$ and

$$\left|\sum_{x\in[-N,2N]} f(x)e(\beta x)\right| \gg \delta^2 N.$$

Similarly, if for some $\beta \in \mathbb{T}$ we have

$$\mathbb{E}_{y\in[N^2]}\left|\sum_d \mu_N(d)f(y+d^2)e(\beta d)\right| \geqslant \delta, \tag{5.4}$$

then there exists $q \ll \delta^{-O(1)}$ such that $\|q\beta\|_{\mathbb{T}} \ll \delta^{-O(1)}/N$ and there exists $\alpha \in \mathbb{T}$ with $\|q\alpha\|_{\mathbb{T}} \ll \delta^{-O(1)}/N^2$ such that

$$\left|\sum_{y\in[2N^2]} f(y)e(\alpha y)\right| \gg \delta^{O(1)}N^2.$$

Proof. Write $h(d) := \mu_N(d)e(\alpha d^2)$. Our assumption (5.3) implies that there exists a phase function $g : \mathbb{Z} \to \{z \in \mathbb{C} : |z| = 1\}$ such that

$$\sum_{x,d} g(x)1_{[N]}(x)f(x+d)1_{[-N,2N]}(x+d)h(d) \geqslant \delta N.$$

By orthogonality and Hölder's inequality, we have

$$\begin{aligned}\delta N &\leqslant \int_{\mathbb{T}} \widehat{g1}_{[N]}(\beta)\widehat{f1}_{[-N,2N]}(-\beta)\hat{h}(\beta)\mathrm{d}\beta \\ &\leqslant \|\widehat{g1}_{[N]}\|_2\|\widehat{f1}_{[-N,2N]}\|_2^{1/2}\|\hat{h}\|_2^{1/2}\|\widehat{f1}_{[-N,2N]}\hat{h}\|_\infty^{1/2} \\ &\leqslant \|1_{[N]}\|_2\,\|1_{[-N,2N]}\|_2^{1/2}\,\|\mu_N\|_2^{1/2}\,\|\widehat{f1}_{[-N,2N]}\hat{h}\|_\infty^{1/2}.\end{aligned}$$

Hence there exists $\beta \in \mathbb{T}$ such that

$$\Big|\sum_d \mu_N(d)e(\alpha d^2+\beta d)\Big| \gg \delta^2 \quad \text{and} \quad \Big|\sum_{x\in[-N,2N]} f(x)e(\beta x)\Big| \gg \delta^2 N.$$

The result now follows from the Weyl bound (Lemma 5.1).

Next we turn to (5.4). This implies that there exists a phase function $g : \mathbb{Z} \to \{z \in \mathbb{C} : |z| = 1\}$ such that

$$\sum_{y,d} g(y)1_{[N^2]}(y)f(y+d^2)1_{[2N^2]}(y+d^2)\mu_N(d)e(\beta d) \geqslant \delta N^2.$$

Write $S_N(\alpha,\beta) := \sum_d \mu_N(d)e(\alpha d^2+\beta d)$. By orthogonality and Hölder's inequality, we have

$$\begin{aligned}\delta N^2 &\leqslant \int_{\mathbb{T}} \widehat{g1}_{[N^2]}(\alpha)\widehat{f1}_{[2N^2]}(-\alpha)S_N(\alpha,\beta)\mathrm{d}\alpha \\ &\leqslant \big\||\widehat{f1}_{[2N^2]}|^{2/7}|S(\cdot,\beta)|^{1/7}\big\|_\infty \int_{\mathbb{T}} |\widehat{g1}_{[N^2]}(\alpha)||\widehat{f1}_{[2N^2]}(-\alpha)|^{5/7}|S_N(\alpha,\beta)|^{6/7}\mathrm{d}\alpha \\ &\leqslant \big\||\widehat{f1}_{[2N^2]}|^{2/7}|S(\cdot,\beta)|^{1/7}\big\|_\infty\|\widehat{g1}_{[N^2]}\|_2\|\widehat{f1}_{[2N^2]}\|_2^{5/7}\|S(\cdot,\beta)\|_6^{6/7}.\end{aligned}$$

Employing a standard estimate for the sixth moment of a quadratic exponential sum (see, for instance, [PP22, Lemma 6.4]), there exists $\alpha \in \mathbb{T}$ such that

$$\Big|\sum_d \mu_N(d)e(\alpha d^2+\beta d)\Big| \gg \delta^7 \quad \text{and} \quad \Big|\sum_{x\in[2N^2]} f(x)e(\alpha x)\Big| \gg \delta^{7/2}N^2.$$

The result now follows from the Weyl bound (Lemma 5.1). □

The following simple observation is used repeatedly in the remainder of the paper, and can be regarded as the key "reason" why Theorem 1.3 has exponential bounds, rather than the tower-type bounds encountered in the popular version of the linear Roth theorem [FPZ22]. It essentially allows us to show that we can pass from *many* "major arc" correlations to a *single* correlation: the major arcs line up. The reason underlying this is that the major arcs form a tiny fraction of all possible frequencies.

Lemma 5.3 (Pigeon-holing major arcs)**.** *Let $f_1,\dots,f_M : [-N,N] \to \mathbb{C}$ be 1-bounded functions. For a fixed positive integer k, let $\alpha,\alpha_1,\dots,\alpha_M \in \mathbb{T}$ be frequencies for which there are positive integers $q_1,\dots,q_M \leqslant Q$ such that $\|q_m(\alpha_m-\alpha)\|_{\mathbb{T}} \leqslant Q/N^k$ for each m. Suppose that*

$$\mathbb{E}_{m\in[M]}\big|\mathbb{E}_{|x|\leqslant N} f_m(x)e(\alpha_m x^k)\big| \geqslant 1/Q.$$

Then there exists $m_0 \in [M]$ such that

$$\mathbb{E}_{m\in[M]}\big|\mathbb{E}_{|x|\leqslant N} f_m(x)e(\alpha_{m_0} x^k)\big| \gg 1/Q^{O(1)}.$$

Moreover, if $\alpha_1,\dots,\alpha_M$ are $\frac{1}{QN^k}$-separated[7]*, then we can ensure that the $m_0 \in [M]$ found above satisfies*

$$\#\{m \in [M] : \alpha_m = \alpha_{m_0}\} \gg MQ^{-O(1)}.$$

[7] By which we mean that, for each i and j, either $\alpha_i = \alpha_j$ or $\|\alpha_i - \alpha_j\| \geqslant \frac{1}{QN^k}$.

Proof. By the popularity principle, there exists a set $\mathcal{M} \subset [M]$ with $|\mathcal{M}| \geqslant \frac{M}{2Q}$ such that for each $m \in \mathcal{M}$ we have

$$\left|\mathbb{E}_{|x|\leqslant N} f_m(x) e(\alpha_m x^k)\right| \geqslant \tfrac{1}{2Q}. \tag{5.5}$$

Set $T := 100QN^k$. If $\alpha_1, \dots, \alpha_M$ are $1/(QN^k)$-separated, then set $\tilde{\alpha}_m := \alpha_m$; otherwise, we round each α_m to the nearest fraction of the form t/T. In either case, we obtain a $1/T$-separated set $\tilde{\alpha}_1, \dots, \tilde{\alpha}_M$ contained in the shifted union

$$\alpha + \bigcup_{1\leqslant a\leqslant q\leqslant Q} \left[\tfrac{a}{q} - \tfrac{2Q}{qN^k}, \tfrac{a}{q} + \tfrac{2Q}{qN^k}\right].$$

Write D for the number of distinct $\tilde{\alpha}_m$ lying in the above set. Since the $\tilde{\alpha}_m$ are $1/T$-separated, a volume packing argument gives

$$D \leqslant \sum_{q\leqslant Q} q(1 + \tfrac{4QT}{qN^k}) \ll Q^2\big(1 + \tfrac{T}{N^k}\big) \ll Q^3.$$

Hence, by the pigeon-hole principle, there exists $\tilde{\alpha}_{m_0}$ such that

$$\#\{m \in \mathcal{M} : \tilde{\alpha}_m = \tilde{\alpha}_{m_0}\} \gg |\mathcal{M}|Q^{-3} \gg MQ^{-4}.$$

Our choice of T ensures that, for each m with $\tilde{\alpha}_m = \tilde{\alpha}_{m_0}$, the frequencies α_m and α_{m_0} are sufficiently close to ensure that

$$\left|\mathbb{E}_{|x|\leqslant N} f_m(x) e(\alpha_{m_0} x^k)\right| \gg 1/Q. \tag{5.6}$$

Summing over $m \in [M]$ with $\tilde{\alpha}_m = \tilde{\alpha}_{m_0}$, we deduce that

$$\mathbb{E}_{m\in[M]}\left|\mathbb{E}_{|x|\leqslant N} f_m(x) e(\alpha_{m_0} x^k)\right| \gg 1/Q^5.$$

□

Our penultimate technical lemma before proving Lemma 5.5 is essentially the main argument underlying Lemma 5.5, but carried out in a notationally simpler context.

