EFFECTIVE BOUNDS FOR ROTH'S THEOREM WITH SHIFTED SQUARE COMMON DIFFERENCE

SARAH PELUSE, ASHWIN SAH, AND MEHTAAB SAWHNEY

Abstract. Let S be a subset of $\{1,\ldots,N\}$ avoiding the nontrivial progressions $x, x+y^2-1, x+2(y^2-1)$. We prove that $|S| \ll N/\log_m N$, where $\log_m$ is the m-fold iterated logarithm and $m \in \mathbf{N}$ is an absolute constant. This answers a question of Green.

1. Introduction

This paper contributes to the program of proving reasonable bounds for sets lacking polynomial progressions, a problem posed by Gowers [14, Problem 11.4] after his proof of the first reasonable bounds in Szemerédi's theorem on arithmetic progressions [13, 15].

In the late 1970's, Furstenberg [11] and Sárközy [46] independently proved that any subset of the natural numbers having positive upper density must contain a nontrivial[1] instance of the progression $x, x+y^2$. Furstenberg's proof, which appeared in the same paper in which he introduced his eponymous correspondence principle and used it to give a proof of Szemerédi's theorem via ergodic theory, produced no quantitative bounds, but Sárközy's proof, which was via the circle method, showed that if $S \subseteq \{1,\ldots,N\}$ contains no nontrivial progressions $x, x+y^2$, then

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

Sárközy [47] extended his argument to all progressions of the form $x, x+y^n$ with bounds of the same quality, which were later improved by Balog, Pelikán, Pintz, and Szemerédi [3] and then Bloom and Maynard [7]. Slijepčević [48] further extended Sárközy's argument to work for all two-term polynomial progressions $x, x+P(y)$ where $P(0)=0$.

Note that it cannot possibly be the case that the Furstenberg–Sárközy theorem holds for every single polynomial progression $x, x+P(y)$ with $P \in \mathbf{Z}[y]$. Indeed, the set of multiples of 3 have positive density in the integers, but contain no progressions of the form $x, x+y^2+1$ because y^2+1 is never divisible by 3 when y is an integer. Polynomials $P \in \mathbf{Z}[y]$ for which any subset of the natural numbers with positive upper density must contain a nontrivial polynomial progression of the form $x, x+P(y)$ are called *intersective*. Kamae and Mendés France [25] showed that a polynomial is intersective if and only if it has a root modulo every natural number. Polynomials $P \in \mathbf{Z}[y]$ with $P(0)=0$ clearly satisfy this criterion, and so does y^2-1 and, more generally, any other polynomial with an integer root. There also exist polynomials, like $(y^3-19)(y^2+y+1)$, that are intersective but have no rational roots. The argument of Kamae and Mendés France produced no quantitative bounds, but Lucier [35] generalized Sárközy's argument to show that if $P \in \mathbf{Z}[y]$ is intersective and $S \subseteq \{1,\ldots,N\}$ contains no nontrivial progressions $x, x+P(y)$, then

$$|S| \ll_P \frac{N}{(\log N)^{1/(\deg P-1)-o(1)}}.$$

The bound has since been improved by Rice [44].

[1] Here, *nontrivial* means that both terms of the progression are distinct.

Bergelson and Leibman [5] proved that if $P_1, \dots, P_m \in \mathbf{Z}[y]$ are any polynomials satisfying $P_1(0) = \cdots = P_m(0) = 0$, then any subset of the natural numbers with positive upper density must contain a nontrivial polynomial progression of the form

$$x, x + P_1(y), \dots, x + P_m(y). \tag{1.1}$$

Their argument, which was via ergodic theory, produced no quantitative bounds. Gowers's proof of Szemerédi's theorem provides quantitative bounds in the case that $P_1, \dots, P_m$ are all linear. Green [17] proved quantitative bounds for subsets of integers avoiding three-term arithmetic progressions with common difference equal to the sum of two squares. This was substantially generalized in work of Prendiville [43] to prove the existence of k-term arithmetic progressions with common difference a perfect d-th power. Both papers [17,43] build on Gowers's seminal work [13,15] and, in particular, crucially rely on the homogeneous nature of these polynomial progressions to proceed via the the density increment strategy using the local inverse theorems for the U^s-norms. The progressions considered by Prendiville are the most general to which Gowers's methods can possibly apply, and no effective results were known for any other progressions of length greater than two until recently.

Progress on effective bounds on the size of sets lacking more polynomial progressions was made first in the finite field setting. Bourgain and Chang [8] proved that any $S \subseteq \mathbf{F}_p$ lacking nontrivial *nonlinear Roth configurations* $x, x+y, x+y^2$ has size $|S| \ll p^{14/15}$. Similar polynomial saving bounds were proven in the case of more general progressions $x, x+P_1(y), x+P_2(y)$ for linearly independent polynomials $P_1(y)$ and $P_2(y)$ by the first author [38] and, independently, Dong, Li, and Sawin [10]. While the proofs of these results avoided the use of the inverse theory of the Gowers norms, the arguments did not extend to longer polynomial patterns. The first result in this direction was due to first author [39], who introduced the degree-lowering method and used it to prove power-saving bounds for sets lacking arbitrarily long progressions (1.1) with linearly independent polynomials $P_1, \dots, P_m$. Degree-lowering was then used by Kuca [27] and Leng [33,34] to give effective bounds for subsets of finite fields avoiding various families of polynomial progressions of complexity[2] 1 or greater.

The first author and Prendiville [41,42] adapted the degree-lowering method to the integer setting to prove that any subset S of $\{1, \dots, N\}$ lacking non-linear Roth configurations must satisfy

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

for some absolute constant $c > 0$. This was extended in work of the first author [40] to arbitrarily long progressions (1.1) where the polynomials $P_1, \dots, P_m$ have all distinct degrees. Proving a fully general quantitative polynomial Szemerédi theorem remains a very challenging open problem, and effective bounds for sets lacking polynomial progressions (1.1) of complexity at least one where the polynomials $P_1, \dots, P_m$ are not homogeneous of the same degree are unknown in the integer setting.

Our work establishes the first effective case of the polynomial Szemerédi theorem over the integers where the underlying pattern has complexity higher than one and the polynomials involved are not homogeneous of the same degree.

Theorem 1.1. *There exists a positive integer $m = m_{1.1}$ such that the following holds. If $S \subseteq \{1, \dots, N\}$ is such that S does not contain a progression of the form*

$$x,\ x + y^2 - 1,\ x + 2(y^2 - 1) \qquad (y \neq \pm 1), \tag{1.2}$$

then

$$|S| \ll \frac{N}{\log_m N}.$$

[2]Here, *complexity* refers to true complexity, as defined in [28].

The problem of proving quantitative bounds for sets lacking (1.2) was explicitly raised by Green [16, Problem 11(i)].

Remark. By tracing through our proof, and the inputs from [39], we can take $m = 200$. By inserting some plausible improvements in the quantitative aspects of the theory of nilsequences, our argument would yield a bound of the form $|S| \ll N \exp(-(\log\log N)^c)$ in the theorem; see the discussion at the end of Section 3.

We give an outline of our key definitions, method, and new techniques in Sections 2 and 3, and describe the structure of the paper in Section 3.1.

Acknowledgments. The first author thanks Sean Prendiville for helpful conversations. The second and third authors thank James Leng for helpful clarifications regarding [34, Lemma 6.1]. The third author thanks Dmitrii Zakharov for help with computations with nilpotent groups. The authors thank Ben Green for useful comments. The first author was supported by the NSF Mathematical Sciences Postdoctoral Research Fellowship Program under Grant No. DMS-1903038. The second author was supported by the PD Soros Fellowship. The second and third authors were supported by NSF Graduate Research Fellowship Program DGE-2141064.

2. Notation and key definitions

We use standard asymptotic notation throughout, as follows. For functions $f = f(n)$ and $g = g(n)$, we write $f = O(g)$ or $f \ll g$ to mean that there is a constant C such that $|f(n)| \le C|g(n)|$ for sufficiently large n. Similarly, we write $f = \Omega(g)$ or $f \gg g$ to mean that there is a constant $c > 0$ such that $f(n) \ge c|g(n)|$ for sufficiently large n. Finally, we write $f \asymp g$ or $f = \Theta(g)$ to mean that $f \ll g$ and $g \ll f$, and we write $f = o(g)$ or $g = \omega(f)$ to mean that $f(n)/g(n) \to 0$ as $n \to \infty$. Subscripts on asymptotic notation indicate quantities that should be treated as constants. Furthermore, throughout the paper, we will use the standard notation $\mathbf{T} = \mathbf{R}/\mathbf{Z}$, $\mathbf{N} = \{1, 2, \ldots\}$, $\mathbf{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, $[X] = \{1, 2\ldots, \lfloor X\rfloor\}$, $[\pm X] = \{-\lfloor X\rfloor, \ldots, \lfloor X\rfloor\}$. Finally given a nonzero real t and a set Q we define $t \cdot Q = \{tq : q \in Q\}$.

One nonstandard piece of notation, following work of Tao and Teräväinen [50], is that we let $\mathrm{poly}_m(Q)$ for $Q \ge 2$ denote a quantity bounded above by $\exp(\exp(m^{O(1)}))Q^{\exp(m^{O(1)})}$. For $0 < \delta \le 1/2$, we let $\mathrm{poly}_m(\delta)$ denote a quantity bounded *below* by $\exp(-\exp(m^{O(1)}))\delta^{\exp(m^{O(1)})}$. Throughout the paper we will always assume $\delta \in (0, 1/2]$.

Given a function $f : \mathbf{Z} \to \mathbf{C}$ with $\|f\|_{\ell^1(\mathbf{Z})} < \infty$, we normalize the Fourier transform by defining

$$\widehat{f}(\theta) = \sum_{x \in \mathbf{Z}} f(x) e(-x\theta),$$

where $e(x) = \exp(2\pi i x)$. Using this normalization, the Fourier inversion formula for f satisfying $\|f\|_{\ell^1(\mathbf{Z})} + \|f\|_{\ell^2(\mathbf{Z})} < \infty$ is

$$f(x) = \int_{\mathbf{T}} \widehat{f}(\Theta) e(x\Theta)\, d\Theta.$$

We define the normalized Fejér kernel on $\mathbf{Z}$ to be

$$\mu_H(h) = \frac{1}{\lfloor H\rfloor}\left(1 - \frac{|h|}{\lfloor H\rfloor}\right)_+$$

and write

$$\mu_H(h) = \mu_H(h_1, \ldots, h_d) = \prod_{i=1}^{d} \mu_H(h_i) \tag{2.1}$$

for $h = (h_1, \ldots, h_d) \in \mathbf{Z}^d$.

We define two types of multiplicative discrete derivatives; for any complex-valued function f on $\mathbf{Z}$ and $h, h_1, h_1' \in \mathbf{Z}$, set

$$\Delta_h f(x) := f(x)\overline{f(x+h)} \qquad \text{and} \qquad \Delta'_{(h_1,h_1')} := \overline{f(x+h_1)}f(x+h_1').$$

We will occasionally write $\Delta_h^{(x)} f(x,y)$ and $\Delta'^{(x)}_{(h,h')} f(x,y)$, for example, if there are multiple possible variables to choose from, so that, for example, $\Delta_h^{(x)} f(x,y) = f(x,y)\overline{f(x+h,y)}$. With the definition of Δ' in hand, we can now define the Gowers box and uniformity norms. We will write expressions such as $\Delta'_{(h_1,h_1'),(h_2,h_2')} f(x)$ as shorthand for $\Delta'_{(h_1,h_1')}\Delta'_{(h_2,h_2')} f(x)$, and so on, where convenient; note the order of these operators does not matter.

Definition 2.1. Let $d \in \mathbf{N}$, $Q_1, \ldots, Q_d \subseteq \mathbf{Z}$ be finite subsets, and $f\colon \mathbf{Z} \to \mathbf{C}$. We define the Gowers box-norm of f with respect to $Q_1, \ldots, Q_d$ to be

$$\|f\|_{\square^d_{Q_1,\ldots,Q_d}}^{2^d} := \sum_{x\in\mathbf{Z}} \mathbf{E}_{\substack{h_i,h_i'\in Q_i \\ i=1,\ldots,d}} \Delta'_{(h_1,h_1'),\ldots,(h_d,h_d')} f(x).$$

For $Q = Q_1 = \cdots = Q_d$ define

$$\|f\|_{U^d_Q} := \|f\|_{\square^d_{Q,\ldots,Q}}.$$

Note that our definition differs from that in [40, Definition 2.1], as the sum is not normalized.

For the entirety of the paper, we define

$$W = \prod_{\substack{2\le p\le w \\ p \text{ prime}}} p, \qquad M = \lfloor N^{1/2}W^{-1/2} \rfloor, \qquad \text{and} \qquad P(y) = Wy^2 + y, \tag{2.2}$$

for some parameter w. Eventually, w will be chosen to be a sufficiently slowly growing function of N; throughout the paper, we ensure that various implied constants are independent of W. It is elementary to prove that $W \le 4^w$. As stated, M is a function of a floating parameter N; N, up to a constant factor, will always denote the size of the support of the sets or functions under consideration.

Next, we define the critical counting operators to be used throughout the paper.

Definition 2.2. Given N and given finitely supported functions $f_1, f_2, f_3\colon \mathbf{Z} \to \mathbf{C}$, we define the trilinear operators Λ^W and Λ^{Model} by

$$\Lambda^W(f_1,f_2,f_3) = \sum_{\substack{x\in\mathbf{Z} \\ |k|\le M}} f_1(x)f_2(x+P(k))f_3(x+2P(k))$$

and

$$\Lambda^{\text{Model}}(f_1,f_2,f_3) = \sum_{\substack{x\in\mathbf{Z} \\ d\in\mathbf{Z}}} f_1(x)f_2(z+d)f_3(z+2d)\nu(d),$$

where

$$\nu(d) = \sqrt{\frac{N}{d}}\,\mathbb{1}_{1\le d\le N}. \tag{2.3}$$

We also define the "difference" counting operator

$$\widetilde{\Lambda}(f_1,f_2,f_3) := (NW)^{1/2}\Lambda^W(f_1,f_2,f_3) - \Lambda^{\text{Model}}(f_1,f_2,f_3).$$

Finally, we will repeatedly encounter the following *dual functions* when carrying out our degree-lowering argument.

Definition 2.3. Given functions $f_1, f_2, f_3 \colon \mathbf{Z} \to \mathbf{C}$, we define

$$\mathcal{D}^1(f_2, f_3)(x) = \mathbf{E}_{y\in[\pm M]} f_2(x + P(y))f_3(x + 2P(y)),$$

$$\mathcal{D}^2(f_3, f_1)(x) = \mathbf{E}_{y\in[\pm M]} f_1(x - P(y))f_3(x + P(y)),$$

and

$$\mathcal{D}^3(f_1, f_2)(x) = \mathbf{E}_{y\in[\pm M]} f_1(x - 2P(y))f_2(x - P(y)).$$

These dual functions arise in a key maneuver in the degree-lowering method known as *stashing*, a term coined by Manners. More discussion on stashing can be found in [36], but for us it will almost always refer to the procedure of noting that if $f_1, f_2, f_3 \colon \mathbf{Z} \to \mathbf{C}$ are 1-bounded functions supported in $[N]$ and

$$\left|\Lambda^W(f_1, f_2, f_3)\right| \geq 2\delta NM,$$

then

$$\left|\Lambda^W\left(\mathcal{D}^1(f_2, f_3), \overline{f_2}, \overline{f_3}\right)\right|, \left|\Lambda^W\left(\overline{f_1}, \mathcal{D}^2(f_1, f_3), \overline{f_3}\right)\right|, \left|\Lambda^W\left(\overline{f_1}, \overline{f_2}, \mathcal{D}^3(f_1, f_2)\right)\right| \gg \delta^2 NM.$$

This is a simple consequence of the Cauchy–Schwarz inequality. For example, we have

$$\Lambda^W(f_1, f_2, f_3) = \sum_{x\in\mathbf{Z}} f_1(x) \cdot \left(\sum_{k\in[\pm M]} f_2(x + P(k))f_3(x + 2P(k))\right),$$

which is bounded above by

$$N^{1/2}\left(\sum_{x\in\mathbf{Z}}\sum_{k,k'\in[\pm M]} f_2(x + P(k))f_3(x + 2P(k))\overline{f_2(x + P(k'))f_3(x + 2P(k'))}\right)^{1/2}$$

$$= N^{1/2}\left(\sum_{\substack{x\in\mathbf{Z}\\ k'\in[\pm M]}}\left(\sum_{k\in[\pm M]} f_2(x + P(k))f_3(x + 2P(k))\right)\overline{f_2(x + P(k'))f_3(x + 2P(k'))}\right)^{1/2}$$

by the Cauchy–Schwarz inequality. Rearranging now yields $\left|\Lambda^W(\mathcal{D}^1(f_2, f_3), \overline{f_2}, \overline{f_3})\right| \gg \delta^2 NM$, and the other two inequalities are proved similarly.

3. Proof sketch

The starting point of our work is to use the W-trick of Green [18] to compare the count of certain three-term arithmetic progressions with shifted square common difference to the count of all three-term arithmetic progressions in a set, and then apply quantitative lower bounds for the number of three-term arithmetic progressions coming from Roth's theorem. This is closely motivated by work of Wooley and Ziegler [51], who proved a version of the polynomial Szemerédi theorem with y restricted to the set of shifted primes via such an approach.

We will show for any 1-bounded functions $f_1, f_2, f_3 \colon \mathbf{Z} \to \mathbf{C}$ with support in $[N]$ that

$$\left|(NW)^{1/2}\Lambda^W(f_1, f_2, f_3) - \Lambda^{\text{Model}}(f_1, f_2, f_3)\right| \ll \frac{N^2}{\log_m N} \tag{3.1}$$

for some absolute constant $m \in \mathbf{N}$ (recall that W will ultimately be chosen to be slowly growing with N). Theorem 1.1 then follows by dividing S into classes modulo $4W$, shifting an appropriately dense congruence class of S and scaling by $(4W)^{-1}$, noting that differences in this rescaled set of the form $Wy^2 + y$ correspond to differences of the form $y^2 - 1$ in the original set, and applying supersaturation results for Roth's theorem.