Lemma 5.4 (Shifted quadratic correlations conspire)**.** *Let $f_1, \dots, f_M : \mathbb{Z} \to \mathbb{C}$ be 1-bounded functions and let μ_N denote the probability measure*

$$\mu_N := \tfrac{1}{N^2} 1_{[N]} * 1_{-[N]} = \tfrac{1}{N}\Big(1 - \tfrac{|\cdot|}{N}\Big)_+.$$

Let T be a positive integer satisfying $T \leqslant \delta^{-1}N^2$. Suppose that there are functions $\phi_1, \dots, \phi_M : \mathbb{Z} \to \{t/T : t \in [T]\}$ and a set $\mathcal{N} \subset [M] \times [N]$ of size at least δMN such that for each $(m, x) \in \mathcal{N}$ we have

$$\left|\sum_d \mu_N(d) f_m(x + d) e\big(\phi_m(x) d^2\big)\right| \geqslant \delta. \tag{5.7}$$

Then there exists $q \ll \delta^{-O(1)}$ and $\phi \in \mathbb{T}$ with $\|q\phi\|_{\mathbb{T}} \ll \delta^{-O(1)}/N^2$ such that

$$\#\{(m, x) \in \mathcal{N} : \phi_m(x) = \phi\} \gg \delta^{O(1)}MN.$$

Proof. We first show that, for fixed m, many of the $\phi_m(x)$ lie in the same shift of the major arcs.

By the popularity principle, there exists a set $\mathcal{M} \subset [M]$ with $|\mathcal{M}| \gg \delta M$ such that for each $m \in \mathcal{M}$ the fibre $\mathcal{N}_m := \{x \in [N] : (m, x) \in \mathcal{N}\}$ satisfies $|\mathcal{N}_m| \gg \delta N$. Summing over $\mathcal{N}_m$, we deduce that

$$\mathbb{E}_{x\in\mathcal{N}_m}\left|\sum_d \mu_N(d) f_m(x + d) e\big(\phi_m(x) d^2\big)\right| \geqslant \delta.$$

Applying the Cauchy–Schwarz inequality to double the d variable and then re-parametrising gives

$$\begin{aligned}\delta^2 &\leqslant \mathbb{E}_{x\in\mathcal{N}_m}\sum_{d,\tilde d}\mu_N(d)\mu_N(\tilde d)f_m(x+d)\overline{f_m(x+\tilde d)}e\Big(\phi_m(x)\Big[d^2-\tilde d^2\Big]\Big)\\ &\ll \frac{1}{\delta N}\sum_{|x|<2N}\sum_{|h|<2N}f_m(x)\overline{f_m(x+h)}\sum_d \Delta_h\mu_N(d)1_{\mathcal{N}_m}(x-d)e\big(-\phi_m(x-d)\big[2hd+h^2\big]\big).\end{aligned}$$

We again apply the Cauchy–Schwarz inequality to double the d variable, yielding

$$\begin{aligned}\delta^6 &\ll \sum_{x,h,d,\tilde d}\Delta_h\mu_N(d)\Delta_h\mu_N(\tilde d)1_{\mathcal{N}_m}(x-d)1_{\mathcal{N}_m}(x-\tilde d)\\ &\qquad\times e\Big(\Big[\phi_m(x-d)-\phi_m(x-\tilde d)\Big]h^2+\phi_m(x-d)2hd-\phi_m(x-\tilde d)2h\tilde d\Big)\\ &\leqslant \sum_{x,d,\tilde d}\mu_N(d)\mu_N(\tilde d)1_{\mathcal{N}_m}(x-d)1_{\mathcal{N}_m}(x-\tilde d)\Bigg|\sum_h\mu_N(d+h)\mu_N(\tilde d+h)\\ &\qquad\times e\Big(\Big[\phi_m(x-d)-\phi_m(x-\tilde d)\Big]h^2+\phi_m(x-d)2hd-\phi_m(x-\tilde d)2h\tilde d\Big)\Bigg|.\end{aligned}$$

Next, we change variables in the above sum, substituting $d=x-n$ and $\tilde d = x-\tilde n$. We also use the pointwise bound $\mu_N\leqslant 1/N$, together with the fact that the only values of x that contribute to the above sum lie in the interval $[-2N,2N]$. Thus taking a maximum over x gives some x_0 such that

$$\begin{aligned}\delta^6 N &\ll \sum_{n,\tilde n}1_{\mathcal{N}_m}(n)1_{\mathcal{N}_m}(\tilde n)\Bigg|\sum_h\mu_N(x_0-n+h)\mu_N(x_0-\tilde n+h)\\ &\qquad\times e\Big([\phi_m(n)-\phi_m(\tilde n)]\,h^2+\phi_m(n)2h(x_0-n)-\phi_m(\tilde n)2h(x_0-\tilde n)\Big)\Bigg|.\end{aligned}$$

Taking a maximum over $\tilde n$ and using the popularity principle in n, there exists $\mathcal{N}_m'\subset\mathcal{N}_m$ with $|\mathcal{N}_m'|\gg\delta^6 N$ and there exists an integer y_0 along with a frequency α_m such that for each $n\in\mathcal{N}_m'$ there exists an integer y_n and a frequency $\beta_m(n)\in\mathbb{T}$ satisfying

$$\frac{\delta^6}{N}\ll\Bigg|\sum_h\mu_N(y_n+h)\mu_N(y_0+h)e\Big([\phi_m(n)-\alpha_m]\,h^2+\beta_m(n)h\Big)\Bigg|.$$

By the Weyl bound (Lemma 5.1), for each $n\in\mathcal{N}_m'$ there exists $q_{m,n}\ll\delta^{-O(1)}$ with

$$\|q_{m,n}[\phi_m(n)-\alpha_m]\|_{\mathbb{T}}\ll\delta^{-O(1)}/N^2.$$

Whilst summing (5.7) over $n\in\mathcal{N}_m'$ gives

$$\sum_{n\in[N]}\Bigg|\sum_d\mu_N(d)f_m(n+d)e\big(\phi_m(n)d^2\big)\Bigg|\gg\delta^{O(1)}N.$$

Applying Lemma 5.3 (using the fact the frequencies are $1/T$-separated) for each $m\in\mathcal{M}$ there exists $n_m\in\mathcal{N}_m'$ such that on setting $\phi_m:=\phi_m(n_m)$ we have

$$\#\{n\in\mathcal{N}_m:\phi_m(n)=\phi_m\}\gg\delta^{O(1)}N.$$

In particular

$$\mathbb{E}_{x\in[N]}\Bigg|\sum_d\mu_N(d)f_m(x+d)e\big(\phi_m d^2\big)\Bigg|\gg\delta^{O(1)}.$$

Applying Lemma 5.2, there exists $q \ll \delta^{-O(1)}$ such that $\|q\phi_m\|_{\mathbb{T}} \ll \delta^{-O(1)}/N^2$. Re-applying Lemma 5.3 gives some $\phi = \phi_{m_0}$ such that

$$\#\{m \in \mathcal{M} : \phi_m = \phi\} \gg \delta^{O(1)} M$$

Hence

$$\#\{(m, x) \in \mathcal{N} : \phi_m(x) = \phi\} \geqslant \sum_{m \in \mathcal{M}: \phi_m = \phi} \#\{x \in \mathcal{N}_m : \phi_m(x) = \phi_m\} \gg \delta^{O(1)} MN.$$

□

We are now in a position to prove our degree lowering lemma.

Lemma 5.5 (Degree lowering). *Let $f_0, f_1 : \mathbb{Z} \to \mathbb{C}$ be 1-bounded functions with*

$$\mathrm{supp}(f_0) \subset [N] \times [N^2] \quad \textit{and} \quad \mathrm{supp}(f_1) \subset [-N, 2N] \times [N^2]$$

Let μ_N denote the probability measure

$$\mu_N := \tfrac{1}{N^2} 1_{[N]} * 1_{-[N]} = \tfrac{1}{N}\Big(1 - \tfrac{|\cdot|}{N}\Big)_+ .$$

Define the dual function

$$F_x(y) = F(x, y) := \sum_d \mu_N(d) f_0(x, y - d^2) f_1(x + d, y - d^2),$$

a 1-bounded function supported on $[N] \times [2N^2]$. Suppose that for some $s \geqslant 2$ we have

$$\mathbb{E}_{x \in [N]} \|F_x\|_{U^s(\mathbb{Z})}^{2^s} \geqslant \delta \big\|1_{[2N^2]}\big\|_{U^s(\mathbb{Z})}^{2^s},$$

where the U^s-norm is given by (2.5). Then, either $N \ll_s \delta^{-O_s(1)}$, or there exists $\alpha \in \mathbb{T}$ and $q \ll_s \delta^{-O_s(1)}$ such that $\|q\alpha\|_{\mathbb{T}} \ll_s \delta^{-O_s(1)}/N^2$ and

$$\mathbb{E}_{x \in [N]} \mathbb{E}_{y \in [N^2]} \Big| \sum_d \mu_N(d) f_1(x + d, y) e\big(\alpha d^2\big) \Big| \gg_s \delta^{O_s(1)}. \tag{5.8}$$