The crux of our proof of Theorem 1.1 is establishing that the "difference" counting operator

$$\widetilde{\Lambda}(f_1, f_2, f_3) = (NW)^{1/2}\Lambda^W(f_1, f_2, f_3) - \Lambda^{\mathrm{Model}}(f_1, f_2, f_3)$$

is controlled by the U^2-norm of the functions f_i (or, more precisely, the $U^2_{W\cdot[N/W]}$-norm). Given such norm control, combining a variant of stashing with the U^2-inverse theorem implies that there exist linear phase functions ψ_1, ψ_2, ψ_3 such that the counting operator $\widetilde{\Lambda}\left(\psi_1 1_{[N]}, \psi_2 1_{[N]}, \psi_3 1_{[N]}\right)$ is large. The existence of such phase functions is ruled out by a direct Fourier analytic computation. Indeed, the weight function $\nu(d)$ is chosen so that the corresponding exponential sums closely matches that of $P(k)$, and the W-trick serves to remove the major arc contributions initially present in the Fourier transform of the squares.

It follows from the triangle inequality that in order to establish U^2-norm control of the counting operator $\widetilde{\Lambda}$, it suffices to establish the result for the counting operators Λ^W and Λ^{Model} separately. The (far) simpler of these two tasks is establishing U^2-norm control for $\Lambda^{\mathrm{Model}}(f_1, f_2, f_3)$. Note that if $\nu(d)$ were absent, then this is precisely the fact that the U^2-norm controls the count of three-term arithmetic progressions weighted by f_1, f_2, and f_3. The result for Λ^{Model} follows by noting that the Fourier transform of $\nu(d)$ (after a bit of smoothing) is appropriately bounded in L^1.

The vast majority of the paper, therefore, is devoted to establishing U^2-control of the operator Λ^W. We will do this by using the degree-lowering method, following work of the first author [39] and the first author and Prendiville [41, 42]. This method, in our setting, can be broken down into two steps. First, we establish that Λ^W is controlled by some high degree Gowers U^s-norm, and then we show (essentially) that U^t-norm control of Λ^W implies U^{t-1}-norm control of Λ^W whenever $t \geq 3$. These two steps taken together imply the desired U^2-norm control. The first step is proven by combining the PET induction scheme of Bergelson and Leibman [5] with the quantitative concatenation results of [40], which we can use as a black box. The majority of our effort, therefore, is concentrated on the second step of the argument.

Via an application of stashing, the key to the second step of our argument is establishing that

$$\|\mathcal{D}^1(f_2, f_3)\|_{U^k_{W\cdot[N/W]}}^{2^k} \geq \delta N \Longrightarrow \|f_i\|_{U^{k-1}_{W\cdot[N/W]}}^{2^{k-1}} \geq \delta' N \tag{3.2}$$

for $k \geq 3$ and $i \in \{2, 3\}$ (and the analogous statement for $\mathcal{D}^3(f_1, f_2)$). This, combined with further applications of stashing, implies U^2-norm control of Λ^W. By dual-difference interchange (Lemma 6.4), it is essentially sufficient to prove the result when $k = 3$, and for the remainder of the sketch we will focus on this special case.

First, let us pretend, for the sake of illustration, that the U^3-inverse theorem implied large correlation with a global quadratic form $e(\alpha x^2 + \beta x)$. This is, of course, a lie due to the existence of bracket-polynomials, but will help to motivate the main technical considerations. Furthermore, suppose for the sake of discussion that $\|\mathcal{D}^1(f_2, f_3)\|_{U^k_{[N]}}$ is large; this is a rather minor technical point that can be handled by splitting into congruence classes modulo W. It then follows from our "fake" U^3-inverse theorem that

$$\left|\frac{1}{N}\sum_{x\in\mathbf{Z}} e(\alpha x^2 + \beta x)\mathbf{E}_{y\in[\pm M]} f_2(x + P(y)) f_3(x + 2P(y))\right| \tag{3.3}$$

is large. Setting

$$\widetilde{f_2}(x) := f_2(x) \cdot e(2\alpha x^2 + 2\beta x)$$

and

$$\widetilde{f_3}(x) := f_3(x) \cdot e(-\alpha x^2 - \beta x),$$

and, as in work of Leng [34], using the polynomial identities

$$x^2 = 2(x + P(y))^2 - (x + 2P(y))^2 + 2P(y)^2$$

and

$$x = 2(x + P(y)) - (x + 2P(y)),$$

we get, by rearranging (3.3), that

$$\left| \frac{1}{N} \sum_{x \in \mathbf{Z}} \mathbf{E}_{y \in [\pm M]} \widetilde{f}_2(x + P(y)) \widetilde{f}_3(x + 2P(y)) e(2\alpha P(y)^2) \right|$$

is large. By applying Fourier inversion to $\widetilde{f}_2$ and $\widetilde{f}_3$ and then using orthogonality of characters and Parseval's identity, it follows that

$$\sup_{\kappa \in \mathbf{T}} \left| \mathbf{E}_{y \in [\pm M]} e(2\alpha P(y)^2 + \kappa P(y)) \right|$$

is large. Using Weyl's inequality, and carefully analyzing various terms in the expansion of $P(y)^2$, shows that α and κ are essentially major arc. More precisely, there exists a positive integer q such that $q \leq \delta^{-O(1)}$ and $\|q\alpha\|_{\mathbf{T}} \leq \delta^{-O(1)}/N^2$ and $\|q\kappa\|_{\mathbf{T}} \leq \delta^{-O(1)}/N$. This computation is a bit delicate; one needs that the coefficients of $P(y)$ are coprime in order to avoiding sacrificing factors of W. To simplify the rest of our discussion, we will pretend that, in fact, Weyl's inequality implies that $\alpha, \kappa = 0$; by passing to intervals of length $\delta^{O(1)}N$ and spacing at most $\delta^{-O(1)}$, one can turn this fantasy into a reality.

Note that if $\alpha = 0$, we would have that

$$\left| \frac{1}{N} \sum_{x \in \mathbf{Z}} e(\beta x) \mathbf{E}_{y \in [\pm M]} f_2(x + P(y)) f_3(x + 2P(y)) \right|$$

is large. Applying the second of our two identities, we may rewrite the above quantity as

$$\left| \frac{1}{N} \sum_{x \in \mathbf{Z}} \mathbf{E}_{y \in [\pm M]} f_2(x + P(y)) e(2\beta(x + P(y)) f_3(x + 2P(y)) e(-\beta(x + 2P(y)) \right|,$$

which, by making the change of variables $x \mapsto x - P(y)$, equals

$$\left| \frac{1}{N} \sum_{x \in \mathbf{Z}} \mathbf{E}_{y \in [\pm M]} f_2(x) e(2\beta x) f_3(x + P(y)) e(-\beta(x + P(y)) \right|.$$

That f_2 and f_3 must have large U^2-norms then follows by U^2-control for the configuration $(x, x + P(y))$, which is implicit in work of Sárközy [46]; this is a simple consequence of Fourier inversion, orthogonality of characters, and the Gowers–Cauchy–Schwarz inequality.

To rigorously prove the implication (3.2) we must use the U^3-inverse theorem of Green and Tao [19] in place of our "fake" U^3-inverse theorem. The Green–Tao inverse theorem produces a Lipschitz function F on a degree 2 nilmanifold G/Γ and a polynomial sequence $g\colon \mathbf{Z} \to G$ (in the sense of Definition A.2) such that

$$\left| \sum_{x \in \mathbf{Z}} F(g(x)) \mathbf{E}_{y \in [\pm M]} f_2(x + P(y)) f_3(x + 2P(y)) \right| \geq \exp(-\delta^{-O(1)}) N.$$

Mimicking our simplified sketch above, we now want to "factor" $F(g(x))$ into terms involving $x + P(y)$, $x + 2P(y)$, and $P(y)$. Leng [34] accomplishes such a maneuver for the pattern $(x, x+P(y), x+Q(y), x + P(y) + Q(y))$ over finite fields via a vertical Fourier expansion of F and noting that the Host-Kra cube of dimension 3 has a constrained orbit for any degree 2 polynomial sequence on a nilmanifold.

In our case, however, the constraints coming from the Host-Kra cube are insufficient, and to proceed directly one would require a suitable understanding of the orbits of the linear forms $(x, y, x+y, x + 2y)$ for a degree 2 polynomial sequence on a nilmanifold. The understanding of such an orbit is rather delicate, as this set of forms does not satisfy the *flag condition*, and the underlying equidistribution theory has only recently been addressed in work of Altman [1]. However, by using an earlier "lifting" trick of Altman [2], which amounts to a simple change of variables in our setting, it instead suffices to constrain the orbit of $6x$ given the images of $(6y, 3(x+y), 2(x+2y))$. As the pattern $(6x, 6y, 3(x+y), 2(x+2y))$ is translation invariant, the flag-equidistribution theory developed in work of Green and Tao [21] applies, and one can then derive the necessary constraint. We do this by following [19, Section 14], which establishes the analogous result for k-term arithmetic progressions, although various related results appear earlier in the ergodic theory literature [4,12,52].

Having obtained a suitable constraint, we will next require a suitable analogue of Weyl's inequality for nilsequences. This can be found in the seminal paper of Green and Tao on the equidistribution of polynomial orbits on nilmanifolds [23]. The main technical result of this work [23, Theorem 1.9] essentially proves that if a polynomial sequence $g(\cdot)$ fails to equidistribute on a nilmanifold, one can identify an abelian reason for it. Using this result we will prove that if the polynomial sequence $g(P(6y))$ fails to equidistribute, then one can factor the polynomial sequence g. By tracking carefully with Mal'cev coordinates (analogously to the sketch with Weyl's inequality earlier), one can prove that the factorization is of the same quality as if one knew that instead $g(y)$ failed to equidistribute. While such a factorization itself is not immediately useful, via iterating the factorization (as in the factorization results of [23]), one can prove that instead of correlating with a degree two nilsequence, one, in fact, correlates with a degree one nilsequence. The form of our result is closely motivated by earlier work of Leng [34, Lemma 6.1]. Given such a result, and then Fourier expanding the degree one nilsequence, we can reduce to dealing with pure polynomial phases, and the analysis follows as sketched earlier.

We end our discussion with a brief remark on bounds in the implication (3.2). Our bounds are of iterated logarithmic type, as δ' is ultimately doubly-exponentially small in δ. The first of these exponential terms is derived from the fact that we use the the U^3-inverse theorem of Green and Tao [19]; given more recent work of Sanders [45] the correlation could be improved to quasi-polynomial. The second source of exponentials comes from the double-exponential dependence on dimension implicit in [23, Theorem 7.1]; this dependence was quantified explicitly in recent work of Tao and Teräväinen [50]. Therefore, even using the results of Sanders [45], our bounds involve a large number of logs. Recently, however, the dimension dependence in results of Green and Tao [23] have been improved to exponential for periodic nilsequences in work of Leng [33]; Leng has also announced analogous results for all nilsequences, and, by inputting such results into our work (along with the necessary quantitative versions of results in [23, Appendix A]), a substantially reduced number of logs would be achieved (likely yielding $\ll N \exp(-(\log\log N)^c)$ in Theorem 1.1).

3.1. **Organization of the paper.** In Section 4, we prove U^2-control for Λ^{Model}. In Section 5, we prove the constraints for degree 2 nilmanifold orbits (in Section 5.1) and the necessary factorization theorem (in Section 5.2). In Section 6 we prove the main degree-lowering statement in this work. In Section 7, we complete proof of Theorem 1.1. In Appendix A, we collect various definitions and basic properties regarding nilmanifolds. In Appendix B, we collect various standard exponential sum estimates for the polynomial $P(y) = Wy^2 + y$. Finally in Appendix C, we collect various basic estimates regarding changing parameters in the box-norm.

4. Control for Λ^{Model}

In this section, we will establish U^2-norm control of Λ^{Model} and deduce a uniform lower bound for Λ^{Model} from the best known bounds in Roth's theorem.

Lemma 4.1. *Let $f_1, f_2, f_3 \colon \mathbf{Z} \to \mathbf{C}$ be 1-bounded functions supported on $[\pm\delta^{-1}N]$. If $N \gg \delta^{-O(1)}$ and*

$$\left|\Lambda^{\text{Model}}(f_1, f_2, f_3)\right| \geq \delta N^2,$$

then

$$\min_{i\in[3]} \|f_i\|_{U^2_{[N]}}^4 \gg \delta^{O(1)}N.$$

Proof. By adjusting implicit constants, we may assume that δ is smaller than an absolute constant. Define

$$\nu^{(1)}(d) = \sqrt{\frac{N}{d}}\mathbb{1}_{\delta^5 N\le d\le N}$$

and

$$\tau(d) = \frac{\mathbb{1}_{|d|\le\delta^{10}N}}{2\delta^{10}N}.$$

Noting that $\nu^{(1)}(d)$ is δ^{-8}/N-Lipschitz away from the boundary of its support and recalling the definition (2.3) of ν, we have

$$\sum_{d\in\mathbf{Z}} |(\tau*\nu^{(1)})(d)-\nu(d)| \le \sum_{d\in\mathbf{Z}}\left(|(\tau*\nu^{(1)})(d)-\nu^{(1)}(d)| + |\nu^{(1)}(d)-\nu(d)|\right) \le \delta^2 N.$$

Therefore, since the f_i are 1-bounded,

$$\sum_{x,d\in\mathbf{Z}} f_1(x)f_2(x+d)f_3(x+2d)(\tau*\nu^{(1)})(d) \ge \delta N^2/2.$$

Furthermore, we have by orthogonality of characters, Cauchy–Schwarz, and Parseval that

$$\begin{aligned}
&\left|\sum_{x,d\in\mathbf{Z}} f_1(x)f_2(x+d)f_3(x+2d)(\tau*\nu^{(1)})(d)\right| \\
&= \left|\int_{\mathbf{T}^2} \widehat{f}_1(\Theta_1)\widehat{f}_2(-2\Theta_1+\Theta_2)\widehat{f}_3(\Theta_1-\Theta_2)(\widehat{\tau*\nu^{(1)}}(\Theta_2))\, d\Theta_1 d\Theta_2\right| \\
&\le \int_{\mathbf{T}} |\widehat{\tau*\nu^{(1)}}(\Theta_2)|\, d\Theta_2 \cdot \sup_{\Theta_2\in\mathbf{T}} \int_{\mathbf{T}} |\widehat{f}_1(\Theta_1)|\cdot|\widehat{f}_2(-2\Theta_1+\Theta_2)|\cdot|\widehat{f}_3(\Theta_1-\Theta_2)|\, d\Theta_1 \\
&\le \int_{\mathbf{T}} |\widehat{\tau}(\Theta_2)||\widehat{\nu^{(1)}}(\Theta_2)|\, d\Theta_2 \cdot \sup_{\mathbf{T}} |\widehat{f}_1(\Theta_1)| \cdot \sup_{\Theta_2\in\mathbf{T}} \int_{\Theta_1\in\mathbf{T}} |\widehat{f}_2(-2\Theta_1+\Theta_2)|\cdot|\widehat{f}_3(\Theta_1-\Theta_2)|\, d\Theta_1 \\
&\le \|\tau\|_{\ell^2(\mathbf{Z})}\|\nu^{(1)}\|_{\ell^2(\mathbf{Z})} \sup_{\Theta_1\in\mathbf{T}} |\widehat{f}_1(\Theta_1)|\cdot\|f_2\|_{\ell^2(\mathbf{Z})}\|f_3\|_{\ell^2(\mathbf{Z})} \\
&\ll (\delta^{-10}N^{-1})^{1/2}\cdot(N\log(1/\delta))^{1/2}\cdot \sup_{\Theta_1\in\mathbf{T}} |\widehat{f}_1(\Theta_1)|\cdot N \\
&\ll \delta^{-5}(\log(1/\delta))^{1/2} N \sup_{\Theta\in\mathbf{T}} \left|\widehat{f}_1(\Theta)\right|.
\end{aligned}$$

An analogous inequality holds for f_2 and f_3, and therefore

$$\inf_{i\in[3]}\sup_{\Theta\in\mathbf{T}} |\widehat{f}_i(\Theta)| \gg \delta^{O(1)}N;$$

the result now follows from the converse of the U^2-inverse theorem (see, e.g., Lemma C.5). ◻

We next establish a uniform lower bound on Λ^{Model} using recent breakthrough work of Kelley and Meka [26].

Lemma 4.2. *Suppose that* $f\colon \mathbf{Z}\to[0,1]$ *with* $\operatorname{supp}(f)\in[N]$ *and* $\sum_{x\in\mathbf{Z}} f(x)\ge\delta N$*. Then*

$$\Lambda^{\text{Model}}(f,f,f) \gg \exp\left(-\log(2/\delta)^{O(1)}\right)N^2.$$

Proof. Noting that $w(d)\ge 1$ for all $1\le d\le N$, the result follows from [26, Theorem 1.2]. ◻

5. Nilmanifold considerations

Throughout this section, we will assume familiarity with standard terminology related to nilsequences and nilmanifolds. All terminology used is defined in Appendix A; our conventions match those in [23, 50]. Furthermore, throughout this section, we will require various quantitative rationality claims from [23, Appendix A], but with explicit dimensional dependencies. As stated in [50, pg. 52], all bounds of the form $Q^{O_m(1)}$ in [23, Appendix A] may in fact be taken to be $\mathrm{poly}_m(Q)$ (where m is the dimension of the underlying nilmanifold). We will cite bounds from [23, Appendix A], but assume this more explicit dimensional quantification.

5.1. **Leibman group considerations.** Throughout this subsection, define

$$\tau(x,y) := (2(x+2y), 3(x+y), 6y, 6x)$$

for all $x, y \in \mathbf{Z}$. We will write $\tau_i(x,y)$, for $i = 1, \ldots, 4$, to refer to the i-th coordinate of $\tau(x,y)$. The key output of this subsection will be Lemma 5.1, which relates the values of a degree 2 polynomial sequence at the first three coordinates of $\tau(x,y)$ to the value at the fourth coordinate.

Lemma 5.1. *Let G/Γ be a filtered nilmanifold of dimension m, degree 2, and complexity at most L. Let F be a function on G/Γ with vertical frequency ξ with $|\xi| \le L$ and $\|F\|_{\mathrm{Lip}} \le L$. Let $g(\cdot)$ be a polynomial sequence with respect to G/Γ (and the corresponding degree 2 filtration). There exist G and $F_{j,\alpha}$ such that for all $x, y \in \mathbf{Z}$,*

$$F(g(\tau_4(x,y))\Gamma) = \sum_{\alpha} \prod_{j\in[3]} F_{j,\alpha}(g(\tau_j(x,y))\Gamma) + G(x,y),$$

where

- *$\|G\|_\infty \le L^{-1}$;*
- *for all α, $F_{1,\alpha}$ has vertical frequency -9ξ, $F_{2,\alpha}$ has vertical frequency 8ξ, and $F_{3,\alpha}$ has vertical frequency 2ξ;*
- *there are $\mathrm{poly}_m(L)$ summand indices α; and*
- *we have $\|F_{j,\alpha}\|_{\mathrm{Lip}} \le \mathrm{poly}_m(L)$ for all α and $j \in \{1,2,3\}$.*

The key input into Lemma 5.1 is that the image of $\tau(x,y)$ under a polynomial sequence on a nilmanifold is constrained. An analogous result k-term arithmetic progressions appears in [19, Lemma 12.7], and for the Host-Kra cube in [22, Proposition 11.5]. Our proof is essentially identical to that of [19, Lemma 12.7] modulo certain algebraic issues regarding the Leibman group [32].