Proof. Let us first assume that $s \geqslant 3$. In this case, we claim that our assumptions imply that

$$\mathbb{E}_{x \in [N]} \|F_x\|_{U^{s-1}(\mathbb{Z})}^{2^{s-1}} \gg_s \delta^{O_s(1)} \big\|1_{[2N^2]}\big\|_{U^{s-1}(\mathbb{Z})}^{2^{s-1}}. \tag{5.9}$$

Given this claim, we intend to iterate in order to arrive at the case in which $s = 2$, so that

$$\mathbb{E}_{x \in [N]} \|F_x\|_{U^2(\mathbb{Z})}^{4} \gg_s \delta^{O_s(1)} \big\|1_{[2N^2]}\big\|_{U^2(\mathbb{Z})}^{4}. \tag{5.10}$$

Returning to the proof of (5.9), since we may assume that $N \ll_s 1$ does not hold, we have

$$\big\|1_{[2N^2]}\big\|_{U^s(\mathbb{Z})}^{2^s} \gg_s N^{2s+2}.$$

Expanding the definition of the U^s-norm (2.5), and using the fact that F_x is supported on $[2N^2]$, gives

$$\mathbb{E}_{|h_1| < 2N^2} \cdots \mathbb{E}_{|h_{s-2}| < 2N^2} \mathbb{E}_{x \in [N]} \big\|\Delta_{h_1, \dots, h_{s-2}} F_x\big\|_{U^2}^4 \gg_s \delta N^6.$$

Applying the U^2-inverse theorem [PP19, Lemma A.1], there exists

$$\mathcal{N} \subset \big((-2N^2, 2N^2)^{s-2} \cap \mathbb{Z}^{s-2}\big) \times [N]$$

of size $|\mathcal{N}| \gg_s \delta N^{2s-3}$ and a function $\phi : \mathbb{Z}^{s-1} \to \mathbb{T}$ such that for every $(\underline{h}, x) \in \mathcal{N}$ we have

$$\Big| \sum_y \Delta_{\underline{h}} F_x(y) e\big(\phi(\underline{h}; x) y\big) \Big| \gg_s \delta N^2. \tag{5.11}$$

Set $T := \lceil C_s \delta^{-1} N^2 \rceil$, with C_s an absolute constant taken sufficiently large to ensure that, on rounding $\phi(\underline{h}; x)$ to the nearest fraction of the form t/T, the inequality (5.11) remains valid.

Summing over $\mathcal{N}$ and applying the dual-difference interchange inequality [PP19, Lemma 6.3], we deduce that

$$\sum_{(\underline{h}^0,x),(\underline{h}^1,x)\in\mathcal{N}}\Bigg|\sum_{y,d}\mu_N(d)\Delta_{(0,\underline{h}^0-\underline{h}^1)}f_0(x,y-d^2)$$
$$\Delta_{(0,\underline{h}^0-\underline{h}^1)}f_1(x+d,y-d^2)e\big(\phi(\underline{h}^0;\underline{h}^1;x)y\big)\Bigg| \gg_s \delta^{O_s(1)}N^{4s-5},$$

where

$$\phi(\underline{h}^0;\underline{h}^1;x) := \sum_{\omega\in\{0,1\}^{s-2}}(-1)^{|\omega|}\phi(\underline{h}^\omega;x) \qquad\text{and}\qquad \underline{h}^\omega := (h_1^{\omega_1},\dots,h_{s-2}^{\omega_{s-2}}).$$

Changing variables from y to $y+d^2$ and taking maxima, there exist y and $\underline{h}^1$ such that

$$\sum_{(\underline{h}^0,x)\in\mathcal{N}}\Bigg|\sum_{d}\mu_N(d)\Delta_{(0,\underline{h}^0-\underline{h}^1)}f_1(x+d,y)e\big(\phi(\underline{h}^0;\underline{h}^1;x)d^2\big)\Bigg| \gg_s \delta^{O_s(1)}N^{2s-3}.$$

Thus, there exists $\mathcal{N}'\subset\mathcal{N}$ of size $\mathcal{N}'\gg_s\delta^{O_s(1)}N^{2s-3}$ such that for each $(\underline{h}^0,x)\in\mathcal{N}'$ we have

$$\Bigg|\sum_{d}\mu_N(d)\Delta_{(0,\underline{h}^0-\underline{h}^1)}f_1(x+d,y)e\big(\phi(\underline{h}^0;\underline{h}^1;x)d^2\big)\Bigg| \gg_s \delta^{O_s(1)}.$$

By Lemma 5.4, there exists $\mathcal{N}''\subset\mathcal{N}'$ of size $\gg_s\delta^{O_s(1)}N^{2s-3}$ and $\phi\in\mathbb{T}$ such that for any $(\underline{h}^0,x)\in\mathcal{N}''$ we have $\phi(\underline{h}^0;\underline{h}^1;x)=\phi$. In particular, when restricted to the set $\mathcal{N}''$, the function ϕ satisfies

$$\phi(\underline{h}^0;x) = \phi - \sum_{\omega\in\{0,1\}^{s-2}\setminus\{0\}}(-1)^{|\omega|}\phi(\underline{h}^\omega;x).$$

For fixed x, the right-hand side of this identity is *low rank* in $\underline{h}^0$ according to the terminology preceding [PP19, Lemma 6.4].

Summing over $\underline{h}$ such that $(\underline{h},x)\in\mathcal{N}''$ in (5.11), we deduce the existence of low-rank functions $\psi_x:\mathbb{Z}^{s-2}\to\mathbb{T}$ such that

$$\mathbb{E}_{x\in[N]}\mathbb{E}_{|h_1|<N^2}\cdots\mathbb{E}_{|h_{s-2}|<N^2}\Bigg|\sum_{y}\Delta_{\underline{h}}F_x(y)e\big(\psi_x(\underline{h})y\big)\Bigg| \gg_s \delta^{O_s(1)}N^2. \tag{5.12}$$

Employing the inequality stated in [PP19, Lemma 6.4] then gives

$$\mathbb{E}_{x\in[N]}\|F_x\|_{U^{s-1}}^{2^{s-1}} \gg_s \delta^{O_s(1)}N^{2s}.$$

This proves the claim (5.9).

By iterating (5.9), our initial assumption of large U^s-norm yields a large U^2-norm, as in (5.10). Moreover, we note that when running the argument with $s=3$, the inequality (5.12) becomes

$$\mathbb{E}_{x\in[N]}\big|\mathbb{E}_{|y|\leqslant 2N^2}F_x(y)e\big(\psi_x y\big)\big| \gg_s \delta^{O_s(1)},$$

where for each x, our application of Lemma 5.4 gives some $q_x\ll_s\delta^{-O_s(1)}$ with $\|q_x\psi_x\|_{\mathbb{T}}\ll_s\delta^{-O_s(1)}/N^2$. Once more applying Lemma 5.3, we may assume that $\psi_x=\psi$ is constant in x, so that

$$\mathbb{E}_{x\in[N]}\Bigg|\sum_{y}F_x(y)e(\psi y)\Bigg| \gg_s \delta^{O_s(1)}N^2.$$

Expanding the dual function F_x, changing variables in y and employing the triangle inequality to bring the sum over y outside the absolute value, we derive (5.8). □

6. Inverse theorems for our counting operator

Degree lowering (Lemma 5.5) can be combined with vertical U^5-control of the dual function (Corollary 4.2), and a little Fourier analysis, in order to prove an inverse theorem for the function weighting the $(x+d,y)$ term in our counting operator.

Theorem 6.1 (Counting operator inverse theorem, I). *Let $f_0, f_1, f_2 : \mathbb{Z}^2 \to \mathbb{C}$ be 1-bounded functions and let μ_N denote the probability measure*

$$\mu_N := \tfrac{1}{N^2} 1_{[N]} * 1_{-[N]} = \tfrac{1}{N}\Big(1 - \tfrac{|\cdot|}{N}\Big)_+ .$$

Write

$$\Lambda_N(f_0, f_1, f_2) := \mathbb{E}_{x\in[N]}\mathbb{E}_{y\in[N^2]} \sum_d \mu_N(d) f_0(x,y) f_1(x+d,y) f_2(x, y+d^2).$$

If $|\Lambda_N(f_0,f_1,f_2)| \geqslant \delta$, then either $N \ll \delta^{-O(1)}$ or there exists $\beta \in \mathbb{T}$ and $q \ll \delta^{-O(1)}$ such that $\|q\beta\|_{\mathbb{T}} \ll \delta^{-O(1)}/N$ and

$$\mathbb{E}_{y\in[N^2]}\big|\mathbb{E}_{x\in[-N,2N]} f_1(x,y) e(\beta x)\big| \gg \delta^{O(1)}. \tag{6.1}$$

Proof. In order to avoid tracking the range of summation, we assume that

$$\operatorname{supp}(f_0) \subset [N]\times[N^2] \quad \text{and} \quad \operatorname{supp}(f_1) \subset [-N, 2N]\times[N^2]$$

Notice that this affects neither the hypothesis nor the conclusion of the theorem.