We first require the notion of being continuous right-invertible.

Definition 5.2. Let N_1, N_2 be compact topological spaces, let $\pi\colon N_1 \to N_2$ be a continuous map, and let $\Sigma \subseteq N_1$. We say that π is *continuously right-invertible* on Σ if, for all $w \in \overline{\pi(\Sigma)}$, there exists a neighborhood $V_w \subseteq N_2$ of w and a continuous map $\pi_w^{-1}\colon V_w \to N_1$ such that $\pi_w^{-1} \circ \pi$ is the identity map on $\Sigma \cap \pi^{-1}(V_w)$.

Now we can precisely described the aforementioned constraint.

Lemma 5.3. *Let G/Γ be a be a filtered nilmanifold of dimension m, degree 2, and complexity at most L. Let $G_\bullet$ denote the degree two filtration $G_0 = G_1 \geqslant G_2 \geqslant \mathrm{Id}_G$ on G and $\mathcal{X}$ denote the chosen Mal'cev basis for G/Γ. Furthermore, define*

$$G^\tau := \{(g_0, g_0 g_1, g_0 g_1^{-2} g_2, g_0 g_1^4 g_2^2)\colon g_i \in G_i\}.$$

Let $\pi\colon (G/\Gamma)^4 \to (G/\Gamma)^3$ denote the standard projection onto the first three coordinates. Then there exists a compact set $\Sigma \subseteq (G/\Gamma)^3$ and a continuous function $Q\colon \Sigma \to G/\Gamma$ such that

- *$\pi(G^\tau\Gamma^4) \subseteq \Sigma$;*
- *$Q(\pi(g\Gamma^4)) = g_4\Gamma$ for all $g = (g_1, g_2, g_3, g_4) \in G^\tau$; and*

- *Q is* $\mathrm{poly}_m(L)$*-Lipschitz, where the metric on Σ is given by restricting*
$$\widetilde{d}((x_1, x_2, x_3), (z_1, z_2, z_3)) = \sum_{i \in [3]} d_{\mathcal{X}}(x_i\Gamma, z_i\Gamma)$$
to Σ.

Remark 5.4. By $G^\tau\Gamma^4$ we mean the image of G^τ under taking Γ-cosets.

Proof. Take $\Sigma = \overline{\pi(G^\tau\Gamma^4)}$, and define
$$\begin{aligned} G_0^\tau &= G^\tau; \\ G_1^\tau &= \{(\mathrm{Id}_G, g_1, g_1^{-2}g_2, g_1^4 g_2^2) \colon g_i \in G_i\}; \\ G_2^\tau &= \{(\mathrm{Id}_G, \mathrm{Id}_G, g_2, g_2^2) \colon g_2 \in G_2\}; \text{ and} \\ G_3^\tau &= \{(\mathrm{Id}_G, \mathrm{Id}_G, \mathrm{Id}_G, \mathrm{Id}_G)\}. \end{aligned}$$
Our argument is identical to [19, Section 14], aside from verifying that G_i^τ are groups. This can be verified using general results of Green and Tao [21]; we provide a short argument specialized to our case. It is trivial to verify that G_2^τ and G_3^τ are groups. That G_0^τ is a group follows from noting that
$$\begin{aligned} \{(g(0), g(1), g(-2), g(4)) \colon g \in \mathrm{Poly}(\mathbf{Z}, G_\bullet)\} &= \big\{(g_0, g_0g_1, g_0g_1^{-2}g_2^3, g_0g_1^4g_2^6) \colon g_i \in G_i \text{ for } i = 0,1,2\big\} \\ &= \big\{(g_0, g_0g_1, g_0g_1^{-2}g_2, g_0g_1^4g_2^2) \colon g_i \in G_i \text{ for } i = 0,1,2\big\} \\ &= G_0^\tau, \end{aligned}$$
where we have used that G_2 is divisible (since G_2, being a connected nilpotent Lie group, has surjective exponential map), and recalling that $\mathrm{Poly}(\mathbf{Z}, G_\bullet)$ is a group. That G_1^τ is a group simply follows from noting that it is the intersection of two groups:
$$G_1^\tau = G_0^\tau \cap (\mathrm{Id}_G \times G_0 \times G_0 \times G_0).$$
Finally, observe that the groups G_i^τ have the nesting property
$$G_3^\tau \subseteq G_2^\tau \subseteq G_1^\tau \subseteq G^\tau.$$

Next, we will prove inductively that the restriction of π is continuously right-invertible on G_i^τ/Γ^4, starting at $i = 3$ and proceeding downwards. The crucial point is that the first non-identity coordinate in a generic element of G_i^τ is g_i and, by inverting the quotient map $G_i \to G_i/\Gamma$ locally, we can "remove" g_i and proceed inductively. We now give a formal proof following [19, Section 14].

Note that G_3^τ is isomorphic to the trivial group, and therefore π is trivially continuously right invertible on G_3^τ/Γ^4. Suppose that the restriction of π to G_{i+1}^τ/Γ^4 is continuously right invertible for some $0 \le i \le 2$; we will show that the same holds for the restriction of π to G_i^τ/Γ^4.

Since Γ acts freely and properly on the manifold G (on the right) and the quotient G/Γ is compact, the quotient maps $\rho_i \colon G_i \to G_i/\Gamma$ are covering maps. Therefore, for any point $z_i \in G_i/\Gamma$, there exists a neighborhood $V_{z_i} \subseteq G_i/\Gamma$ and a continuous function $f \colon V_{z_i} \to G_i$ such that $\rho_i \circ f$ is the identity map on V_{z_i}.

Now, consider a point $\pi(z) \in \overline{\pi(G_i^\tau/\Gamma^4)}$, with $z = (z_1, z_2, z_3, z_4)$. Note that the first i coordinates of $\pi(z)$ are $\mathrm{Id}_G\Gamma$. Consider the $(i+1)$-st coordinate, $z_{i+1} \in G_i/\Gamma$, of $\pi(z)$, and let $x = (x_1, x_2, x_3, x_4) \in G_i^\tau/\Gamma^4$ be such that $x_{i+1} \in V_{z_{i+1}}$ (with $V_{z_{i+1}}$ defined as in the previous paragraph). This implies that $(\rho_{i+1} \circ f)(x_{i+1}) = x_{i+1}$, i.e., $x_{i+1} = f(x_{i+1})\Gamma$. Define
$$F_i(x_{i+1}) = \begin{cases} (f(x_1), f(x_1), f(x_1), f(x_1)) & \text{if } i = 0 \\ (\mathrm{Id}_G, f(x_2), f(x_2)^{-2}, f(x_2)^4) & \text{if } i = 1 \\ (\mathrm{Id}_G, \mathrm{Id}_G, f(x_3), f(x_3)^2) & \text{if } i = 2 \end{cases}$$

for all such x. We write $F = F_i$ as shorthand. Note that $F(x_{i+1})$ is continuous as a function of x_{i+1}, and hence of $\pi(x)$ (as $0 \le i \le 2$), which means that F defines a continuous function in an open neighborhood of $\pi(z)$. By definition, if $x \in G_i^\tau/\Gamma^4$, then there exists $g \in G_i^\tau$ such that $g\Gamma^4 = (g_1, g_2, g_3, g_4)\Gamma^4 = x$ (which we choose arbitrarily). Observe that $f(x_{i+1})^{-1}g_{i+1}$ is an element of both G_i and Γ. Now, let

$$\widetilde{g}_i = \begin{cases} (f(x_1)^{-1}g_1, f(x_1)^{-1}g_1, f(x_1)^{-1}g_1, f(x_1)^{-1}g_1) & \text{if } i = 0 \\ (\mathrm{Id}_G, f(x_2)^{-1}g_2, (f(x_2)^{-1}g_2)^{-2}, (f(x_2)^{-1}g_2)^4) & \text{if } i = 1 \\ (\mathrm{Id}_G, \mathrm{Id}_G, f(x_3)^{-1}g_3, (f(x_3)^{-1}g_3)^2) & \text{if } i = 2 \end{cases}$$

if x is in a sufficiently small open neighborhood of z. Again let $\widetilde{g} = \widetilde{g}_i$ as shorthand. Observe that $\widetilde{g} \in G_i^\tau$ and $\widetilde{g} \in \Gamma^4$ by construction. Define h to be such that

$$g = F(x_{i+1})h\widetilde{g}.$$

The $(i+1)$-st coordinate of h is the identity (as are the first i coordinates). So, h must lie in G_{i+1}^τ. Therefore,

$$x = g\Gamma^4 = F(x_{i+1})h\widetilde{g}\Gamma^4 = F(x_{i+1})h\Gamma^4,$$

since $\widetilde{g} \in \Gamma^4$. Thus,

$$F(x_{i+1})^{-1}x = h\Gamma^4.$$

Note that, as $F(x_{i+1})$ depends continuously on $\pi(x)$ in a neighborhood of $\pi(z)$, and is defined via a local continuous right-inverse, we have that $\pi(F(x_{i+1})^{-1}x)$ is within a neighborhood of $\pi(F(z_{i+1})^{-1}z)$. Furthermore, note that, as $h \in G_{i+1}^\tau$, the $(i+1)$-st coordinate of $F(x_{i+1})^{-1}x$ is $\mathrm{Id}_G\Gamma$, and therefore we are in position to apply induction. By induction, we may write

$$F(x_{i+1})^{-1}x = h\Gamma^4 = \pi^{-1}_{(F(z_{i+1})^{-1}z)}(\pi(F(x_{i+1})^{-1}x)), \tag{5.1}$$

and therefore

$$x = F(x_{i+1})\pi^{-1}_{F(z_{i+1})^{-1}z}(\pi(F(x_{i+1})^{-1}x)),$$

where $\pi^{-1}_{F(z_{i+1})^{-1}z}$ is the (localized) continuous right-inverse we have constructed for G_{i+1}^τ/Γ^4. Note that $\pi(F(x_{i+1})x) = \widetilde{\pi}(F(x_{i+1}))\pi(x)$, where $\widetilde{\pi}$ is the projection onto the first three coordinates in G^4. By the previous discussion, the right-hand-side of (5.1) depends continuously on $\pi(x)$ and is defined in a sufficiently small neighborhood of $\pi(x)$. Thus, the right-hand-side of (5.1) provides the desired continuous right-inverse and the result follows.

We now glue these local right-inverses into a global continuous right-inverse $\Pi\colon \Sigma \to G^\tau\Gamma^4$ satisfying $(\Pi \circ \pi)(x) = x$ for all $x \in G^\tau\Gamma^4$. We can perform such gluing as long as all our local right-inverses agree on intersections. To see this, it suffices to show that π is injective on $G^\tau\Gamma^4$. Suppose $\pi(x) = \pi(y)$ for $x, y \in G^\tau\Gamma^4$. We can find $g \in G^\tau$ such that $g^{-1}y \in \Gamma^4$, so $\pi(g^{-1}x) = (\mathrm{Id}_G\Gamma, \mathrm{Id}_G\Gamma, \mathrm{Id}_G\Gamma)$ since the right-action of G on G/Γ is compatible with π. Now $g^{-1}x \in G^\tau\Gamma^4$ and has first three coordinates being $\mathrm{Id}_G\Gamma$. This implies that if we write $g^{-1}x = (g_0\Gamma, g_0g_1\Gamma, g_0g_1^{-2}g_2\Gamma, g_0g_1^4g_2^2\Gamma)$, then $g_0 \in \Gamma$ hence $g_1 \in \Gamma$ hence $g_2 \in \Gamma$ (since Γ is a subgroup of G). Thus the final coordinate of $g^{-1}x$ is also $\mathrm{Id}_G\Gamma$, and hence $g^{-1}x, g^{-1}y \in \Gamma^4$. This implies $x = y$ as cosets, completing the proof of injectivity (and hence existence of a global inverse).

We now define Q to be the fourth coordinate of this global right-inverse of π on $G^\tau\Gamma^4$. By the above arguments, the first two bullet points are satisfied.

We finally briefly sketch how to obtain the necessary Lipschitz bound on Q. First, note from above that Q is unique and, as $(G/\Gamma)^4$ has diameter bounded by $\mathrm{poly}_m(L)$ [23, Lemma A.16], it suffices to consider points which are within distance $\mathrm{poly}_m(L^{-1})$ of each other to prove Lipschitz bounds on Q. Furthermore, looking at our inductive construction, it suffices to show we can invert ρ_i for $x' \in G_i/\Gamma$ such that the preimage in G_i is suitably bounded and in a Lipschitz manner. The remainder of the analysis then consists of left multiplication by bounded group elements, which is Lipschitz by

[23, Lemma A.5] for left-multiplication and right-multiplication is always Lipschitz due to right-invariance of the metric on G. Note here the fact that if $g \in G$ is bounded, then g^k for bounded k and g^{-1} are as well since $d(g^k, \mathrm{Id}_G) \leq \sum_{i=1}^k d(g^i, g^{i-1}) = kd(g, \mathrm{Id}_G)$ and $d(g, \mathrm{Id}_G) = d(g^{-1}, \mathrm{Id}_G)$.

To invert ρ_i in the neighborhood of a point $x' \in G_i/\Gamma$, first note that G_i is a closed rational subgroup of G and the last $\dim(G_i)$ elements of $\mathcal{X}$ are a valid Mal'cev basis for G_i. Therefore, by combining [23, Lemmas A.16 and A.17], there exists $g' \in G_i$ such that $g'\Gamma = x'$ and $d(g', \mathrm{Id}_G) \leq \mathrm{poly}_m(L)$. Taking a sufficiently small neighborhood around g', of size $\mathrm{poly}_m(L^{-1})$, for any points $g^{(1)}$ and $g^{(2)}$ in this neighborhood, we have

$$\begin{aligned}\inf_{\gamma \in \Gamma \setminus \{0\}} d(g^{(1)}, g^{(2)}\gamma) &\geq \mathrm{poly}_m(L^{-1}) \cdot \inf_{\gamma \in \Gamma \setminus \{0\}} d((g^{(2)})^{-1} g^{(1)}, \gamma) \\ &\geq \mathrm{poly}_m(L^{-1}) \cdot \left(\inf_{\gamma \in \Gamma \setminus \{0\}} d(\mathrm{Id}_G, \gamma) - d((g^{(2)})^{-1} g^{(1)}, \mathrm{Id}_G) \right) \\ &\geq \mathrm{poly}_m(L^{-1}).\end{aligned}$$

The last inequality comes from the fact that $\psi(\gamma) \in \mathbf{Z}^m$, where ψ are Mal'cev coordinates of the second kind (with respect to an implicit Mal'cev basis $\mathcal{X}$ giving the complexity bound). As $\psi(\gamma)$ is nonzero, [23, Lemma A.4] then gives the lower bound.

Thus, in a small neighborhood of g', we have that $d_G(g^{(1)}, g^{(2)}) = d_{G/\Gamma}(g^{(1)}\Gamma, g^{(2)}\Gamma)$. Furthermore, the pushforward under ρ_i of the neighborhood of g' in G_i surjects onto a small neighborhood of $x' = g'\Gamma$ in G_i/Γ. Therefore, given $z' \in G_i/\Gamma$ near x', the map $f(z')$ can be defined by taking the closest point to g'_z to g' in G_i such that $g'_z\Gamma = z'$. This gives the desired inverse map in the neighborhood of x' which is Lipschitz by the above equality of metrics and, furthermore, we have that the inverse image of x' in G_i is $\mathrm{poly}_m(L)$-bounded, as desired. ☐

We are now in position to prove Lemma 5.1.

Proof sketch of Lemma 5.1. Let the filtration $G_\bullet$ be denoted by $G = G_0 = G_1 \geqslant G_2 \geqslant \mathrm{Id}_G$. Let $\tau^{[i]}$ denote the span in $\mathbf{R}^4$ of the set of vectors

$$\left\{ (\tau_1(x,y)^i, \tau_2(x,y)^i, \tau_3(x,y)^i, \tau_4(x,y)^i) \colon x, y \in \mathbf{Z} \right\}.$$

We find that

$$\begin{aligned}\tau^{[1]} &= \mathbf{R}(1,1,1,1) \oplus \mathbf{R}(0,1,-2,4) \\ \tau^{[2]} &= \mathbf{R}(1,1,1,1) \oplus \mathbf{R}(0,1,-2,4) \oplus \mathbf{R}(0,0,1,2) \\ \tau^{[3]} &= \mathbf{R}(1,1,1,1) \oplus \mathbf{R}(0,1,-2,4) \oplus \mathbf{R}(0,0,1,2) \oplus \mathbf{R}(0,0,0,1) = \mathbf{R}^4.\end{aligned}$$

Therefore τ satisfies the *flag condition* (which also follows from the fact that $\tau(x,y)$ is translation-invariant) and by [21, Lemma 3.2] we have that $g(\tau(x,y))$ takes values within G^τ (abusively extending g to vectors coordinate-wise).

Furthermore, by Lemma 5.3, for $(g_1, g_2, g_3, g_4) \in G^\tau$ we have

$$F(g_4\Gamma) = F(Q(g_1\Gamma, g_2\Gamma, g_3\Gamma)),$$

with Q as in Lemma 5.3. Using the partition of unity argument suggested in [50, Footnote 10] and the quantitative bounds on Q proven in Lemma 5.3, we have that for $(g_1, g_2, g_3, g_4) \in G^\tau$,

$$F(g_4\Gamma) = \sum_{\alpha \in A} \prod_{j \in [3]} F_{j,\alpha}(g_j\Gamma) + G((g_1, g_2, g_3, g_4)\Gamma),$$

where

- $\|G((g_1, g_2, g_3, g_4)\Gamma)\|_\infty \leq L^{-1}/2$ for all $(g_1, g_2, g_3, g_4) \in G^\tau$;
- there are $\mathrm{poly}_m(L)$ terms in the sum over α; and

- the functions $F_{j,\alpha}(g_j\Gamma)$ are $\mathrm{poly}_m(L)$-Lipschitz.

Note here that a qualitative version follows simply by applying the Stone–Weierstrass theorem (and noting that $(G/\Gamma)^3$ is compact).

This procedure, however, does not immediately yield that the $F_{j,\alpha}$'s have the desired vertical frequencies. Applying Lemma A.9 (vertical expansion), we have may assume that the $F_{j,\alpha}$'s each have vertical frequencies $\xi_{j,\alpha}$ bounded by $\mathrm{poly}_m(L)$. The crucial idea at this point (due to Leng [34, Lemma A.3]) is noting that for $(g_1, g_2, g_3, g_4) \in G^\tau$ and $g' \in G_2$, we have

$$(g_1g', g_2g', g_3g', g_4g') \in G^\tau,\ (g_1, g_2g', g_3g'^{-2}, g_4g'^4) \in G^\tau,\ \text{and } (g_1, g_2, g_3g', g_4g'^2) \in G^\tau.$$

Thus, if $g_1', g_2', g_3' \in G_2$, then $\widetilde{g} = (g_1g_1', g_2g_1'g_2', g_3g_1'g_2'^{-2}g_3', g_4g_1'g_2'^4g_3'^2) \in G^\tau$, and

$$\begin{aligned} F(g_4\Gamma) &= e(-\xi(g_1'))e(-4\xi(g_2'))e(-2\xi(g_3'))F(g_4g_1'g_2'^4g_3'^2\Gamma) \\ &= e(-\xi(g_1'))e(-4\xi(g_2'))e(-2\xi(g_3'))\sum_\alpha \prod_{j\in[3]} F_{j,\alpha}(\widetilde{g}_j\Gamma) + \widetilde{G}((g_1, g_2, g_3, g_4)\Gamma, g_1', g_2', g_3'). \end{aligned}$$

We now integrate over each $g_i' \in G_2/(\Gamma \cap G_2)$ (this is well-defined because $G_2/(\Gamma \cap G_2)$ is a torus onto which $e(\cdot)$ descends). Note that the integral of a nontrivial character ξ over $G_2/(\Gamma \cap G_2)$ is zero, and therefore a term α only remains if the vertical frequencies solve the following system of linear equations:

$$\begin{aligned} 0 &= -\xi + \xi_{1,\alpha} + \xi_{2,\alpha} + \xi_{3,\alpha}, \\ 0 &= -4\xi + \xi_{2,\alpha} - 2\xi_{3,\alpha}, \\ 0 &= -2\xi + \xi_{3,\alpha}, \end{aligned}$$

using the formulas for $\widetilde{g}_j$ and the vertical frequencies of the $F_{j,\alpha}$. The unique solution is $\xi_{1,\alpha} = -9\xi$, $\xi_{2,\alpha} = 8\xi$, and $\xi_{3,\alpha} = 2\xi$. Thus, after performing this integration, we find

$$F(g_4\Gamma) = \sum_{\alpha\in A^*} \prod_{j\in[3]} F_{j,\alpha}(g_j\Gamma) + \int_{(G_2/(\Gamma\cap G_2))^3} \widetilde{G}((g_1, g_2, g_3, g_4)\Gamma, g_1', g_2', g_3')\ dg_1'dg_2'dg_3'.$$

where all $\alpha \in A^*$ are such that $F_{j,\alpha}$ has vertical frequencies $-9\xi, 8\xi, 2\xi$ for $j = 1, 2, 3$ respectively. This is valid for all $(g_1, g_2, g_3, g_4) \in G^\tau$, hence it applies to $g(\tau(x, y)) \in G^\tau$ and we have the desired expression. □

5.2. **Factorization result.** The next lemma serves as the crucial analogue of Weyl's inequality for degree 2 nilsequences. Although the statement is motivated by work of Leng [34, Lemma 6.1], our proof mimics the factorization of polynomial sequences on nilmanifolds due to Green and Tao [23, Theorem 1.19]. However, our analogue of the basic decomposition result [23, Proposition 9.2] assumes that the polynomial sequence $g(P(6y))$ is not equidistributed, instead of the sequence $g(y)$. The crucial point, analogous to the case of polynomial phases sketched in Section 3, is that one can still deduce a useful factorization of $g(y)$ from this.

The key input into our argument is the following result on equidistribution of polynomial orbits in nilmanifolds due to Green and Tao [23, Theorem 2.9] with the explicit dimension dependencies given in work of Tao and Teräväinen [50].

Theorem 5.5 ([50, Theorem A.3])**.** *Let $m \geq 0$, $\delta \in (0, 1/2)$, and $N \geq 1$. Let G/Γ be a filtered nilmanifold of degree d with complexity at most $1/\delta$. Let $g\colon \mathbf{Z} \to G$ be a polynomial sequence. If $(g(n)\Gamma)_{n\in[N]}$ is not δ-equidistributed (Definition A.5), then there exists a horizontal character $0 < |\eta| \leq \delta^{-\exp((2m)^{O_d(1)})}$ such that*

$$\|\eta \circ g\|_{C^\infty[N]} \leq \delta^{-\exp((2m)^{O_d(1)})},$$

where the implicit constant $O_d(1)$ only depends on d.

Recall the $C^\infty[N]$-norm from Definition A.7. We now state our analogue of [23, Proposition 9.2].

Proposition 5.6. *Fix $\delta \in (0,1/2)$ and P and W as in (2.2). Let G/Γ be an m-dimensional filtered nilmanifold of degree 2 and complexity L with filtration $G = G_0 = G_1 \geqslant G_2 \geqslant \mathrm{Id}_G$ denoted by $G_\bullet$. Furthermore, let $\mathcal{X}$ denote the Mal'cev basis of G/Γ and let $F\colon G/\Gamma \to \mathbf{C}$ be such that $\|F\|_{\mathrm{Lip}} \le L$ and F has a nonzero vertical frequency $2\cdot\xi$ such that $\|\xi\|_\infty \le L$. Let $g\colon \mathbf{Z} \to \mathbf{G}$ be a polynomial sequence with respect to $G_\bullet$ with $g(0) = \mathrm{Id}_G$. For all $r \in [W]$, define*

$$P_r(y) := \frac{P(Wy+r) - P(r)}{W}.$$

Let $\mathcal{I} \subseteq [\pm\delta^{-1}N^{1/2}W^{-1}]$ be an arithmetic progression of difference at most δ^{-1}. Suppose that $W \le N^{1/10^4}$, $N \ge \mathrm{poly}_m(\delta^{-1}L)$, and

$$\left|\sum_{y\in\mathcal{I}} F(g(6P_r(y))\Gamma)\right| \ge \delta N^{1/2}W^{-1}.$$

Then, there exists a factorization $g = \varepsilon g'\gamma$ with polynomial sequences $\varepsilon, g', \gamma\colon \mathbf{Z} \to G$ such that

- *for all $n \in [\pm\delta^{-1}N]$, $d(\varepsilon(n), \varepsilon(n-1)) \le \mathrm{poly}_m(L\delta^{-1})/N$ and $d(\varepsilon(n), \mathrm{Id}_G) \le \mathrm{poly}_m(L\delta^{-1})$;*
- *γ is $\mathrm{poly}_m(L\delta^{-1})$-rational and $\gamma(n)\Gamma$ is periodic with period at most $\mathrm{poly}_m(L\delta^{-1})$; and*
- *g' takes values only in G', a simply connected proper $\mathrm{poly}_m(L\delta^{-1})$-rational subgroup with respect to $\mathcal{X}$, and may be viewed as a polynomial sequence with respect to the filtration $G'_\bullet$ where $G_i = G' \cap G_i$.*

Here the condition $N \ge \mathrm{poly}_m(\delta^{-1}L)$ is used abusively to express that there is some such polynomial expression such that this condition on N is sufficient; we use similar conventions later without comment. Now, we first state the following explicit binomial coefficient identities. While the precise constant coefficients are unimportant, various powers of A and B will indeed be used in our analysis.

Claim 5.7. *We have*

$$\binom{An^2+Bn}{1} = 2A\binom{n}{2} + (A+B)\binom{n}{1}$$

and

$$\binom{An^2+Bn}{2} = 12A^2\binom{n}{4} + (18A^2+6AB)\binom{n}{3} + (7A^2+6AB-A+B^2)\binom{n}{2} + \left(AB + \binom{A}{2} + \binom{B}{2}\right)\binom{n}{1}.$$

We also require the following claim regarding polynomial sequences and the $C^\infty[N]$-norm.

Claim 5.8. *Fix a constant $C \ge 1$. Suppose that S is a nonzero integer such that $|S| \le C$ and I is an integer such that $|I| \le CN$. If p is a polynomial of degree at most d, there exists $S' \le C^{O_d(1)}$ such that*

$$\|S'p(x)\|_{C^\infty[N]} \le C^{O_d(1)}\|p(Sx+I)\|_{C^\infty[N]}.$$

Proof. Let $q(x) = p(Sx+I)$ and $I' \in [S]$ be such that $I' \equiv I \mod S$. Note that

$$\|p(Sx+I')\|_{C^\infty[N]} = \|q(x+(I'-I)/S)\|_{C^\infty[N]}.$$

Vandermonde's identity implies

$$\binom{n+I}{j} = \sum_{0\le t\le j}\binom{n}{t}\binom{I}{j-t}.$$

As $|(I-I')/S| \le CN$, using Vandermonde's identity we have by expansion that

$$\|q(x+(I-I')/S)\|_{C^\infty[N]} \le C^{O_d(1)}\|q(x)\|_{C^\infty[N]}.$$

Putting it together, we have

$$\|p(Sx+I')\|_{C^\infty[N]} \le C^{O_d(1)}\|p(Sx+I)\|_{C^\infty[N]}.$$

Finally, by [23, Lemma 8.4] (applicable since the heights of S, I' are bounded by C) we can find appropriate S' so that $\|S'p(x)\|_{C^\infty[N]} \le C^{O_d(1)}\|p(Sx+I')\|$. This completes the proof. □

Proof of Proposition 5.6. Let $\min(\mathcal{I}) = I$, S denote the difference of the progression $\mathcal{I}$, and T the length of $\mathcal{I}$. By assumption, we have that

$$\Bigg|\sum_{y\in[T]} F(g(6P_r(Sy+I))\Gamma)\Bigg| \ge \delta N^{1/2}W^{-1}.$$

Next, as $2\cdot\xi$ is a nonzero vertical frequency for F, we have

$$\int_{y\in G/\Gamma} F(y)\ dy = 0.$$

Therefore, by definition we see the polynomial sequence $g(6P_r(Sy+I))$ is not $3\delta^2L^{-1}$-equidistributed. (Notice that $g(6P_r(Sy+I))$ is a polynomial sequence with respect to the filtration $\widetilde{G}_\bullet$ defined by $G = \widetilde{G}_0 = \widetilde{G}_1 = \widetilde{G}_2 = \widetilde{G}_3 = \widetilde{G}_4 \geqslant \widetilde{G}_5 = G_2 \geqslant \mathrm{Id}_G$.)

Let $\psi(g)$ denote the Mal'cev coordinates of $g \in G$ with respect to $\mathcal{X}$. By the classification of polynomial sequences in terms of Mal'cev coordinates [21, Lemma 6.7] and the assumption that $g(0) = \mathrm{Id}_G$, we have

$$\psi(g(n)) = \binom{n}{2}t_2 + \binom{n}{1}t_1,$$

where $t_i \in \mathbf{R}^m$ and the first $m - \dim(G_2)$ coordinates of t_2 are zero. As $g(6P_r(Sy+I))$ is not $3\delta^2L^{-1}$-equidistributed, by Theorem 5.5 there exists a nonzero horizontal character η such that

$$\|(\eta\circ g)(6P_r(Sy+I))\|_{C^\infty[N^{1/2}W^{-1}]} \le \mathrm{poly}_m(\delta^{-1}L).$$

The implied constants in $\mathrm{poly}_m(\cdot)$ are absolute, as the degree of the filtration under consideration is always bounded by 5. By Claim 5.8 there exists a positive integer $Q \le \delta^{-O(1)}$ such that

$$\|Q(\eta\circ g)(6P_r(y))\|_{C^\infty[N^{1/2}W^{-1}]} \le \mathrm{poly}_m(\delta^{-1}L). \tag{5.2}$$

By a direct computation, we have

$$6P_r(y) = Ay^2 + By,$$

where

$$A = 6W^2 \quad\text{ and }\quad B = 6(2Wr+1).$$

Now let the horizontal character η be represented by $k \in \mathbf{Z}^m$ in Mal'cev coordinates. Thus $(\eta\circ g)(n) = k\cdot(\binom{n}{2}t_2 + \binom{n}{1}t_1)$. Plugging into (5.2) and using Claim 5.7, and unwrapping the definition of the $C^\infty[N^{1/2}W^{-1}]$-norm, we can initially deduce that

$$\|12QA^2\cdot(t_2\cdot k)\|_{\mathbf{T}} \le \frac{\mathrm{poly}_m(\delta^{-1}L)W^4}{N^2}, \qquad \|(18A^2+6AB)Q\cdot(t_2\cdot k)\|_{\mathbf{T}} \le \frac{\mathrm{poly}_m(\delta^{-1}L)W^3}{N^{3/2}}.$$

Therefore, there exists a positive integer $Q_1 \le \delta^{-O(1)}$ such that

$$\|Q_1W^4\cdot(t_2\cdot k)\|_{\mathbf{T}} \le \frac{\mathrm{poly}_m(\delta^{-1}L)W^4}{N^2}, \qquad \|Q_1(3W^4+W^2(2Wr+1))\cdot(t_2\cdot k)\|_{\mathbf{T}} \le \frac{\mathrm{poly}_m(\delta^{-1}L)W^3}{N^{3/2}}. \tag{5.3}$$

Combining these bounds yields

$$\|Q_1W^2(2Wr+1)\cdot(t_2\cdot k)\|_{\mathbf{T}}\le\frac{\mathrm{poly}_m(\delta^{-1}L)W^3}{N^{3/2}}.$$

We now, crucially, use that $\gcd(W,2Wr+1)=1$. Note that

$$Q_1(t_2\cdot k)=\frac{T_1}{W^4}+E_1=\frac{T_2}{W^2(2Wr+1)}+E_2$$

with $|E_1|,|E_2|\le N^{-1}$ (say) and $T_1,T_2\in\mathbf{Z}$. However,

$$\left|\frac{T_1}{W^4}-\frac{T_2}{W^2(2Wr+1)}\right|\ge N^{-1/2}$$

unless $T_1\cdot W^2(2Wr+1)-T_2\cdot(W^4)=0$. It follows that $W^2\mid T_1$ and, therefore,

$$\|Q_1W^2\cdot(t_2\cdot k)\|_{\mathbf{T}}\le\frac{\mathrm{poly}_m(\delta^{-1}L)W^2}{N^2},$$

using the first bound in (5.3). Noting that $W^2\mid A$ and using that $|B|\ll W^2$, we have

$$\|Q_1(7A^2+6AB-A)(t_2\cdot k)\|_{\mathbf{T}}+\left\|2Q_1\left(AB+\binom{A}{2}\right)(t_2\cdot k)\right\|_{\mathbf{T}}\le\frac{\mathrm{poly}_m(\delta^{-1}L)W^4}{N^2}.\tag{5.4}$$

Now we use (5.2) again but applied to the lower coefficients, and we appropriately cancel out the contributions from the terms in (5.4). We find

$$\|2Q_1(B^2(t_2\cdot k)+2A(t_1\cdot k))\|_{\mathbf{T}}\le\frac{\mathrm{poly}_m(\delta^{-1}L)W^2}{N}$$

and

$$\left\|2Q_1\left(\binom{B}{2}(t_2\cdot k)+(A+B)(t_1\cdot k)\right)\right\|_{\mathbf{T}}\le\frac{\mathrm{poly}_m(\delta^{-1}L)W}{N^{1/2}}.$$

Multiplying the first equation by $A+B$ and the second equation by $2A$ and subtracting, we find that

$$\|2Q_1(B^2(A+B)-AB(B-1))(t_2\cdot k)\|_{\mathbf{T}}\le\frac{\mathrm{poly}_m(\delta^{-1}L)W^3}{N^{1/2}}.$$

As $W^2\mid A$, we find by similar argumentation that

$$\|2Q_1B^3(t_2\cdot k)\|_{\mathbf{T}}\le\frac{\mathrm{poly}_m(\delta^{-1}L)W^3}{N^{1/2}}.$$

Again, crucially, $\gcd((2Wr+1)^3,W)=1$. As $Q_1(t_2\cdot k)$ is near a fraction with denominator W^2, repeating fraction comparison arguments similar to above we find that for $Q_2=4Q_1\le\delta^{-O(1)}$ we have

$$\|Q_2(t_2\cdot k)\|_{\mathbf{T}}\le\frac{\mathrm{poly}_m(\delta^{-1}L)}{N^2}.$$

We may substitute this bound into earlier equations, and using the size bounds on B deduce that

$$\|Q_2A(t_1\cdot k)\|_{\mathbf{T}}\le\frac{\mathrm{poly}_m(\delta^{-1}L)W^2}{N},\qquad\|Q_2(A+B)(t_1\cdot k)\|_{\mathbf{T}}\le\frac{\mathrm{poly}_m(\delta^{-1}L)W}{N^{1/2}}.$$

As $\gcd(A,A+B)=\gcd(A,B)=6$, another fraction comparison argument shows

$$\|6Q_2(t_1\cdot k)\|_{\mathbf{T}}\le\frac{\mathrm{poly}_m(\delta^{-1}L)}{N}.$$

Thus for $Q_3=6Q_2\le\delta^{-O(1)}$ we have

$$\|Q_3(t_2\cdot k)\|_{\mathbf{T}}\le\frac{\mathrm{poly}_m(\delta^{-1}L)}{N^2},\qquad\|Q_3(t_1\cdot k)\|_{\mathbf{T}}\le\frac{\mathrm{poly}_m(\delta^{-1}L)}{N}.$$

Note that the quality of major arc control here is comparable to a situation where we knew that $g(y)$ itself were poorly equidistributed on $[N]$.

The remaining proof is now essentially identical to the argument in [23, Proposition 9.2], as we are in the same essential position. We will define G' to be the connected component of $\ker(\eta)$ (as a subgroup of G) and, due to the size bounds on η, we have that G' is a $\mathrm{poly}_m(\delta^{-1}L)$-rational subgroup. It is seen to be simply connected by considering the Mal'cev coordinate representation for η.