By vertical U^5-control of the dual function (Corollary 4.2), our assumptions give that

$$\mathbb{E}_{x\in[N]} \|F_x\|_{U^5(\mathbb{Z})}^{2^5} \gg \delta^{O(1)} \big\|1_{[2N^2]}\big\|_{U^5(\mathbb{Z})}^{2^5},$$

where

$$F_x(y) = F(x,y) := \sum_d \mu_N(d) f_0(x, y-d^2) f_1(x+d, y-d^2).$$

By degree lowering (Lemma 5.5), either $N \ll \delta^{-O(1)}$ or there exists $\alpha \in \mathbb{T}$ and $q \ll \delta^{-O(1)}$ with $\|q\alpha\|_{\mathbb{T}} \ll \delta^{-O(1)}/N^2$ such that

$$\mathbb{E}_{x\in[N]}\mathbb{E}_{y\in[N^2]}\left|\sum_d \mu_N(d) f_1(x+d,y) e(\alpha d^2)\right| \gg \delta^{O(1)}.$$

By the popularity principle and Lemma 5.2, for each $y \in [N^2]$ there exist $q_y \ll \delta^{-O(1)}$ and $\beta_y \in \mathbb{T}$ with $\|q_y\beta_y\|_{\mathbb{T}} \ll \delta^{-O(1)}/N$ such that

$$\mathbb{E}_{y\in[N^2]}\big|\mathbb{E}_{x\in[-N,2N]} f_1(x,y) e(\beta_y x)\big| \gg \delta^{O(1)}.$$

Pigeon-holing in the major arcs (Lemma 5.3), there exists $\beta \in \mathbb{T}$ and $q \ll \delta^{-O(1)}$ with $\|q\beta\|_{\mathbb{T}} \ll \delta^{-O(1)}/N$ such that

$$\mathbb{E}_{y\in[N^2]}\big|\mathbb{E}_{x\in[-N,2N]} f_1(x,y) e(\beta x)\big| \gg \delta^{O(1)}.$$

The result follows. □

We now use the inverse theorem for the function weighting the $(x+d,y)$ term to obtain an inverse theorem for the function weighting the $(x, y+d^2)$ term.

Theorem 6.2 (Counting operator inverse theorem, II). *Let $f_0, f_1, f_2 : \mathbb{Z}^2 \to \mathbb{C}$ be 1-bounded functions and let μ_N denote the probability measure*

$$\mu_N := \tfrac{1}{N^2} 1_{[N]} * 1_{-[N]} = \tfrac{1}{N}\Big(1 - \tfrac{|\cdot|}{N}\Big)_+ .$$

Write

$$\Lambda_N(f_0, f_1, f_2) := \mathbb{E}_{x\in[N]}\mathbb{E}_{y\in[N^2]}\sum_d \mu_N(d) f_0(x, y) f_1(x + d, y) f_2(x, y + d^2).$$

If $|\Lambda_N(f_0, f_1, f_2)| \geqslant \delta$*, then either* $N \ll \delta^{-O(1)}$ *or there exists* $\alpha \in \mathbb{T}$ *and* $q \ll \delta^{-O(1)}$ *such that* $\|q\alpha\|_{\mathbb{T}} \ll \delta^{-O(1)}/N^2$ *and*

$$\mathbb{E}_{x\in[N]}\big|\mathbb{E}_{y\in[2N^2]} f_2(x, y)e(\alpha y)\big| \gg \delta^{O(1)}.$$

Proof. We may assume that

$$\mathrm{supp}(f_0) \subset [N] \times [N^2], \quad \mathrm{supp}(f_1) \subset [-N, 2N] \times [N^2], \quad \mathrm{supp}(f_2) \subset [N] \times [2N^2].$$

Define the dual function

$$G(x, y) := \sum_d \mu_N(d) f_0(x - d, y) f_2(x - d, y + d^2).$$

Then, by the Cauchy–Schwarz inequality:

$$\delta \leqslant \left|\frac{1}{N^3}\sum_{x,y} f_1(x, y)G(x, y)\right| \ll \left(\frac{1}{N^3}\sum_{x,y} G(x, y)\overline{G(x, y)}\right)^{1/2} = \Lambda_N\big(\overline{f}_0, G, \overline{f}_2\big)^{1/2}.$$

Hence, by our first inverse theorem (Theorem 6.1), either $N \ll \delta^{-O(1)}$ or there exists $\beta \in \mathbb{T}$ and $q \ll \delta^{-O(1)}$ such that $\|q\beta\|_{\mathbb{T}} \ll \delta^{-O(1)}/N$ and

$$\mathbb{E}_{y\in[N^2]}\big|\mathbb{E}_{x\in[-N,2N]} G(x, y)e(\beta x)\big| \gg \delta^{O(1)}.$$

Expanding the dual function, changing variables and employing the triangle inequality, we have

$$\mathbb{E}_{x\in[N]}\mathbb{E}_{y\in[N^2]}\left|\sum_d \mu_N(d) f_2(x, y + d^2)e(\beta d)\right| \gg \delta^{O(1)}.$$

By the popularity principle and Lemma 5.2, for each $x \in [N]$ there exist $q_x \ll \delta^{-O(1)}$ and $\alpha_x \in \mathbb{T}$ with $\|q_x\alpha_x\|_{\mathbb{T}} \ll \delta^{-O(1)}/N^2$ such that

$$\mathbb{E}_{x\in[N]}\big|\mathbb{E}_{y\in[2N^2]} f_2(x, y)e(\alpha_x y)\big| \gg \delta^{O(1)}.$$

Pigeon-holing in the major arcs (Lemma 5.3), there exists $\alpha \in \mathbb{T}$ and $q \ll \delta^{-O(1)}$ with $\|q\alpha\|_{\mathbb{T}} \ll \delta^{-O(1)}/N^2$ such that

$$\mathbb{E}_{x\in[N]}\big|\mathbb{E}_{y\in[2N^2]} f_2(x, y)e(\alpha y)\big| \gg \delta^{O(1)}.$$

The result follows. □

The previous two inverse theorems concern the global count of configurations. As indicated in §3, our strategy for proving Theorem 1.2 is to focus on counting configurations where the difference parameter is divisible by a small positive integer q and is localised to an interval shorter than N. We therefore require an inverse theorem for this localised counting operator. We also take this opportunity to replace correlation with major arcs by convolution with a Fejér kernel, which facilitates the energy increment argument in Theorem 7.1. Making these adjustments to our inverse theorem is the purpose of the next result.

Theorem 6.3 (Large localised count gives a large convolution)**.** *Let* $f_0, f_1, f_2 : \mathbb{Z}^2 \to \mathbb{C}$ *be 1-bounded functions with support contained in* $[-N_1, N_1] \times [-N_2, N_2]$ *where* $N_1 \geqslant N_2^{1/2}$*. Suppose that for some positive integers* q, M *with* $qM \leqslant N_2^{1/2}$ *we have*

$$\left|\sum_{x,y,d} \mu_M(d) f_0(x, y) f_1(x + qd, y) f_2(x, y + q^2d^2)\right| \geqslant \delta N_1 N_2.$$

Then there exists a positive integer $\tilde{q}$ with $\tilde{q} \ll \delta^{-O(1)}$ such that for any positive integer $\widetilde{M}$, either $\widetilde{M} \gg \delta^{O(1)} M$ or

$$\sum_{x,y} \left| \sum_{\tilde{d}} \mu_{\widetilde{M}^2}(\tilde{d}) f_2(x, y + q^2\tilde{q}^2\tilde{d}) \right| \gg \delta^{O(1)} N_1 N_2. \tag{6.2}$$

Here μ_M and $\mu_{\widetilde{M}^2}$ denote the Fejér kernels defined in (2.7).

Proof. By a change of variables, we may average our operator over additional shifts, to deduce that

$$\Big|\sum_{x,y} \mathbb{E}_{x'\in[M]}\mathbb{E}_{y'\in[M^2]} \sum_{d} \mu_M(d) f_0(x+qx', y+q^2y') f_1(x+q(x'+d), y+q^2y') f_2(x+qx', y+q^2(y'+d^2))\Big| \geqslant \delta N_1 N_2.$$

Hence, on setting

$$f_i^{x,y,q}(x', y') := f_i(x+qx', y+q^2y'),$$

we have that

$$\Big|\sum_{x,y} \Lambda_M(f_0^{x,y,q}, f_1^{x,y,q}, f_2^{x,y,q})\Big| \geqslant \delta N_1 N_2. \tag{6.3}$$

We note that, since $qM \leqslant N_2^{1/2} \leqslant N_1$ and each f_i is supported on $[-N_1, N_1] \times [-N_2, N_2]$, the set of pairs $(x, y) \in \mathbb{Z}^2$ that contribute to (6.3) is contained in $[-2N_1, 2N_1] \times [-2N_2, 2N_2]$. Thus, by the popularity principle, there exists a set $\mathcal{P} \subset [-2N_1, 2N_1] \times [-2N_2, 2N_2]$ of size $\gg \delta N_1 N_2$ such that

$$|\Lambda_M(f_0^{x,y,q}, f_1^{x,y,q}, f_2^{x,y,q})| \gg \delta \quad \text{for all } (x, y) \in \mathcal{P}.$$

By our second counting operator inverse theorem (Theorem 6.2), either $M \ll \delta^{-O(1)}$ or, for each pair $(x, y) \in \mathcal{P}$, there exists $\alpha_{x,y} \in \mathbb{T}$ and $q_{x,y} \ll \delta^{-O(1)}$ with $\|q_{x,y}\alpha_{x,y}\|_{\mathbb{T}} \ll \delta^{-O(1)}/M^2$ and such that

$$\mathbb{E}_{x'\in[M]}\big|\mathbb{E}_{y'\in[2M^2]} f_2^{x,y,q}(x', y') e(\alpha_{x,y} y')\big| \gg \delta^{O(1)}.$$

Notice that we can ignore the possibility that $M \ll \delta^{-O(1)}$, since in this case *all* positive integers $\widetilde{M}$ satisfy $\widetilde{M} \gg \delta^{O(1)} M$. Pigeon-holing in the major arcs, as in Lemma 5.3, we obtain $\alpha \in \mathbb{T}$ and $\tilde{q} \ll \delta^{-O(1)}$ with $\|\tilde{q}\alpha\|_{\mathbb{T}} \ll \delta^{-O(1)}/M^2$ and such that

$$\sum_{x,y} \mathbb{E}_{x'\in[M]}\big|\mathbb{E}_{y'\in[2M^2]} f_2^{x,y,q}(x', y') e(\alpha y')\big| \gg \delta^{O(1)} N_1 N_2. \tag{6.4}$$

We note that

$$\Big|e(\alpha y') - e\Big(\alpha(y' + \tilde{q}^2\tilde{d})\Big)\Big| \ll \delta^{-O(1)}|\tilde{d}|/M^2.$$

Combining this with the fact that the set of pairs (x, y) that contribute to (6.4) is contained in $[-2N_1, 2N_1] \times [-2N_2, 2N_2]$, we deduce that for any positive integer $\widetilde{M}$, we either have $\widetilde{M} \gg \delta^{O(1)} M$ or

$$\sum_{x,y} \mathbb{E}_{x'\in[M]}\Big|\mathbb{E}_{y'\in[2M^2]} f_2^{x,y,q}(x', y') \mathbb{E}_{\tilde{d}\in[\widetilde{M}^2]} e\Big(\alpha(y' + \tilde{q}^2\tilde{d})\Big)\Big| \gg \delta^{O(1)} N_1 N_2. \tag{6.5}$$

Instead of averaging the phase in (6.5) over shifts of the form $\tilde{q}^2\tilde{d}$, we would like to change variables and average $f_2^{x,y,q}(x', \cdot)$ over these shifts. We have the identity

$$\mathbb{E}_{y'\in[2M^2]} f_2^{x,y,q}(x', y') \mathbb{E}_{\tilde{d}\in[\widetilde{M}^2]} e\Big(\alpha(y' + \tilde{q}^2\tilde{d})\Big) = \mathbb{E}_{\tilde{d}\in[\widetilde{M}^2]} \mathbb{E}_{y'-\tilde{q}^2\tilde{d}\in[2M^2]} f_2^{x,y,q}(x', y' - \tilde{q}^2\tilde{d}) e(\alpha y')\,.$$

Hence, estimating the symmetric difference of the sets of summands gives that

$$\mathbb{E}_{y'\in[2M^2]} f_2^{x,y,q}(x',y')\mathbb{E}_{\tilde d\in[\widetilde{M}^2]} e\Big(\alpha(y'+\tilde q^2\tilde d)\Big) \\ = \mathbb{E}_{y'\in[2M^2]}\mathbb{E}_{\tilde d\in[\widetilde{M}^2]} f_2^{x,y,q}(x',y'-\tilde q^2\tilde d)e(\alpha y') + O\left(\frac{\tilde q^2\widetilde{M}^2}{M^2}\right).$$

It follows that either $\widetilde{M} \gg \delta^{O(1)}M$ or we have

$$\sum_{x,y}\mathbb{E}_{x'\in[M]}\Big|\mathbb{E}_{y'\in[2M^2]}\mathbb{E}_{\tilde d\in[\widetilde{M}^2]} f_2^{x,y,q}(x',y'-\tilde q^2\tilde d)e(\alpha y')\Big| \gg \delta^{O(1)}N_1N_2.$$

Applying the triangle inequality gives

$$\mathbb{E}_{x'\in[M]}\mathbb{E}_{y'\in[2M^2]}\sum_{x,y}\Big|\mathbb{E}_{\tilde d\in[\widetilde{M}^2]} f_2^{x,y,q}(x',y'-\tilde q^2\tilde d)\Big| \gg \delta^{O(1)}N_1N_2.$$

Taking a maximum over x' and y', then expanding the definition of each $f_2^{x,y,q}$ and changing variables, we have

$$\sum_{x,y}\Big|\mathbb{E}_{\tilde d\in[\widetilde{M}^2]} f_2(x,y-q^2\tilde q^2\tilde d)\Big| \gg \delta^{O(1)}N_1N_2.$$

An application of Cauchy–Schwarz to double the $\tilde d$ variables then yields

$$\delta^{O(1)}N_1N_2 \ll \sum_{x,y}\mathbb{E}_{\tilde d_1,\tilde d_2\in[\widetilde{M}^2]} f_2(x,y-q^2\tilde q^2\tilde d_1)\overline{f_2(x,y-q^2\tilde q^2\tilde d_2)} \\ \leqslant \sum_{x,y}\Big|\mathbb{E}_{\tilde d_1,\tilde d_2\in[\widetilde{M}^2]} f_2\Big(x,y+q^2\tilde q^2(\tilde d_1-\tilde d_2)\Big)\Big|$$

This gives (6.2), on recalling the definition (2.7) of the Fejér kernel μ_H. □

We end this section by deriving the inverse theorem claimed in our introduction (Theorem 1.4), where there is no occurrence of the Fejér kernel μ_N in the counting operator.

Proof of Theorem 1.4. Set $M := \lceil 2\delta^{-1}N\rceil$. Due to the support of the f_i, if

$$f_0(x,y)f_1(x+d,y)f_2(x,y+d^2) \neq 0$$

then $x\in[N]$, $y\in[N^2]$ and $|d|<N$. It follows that

$$\big|f_0(x,y)f_1(x+d,y)f_2(x,y+d^2) - f_0(x,y)f_1(x+d,y)f_2(x,y+d^2)M\mu_M(d)\big| \\ \leqslant \tfrac{\delta}{2}1_{[N]}(x)1_{[N^2]}(y).$$

Thus the hypothesis (1.2) implies that

$$|\Lambda_M(f_0,f_1,f_2)| \gg \delta\tfrac{N^4}{M^4} \gg \delta^{O(1)}.$$

Applying Theorem 6.1 and Theorem 6.2, either $N \ll \delta^{-O(1)}$ or there exist $\alpha,\beta\in\mathbb{T}$ and $q_1,q_2 \ll \delta^{-O(1)}$ such that $\|q_1\alpha\| \ll \delta^{-O(1)}/N$ and $\|q_2\beta\| \ll \delta^{-O(1)}/N^2$, as well as (1.3) holds. The theorem follows on taking $q := q_1q_2$. □

7. ENERGY INCREMENT

We are now in a position to run an energy increment argument in order to obtain a modified counting operator that is close to our original count: see our outline in §3 for more on this. Green and Tao [GT12] call the argument underlying the following result a *local Koopman-von Neuman theorem.*

Theorem 7.1 (Energy increment to a subprogression)**.** *Let* $0 < \varepsilon \leqslant 1/2$ *and let* $f_0, f_1, f_2 : \mathbb{Z}^2 \to \mathbb{C}$ *be 1-bounded functions, each with support contained in* $[N_1] \times [N_2]$ *where* $N_1 \geqslant N_2^{1/2}$. *Then either* $N_2 \leqslant \exp\big(\varepsilon^{-O(1)}\big)$, *or there exist* $q \leqslant \exp\big(\varepsilon^{-O(1)}\big)$ *and* $M \geqslant N_2^{1/2}/\exp\big(\varepsilon^{-O(1)}\big)$ *such that*

$$\left|\sum_{x,y,d} \mu_{\varepsilon M}(d) f_0(x,y) f_1(x+qd,y)\Big[f_2\big(x,y+q^2d^2\big) - \sum_{\tilde d} \mu_{M^2}(\tilde d) f_2\Big(x, y+q^2\Big(\tilde d + d^2\Big)\Big)\Big]\right| \leqslant \varepsilon N_1 N_2. \quad (7.1)$$

Moreover, we may choose M *sufficiently small to ensure that* $qM \leqslant \varepsilon N_2^{1/2}$. *Here* $\mu_{\varepsilon M}$ *and* μ_{M^2} *denote Fejér kernels, as defined in* (2.7).