We choose vectors $u_1, u_2 \in \mathbf{R}^m$ such that $\|t_j - u_j\|_\infty \le \mathrm{poly}_m(\delta^{-1}L)N^{-j}$ for $j = 1, 2$, such that $Q_3(u_1 \cdot k)$ and $Q_3(u_2 \cdot k)$ are integers, and such that the first $m - \dim(G_2)$ coordinates of u_2 are zero. We then choose vectors v_1 and v_2 with coordinates rationals with denominator bounded by $\mathrm{poly}_m(\delta^{-1}L)$ and such that $k \cdot u_j = k \cdot v_j$ for $j = 1, 2$.

Let ε and γ be the polynomial sequences $\mathbf{Z} \to G$ for which

$$\psi(\varepsilon(n)) = \binom{n}{2}(t_2 - u_2) + \binom{n}{1}(t_1 - u_1)$$

and

$$\psi(\gamma(n)) = \binom{n}{2}v_2 + \binom{n}{1}v_1,$$

and set

$$g' = \varepsilon^{-1} g \gamma^{-1}.$$

By construction, g' takes values in G' since η is a horizontal character. We have that γ is rational, as the denominators of v_i are $\mathrm{poly}_m(\delta^{-1}L)$-bounded, and therefore by [23, Lemma A.11(iv), A.12(ii)] we have that $\gamma(\cdot)$ is $\mathrm{poly}_m(\delta^{-1}L)$-rational and periodic of period at most $\mathrm{poly}_m(\delta^{-1}L)$. The claimed smoothness bounds for ε follow using that $\|t_j - u_j\|_\infty \le \mathrm{poly}_m(\delta^{-1}L)N^{-j}$ and [23, Lemma A.4], which converts between distance in the metric $d_{\mathcal{X}}$ and differences in Mal'cev coordinates. This completes the proof. □

Note that subgroup G' obtained from Proposition 5.6 is not dependent on the vertical character ξ in any manner; we only needed that the mean of F on G/Γ is 0. However, we may iterate Proposition 5.6 until ξ is trivial on $G_2' = G_2 \cap G'$.

Lemma 5.9. *Fix $\delta \in (0, 1/2)$ and P and W as in (2.2). Let G/Γ be an m-dimensional filtered nilmanifold of degree 2 and complexity L. Furthermore, let $\mathcal{X}$ denote the Mal'cev basis of G/Γ and let $F\colon G/\Gamma \to \mathbf{C}$ be such that $\|F\|_{\mathrm{Lip}} \le L$ and F has vertical frequency $2 \cdot \xi$ such $\|\xi\|_\infty \le L$. Let $g\colon \mathbf{Z} \to \mathbf{G}$ be a polynomial sequence with respect to the filtration $G = G_0 = G_1 \geqslant G_2 \geqslant \mathrm{Id}_G$, denoted by $G_\bullet$, and $g(0) = \mathrm{Id}_G$. Finally, for $r \in [W]$, define*

$$P_r(y) = \frac{P(Wy + r) - P(r)}{W}.$$

Suppose that $W \le N^{1/10^4}$, $N \ge \mathrm{poly}_m(\delta^{-1}L)$, and that

$$\left|\sum_{y \in [\pm T]} F(g(6P_r(y))\Gamma)\right| \ge \delta N^{1/2} W^{-1}$$

for some $T \in [\delta^{-1}N^{1/2}W^{-1}]$. Then there exists a factorization $g = \varepsilon g' \gamma$ and subgroup G' with polynomial sequences $\varepsilon, g', \gamma\colon \mathbf{Z} \to G$ such that

- *for all $n \in [\pm\delta^{-2}N]$, $d(\varepsilon(n), \varepsilon(n-1)) \le \mathrm{poly}_m(L\varepsilon^{-1})/N$ and $d(\varepsilon(n), \mathrm{Id}_G) \le \mathrm{poly}_m(L\delta^{-1})$;*
- *γ is $\mathrm{poly}_m(L\delta^{-1})$-rational and $\gamma(n)\Gamma$ is periodic with period at most $\mathrm{poly}_m(L\delta^{-1})$;*
- *g' takes values in a connected $\mathrm{poly}_m(L\delta^{-1})$-rational subgroup G' and is a polynomial sequence with respect to the filtration $G'_\bullet$, where $G'_j = G_j \cap G'$; and*
- *ξ is trivial on $G'_2 = G_2 \cap G'$.*

Proof. We first handle the trivial case where $\xi = 0$. This case is dispatched via setting $G' = G$, $g' = g$, and ε and γ to both be identically Id_G.

Otherwise, we iteratively define a sequence of parameters (δ_i) with $\delta_1^{-1} = \delta^{-1}L$ and $\delta_{i+1}^{-1} = \mathrm{poly}_m(\delta_i^{-1}L)$ and a sequence of rational connected subgroups $(G^{(i)})$ with $G^{(1)} = G$ and $G^{(i)}$ being δ_i^{-1}-rational with respect to G. We write $G_j^{(i)} = G_j \cap G^{(i)}$. At each stage, we have the factorization

$$g = \varepsilon_i g_i \gamma_i$$

with $g_i(0) = \mathrm{Id}_G$, g_i taking values in $G^{(i)}$, ε_i satisfying, for $n \in [\pm\delta^{-2}N]$, that $d(\varepsilon_i(n), \varepsilon_i(n-1)) \le \delta_i^{-1}N^{-1}$ and $d(\varepsilon_i(n), \mathrm{Id}_G) \le \delta_i^{-1}$, and γ_i being δ_i^{-1}-rational and periodic with period at most δ_i^{-1}. We let $\varepsilon_1 = \gamma_1 = \mathrm{Id}_G$ and $g_1 = g$ to start.

Now given i, we define the next factorization data. If ξ is trivial on $G_2^{(i)}$ then we terminate, providing our desired final factorization. Else, decompose $[\pm T]$ into arithmetic progressions that are $N^{1/2}W^{-1}\,\mathrm{poly}_m(\delta_i L^{-1})$ in length and with common difference divisible by the period of γ_i. Then, by the pigeonhole principle, there exists such a progression Q for which

$$\left|\sum_{y\in Q} F(g(6P_r(y))\Gamma)\right| \ge \mathrm{poly}_m(\delta_i L^{-1})N^{1/2}W^{-1}.$$

By the smoothness of ε_i, the rationality of γ_i, and the Lipschitz bound for F, there exist group elements ε_Q and γ_Q, each of size $\mathrm{poly}_m(\delta_i^{-1}L)$, with γ_Q being $\mathrm{poly}_m(\delta_i^{-1}L)$-rational, such that

$$\left|\sum_{y\in Q} F(\varepsilon_Q g_i(6P_r(y))\gamma_Q\Gamma)\right| \ge \mathrm{poly}_m(\delta_i L^{-1})N^{1/2}W^{-1}.$$

Note here that γ_Q is essentially a "representative" for γ_i in this modular class that is bounded, and not the value of γ_i itself. Such a representative exists, as any group element can be made bounded by right-multiplying by an element of the Γ [23, Lemma A.14] and the product of two rational elements is rational with appropriate height bounds [23, Lemma A.11].

Set $F_i(x) = F(\varepsilon_Q\gamma_Q x)$. Note that F_i is $\mathrm{poly}_m(\delta_i^{-1}L)$-Lipschitz, as left-multiplication by bounded elements approximately preserves the metric [23, Lemma A.5]. Furthermore, letting $g_i' = \gamma_Q^{-1}g_i\gamma_Q$, we have

$$\left|\sum_{y\in Q} F_i(g_i'(6P_r(y))\Gamma)\right| \ge \mathrm{poly}_m(\delta_i L^{-1})N^{1/2}W^{-1}.$$

Since $G^{(i)}$ is a $\mathrm{poly}_m(\delta_i^{-1}L)$-rational subgroup of G, the conjugate subgroup $\gamma_Q^{-1}G^{(i)}\gamma_Q$ is similarly rational by [23, Lemma A.13]. Furthermore, note that $\gamma_Q^{-1}G_2^{(i)}\gamma_Q = G_2^{(i)}$, as $G_2^{(i)} \subseteq G_2$ is in the center of G because we have a degree 2 filtration on G. Therefore, as ξ is nonzero on $\gamma_Q^{-1}G_2^{(i)}\gamma_Q = G_2^{(i)}$, and since $G_2^{(i)}$ being simply connected implies that if ξ is nonzero then $2\cdot\xi$ is nonzero, we can apply Proposition 5.6 to obtain

$$g_i' = \widetilde{\varepsilon}_{i+1} g_{i+1}\widetilde{\gamma}_{i+1}$$

where $\widetilde{\gamma}_{i+1}$ is $\mathrm{poly}_m(\delta_i^{-1}L)$-rational and periodic, $d(\widetilde{\varepsilon}_{i+1}(n), \widetilde{\varepsilon}_{i+1}(n-1)) \le \mathrm{poly}_m(\delta_i^{-1}L)N^{-1}$ and $d(\widetilde{\varepsilon}_{i+1}(n), \mathrm{Id}_G) \le \mathrm{poly}_m(\delta_i^{-1}L)$ for $n \in [\pm\delta^{-2}N]$, and g_{i+1} lives in a subgroup $G^{(i+1)}$ that is $\mathrm{poly}_m(\delta_i^{-1}L)$-rational with respect to $G^{(i)}$. Thus,

$$g_i = \gamma_Q\widetilde{\varepsilon}_{i+1} g_{i+1}\widetilde{\gamma}_{i+1}\gamma_Q^{-1}$$

and, so,

$$g = \varepsilon_i\gamma_Q\widetilde{\varepsilon}_{i+1} g_{i+1}\widetilde{\gamma}_{i+1}\gamma_Q^{-1}\gamma_i.$$

Taking $\varepsilon_{i+1} = \varepsilon_i\gamma_Q\widetilde{\varepsilon}_{i+1}$ and $\gamma_{i+1} = \widetilde{\gamma}_{i+1}\gamma_Q^{-1}\gamma_i$ completes the iteration. In particular, $\gamma_Q\widetilde{\varepsilon}_{i+1}$ is seen to be sufficiently smooth as left-multiplication by bounded elements approximately preserves distances [23, Lemma A.5], and ε_i is sufficiently smooth as the product of smooth sequences is sufficiently smooth by [23, Lemma 10.1]. The rationality claims for γ_{i+1} follow immediately from [23, Lemma A.11, A.12].

Note that at each step of the iteration we have $\delta_{i+1}^{-1} = \mathrm{poly}_m(L\delta_i^{-1})$, where the implied constants in $\mathrm{poly}_m(\cdot)$ are absolute. Note also that there are at most m iterations, as each iteration decreases the dimension of $G^{(i)}$ (since the G' produced by Proposition 5.6 is a connected proper subgroup), and therefore we obtain the desired result (up to slightly increasing the implicit constants in the underlying notation). □

6. Degree-lowering

The main purpose of this section is to deduce the following key degree-lowering result.

Proposition 6.1. *Fix a positive integer $k \geq 3$, let w, W, and P be as in (2.2), and let $\delta \in (0, 1/2)$. Let $f_1, f_2, f_3 \colon \mathbf{Z} \to \mathbf{C}$ be 1-bounded functions such that $\mathrm{supp}(f_i) \subseteq [\pm\delta^{-1}N]$ for $i = 1, 2, 3$. Suppose that*

$$\|\mathcal{D}^1(f_2, f_3)\|_{U^k_{W\cdot[N/W]}}^{2^k} \geq \delta N.$$

Furthermore, suppose that $N \geq W^{\Omega(1)} \cdot \exp(\exp(\delta^{-\Omega_k(1)}))$. Then,

$$\min_{i=2,3} \|f_i\|_{U^{k-1}_{W\cdot[N/W]}}^{2^{k-1}} \gg \exp(-\exp(\delta^{-O_k(1)})) \cdot N.$$

We will also require the following variant of the above result; the proof is identical, just replacing the polynomial $P(y) = Wy^2 + y$ with $-Wy^2 - y$.

Proposition 6.2. *Fix a positive integer $k \geq 3$, and let w, W, and P be as in (2.2), and let $\delta \in (0, 1/2)$. Let $f_1, f_2, f_3 \colon \mathbf{Z} \to \mathbf{C}$ be 1-bounded functions such that $\mathrm{supp}(f_i) \subseteq [\delta^{-1}N]$ for $i = 1, 2, 3$. Suppose that*

$$\|\mathcal{D}^3(f_1, f_2)\|_{U^k_{W\cdot[N/W]}}^{2^k} \geq \delta N.$$

Furthermore, suppose that $N \geq W^{\Omega(1)} \cdot \exp(\exp(\delta^{-\Omega_k(1)}))$. Then,

$$\min_{i=1,2} \|f_i\|_{U^{k-1}_{W\cdot[N/W]}}^{2^{k-1}} \gg \exp(-\exp(\delta^{-O_k(1)})) \cdot N.$$

Remark. The methods in this paper do *not* prove the analogous statement for $\mathcal{D}^2(f_3, f_1)$, as our methods do not prove the needed statement corresponding to Lemma 5.3. By symmetry, the constraints required for $\mathcal{D}^1(f_2, f_3)$ and $\mathcal{D}^3(f_1, f_2)$ are identical.

6.1. U^2-control for Sárközy-type configurations.

We first require U^2-control for Sárközy-type configurations. The proof we give is identical to that of Green [17, Section 3], modulo standard circle method computations that we place in Appendix B.

Lemma 6.3. *There exists a constant $c = c_{6.3} > 0$ such that the following holds. Let W be as in (2.2) with $W \leq N^c$ and let $f_i \colon \mathbf{Z} \to \mathbf{C}$ be 1-bounded with $\mathrm{supp}(f_i) \subseteq [\pm\delta^{-1}N]$. Define*

$$P_k(y) = \frac{P(Wy + k) - P(k)}{W}$$

for $k \in [W]$, and suppose that

$$\left| \sum_{x \in \mathbf{Z}} \mathbf{E}_{y \in [\pm N^{1/2}W^{-1}]} f_1(x + P_k(y)) f_2(x + 2P_k(y)) \right| \geq \delta N.$$

Then,

$$\min_{i \in \{1,2\}} \sup_{\Theta \in \mathbf{T}} |\widehat{f_i}(\Theta)| \gg \delta^{O(1)} N.$$

Proof. We have
$$\Bigg|\sum_{x\in\mathbf{Z}}\sum_{y\in[\pm N^{1/2}W^{-1}]} f_1(x+P_k(y))f_2(x+2P_k(y))\Bigg| \geq \delta N^{3/2}W^{-1}.$$
Let $F(t)$ denote the indicator of the set $\{P_k(y)\colon y\in[\pm N^{1/2}W^{-1}]\}$[3], and thus we have
$$\Bigg|\sum_{x\in\mathbf{Z}}\sum_{y\in\mathbf{Z}} f_1(x+t)f_2(x+2t)F(t)\Bigg| \geq \delta N^{3/2}W^{-1}.$$
Applying Fourier inversion, this is equivalent to
$$\Bigg|\int_{\mathbf{T}} \widehat{f_1}(\Theta)\widehat{f_2}(-\Theta)\widehat{F}(\Theta)\ d\Theta\Bigg| \geq \delta N^{3/2}W^{-1}.$$
We now prove the result for $i=1$; the result for $i=2$ is analogous. Note that
$$\begin{aligned}\delta N^{3/2}W^{-1} &\leq \Bigg|\int_{\mathbf{T}} \widehat{f_1}(\Theta)\widehat{f_2}(-\Theta)\widehat{F}(\Theta)\ d\Theta\Bigg| \leq \sup_{\Theta\in\mathbf{T}}|\widehat{f_1}(\Theta)|^{1/3}\cdot\int_{\mathbf{T}}|\widehat{f_1}(\Theta)|^{2/3}|\widehat{f_2}(-\Theta)||\widehat{F}(\Theta)|\ d\Theta\\
&\leq \sup_{\Theta\in\mathbf{T}}|\widehat{f_1}(\Theta)|^{1/3}\cdot\Bigg(\int_{\mathbf{T}}|\widehat{f_1}(\Theta)|^2\ d\Theta\Bigg)^{1/3}\Bigg(\int_{\mathbf{T}}|\widehat{f_2}(-\Theta)|^2\ d\Theta\Bigg)^{1/2}\Bigg(\int_{\mathbf{T}}|\widehat{F}(\Theta)|^6\ d\Theta\Bigg)^{1/6}\\
&\ll \delta^{-O(1)}N^{5/6}\sup_{\Theta\in\mathbf{T}}|\widehat{f_1}(\Theta)|^{1/3}\Bigg(N^2W^{-6}\Bigg)^{1/6}\\
&\ll \delta^{-O(1)}N^{7/6}W^{-1}\sup_{\Theta\in\mathbf{T}}|\widehat{f_1}(\Theta)|^{1/3},\end{aligned}$$
where we have used Lemma B.7 (with N replaced by $N^{1/2}W^{-1}$) to bound the L^6-norm of $\widehat{F}$. □

6.2. **Dual-difference interchange.** The version of dual-difference interchange we use is a minor variant of [40, Lemma 7.4]; we include a proof for completeness.