Proof. We perform an iterative procedure. At stage 0, we take $q = q_0 := 1$ and $M = M_0 := \lfloor \varepsilon N_2^{1/2} \rfloor$. Suppose that at stage n of our iteration we have a positive integers $q = q_n \leqslant \varepsilon^{-O(n)}$ and $M = M_n \in [\varepsilon^{O(n+1)} N_2^{1/2}, \varepsilon N_2^{1/2}/q]$ that satisfy the energy lower bound

$$\sum_{x,y}\left|\sum_{\tilde d} f_2(x, y+q^2\tilde d)\mu_{M^2}(\tilde d)\right|^2 \gg n\varepsilon^{O(1)} N_1 N_2. \quad (7.2)$$

Given this, we query whether (7.1) holds or not. **If** (7.1) **holds, then the process terminates.** Therefore suppose that (7.1) does not hold.

Write

$$g_2(x,y) := f_2(x,y) - \sum_{\tilde d} \mu_{M^2}(\tilde d) f_2\Big(x, y+q^2\tilde d\Big),$$

which is a function with support contained in $[-N_1, N_1] \times [-N_2, N_2]$, since $q^2M^2 \leqslant N_2$. Applying the inverse theorem for our asymmetric counting operator (Theorem 6.3), there exists a positive integer $\tilde q$ with $\tilde q \leqslant \varepsilon^{-O(1)}$ such that for any positive integer $\widetilde M$, either $\widetilde M \geqslant \varepsilon^{O(1)} M$ or

$$\sum_{x,y}\left|\sum_{\tilde d} \mu_{\widetilde{M}^2}(\tilde d) g_2(x, y+q^2\tilde q^2 \tilde d)\right| \gg \varepsilon^{O(1)} N_1 N_2. \quad (7.3)$$

Provided that it is not the case that $M \ll \varepsilon^{-O(1)}$, we can find a positive integer $\widetilde M \geqslant \varepsilon^{O(1)} M$ sufficiently small to guarantee (7.3). **If** $M \ll \varepsilon^{-O(1)}$ **then the process terminates**.

On writing $f_2^x(y) := f_2(x,y)$ and $\mu_{q,H}(y) := 1_{q\cdot\mathbb{Z}}(y)\mu_H(y/q)$, we may expand the definition of g_2 to re-formulate (7.3) as[8]

$$\sum_x \left\| f_2^x * \mu_{q^2\tilde q^2, \widetilde{M}^2} - f_2^x * \mu_{q^2,M^2} * \mu_{q^2\tilde q^2,\widetilde{M}^2}\right\|_1 \gg \varepsilon^{O(1)} N_1 N_2.$$

One can check that we have the bound

$$\left\| \mu_{q^2,M^2} * \mu_{q^2\tilde q^2,\widetilde{M}^2} - \mu_{q^2,M^2}\right\|_1 \ll \frac{\tilde q^2 \widetilde{M}^2}{M^2} + \frac{\tilde q^4 \widetilde{M}^4}{M^4}.$$

[8]Here we have normalised the $L^1(\mathbb{Z})$ norm as in (2.1).

Decreasing $\widetilde{M}$ if necessary, but still retaining a lower bound of the form $\widetilde{M} \geqslant \varepsilon^{O(1)}M$, we have

$$\sum_x \Big\| f_2^x * \mu_{q^2\tilde{q}^2,\widetilde{M}^2} - f_2^x * \mu_{q^2,M^2} \Big\|_1 \gg \varepsilon^{O(1)} N_1 N_2.$$

Applying the Cauchy–Schwarz inequality then gives

$$\sum_x \Big\| f_2^x * \mu_{q^2\tilde{q}^2,\widetilde{M}^2} - f_2^x * \mu_{q^2,M^2} \Big\|_2^2 \gg \varepsilon^{O(1)} N_1 N_2.$$

We wish to replace the above, which measures the energy of a difference of two functions $f-g$, with the difference of their energies $\|f\|_2^2 - \|g\|_2^2$. From Pythagoras' theorem, we know that to make this substitution we must show that $f-g$ and g are (approximately) orthogonal. More concretely, writing L^2-norms in terms of inner products, we have

$$\begin{aligned} \big\| f_2^x * \mu_{q^2\tilde{q}^2,\widetilde{M}^2} \big\|_2^2 - \big\| f_2^x * \mu_{q^2,M^2} \big\|_2^2 &= \big\| f_2^x * \mu_{q^2\tilde{q}^2,\widetilde{M}^2} - f_2^x * \mu_{q^2,M^2} \big\|_2^2 \\ &\quad - 2\,\mathrm{Re}\big\langle f_2^x * \mu_{q^2,M^2}, f_2^x * \big(\mu_{q^2,M^2} - \mu_{q^2\tilde{q}^2,\widetilde{M}^2}\big)\big\rangle. \end{aligned} \tag{7.4}$$

So we wish to show that the latter inner product is small.

Suppose otherwise, so that the inner product in (7.4) is at least $\delta\|f_2^x\|_2^2$. Then Parseval's identity and the convolution identity give

$$\begin{aligned} \delta\|f_2^x\|_2^2 &\leqslant \mathrm{Re}\big\langle f_2^x * \mu_{q^2,M^2}, f_2^x * \big(\mu_{q^2,M^2} - \mu_{q^2\tilde{q}^2,\widetilde{M}^2}\big)\big\rangle \\ &= \int_{\mathbb{T}} \big|\widehat{f_2^x}(\alpha)\big|^2 \widehat{\mu}_{q^2,M^2}(\alpha)\Big[\widehat{\mu}_{q^2,M^2}(\alpha) - \widehat{\mu}_{q^2\tilde{q}^2,\widetilde{M}^2}(\alpha)\Big]\,\mathrm{d}\alpha \\ &\leqslant \|f_2^x\|_2^2 \sup_\alpha \Big(\widehat{\mu}_{q^2,M^2}(\alpha)\Big[\widehat{\mu}_{q^2,M^2}(\alpha) - \widehat{\mu}_{q^2\tilde{q}^2,\widetilde{M}^2}(\alpha)\Big]\Big). \end{aligned}$$

By non-negativity of $\widehat{\mu}_{q^2,M^2}$, there exists $\alpha \in \mathbb{T}$ such that both of the following hold

$$\widehat{\mu}_{q^2,M^2}(\alpha) \geqslant \delta \quad \text{and} \quad \widehat{\mu}_{q^2,M^2}(\alpha) - \widehat{\mu}_{q^2\tilde{q}^2,\widetilde{M}^2}(\alpha) \geqslant \delta.$$

Yet, we know from summing the geometric series that if

$$M^{-4}\big|\hat{1}_{[M^2]}(q^2\alpha)\big|^2 = \widehat{\mu}_{q^2,M^2}(\alpha) \geqslant \delta,$$

then $\|q^2\alpha\|_{\mathbb{T}} \leqslant \delta^{-1/2}/M^2$. Yet, for such values of α, we have

$$\widehat{\mu}_{q^2\tilde{q}^2,\widetilde{M}^2}(\alpha) = \big|\mathbb{E}_{\tilde{d}\in[\widetilde{M}^2]} e(\alpha q^2\tilde{q}^2\tilde{d})\big|^2 \geqslant \big(1 - \frac{2\pi\tilde{q}^2\widetilde{M}^2}{\delta^{1/2}M^2}\big)^2 \geqslant 1 - \frac{4\pi\tilde{q}^2\widetilde{M}^2}{\delta^{1/2}M^2}.$$

Thus, taking $\delta = (5\pi\tilde{q}^2\widetilde{M}^2/M^2)^{2/3}$ gives the contradiction that such values of α satisfy $\widehat{\mu}_{q^2\tilde{q}^2,\widetilde{M}^2}(\alpha) > 1-\delta$ and $\widehat{\mu}_{q^2,M^2}(\alpha) > 1$.