Lemma 6.4. *Consider a 1-bounded function $f\colon\mathbf{Z}\times S\to\mathbf{C}$ such that $\operatorname{supp} f(\cdot,y)\subseteq[-CN,CN]$ for all $y\in S$, and integers T_1,T_2 such that $T_1\cdot T_2\leq CN$. Set $F(x):=\mathbf{E}_{y\in S}f(x,y)$, fix integers $1\leq\ell\leq k$, and suppose that*
$$\sum_{x\in\mathbf{Z}}\mathbf{E}_{\substack{h_i,h_i'\in T_1\cdot[T_2]\\1\leq i\leq k}}\Delta'_{(h_i,h_i')_{i=1}^k}F(x)\geq\delta N.$$
Then, we have that
$$\mathbf{E}_{\substack{h_i,h_i'\in T_1\cdot[T_2]\\1\leq i\leq \ell}}\|\mathbf{E}_{y\in S}\Delta'^{(x)}_{(h_i,h_i')_{i=1}^{\ell}}f(x,y)\|^{2^{k-\ell}}_{U^{k-\ell}_{T_1\cdot[T_2]}}\gg(C^{-1}\delta)^{O_k(1)}N.$$

Proof. The proof is exactly as in [40, Lemma 7.4], noting that the properties of the dual function are used only in the form of F given above. For the computation below, let $\vec{h}=(h_1,\ldots,h_{k-1})$ and $\vec{h'}=(h_1',\ldots,h_{k-1}')$, and let $\mathcal{C}^t$ denote complex conjugation t times (which depends only on the parity of t). We have, using Cauchy–Schwarz to duplicate h_k' in the middle,
$$\begin{aligned}&\Bigg(\sum_{x\in\mathbf{Z}}\mathbf{E}_{\substack{h_i,h_i'\in T_1\cdot[T_2]\\1\leq i\leq k}}\Delta'_{(h_i,h_i')_{i=1}^k}F(x)\Bigg)^2\\
&=\Bigg(\mathbf{E}_{\substack{y_{\omega0},y_{\omega1}\in S\\\omega\in\{0,1\}^{k-1}}}\sum_{x\in\mathbf{Z}}\mathbf{E}_{\substack{h_i,h_i'\in T_1\cdot[T_2]\\1\leq i\leq k}}\prod_{\omega\in\{0,1\}^{k-1}}\mathcal{C}^{|\omega|-1}(f(x+\vec{h}\cdot\omega+\vec{h'}\cdot(1-\omega)+h_k,y_{\omega0})\end{aligned}$$

[3]Note that $P(y_1)=P(y_2)$ implies $(y_1-y_2)(W(y_1+y_2)+1)=0$. Therefore, every element in the set occurs with multiplicity 1.

$$\times \overline{f(x+\vec{h}\cdot\omega+\vec{h'}\cdot(1-\omega)+h'_k,y_{\omega 1}))}\Bigg)^2$$

$$\leq \Bigg(\mathbf{E}_{\substack{y_{\omega 0},y_{\omega 1}\in S\\ \omega\in\{0,1\}^{k-1}}}\sum_{x\in\mathbf{Z}}\mathbf{E}_{\substack{h_i,h'_j\in T_1\cdot[T_2]\\ 1\leq i\leq k\\ 1\leq j\leq k-1}}\prod_{\omega\in\{0,1\}^{k-1}}|f(x+\vec{h}\cdot\omega+\vec{h'}\cdot(1-\omega)+h_k,y_{\omega 0})|^2\Bigg)$$

$$\cdot\Bigg(\mathbf{E}_{\substack{y_{\omega 0},y_{\omega 1}\in S\\ \omega\in\{0,1\}^{k-1}}}\sum_{x\in\mathbf{Z}}\mathbf{E}_{\substack{h_i,h'_j\in T_1\cdot[T_2]\\ 1\leq i\leq k\\ 1\leq j\leq k-1}}\mathbf{E}_{h'_{k,1},h'_{k,2}\in T_1\cdot[T_2]}\prod_{\omega\in\{0,1\}^{k-1}}\mathcal{C}^{|\omega|-1}(f(x+\vec{h}\cdot\omega+\vec{h'}\cdot(1-\omega)+h'_{k,1},y_{\omega 1})$$

$$\times\overline{f(x+\vec{h}\cdot\omega+\vec{h'}\cdot(1-\omega)+h'_{k,2},y_{\omega 1}))}\Bigg)$$

$$\ll CN\cdot\Bigg(\mathbf{E}_{\substack{y_{\omega 0},y_{\omega 1}\in S\\ \omega\in\{0,1\}^{k-1}}}\sum_{x\in\mathbf{Z}}\mathbf{E}_{\substack{h_i,h'_i\in T_1\cdot[T_2]\\ 1\leq i\leq k}}\prod_{\omega\in\{0,1\}^{k-1}}\mathcal{C}^{|\omega|-1}\Delta'^{(x)}_{(h_k,h'_k)}f(x+\vec{h}\cdot\omega+\vec{h'}\cdot(1-\omega),y_{\omega 1})\Bigg)$$

$$\ll CN\cdot\Bigg(\mathbf{E}_{\substack{y_{\omega}\in S\\ \omega\in\{0,1\}^{k-1}}}\sum_{x\in\mathbf{Z}}\mathbf{E}_{\substack{h_i,h'_i\in T_1\cdot[T_2]\\ 1\leq i\leq k}}\prod_{\omega\in\{0,1\}^{k-1}}\mathcal{C}^{|\omega|-1}\Delta'^{(x)}_{(h_k,h'_k)}f(x+\vec{h}\cdot\omega+\vec{h'}\cdot(1-\omega),y_{\omega})\Bigg).$$

The result follows by replacing replacing f by $\Delta'^{(x)}_{(h_k,h'_k)}f$ and applying iterating, for a total of ℓ times. We use that $T_1\cdot T_2$ is smaller than CN in order to guarantee appropriate support conditions and bounds. ☐

6.3. **Hensel's lemma.** We will also require an elementary result number-theoretic result; this is ultimately why the W-trick can be used to treat arithmetic progressions with common difference of the form y^2-1, but not y^2.

Proposition 6.5. *Let $Q(y)=ay^2+by$ and fix a prime p such that $p\mid a$ but $p\nmid b$. Then, for all $k\geq 1$, $Q(y)$ gives a bijective map $\mathbf{Z}/p^k\mathbf{Z}\to\mathbf{Z}/p^k\mathbf{Z}$.*

Proof. Note that for $k=1$ this is immediate, as $P(y)$ reduces to a nontrivial linear function on $\mathbf{Z}/p\mathbf{Z}$. Furthermore, note that $P'(y)=2ay+b$ is always nonzero when viewed modulo p. Therefore, the desired result follows from Hensel's lemma. ☐

6.4. **Completing the proof of Proposition 6.1.** Before proceeding with the main proof, we require the U^3-inverse theorem. The result stated follows by embedding the interval $[N]$ into a slightly larger cyclic group and using the U^3-inverse theorem of Green and Tao [19, Theorem 12.8][4]. We give a brief deduction of the inverse theorem stated below from [19, Theorem 12.8], since the definition of U^3-norm we use is slightly different from the standard version.

Theorem 6.6. *Suppose that $f\colon\mathbf{Z}\to\mathbf{C}$ is a 1-bounded function such that* $\operatorname{supp}(f)\subseteq[\pm N]$ *and*

$$\|f\|^8_{U^3_{[5N]}}\geq\delta N.$$

Then, there exists a degree 2 nilmanifold G/Γ with dimension $\delta^{-O(1)}$ and complexity $\exp(\delta^{-O(1)})$, a function $F\colon G/\Gamma\to\mathbf{C}$ with $\|F\|_{\mathrm{Lip}}\leq\exp(\delta^{-O(1)})$, and a polynomial sequence $g\colon\mathbf{Z}\to G$ such that

$$\Bigg|\sum_{n\in\mathbf{Z}}f(n)F(g(n)\Gamma)\Bigg|\geq\exp(-\delta^{O(1)})N.$$

[4] Note that the theorem stated in [19, Theorem 12.8] produces correlation of a shifted version of f with a nilsequence but, as remarked after the theorem, the shift can be removed.

Proof. Note that

$$\delta N \le \|f\|_{U^3_{[5N]}}^8 = \sum_h \mu_{5N}\|\Delta_h f\|_{U^2_{[5N]}}^2 \asymp \mathbf{E}_{h\in[\pm 2N]}\|\Delta_h f\|_{U^2_{[5N]}}^2,$$

where we have used that $\Delta_h f = 0$ for $|h| > 2N$ and that there exist absolute constants $c, C > 0$ such that $cN^{-1} \le \mu_{5N}(h) \le CN^{-1}$ for all $|h| \le 2N$. By Markov and Lemma C.4, we find that

$$\mathbf{E}_{h\in[\pm 2N]}\sup_{\beta\in\mathbf{T}}\left|\sum_{x\in\mathbf{Z}}\Delta_h f(x)e(\beta x)\right| \gg \delta^{O(1)}N.$$

Note that if the β Fourier sum is large then so will the $[\beta \pm \delta^{O(1)}/N]$ Fourier sums. So, by the identity

$$\sum_{x,h_1,h_2} f(x)\overline{f(x+h_1)f(x+h_2)}f(x+h_1+h_2) = \int_{\mathbf{T}}|\widehat{f}(\Theta)|^4\, d\Theta$$

and Markov, it follows that

$$\mathbf{E}_{h_3\in[\pm 2N]}\sum_{x,h_1,h_2}(\Delta_{h_3}f)(x)\overline{(\Delta_{h_3}f)(x+h_1)(\Delta_{h_3}f)(x+h_2)}(\Delta_{h_3}f)(x+h_1+h_2) \gg \delta^{O(1)}N^3.$$

This implies

$$\sum_{x,h_1,h_2,h_3}\Delta_{h_1,h_2,h_3}f(x) \gg \delta^{O(1)}N^4.$$

Now treat f as a function on the cyclic group $\mathbf{Z}/(2L+1)\mathbf{Z}$, where $L \in [25N, 50N]$, $2L+1$ is prime, and we identify $\mathbf{Z}/(2L+1)\mathbf{Z}$ with $[-L, L]$. The above lower bound implies that f, viewed as a function on $\mathbf{Z}/(2L+1)\mathbf{Z}$, has large U^3-norm in the sense of [19, Theorem 12.8], and therefore the desired result follows from [19, Theorem 12.8]. □

We now perform a preliminary transformation of Theorem 6.6 that allow us to assume that $g(0) = \mathrm{Id}_G$ and that F has a vertical frequency.

Theorem 6.7. *Suppose that* $f\colon \mathbf{Z} \to \mathbf{C}$ *is a* 1*-bounded function such that* $\mathrm{supp}(f) \subseteq [\pm N]$ *and*

$$\|f\|_{U^3_{[5N]}}^8 \ge \delta N.$$

Then there exists a degree 2 *nilmanifold* G/Γ *with dimension* $\delta^{-O(1)}$ *and complexity* $\exp(\delta^{-O(1)})$*, a function* $F\colon G/\Gamma \to \mathbf{C}$ *with* $\|F\|_{\mathrm{Lip}} \le \mathrm{poly}_{\delta^{-1}}(\delta^{-1})$ *possessing a vertical frequency* ξ *with* $\|\xi\| \le \mathrm{poly}_{\delta^{-1}}(\delta^{-1})$*, and a polynomial sequence* $g\colon \mathbf{Z} \to G$ *with* $g(0) = \mathrm{Id}_G$ *such that*

$$\left|\sum_{n\in\mathbf{Z}} f(n)F(g(n)\Gamma)\right| \ge \mathrm{poly}_{\delta^{-1}}(\delta)N.$$

Proof. First apply Theorem 6.6 to find some $G/\Gamma, F, g$ which appropriately correlated with f. We may replace F by some F' which has a vertical frequency $\|\xi\| \le \mathrm{poly}_{\delta^{-1}}(\delta^{-1})$ by applying Lemma A.9 with error parameter ε taken to be $\exp(-\delta^{-O(1)})$ and using the pigeonhole principle. The Lipschitz constant is now of quality $\mathrm{poly}_{\delta^{-1}}(\delta^{-1})$. Note here we are using that $\mathrm{poly}_{\delta^{-1}}(\exp(\delta^{-O(1)})) \le \mathrm{poly}_{\delta^{-1}}(\delta^{-1})$ up to changing the implicit constants.

To force $g(0) = \mathrm{Id}_G$, by using [23, Lemma A.14] we can factor $g(0) = \{g(0)\}[g(0)]$ with $\|\psi(\{g(0)\})\|_\infty \le 1$ and $[g(0)] \in \Gamma$. Then, we have that

$$\begin{aligned} F'(g(n)\Gamma) &= F'(g(n)g(0)^{-1}g(0)\Gamma) \\ &= F'(g(n)g(0)^{-1}\{g(0)\}\Gamma) \\ &= F'(\{g(0)\}(\{g(0)\}^{-1}g(n)g(0)^{-1}\{g(0)\})\Gamma), \end{aligned}$$

and taking $\widetilde{F}(x) = F'(\{g(0)\}^{-1}x)$ and $\widetilde{g}(n) = \{g(0)\}^{-1}g(n)g(0)^{-1}\{g(0)\}$ gives the desired. □

We are now in position to complete the proof of Proposition 6.1.

Proof of Proposition 6.1. Throughout the proof δ will be assumed to be smaller than an appropriate absolute constant.

Step 1: Applying dual-difference interchange. By the definition of the box-norm, we have that

$$\sum_{x\in\mathbf{Z}}\mathbf{E}_{\substack{h_i,h_i'\in W\cdot[N/W]\\1\le i\le k}}\Delta'_{(h_i,h_i')_{i=1}^k}\mathcal{D}^1(f_2,f_3)\ge\delta N.$$

Recall that $\mathcal{D}^1(f_2,f_3)(x)=\mathbf{E}_{y\in[\pm M]}f_2(x+P(y))f_3(x+2P(y))$ with $M=\lfloor N^{1/2}W^{-1/2}\rfloor$, and define

$$g(x,y)=f_2(x+P(y))f_3(x+2P(y))\mathbb{1}_{|y|\le M}\mathbb{1}_{|x|\le 100\delta^{-1}N}.$$

It follows via the support conditions on the f_i that we immediately have

$$\sum_{x\in\mathbf{Z}}\mathbf{E}_{\substack{h_i,h_i'\in W\cdot[N/W]\\1\le i\le k}}\Delta'_{(h_i,h_i')_{i=1}^k}(\mathbf{E}_{y\in[\pm M]}g(x,y))\ge\delta N.$$

Applying Lemma 6.4, we deduce that

$$\mathbf{E}_{\substack{h_i,h_i'\in W\cdot[N/W]\\1\le i\le k-3}}\|\mathbf{E}_{y\in[\pm M]}\Delta'^{(x)}_{(h_i,h_i')_{i=1}^{k-3}}g(x,y)\|_{U^3_{W\cdot[N/W]}}^8\gg\delta^{O_k(1)}N.$$

Therefore, there are at least $\delta^{O_k(1)}(N/W)^{2(k-3)}$ shifts $(h_i,h_i')_{i=1}^{k-3}\in(W\cdot[N/W])^{2\times(k-3)}$ such that

$$\|\mathbf{E}_{y\in[\pm M]}\Delta'^{(x)}_{(h_i,h_i')_{i=1}^{k-3}}g(x,y)\|_{U^3_{W\cdot[N/W]}}^8\gg\delta^{O_k(1)}N.$$

Step 2: Setup for applying the U^3-inverse theorem. For the next few labeled steps, we fix shifts $(h_i,h_i')_{i=1}^{k-3}\in(W\cdot[N/W])^{2\times(k-3)}$ such that

$$\|\mathbf{E}_{y\in[\pm M]}\Delta'^{(x)}_{(h_i,h_i')_{i=1}^{k-3}}g(x,y)\|_{U^3_{W\cdot[N/W]}}^8\gg\delta^{O_k(1)}N.$$

Furthermore, denote

$$f_j^{(1)}(x)=\Delta'^{(x)}_{(h_i,h_i')_{i=1}^{k-3}}f_j(x)$$

for $j\in\{2,3\}$.

Since all the differences defining the box-norm are divisible by W, we have

$$\sum_{j\in[W]}\|\mathbf{E}_{y\in[\pm M]}f_2^{(1)}(Wx+j+P(y))f_3^{(1)}(Wx+j+2P(y))\|_{U^3_{[N/W]}}^8\gg\delta^{O_k(1)}N.$$

By the triangle inequality, we have

$$\sum_{\substack{j\in[W]\\k\in[W]}}\|\mathbf{E}_{y\in[\pm MW^{-1}]}f_2^{(1)}(Wx+j+P(Wy+k))f_3^{(1)}(Wx+j+2P(Wy+k)))\|_{U^3_{[N/W]}}^8\gg\delta^{O_k(1)}NW.$$

Defining

$$f_{i,t}^{(2)}(x)=f_i^{(1)}(Wx+t),\quad\text{and}\quad P_r(y)=\frac{P(Wy+r)-P(r)}{W},$$

we therefore have

$$\sum_{\substack{j\in[W]\\k\in[W]}}\|\mathbf{E}_{y\in[\pm MW^{-1}]}f_{2,j+P(k)}^{(2)}(x+P_k(y))f_{3,j+2P(k)}^{(2)}(x+2P_k(y))\|_{U^3_{[N/W]}}^8\gg\delta^{O_k(1)}NW.$$

Thus, for at least a $\delta^{O_k(1)}$ fraction of pairs $(j,k)\in[W]^2$, we have that

$$\|\mathbf{E}_{y\in[\pm MW^{-1}]}f_{2,j+P(k)}^{(2)}(x+P_k(y))f_{3,j+2P(k)}^{(2)}(x+2P_k(y))\|_{U^3_{[N/W]}}^8\gg\delta^{O_k(1)}NW^{-1}.$$

We fix such j and k for the next few labeled steps within the argument. We now perform a certain set of artificial changes of variables; this change of variable is directly inspired by work of Altman [2], and is used to reduce to considering the "flag" set of forms $\{2(x+2y), 3(x+y), 6y, 6x\}$ that was considered in Section 5.

Define $f_{2,t}^{(3)}(x) = f_{2,t}^{(2)}(x/3)\mathbb{1}_{3|x}$, $f_{3,t}^{(3)}(x) = f_{3,t}^{(2)}(x/2)\mathbb{1}_{2|x}$,

$$H(x) = \mathbf{E}_{y\in[\pm MW^{-1}]} f_{2,j+P(k)}^{(3)}(3(x+P_k(y))) f_{3,j+2P(k)}^{(3)}(2(x+2P_k(y))),$$

and

$$H^*(x) = H(x/6)\mathbb{1}_{6|x}.$$

Note that

$$\|H^*(x)\|_{U^3_{[6N/W]}}^8 \gg \delta^{O_k(1)} NW^{-1};$$

this follows via expanding the definition of the box-norm and noting that H^* is only supported on multiples of 6.

Step 3: Applying the U^3-inverse theorem and reduction to Lemma 5.9. Note that, by Corollary C.6, we have

$$\|H^*(x)\|_{U^3_{[10^3\delta^{-1}N/W]}}^8 \gg \delta^{O_k(1)} NW^{-1}.$$

Therefore, by Theorem 6.7 (applied noting that $H^*(x)$ has support contained in $[\pm 30\delta^{-1}N/W]$), we have

$$\left|\sum_{x\in\mathbf{Z}} H^*(x) F(g(x)\Gamma)\right| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)}) NW^{-1},$$

where G/Γ, $F\colon G/\Gamma \to \mathbf{C}$, and g are as in Theorem 6.7. Let the vertical character of F be ξ. Unwinding the definition of $H^*(x)$, we in fact that have that

$$\left|\sum_{x\in\mathbf{Z}} H(x) F(g(6x)\Gamma)\right| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)}) NW^{-1}.$$

Inserting the definition of $H(x)$ yields

$$\left|\sum_{x\in\mathbf{Z}} F(g(6x)\Gamma)\mathbf{E}_{y\in[\pm MW^{-1}]} f_{2,j+P(k)}^{(3)}(3(x+P_k(y))) f_{3,j+2P(k)}^{(3)}(2(x+2P_k(y)))\right| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)}) NW^{-1}. \tag{6.1}$$

We will return to (6.1) eventually; we first deduce a series of structural claims regarding the polynomial sequence $g(\cdot)$.