We have therefore deduced that

$$\mathrm{Re}\Big\langle f_2^x * \mu_{q^2,M^2}, f_2^x * \Big(\mu_{q^2,M^2} - \mu_{q^2\tilde{q}^2,\widetilde{M}^2}\Big)\Big\rangle \ll \Big(\tfrac{\tilde{q}\widetilde{M}}{M}\Big)^{4/3} \|f_2^x\|_2^2.$$

Putting all of the above together, we have that

$$\sum_x \Big(\big\|f_2^x * \mu_{q^2\tilde{q}^2,\widetilde{M}^2}\big\|_2^2 - \|f_2^x * \mu_{q^2,M^2}\|_2^2\Big) \geqslant \Omega\big(\varepsilon^{O(1)}N_1N_2\big) - O\bigg(\Big(\tfrac{\tilde{q}\widetilde{M}}{M}\Big)^{4/3} N_1N_2\bigg).$$

Decreasing $\widetilde{M}$ if necessary, but still retaining a lower bound of the form $\widetilde{M} \geqslant \varepsilon^{O(1)}M$, we obtain the energy increment

$$\sum_x \big\|f_2^x * \mu_{q^2\tilde{q}^2,\widetilde{M}^2}\big\|_2^2 \geqslant \sum_x \big\|f_2^x * \mu_{q^2,M^2}\big\|_2^2 + \Omega\big(\varepsilon^{O(1)}N_1N_2\big).$$

With this bound in hand, our inductive hypotheses are satisfied (see (7.2)), and our procedure iterates to the next stage.

Since the energy (7.2) is bounded above by N_1N_2, our procedure must terminate at some stage $n \ll \varepsilon^{O(1)}$, yielding either $M = M_n \ll \varepsilon^{-O(1)}$ or (7.1). Unpacking the rate at which $q = q_n$ and $M = M_n$ change, the former possibility leads to $N \leqslant \exp(\varepsilon^{-O(1)})$, whilst the latter gives our other desired outcome. □

8. Obtaining a popular difference

Finally we prove the remaining results claimed in our introduction §1.

Theorem 8.1 (Existence of a popular difference)**.** *Let $A \subset [N_1] \times [N_2]$ with $|A| \geqslant \delta N_1 N_2$ and $N_1 \geqslant N_2^{1/2}$. Given $0 < \varepsilon \leqslant 1/2$, either $N_2 \leqslant \exp\big(\varepsilon^{-O(1)}\big)$ or there exist $q \leqslant \exp\big(\varepsilon^{-O(1)}\big)$ and $M \geqslant N_2^{1/2}/\exp\big(\varepsilon^{-O(1)}\big)$ such that*

$$\sum_{x,y,d} \mu_M(d) 1_A(x,y) 1_A(x+qd,y) 1_A(x, y+q^2d^2) \geqslant \big(\delta^3 - \varepsilon\big) N_1 N_2.$$

Proof. Let us apply our energy increment argument to 1_A, so we take $f_0 = f_1 = f_2 = 1_A$ in Theorem 7.1. This tells us that we either have $N_2 \leqslant \exp\big(\varepsilon^{-O(1)}\big)$ or there exist $q \leqslant \exp\big(\varepsilon^{-O(1)}\big)$ and $M \geqslant N_2^{1/2}/\exp\big(\varepsilon^{-O(1)}\big)$ such that $qM \leqslant \varepsilon N_2^{1/2}$ and

$$\Bigg|\sum_{x,y,d} \mu_{\varepsilon M}(d) 1_A(x,y) 1_A(x+qd,y) \Bigg[1_A\big(x, y+q^2d^2\big) - \sum_{\tilde d} \mu_{M^2}(\tilde d) 1_A\Big(x, y + q^2\Big(d^2+\tilde d\Big)\Big)\Bigg]\Bigg| \leqslant \varepsilon N_1 N_2.$$

Focusing on the second counting operator, we have

$$\sum_{x,y,d} \mu_{\varepsilon M}(d) 1_A(x,y) 1_A(x+qd,y) \sum_{\tilde d} \mu_{M^2}(\tilde d) 1_A\Big(x, y+q^2\Big(d^2+\tilde d\Big)\Big) = \sum_{x,y,d} \mu_{\varepsilon M}(d) 1_A(x,y) 1_A(x+qd,y) \sum_{\tilde d} \mu_{M^2}\Big(\tilde d - d^2\Big) 1_A\Big(x, y+q^2\tilde d\Big).$$

The Lipschitz properties of μ_{M^2} give that

$$\Big|\mu_{M^2}\Big(\tilde d - d^2\Big) - \mu_{M^2}(\tilde d)\Big| \ll d^2 M^{-4}.$$

Hence our restriction to $|d| < \varepsilon M$ gives

$$\sum_{x,y,d} \mu_{\varepsilon M}(d) 1_A(x,y) 1_A(x+qd,y) \sum_{\tilde d} \mu_{M^2}(\tilde d) 1_A\Big(x, y+q^2\Big(d^2+\tilde d\Big)\Big) \geqslant \sum_{x,y,d} \mu_{\varepsilon M}(d) 1_A(x,y) 1_A(x+qd,y) \sum_{\tilde d} \mu_{M^2}(\tilde d) 1_A\Big(x, y+q^2\tilde d\Big) - O(\varepsilon N_1 N_2).$$

Expanding the definition (2.7) of the Fejér kernel, then changing variables, we have

$$\sum_{x,y,d} \mu_{\varepsilon M}(d) 1_A(x,y) 1_A(x+qd,y) \sum_{\tilde d} \mu_{M^2}(\tilde d) 1_A\Big(x, y+q^2\tilde d\Big) \geqslant \sum_{x,y} \mathbb{E}_{d_1,d_2 \in [\varepsilon M]} \mathbb{E}_{\tilde d_1, \tilde d_2 \in [M^2]} 1_A\big(x+qd_1, y+q^2\tilde d_1\big) 1_A\big(x+qd_2, y+q^2\tilde d_1\big) 1_A\big(x+qd_1, y+q^2\tilde d_2\big).$$

By an application of Sidorenko's inequality for paths of length three (3.3), sometimes termed the Blakley-Roy inequality, we have

$$\mathbb{E}_{d_1,d_2\in[\varepsilon M]}\mathbb{E}_{\tilde{d}_1,\tilde{d}_2\in[M^2]}1_A(x+qd_1,y+q^2\tilde{d}_1)1_A(x+qd_2,y+q^2\tilde{d}_1)1_A(x+qd_1,y+q^2\tilde{d}_2)$$
$$\geqslant \Big(\mathbb{E}_{d\in[\varepsilon M]}\mathbb{E}_{\tilde{d}\in[M^2]}1_A(x+qd,y+q^2\tilde{d})\Big)^3.$$

If the sum inside the cube is non-zero, then (because $A\subset[N_1]\times[N_2]$) we have

$$(x,y)\in(-q\varepsilon M,N_1)\times(-q^2M^2,N_2)\subset(-\varepsilon N_1,N_1)\times(-\varepsilon N_2,N_2).$$

So, again by Hölder's inequality,

$$\sum_{x,y}\Big(\mathbb{E}_{d\in[\varepsilon M]}\mathbb{E}_{\tilde{d}\in[M^2]}1_A(x+qd,y+q^2\tilde{d})\Big)^3$$
$$\geqslant N_1^{-2}N_2^{-2}(1+\varepsilon)^{-4}\left(\sum_{x,y}\mathbb{E}_{d\in[M]}\mathbb{E}_{\tilde{d}\in[M^2]}1_A(x+qd,y+q^2\tilde{d})\right)^3$$
$$\geqslant \delta^3N_1N_2(1-\varepsilon)^4.$$

Putting everything together gives

$$\sum_{x,y,d}\mu_{\varepsilon M}(d)1_A(x,y)1_A(x+qd,y)1_A(x,y+q^2d^2)\geqslant\big(\delta^3-O(\varepsilon)\big)N_1N_2,$$

as required. □

Taking $N_1=N_2$ in Theorem 8.1 yields Theorem 1.2. Taking $N_2=N_1^2$ yields the following, which is perhaps more natural.

Corollary 8.2. *Let $A\subset[N]\times[N^2]$ with $|A|\geqslant\delta N^3$ and let $0<\varepsilon\leqslant1/2$. Either $N\leqslant\exp\big(\varepsilon^{-O(1)}\big)$ or there exists $d\neq0$ such that*

$$\#\big\{(x,y)\in A:(x+d,y),(x,y+d^2)\in A\big\}\geqslant\big(\delta^3-\varepsilon\big)N^3.$$

From this we derive the existence of a one-dimensional popular common difference, which is claimed in Theorem 1.3.