Applying Lemma 5.1 with $\varepsilon = \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})$ and using the pigeonhole principle to choose a single α, there exist functions F_1 with vertical character -9ξ, F_2 with vertical character 8ξ, and F_3 with vertical character 2ξ such that

$$\Bigg|\sum_{x\in\mathbf{Z}} \mathbf{E}_{y\in[\pm MW^{-1}]} F_3(g(6P_k(y))\Gamma) f_{2,j+P(k)}^{(3)}(3(x+P_k(y))) F_2(g(3(x+P_k(y)))\Gamma)$$
$$f_{3,j+2P(k)}^{(3)}(2(x+2P_k(y))) F_1(g(2(x+2P_k(y)))\Gamma)\Bigg| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)}) NW^{-1}$$

and $\|F_i\|_{\mathrm{Lip}} \leq \mathrm{poly}_{\delta^{-1}}(\delta^{-O_k(1)})$ for each $i = 1,2,3$. Set $\widetilde{F}_1(z) := f_{2,j+P(k)}^{(3)}(3z)F_2(g(3z)\Gamma)$ and $\widetilde{F}_2(z) := f_{3,j+2P(k)}^{(3)}(2z)F_1(g(2z)\Gamma)$. By Parseval's identity, we have

$$\left|\mathbf{E}_{y\in[\pm MW^{-1}]} F_3(g(6P_k(y))\Gamma)\int_{\mathbf{T}} \widehat{\widetilde{F}_1}(\eta)\widehat{\widetilde{F}_2}(\eta) e(\eta P_k(y))\,d\eta\right| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)}) NW^{-1}$$

or equivalently

$$\left|\int_{\mathbf{T}} \widehat{\widetilde{F_1}}(\eta)\widehat{\widetilde{F_2}}(\eta)\mathbf{E}_{y\in[\pm MW^{-1}]}F_3(g(6P_k(y))\Gamma)e(\eta P_k(y))d\eta\right| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})NW^{-1}.$$

Thus, there exists $\beta \in \mathbf{T}$ such that

$$\left|\mathbf{E}_{y\in[\pm MW^{-1}]}F_3(g(6P_k(y))\Gamma)e(6\beta P_k(y))\right| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)}),$$

since $\max\{\|\widetilde{F_1}\|_2^2, \|\widetilde{F_2}\|_2^2\} \ll \delta^{-1}NW^{-1}$. Now, fix a choice of $\beta^* \in \mathbf{T}$ such that $2\beta^* = \beta$.

Step 4: Applying Lemma 5.9. We now use Lemma 5.9 to reduce the degree of the polynomial sequence g. The argument splits into two cases. In the case when ξ is zero, we will be able to directly reduce the degree of the nilsequence; we defer this case until later.

If ξ is nonzero, let $G_\bullet$ denote the degree 2 filtration $G = G_0 = G_1 \geqslant G_2 \geqslant \mathrm{Id}_G$ relative to which g is a polynomial sequence. As ξ is nonzero, there exists $h \in G_2$ such that $\xi(h) = \beta^*$. Define $\widetilde{g}(n) = g(n)h^n$. Then $\widetilde{g}(0) = \mathrm{Id}_G$ and $\widetilde{g}$ is a polynomial sequence with respect to $G_\bullet$. By construction, we have that

$$\left|\mathbf{E}_{y\in[\pm MW^{-1}]}F_3(\widetilde{g}(6P_k(y))\Gamma)\right| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)}).$$

This is exactly the setup of Lemma 5.9. We may thus factor $\widetilde{g}(n)$ as $\widetilde{g} = \varepsilon \cdot g' \cdot \gamma$ with $\varepsilon, g', \gamma \in \mathrm{Poly}(\mathbf{Z}, G_\bullet)$, where

- for all $t \in [\pm 100\delta^{-1} \cdot N/W]$, $d(\varepsilon(t), \varepsilon(t-1)) \leq W\,\mathrm{poly}_{\delta^{-1}}(\delta^{-O_k(1)})/N$ and $d(\varepsilon(t), \mathrm{Id}_G) \leq \mathrm{poly}_{\delta^{-1}}(\delta^{-O_k(1)})$,
- γ is $\mathrm{poly}_{\delta^{-1}}(\delta^{-O_k(1)})$-rational and $\gamma(n)\Gamma$ is periodic with period at most $\mathrm{poly}_{\delta^{-1}}(\delta^{-O_k(1)})$,
- g' takes values only G', a connected proper $\mathrm{poly}_{\delta^{-1}}(\delta^{-O_k(1)})$-rational subgroup with respect to $\mathcal{X}$, and may be viewed as a polynomial sequence with respect to the filtration $G'_\bullet$, where $G'_i = G' \cap G_i$,
- ξ is trivial on $G'_2 = G' \cap G_2$.

Step 5: Setup for degree-reduction. Recall from (6.1) that

$$\left|\sum_{x\in\mathbf{Z}} F(g(6x)\Gamma)\mathbf{E}_{y\in[\pm MW^{-1}]}f^{(3)}_{2,j+P(k)}(3(x+P_k(y))f^{(3)}_{3,j+2P(k)}(2(x+2P_k(y)))\right| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})NW^{-1}.$$

By the definitions of $f^{(3)}_{2,t}$, $f^{(3)}_{3,t}$, and $\widetilde{g}$ and since F has vertical character ξ, it follows that

$$\left|\sum_{x\in\mathbf{Z}} e(-6\beta^* x)F(\widetilde{g}(6x)\Gamma)\mathbf{E}_{y\in[\pm MW^{-1}]}f^{(2)}_{2,j+P(k)}(x+P_k(y))f^{(2)}_{3,j+2P(k)}(x+2P_k(y))\right|$$
$$\geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})NW^{-1}.$$

We next break $[\pm 100\delta^{-1}N/W]$ into nearly-equal length arithmetic progressions $\{Q_1, \ldots, Q_t\}$ of length $\mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})N/W$, with difference equal to the period of γ. By the pigeonhole principle, there exists $Q \in \{Q_1, \ldots, Q_t\}$ such that

$$\left|\sum_{x\in Q} e(-6\beta^* x)F(\widetilde{g}(6x)\Gamma)\mathbf{E}_{y\in[\pm MW^{-1}]}f^{(2)}_{2,j+P(k)}(x+P_k(y))f^{(2)}_{3,j+2P(k)}(x+2P_k(y))\right|$$
$$\geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})|Q|.$$

By choosing the implicit constant in the length of Q sufficiently large (so that the length is small), we get, in fact, that

$$\Big|\sum_{x\in Q} e(-6\beta^* x)F(\varepsilon_Q g'(6x)\gamma_Q\Gamma)\mathbf{E}_{y\in[\pm MW^{-1}]} f^{(2)}_{2,j+P(k)}(x+P_k(y))f^{(2)}_{3,j+2P(k)}(x+2P_k(y))\Big|$$
$$\geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})|Q|,$$

where ε_Q is a $\mathrm{poly}_{\delta^{-1}}(\delta^{-O_k(1)})$-bounded element and γ_Q is $\mathrm{poly}_{\delta^{-1}}(\delta^{-O_k(1)})$-bounded and rational element. (A similar argument appears in the proof of Lemma 5.9.) Let $\widetilde{F}(x) = F(\varepsilon_Q\gamma_Q x)$ and $g^{(2)} = \gamma_Q^{-1}g'\gamma_Q$, so that

$$\Big|\sum_{x\in Q} e(-6\beta^* x)\widetilde{F}(g^{(2)}(6x)\Gamma)\mathbf{E}_{y\in[\pm MW^{-1}]} f^{(2)}_{2,j+P(k)}(x+P_k(y))f^{(2)}_{3,j+2P(k)}(x+2P_k(y))\Big|$$
$$\geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})|Q|.$$

Note that $g^{(2)}$ is a polynomial sequence with respect to the filtration $\gamma_Q^{-1}G'_\bullet\gamma_Q$. We now claim that ξ is trivial on $\gamma_Q^{-1}G'_2\gamma_Q$. Indeed, $G'_2 \subseteq G_2$ and $[G,G_2] = \mathrm{Id}_G$, and thus $G'_2 \subseteq Z(G)$ (the center of G). It follows that $\gamma_Q^{-1}G'_2\gamma_Q = G'_2$ and thus ξ is trivial on $\gamma_Q^{-1}G'_2\gamma_Q$. Furthermore, note that $\widetilde{F}$ has vertical frequency ξ, is $\mathrm{poly}_{\delta^{-1}}(\delta^{-O_k(1)})$-Lipschitz (with respect to a suitable Mal'cev basis on $\gamma_Q^{-1}G'\gamma_Q$), and that $\gamma_Q^{-1}G'\gamma_Q$ is $\mathrm{poly}_{\delta^{-1}}(\delta^{-O_k(1)})$-rational (see [23, Lemma A.13]) with respect to G.

Let $\Gamma' = (\gamma_Q^{-1}G'\gamma_Q)\cap\Gamma$. Since $(\gamma_Q^{-1}G'\gamma_Q)\Gamma/\Gamma \cong (\gamma_Q^{-1}G'\gamma_Q)/\Gamma'$, we have

$$\Big|\sum_{x\in Q} e(-6\beta^* x)\widetilde{F}(g^{(2)}(6x)\Gamma')\mathbf{E}_{y\in[\pm MW^{-1}]} f^{(2)}_{2,j+P(k)}(x+P_k(y))f^{(2)}_{3,j+2P(k)}(x+2P_k(y))\Big|$$
$$\geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})|Q|.$$

As $\gamma_Q^{-1}G'\gamma_Q$ is a sufficiently rational subgroup, one may put a Mal'cev basis $\mathcal{X}'$ on Γ' such that the Lipschitz bounds on F transfer to $\mathcal{X}'$. We note that until this point, we have been operating under the assumption that ξ is nonzero. When ξ is zero, by taking $\beta^* = 0$, we can immediately find ourselves in the same situation by taking $\varepsilon_Q = \gamma_Q = \mathrm{Id}_G$, $G' = G$, and $\Gamma' = \Gamma$.

From these last couple steps, the key extra property we have guaranteed compared to (6.1) is that we know $g^{(2)}$ lives in $\gamma_Q^{-1}G'\gamma_Q$ and also ξ is trivial on $G_2 \cap (\gamma_Q^{-1}G'\gamma_Q)$.

Step 6: Degree-reduction. We are finally in a position to obtain the necessary degree reduction. Given the above setup, we define $G^* := \gamma_Q^{-1}G'\gamma_Q/(\gamma_Q^{-1}G'_2\gamma_Q)$ and take $g^{(3)} \equiv g^{(2)} \mod \gamma_Q^{-1}G'_2\gamma_Q$ to be a polynomial sequence in G^*. Furthermore, let $\Gamma^* = \Gamma'/(\Gamma'\cap\gamma_Q^{-1}G'_2\gamma_Q)$ and F^* be the projection of $\widetilde{F}$ from the domain G'/Γ' to the domain G^*/Γ^* (which is well defined, as $\widetilde{F}$ is invariant under $\gamma_Q^{-1}G'_2\gamma_Q$). We have

$$\Big|\sum_{x\in Q} e(-6\beta^* x)F^*(g^{(3)}(6x)\Gamma^*)\mathbf{E}_{y\in[\pm MW^{-1}]} f^{(2)}_{2,j+P(k)}(x+P_k(y))f^{(2)}_{3,j+2P(k)}(x+2P_k(y))\Big|$$
$$\geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})|Q|.$$

Note, however, that now $g^{(3)}$ is a polynomial sequence of degree 1. Combining Lemma A.9, the fact that the functions $f^{(2)}_{2,j+P(k)}$, $f^{(2)}_{3,j+2P(k)}$ are 1-bounded, and the fact that Q is an arithmetic progression of appropriate length and common difference, it follows using Lemma C.7 that

$$\sup_{\alpha\in\mathbf{T}}\left|\sum_{x\in\mathbf{Z}} e(-\alpha x)\mathbf{E}_{y\in[\pm MW^{-1}]} f^{(2)}_{2,j+P(k)}(x+P_k(y)) f^{(2)}_{3,j+2P(k)}(x+2P_k(y))\right| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})NW^{-1}.$$

Step 7: U^2-control of Sárközy-type configurations Fix α such that

$$\left|\sum_{x\in\mathbf{Z}} e(-\alpha x)\mathbf{E}_{y\in[\pm MW^{-1}]} f^{(2)}_{2,j+P(k)}(x+P_k(y)) f^{(2)}_{3,j+2P(k)}(x+2P_k(y))\right| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})NW^{-1}.$$

This is equivalent to

$$\Bigg|\sum_{x\in\mathbf{Z}} \mathbf{E}_{y\in[\pm MW^{-1}]} f^{(2)}_{2,j+P(k)}(x+P_k(y)) e(-2\alpha(x+P_k(y))) \cdot f^{(2)}_{3,j+2P(k)}(x+2P_k(y)) e(\alpha(x+2P_k(y)))\Bigg| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})NW^{-1}.$$

This immediately implies, by Lemma 6.3, that

$$\min_{i\in\{2,3\}}\sup_{\alpha\in\mathbf{T}}\left|\sum_{x\in\mathbf{Z}} e(\alpha x) f^{(2)}_{i,j+(i-1)P(k)}(x)\right| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})NW^{-1}$$

for our original choice of $(j,k)\in[W]^2$.

Step 8: Unwinding and deducing the final result. Note that if one samples $j\in[W]$ and $k\in[W]$ uniformly, then $j+(i-1)P(k)$ is uniformly distributed modulo W for each i. Also, recall that the correlation was deduced for a positive portion of j and k. So, we can deduce

$$\min_{i\in\{2,3\}}\sum_{j\in[W]}\sup_{\alpha\in\mathbf{T}}\left|\sum_{x\in\mathbf{Z}} e(\alpha x) f^{(2)}_{i,j}(x)\right| \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})N.$$

By the converse to the U^2-inverse theorem (see, e.g., Lemma C.5), it follows that

$$\min_{i\in\{2,3\}}\sum_{j\in[W]} \|f^{(2)}_{i,j}(x)\|^4_{U^2_{[N/W]}} \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})N.$$

Inserting the definition of $f^{(2)}_{i,j}$ yields

$$\min_{i\in\{2,3\}}\|f^{(1)}_i(x)\|^4_{U^2_{W\cdot[N/W]}} \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})N.$$

We now unwind the definition of $f^{(1)}_i$. Recall from Step 1 that a positive proportion of shifts $(h_i,h_i')_{i=1}^{k-3}\in(W\cdot[N/W])^{k-3}$ were satisfied conditions sufficient for the analysis in Step 2 (and thus subsequent steps) to follow. Therefore, using that the box-norm is always nonnegative, we obtain

$$\min_{i\in\{2,3\}}\mathbf{E}_{\substack{h_j,h_j'\in W\cdot[N/W]\\ 1\leq j\leq k-3}}\|\Delta_{(h_j,h_j')_{i=1}^{k-3}} f_i(x)\|^4_{U^2_{W\cdot[N/W]}} \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})N.$$

By the definition of the box-norm, this is equivalent to

$$\min_{i\in\{2,3\}}\|f_i(x)\|^{2^{k-1}}_{U^{k-1}_{W\cdot[\pm N/W]}} \geq \mathrm{poly}_{\delta^{-1}}(\delta^{O_k(1)})N.$$

This (finally) completes the proof. □

7. Proof of Theorem 1.1

7.1. **Initial U^s-norm control and degree-lowering output.** To obtain our initial U^s-norm control for the counting operator Λ^W, we can, essentially, apply [40, Theorem 6.1] as a black-box.

Proposition 7.1. *There exists a positive integer $s = s_{7.1}$ such that the following holds. Fix 1-bounded functions $f_1, f_2, f_3\colon \mathbf{Z} \to \mathbf{C}$ with $\operatorname{supp}(f_i) \subseteq [\pm CN]$ for $i = 1, 2, 3$, W, M, and P as in (2.2), and $N \geq (W\delta^{-1})^{\Omega(1)}$. If*

$$\left|\Lambda^W(f_1, f_2, f_3)\right| \geq \delta MN,$$

then

$$\min_{i\in[3]} \|f_i\|_{U^s_{W\cdot[N/W]}}^{2^s} \gg_C \delta^{O(1)} N.$$

Proof. By shifting the f_i, we may assume that they are supported in $[2CN]$ instead. The result is then, essentially, an immediate consequence of [40, Theorem 6.1]. For f_1, apply the result with $P_1(y) = 2Wy^2 + y$ and $P_2(y) = 4Wy^2 + 2y$; for f_2, apply the result with $P_1(y) = -2Wy^2 - y$ and $P_2(y) = 2Wy^2 + y$; and for f_3, apply the result with $P_1(y) = -2Wy^2 - y$ and $P_2(y) = -4Wy^2 - 2y$. In each case, we take $M = \sqrt{N/W}$ and the desired result follows, except that the box-norm may have shift parameters lying in $qW \cdot [\delta^{O(1)} N/W]$ with $q \ll 1$. By applying Lemmas C.2 and C.3, we may assume that the shift parameters are the same and thus instead a Gowers norm with parameter $qW \cdot [\delta^{O(1)} N/W]$ with $q \ll 1$. This Gowers norm can be upgraded to the one in the conclusion of the proposition using Corollaries C.6 and C.8. □

By combining Proposition 7.1 with our key degree-lowering result, we can deduce that Λ^W is controlled by the U^2-norm. For the statements below, we let $\exp^k$ denote the k-fold iterated exponential.