Corollary 8.3 (Theorem 1.3, re-stated)**.** *Let $A\subset[N]$ with $|A|\geqslant\delta N$ and let $0<\varepsilon\leqslant1/2$. Either $N\leqslant\exp\big(\varepsilon^{-O(1)}\big)$ or there exists $d\neq0$ such that*

$$\#\big\{x\in A:x+d,x+d^2\in A\big\}\geqslant\big(\delta^3-\varepsilon\big)N.$$

Proof. Define

$$\tilde{A}:=\big\{(x,y)\in[N^{1/2}]\times[N]:x+y\in A\big\}.$$

Let us bound the size of $\tilde{A}$. We have:

$$|\tilde{A}|\geqslant\sum_{a\in A\cap(N^{1/2},N]}\sum_{x\in[N^{1/2}]}1_{[N]}(a-x)\geqslant\big(\delta N-N^{1/2}\big)N^{1/2}\geqslant(\delta-\varepsilon)N^{3/2},$$

or else $N\ll\varepsilon^{-O(1)}$.

Applying Theorem 8.1, either $N\leqslant\exp\big(\varepsilon^{-O(1)}\big)$ or there exists an integer $d\neq0$ such that

$$\#\big\{(x,y)\in[N^{1/2}]\times[N]:x+y,\ x+y+d,\ x+y+d^2\in A\big\}\geqslant\big((\delta-\varepsilon)^3-\varepsilon\big)N^{3/2}.$$

Taking a maximum over $x\in[N^{1/2}]$, we obtain some x_0 such that

$$\#\big\{y\in[N]:x_0+y,\ x_0+y+d,\ x_0+y+d^2\in A\big\}\geqslant\big((\delta-\varepsilon)^3-\varepsilon\big)N.$$

The result follows on observing that (by the binomial theorem) $(\delta-\varepsilon)^3\geqslant\delta^3-4\varepsilon$. □

REFERENCES

[BHK05] Vitaly Bergelson, Bernard Host, and Bryna Kra. Multiple recurrence and nilsequences. *Invent. Math.*, 160(2):261–303, 2005. With an appendix by Imre Ruzsa. ↑4

[BL96] Vitaly Bergelson and Alexander Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.*, 9(3):725–753, 1996. ↑1, ↑8

[BM22] Thomas F. Bloom and James Maynard. A new upper bound for sets with no square differences. *Compos. Math.*, 158(8):1777–1798, 2022. ↑3

[BSST22] Aaron Berger, Ashwin Sah, Mehtaab Sawhney, and Jonathan Tidor. Popular differences for matrix patterns. *Trans. Amer. Math. Soc.*, 375(4):2677–2704, 2022. ↑4

[CFH11] Qing Chu, Nikos Frantzikinakis, and Bernard Host. Ergodic averages of commuting transformations with distinct degree polynomial iterates. *Proc. Lond. Math. Soc. (3)*, 102(5):801–842, 2011. ↑4

[Chu11] Qing Chu. Multiple recurrence for two commuting transformations. *Ergodic Theory Dynam. Systems*, 31(3):771–792, 2011. ↑4

[DLMS21] Sebastián Donoso, Anh Ngoc Le, Joel Moreira, and Wenbo Sun. Optimal lower bounds for multiple recurrence. *Ergodic Theory Dynam. Systems*, 41(2):379–407, 2021. ↑4

[FK78] Hillel Furstenberg and Yitzhak Katznelson. An ergodic Szemerédi theorem for commuting transformations. *J. Analyse Math.*, 34:275–291 (1979), 1978. ↑1

[FP19] Jacob Fox and Huy Tuan Pham. Popular progression differences in vector spaces II. *Discrete Anal.*, pages Paper No. 16, 39, 2019. ↑4

[FP21] Jacob Fox and Huy Tuan Pham. Popular progression differences in vector spaces. *Int. Math. Res. Not. IMRN*, (7):5261–5289, 2021. ↑4

[FPZ22] Jacob Fox, Huy Tuan Pham, and Yufei Zhao. Tower-type bounds for Roth's theorem with popular differences. *J. Eur. Math. Soc. (to appear)*, 2022. ↑2, ↑4, ↑13

[Fra08] Nikos Frantzikinakis. Multiple ergodic averages for three polynomials and applications. *Trans. Amer. Math. Soc.*, 360(10):5435–5475, 2008. ↑4

[FSS+20] Jacob Fox, Ashwin Sah, Mehtaab Sawhney, David Stoner, and Yufei Zhao. Triforce and corners. *Math. Proc. Cambridge Philos. Soc.*, 169(1):209–223, 2020. ↑4

[Gow98] W. Timothy Gowers. A new proof of Szemerédi's theorem for arithmetic progressions of length four. *Geom. Funct. Anal.*, 8(3):529–551, 1998. ↑1

[Gow01] W. Timothy Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001. ↑3

[Gow22] W. Timothy Gowers. The enduring appeal of Szemerédi's theorem. *Newsletter of the London Mathematical Society*, 500:37–44, May 2022. ↑1

[Gre05a] Ben Green. Finite field models in additive combinatorics. In *Surveys in combinatorics 2005*, volume 327 of *London Math. Soc. Lecture Note Ser.*, pages 1–27. Cambridge Univ. Press, Cambridge, 2005. ↑7

[Gre05b] Ben Green. A Szemerédi-type regularity lemma in abelian groups, with applications. *Geom. Funct. Anal.*, 15(2):340–376, 2005. ↑4

[GT10] Ben Green and Terence Tao. An arithmetic regularity lemma, an associated counting lemma, and applications. In *An irregular mind*, volume 21 of *Bolyai Soc. Math. Stud.*, pages 261–334. János Bolyai Math. Soc., Budapest, 2010. ↑4

[GT12] Ben Green and Terence Tao. New bounds for Szemerédi's theorem, Ia: Progressions of length 4 in finite field geometries revisited. *ArXiv e-prints*, 2012. ↑22

[GT12nil] Ben Green and Terence Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)*, 175(2):465–540, 2012. ↑12

[HLY21] Rui Han, Michael T. Lacey, and Fan Yang. A polynomial Roth theorem for corners in finite fields. *Mathematika*, 67(4):885–896, 2021. ↑4, ↑7

[KKL] Noah Kravitz, Borys Kuca, and James Leng. Corners with polynomial side length. *preprint*, 2024. ↑1, ↑3

[KMT22] Ben Krause, Mariusz Mirek, and Terence Tao. Pointwise ergodic theorems for non-conventional bilinear polynomial averages. *Ann. of Math. (2)*, 195(3):997–1109, 2022. ↑2

[Kuc24] Borys Kuca. Multidimensional polynomial Szemerédi theorem in finite fields for polynomials of distinct degrees. *Israel J. Math.*, 259(2):589–620, 2024. ↑4, ↑7

[LM09] Neil Lyall and Ákos Magyar. Polynomial configurations in difference sets. *J. Number Theory*, 129(2):439–450, 2009. ↑3

[LM13] Neil Lyall and Ákos Magyar. Optimal polynomial recurrence. *Canad. J. Math.*, 65(1):171–194, 2013. ↑5

[Man21] Matei Mandache. A variant of the Corners theorem. *Math. Proc. Cambridge Philos. Soc.*, 171(3):607–621, 2021. ↑4

[Pel20] Sarah Peluse. Bounds for sets with no polynomial progressions. *Forum Math. Pi*, 8:e16, 55, 2020. ↑4, ↑10

[Pel24] Sarah Peluse. Subsets of $\mathbb{F}_p^n \times \mathbb{F}_p^n$ without L-shaped configurations. *Compos. Math.*, 160(1):176–236, 2024. ↑4

[PP19] Sarah Peluse and Sean Prendiville. Quantitative bounds in the non-linear Roth theorem. *ArXiv e-prints*, 2019. ↑2, ↑4, ↑8, ↑9, ↑10, ↑16, ↑17

[PP22] Sarah Peluse and Sean Prendiville. A polylogarithmic bound in the nonlinear Roth theorem. *Int. Math. Res. Not. IMRN*, (8):5658–5684, 2022. ↑2, ↑10, ↑13

[Pre15] Sean Prendiville. Matrix progressions in multidimensional sets of integers. *Mathematika*, 61(1):14–48, 2015. ↑3, ↑4

[Pre17] Sean Prendiville. Quantitative bounds in the polynomial Szemerédi theorem: the homogeneous case. *Discrete Anal.*, pages 34, Paper No. 5, 2017. ↑3

[Sár78] András Sárközy. On difference sets of sequences of integers. I. *Acta Math. Acad. Sci. Hungar.*, 31(1–2):125–149, 1978. ↑3

[Shk06] Ilya D. Shkredov. On a generalization of Szemerédi's theorem. *Proc. London Math. Soc. (3)*, 93(3):723–760, 2006. ↑3

[SSZ21] Ashwin Sah, Mehtaab Sawhney, and Yufei Zhao. Patterns without a popular difference. *Discrete Anal.*, Paper No. 8, 30, 2021. ↑4

[Tao12] Terence Tao. *Higher order Fourier analysis*, volume 142 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012. ↑12

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY
Email address: speluse@stanford.edu

SCHOOL OF MATHEMATICAL SCIENCES, LANCASTER UNIVERSITY
Email address: s.prendiville@lancaster.ac.uk

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF KENTUCKY
Email address: xuancheng.shao@uky.edu