Proposition 7.2. *There exists a positive integer $K = K_{7.2}$ such that the following holds. Suppose that $f_1, f_2, f_3\colon \mathbf{Z} \to \mathbf{C}$ are 1-bounded functions with $\operatorname{supp}(f_i) \subseteq [CN]$ for $i = 1, 2, 3$, W, M, and P are as in (2.2), and $N \geq W^{\Omega(1)} \exp^K(\delta^{-1})$. If*

$$\left|\Lambda^W(f_1, f_2, f_3)\right| \geq \delta MN,$$

then

$$\min_{i\in[3]} \|f_i\|_{U^2_{W\cdot[N/W]}}^{4} \gg_C (\exp^K(\delta^{-1}))^{-1} N.$$

Proof. Let $s = s_{7.1}$. We prove by downwards induction on $k \in \{2, \dots, s\}$ that given appropriate support and boundedness conditions on functions $h_i\colon \mathbf{Z} \to \mathbf{C}$, we have that $|\Lambda^W(h_1, h_2, h_3)| \geq \delta MN$ implies

$$\min_{i\in[3]} \|h_i\|_{U^2_{W\cdot[N/W}}^{4} \gg_C \exp^{2(s-k)}(\delta^{-O(1)})^{-1} N.$$

For $k = s$, this is Proposition 7.1. The result for $k = 2$ with $h_i = f_i$ is the desired.

Now suppose that we have established the result for $k \geq 3$ and wish to prove it for $k - 1$. Note that

$$\begin{aligned}
\delta MN \ll \Lambda^W(h_1, h_2, h_3) &= (2M+1) \sum_{x\in\mathbf{Z}} h_i(x) \mathcal{D}^1(h_2, h_3)(x) \\
&\leq (2M+1) \Big(\sum_{x\in\mathbf{Z}} |h_i(x)|^2\Big)^{1/2} \Big(\sum_{x\in\mathbf{Z}} |\mathcal{D}^1(h_2, h_3)(x)|^2\Big)^{1/2} \\
&\ll M \cdot N^{1/2} \cdot \Lambda^W(\overline{\mathcal{D}^1(h_2, h_3)}, h_2, h_3)^{1/2}
\end{aligned}$$

and, therefore,

$$\Lambda^W(\overline{\mathcal{D}^1(h_2,h_3)},h_2,h_3) \gg \delta^2 MN.$$

Now apply the inductive hypothesis with h_1 replaced by $\overline{\mathcal{D}^1(h_2,h_3)}$ (which still is bounded and with appropriate support) and δ replaced by $\Omega(\delta^2)$. We deduce

$$\|\mathcal{D}^1(h_2,h_3)\|_{U^k_{W\cdot[N/W]}}^{2^k} \gg_C \exp^{2(s-k)}(\delta^{-O(1)})^{-1}N.$$

Similarly, we have

$$\|\mathcal{D}^3(h_1,h_2)\|_{U^k_{W\cdot[N/W]}}^{2^k} \gg_C \exp^{2(s-k)}(\delta^{-O(1)})^{-1}N.$$

(Note that the $O(1)$ exponents here may decay with each induction step, but $s = s_{7.1}$ is an absolute constant so this will remain bounded at the end.)

Now using Propositions 6.1 and 6.2, it follows that

$$\min_{i\in[3]}\|h_i\|_{U^{k-1}_{W\cdot[N/W]}}^{2^{k-1}} \gg_C \exp^{2(s-k)+2}(\delta^{-O(1)})^{-1}N,$$

using that $O_k(1) = O(1)$ as k is bounded, which completes the induction. □

7.2. **Completing the proof.** We are now in position to complete the proof. The following result states that, for 1-bounded functions, the counting operators Λ^W and Λ^{Model} agree up to a universal scaling factor.

Proposition 7.3. *There exists an integer $K = K_{7.3} > 0$ such that the following holds. Suppose $f_1, f_2, f_3\colon \mathbf{Z} \to \mathbf{C}$ are 1-bounded functions such that $\mathrm{supp}(f_i) \subseteq [N]$ for $i = 1,2,3$, W, M, w, and P are as in (2.2), and $N \gg W^{\Omega(1)}$ and $W \gg \exp^K(\delta^{-1})$. Then,*

$$\left|(NW)^{1/2}\Lambda^W(f_1,f_2,f_3) - \Lambda^{\mathrm{Model}}(f_1,f_2,f_3)\right| \le \delta N^2.$$

Proof. Assume for the sake of contradiction that

$$\left|(NW)^{1/2}\Lambda^W(f_1,f_2,f_3) - \Lambda^{\mathrm{Model}}(f_1,f_2,f_3)\right| \ge \delta N^2,$$

and define

$$\widetilde{\Lambda}(f_1,f_2,f_3) = (NW)^{1/2}\Lambda^W(f_1,f_2,f_3) - \Lambda^{\mathrm{Model}}(f_1,f_2,f_3).$$

For this proof, define the modified dual functions

$$D^{1,*}(f_2,f_3)(x) = N^{-1}\Big((NW)^{1/2}\sum_{|k|\le M} f_2(x+P(k))f_3(x+2P(k)) - \sum_{d\in\mathbf{Z}} f_2(z+d)f_3(z+2d)\nu(d)\Big),$$

$$D^{2,*}(f_3,f_1)(x) = N^{-1}\Big((NW)^{1/2}\sum_{|k|\le M} f_1(x-P(k))f_3(x+P(k)) - \sum_{d\in\mathbf{Z}} f_1(z-d)f_3(z+d)\nu(d)\Big),$$

$$D^{3,*}(f_1,f_2)(x) = N^{-1}\Big((NW)^{1/2}\sum_{|k|\le M} f_1(x-2P(k))f_2(x-P(k)) - \sum_{d\in\mathbf{Z}} f_1(z-2d)f_2(z-d)\nu(d)\Big).$$

By an application of the Cauchy–Schwarz inequality analogous to that used in Proposition 7.2 (and at the end of Section 2), we have

$$\left|\widetilde{\Lambda}(D^{1,*}(f_2,f_3),f_2,f_3)\right| \gg \delta^2 N^2.$$

By the triangle inequality, we have that

$$\left|\Lambda^{\mathrm{Model}}(D^{1,*}(f_2,f_3),f_2,f_3)\right| \gg \delta^2 N^2 \quad \text{or} \quad \left|(NW)^{1/2}\Lambda^W(D^{1,*}(f_2,f_3),f_2,f_3)\right| \gg \delta^2 N^2.$$

Therefore, by Lemma 4.1, Proposition 7.2, and Lemma C.3, we have

$$\|D^{1,*}(f_2,f_3)\|_{U^2_{W\cdot[N/W]}}^4 \gg \exp^{K}7.2(\delta^{-O(1)})N.$$

Applying the U^2-inverse theorem Lemma C.4 to each progression of spacing W and passing to the interval $[\pm C_1N]$ (which contains the support of $D^{1,*}(f_2,f_3)$), there exist constants $\alpha_{j,1},\beta_{j,1}$ for each $j\in[W]$ such that

$$\widetilde{f}_1(x) := \mathbb{1}_{x\in[\pm C_1N]}\sum_{j\in[W]}\mathbb{1}_{W|(x-j)}e(\alpha_{j,1}x+\beta_{j,1})$$

satisfies

$$\sum_{x\in\mathbf{Z}}\widetilde{f}_1(x)D^{1,*}(f_2,f_3)(x) \gg \exp^{K}7.2(\delta^{-O(1)})^{-1}N.$$

By construction, this implies that

$$|\widetilde{\Lambda}(\widetilde{f}_1,f_2,f_3)| \geq \exp^{K}7.2(\delta^{-O(1)})^{-1}N^2.$$

Repeating this procedure, we find $\alpha_{j,i},\beta_{j,i}$ for each $j\in[W]$ such that defining

$$\widetilde{f}_i(x) := \mathbb{1}_{x\in[\pm C_iN]}\sum_{j\in[W]}\mathbb{1}_{W|(x-j)}e(\alpha_{j,i}x+\beta_{j,i})$$

for $i\in\{2,3\}$ (with appropriate absolute constants C_i), we have

$$|\widetilde{\Lambda}(\widetilde{f}_1,\widetilde{f}_2,\widetilde{f}_3)| \gg \exp^{3K}7.2(\delta^{-O(1)})^{-1}N^2.$$

Unwinding the definition of Λ^W and Λ^{Model} (recall ν is supported on $[1,N]$), we have

$$\Bigg|\sum_{\substack{x\in\mathbf{Z}\\ d\in[N]}}\Big((NW)^{1/2}\widetilde{f}_1(x)\widetilde{f}_2(x+d)\widetilde{f}_3(x+2d)\mathbb{1}_{d\in\{P(k):k\in\mathbf{Z}\}\cap[N]} - \widetilde{f}_1(x)\widetilde{f}_2(x+d)\widetilde{f}_3(x+2d)\nu(d)\Big)\Bigg|$$
$$\gg \exp^{3K}7.2(\delta^{-O(1)})^{-1}N^2.$$

Define $\nu^*(d) = (NW)^{1/2}\cdot|\{d=P(k)\colon k\in\mathbf{Z}\text{ and }|k|\leq M\}|$. Since P is injective on $\mathbf{Z}$ so this set is only size 0 or 1. We have

$$\Bigg|\sum_{\substack{x\in\mathbf{Z}\\ d\in\mathbf{Z}}}\widetilde{f}_1(x)\widetilde{f}_2(x+d)\widetilde{f}_3(x+2d)(\nu^*(d)-\nu(d))\Bigg| \gg \exp^{3K}7.2(\delta^{-O(1)})^{-1}N^2.$$

Note that

$$\Bigg|\sum_{\substack{x\in\mathbf{Z}\\ d\in\mathbf{Z}}}\widetilde{f}_1(x)\widetilde{f}_2(x+d)\widetilde{f}_3(x+2d)(\nu^*(d)-\nu(d))\Bigg|$$
$$\leq \sum_{k,\ell\in[W]}\Bigg|\sum_{\substack{x\in\mathbf{Z}\\ d\in\mathbf{Z}\\ W|(x-\ell)\\ W|(d-k)}}\widetilde{f}_1(x)\widetilde{f}_2(x+d)\widetilde{f}_3(x+2d)(\nu^*(d)-\nu(d))\Bigg|$$
$$\leq \sum_{k,\ell\in[W]}\Bigg|\sum_{\substack{x\in\mathbf{Z}\\ d\in\mathbf{Z}}}\widetilde{f}_1(Wx+\ell)\widetilde{f}_2(W(x+d)+\ell+k)\widetilde{f}_3(W(x+d)+\ell+2k)(\nu^*(Wd+k)-\nu(Wd+k))\Bigg|$$
$$\leq W^2\sup_{\substack{\alpha_1,\alpha_2,\alpha_3\in\mathbf{T}\\ k,\ell\in[W]}}\Bigg|\sum_{\substack{x\in\mathbf{Z}\\ d\in\mathbf{Z}}}e(\alpha_1x)\mathbb{1}_{|Wx+\ell|\leq C_1N}e(\alpha_2(x+d))\mathbb{1}_{|W(x+d)+\ell+k|\leq C_2N}$$

$$e(\alpha_3(x+2d))\mathbb{1}_{|W(x+2d)+\ell+2k|\le C_3N}(\nu^*(Wd+k)-\nu(Wd+k))\Big|.$$

Letting $\tau_{i,\alpha_i}(x)=e(\alpha_i x)\mathbb{1}_{|x|\le C_iNW^{-1}}$, we have

$$\sup_{\substack{\alpha_1,\alpha_2,\alpha_3\in\mathbf{T}\\ k\in[W]}}\Big|\sum_{\substack{x\in\mathbf{Z}\\ d\in\mathbf{Z}}}\tau_{1,\alpha_1}(x)\tau_{2,\alpha_2}(x+d)\tau_{3,\alpha_3}(x+2d)(\nu^*(Wd+k)-\nu(Wd+k))\Big|\gg\exp^{3K7.2}(\delta^{-O(1)})^{-1}N^2W^{-2}.$$

We now take a Fourier transform. Defining $\widetilde{\nu}_k(d)=(\nu^*(Wd+k)-\nu(Wd+k))$ we have

$$\begin{aligned}
&\Big|\sum_{\substack{x\in\mathbf{Z}\\ d\in\mathbf{Z}}}\tau_{1,\alpha_1}(x)\tau_{2,\alpha_2}(x+d)\tau_{3,\alpha_3}(x+2d)(\nu^*(Wd+k)-\nu(Wd+k))\Big|\\
&=\Big|\int_{\mathbf{T}^2}\widehat{\tau_{1,\alpha_1}}(\Theta_1)\widehat{\tau_{2,\alpha_2}}(\Theta_2)\widehat{\tau_{3,\alpha_3}}(-\Theta_1-\Theta_2)\widehat{\widetilde{\nu}}_k(\Theta_2)\ d\Theta_1 d\Theta_2\Big|\\
&\le\|\widehat{\widetilde{\nu}}_k\|_\infty\cdot\int_{\mathbf{T}^2}|\widehat{\tau_{1,\alpha_1}}(\Theta_1)|\cdot|\widehat{\tau_{2,\alpha_2}}(\Theta_2)|\cdot|\widehat{\tau_{3,\alpha_3}}(-\Theta_1-\Theta_2)|\ d\Theta_1 d\Theta_2\\
&\le\|\widehat{\widetilde{\nu}}_k\|_\infty\Big(\int_{\mathbf{T}^2}|\widehat{\tau_{1,\alpha_1}}(\Theta_1)|^{3/2}\cdot|\widehat{\tau_{2,\alpha_2}}(\Theta_2)|^{3/2}\ d\Theta_1 d\Theta_2\Big)^{1/3}\\
&\qquad\Big(\int_{\mathbf{T}^2}|\widehat{\tau_{1,\alpha_1}}(\Theta_1)|^{3/2}\cdot|\widehat{\tau_{3,\alpha_3}}(-\Theta_1-\Theta_2)|^{3/2}\ d\Theta_1 d\Theta_2\Big)^{1/3}\\
&\qquad\Big(\int_{\mathbf{T}^2}|\widehat{\tau_{2,\alpha_2}}(\Theta_2)|^{3/2}\cdot|\widehat{\tau_{3,\alpha_3}}(-\Theta_1-\Theta_2)|^{3/2}\ d\Theta_1 d\Theta_2\Big)^{1/3}\\
&=\|\widehat{\widetilde{\nu}}_k\|_\infty\prod_{i\in[3]}\Big(\int_{\mathbf{T}}|\widehat{\tau_{i,\alpha_i}}(\Theta)|^{3/2}d\Theta\Big)^{2/3}\ll\|\widehat{\widetilde{\nu}}_k\|_\infty\cdot(N/W),
\end{aligned}$$

where in the final line we have used standard fact that the L^p-norm of the Fourier transform of an interval of length N is $\ll_p N^{(p-1)/p}$ for $p>1$. However, by Lemma B.8, we have

$$\sup_{k\in[W]}\|\widehat{\widetilde{\nu}}_k\|_\infty\ll\frac{N}{W}\cdot\frac{1}{\sqrt{w}}.$$

We have our desired contradiction if w (i.e., W) is sufficiently large with respect to δ^{-1}. □

The main result now follows in a straightforward manner.

Proof of Theorem 1.1. Let S be a subset of density δ in $[N]$ and W be a sufficiently large parameter to be chosen at the end of the proof. By the pigeonhole principle, there exists $j\in[4W]$ such that $S_j=S\cap(4W\mathbf{Z}+j)$ has size at least $\delta N/(4W)$. Set $S_j^*=(S_j-j)/(4W)\subseteq[N/(4W)]$. Note that, since $4W^2y^2+4Wy=(2Wy+1)^2-(2Wy+1)$, differences of the form $((2Wy+1)^2-1)/(4W)=Wy^2+y$ in the set S_j^* lift to differences of the form z^2-1 in S. By Lemma 4.2, we have

$$\Lambda^{\mathrm{Model}}(\mathbb{1}_{S_j^*},\mathbb{1}_{S_j^*},\mathbb{1}_{S_j^*})\ge\exp(-\log(2/\delta)^{O(1)})N^2W^{-2}.$$

Taking $N\ge W^{\Omega(1)}\exp^{K+1}(\delta^{-1})$ and $W\ge\exp^{K+1}(\delta^{-1})$ with $K=K_{7.3}$, Proposition 7.3 with N replaced by N/W (note this alters the value of M) and δ appropriately changed implies

$$\Lambda^{W}(\mathbb{1}_{S_j^*},\mathbb{1}_{S_j^*},\mathbb{1}_{S_j^*})\ge\exp(-\log(2/\delta)^{O(1)})N^{3/2}W^{-2}.$$

However, if S is free of nontrivial progressions of the form $x,x+y^2-1,x+2(y^2-1)$, we have

$$\Lambda^{W}(\mathbb{1}_{S_j^*},\mathbb{1}_{S_j^*},\mathbb{1}_{S_j^*})\ll N/W.$$

Therefore, if $\delta \geq \log_{K+2}(N)^{-1}$, taking $N = W^{\Omega(1)} \exp^{K+1}(\delta^{-1})$ and $W = \exp^{K+1}(\delta^{-1})$, we obtain a nontrivial progression of the form $x, x + y^2 - 1, x + 2(y^2 - 1)$ in S, as desired. □

References

[1] D. Altman, *A non-flag arithmetic regularity lemma and counting lemma*, arXiv:2209.14083. 8

[2] D. Altman, *On a conjecture of Gowers and Wolf*, Discrete Analysis (2022), Paper No. 10, 13. 8, 25, 46

[3] A. Balog, J. Pelikán, J. Pintz, and E. Szemerédi, *Difference sets without κth powers*, Acta Mathematica Hungarica **65** (1994), 165–187. 1

[4] V. Bergelson, B. Host, and B. Kra, *Multiple recurrence and nilsequences*, Inventiones Mathematicae **160** (2005), 261–303, With an appendix by Imre Ruzsa. 8

[5] V. Bergelson and A. Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems*, Journal of the American Mathematical Society **9** (1996), 725–753. 2, 6

[6] Bruce C. Berndt, Ronald J. Evans, and Kenneth S. Williams, *Gauss and Jacobi sums*, Canadian Mathematical Society Series of Monographs and Advanced Texts, John Wiley & Sons, Inc., New York, 1998, A Wiley-Interscience Publication. 39

[7] T. F. Bloom and J. Maynard, *A new upper bound for sets with no square differences*, Compositio Mathematica **158** (2022), 1777–1798. 1

[8] J. Bourgain and M.-C. Chang, *Nonlinear Roth type theorems in finite fields*, Israel Journal of Mathematics **221** (2017), 853–867. 2

[9] T. D. Browning and S. M. Prendiville, *A transference approach to a Roth-type theorem in the squares*, International Mathematics Research Notices. IMRN (2017), 2219–2248. 38

[10] D. Dong, X. Li, and W. Sawin, *Improved estimates for polynomial Roth type theorems in finite fields*, Journal d'Analyse Mathématique **141** (2020), 689–705. 2

[11] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, Journal d'Analyse Mathématique **31** (1977), 204–256. 1

[12] H. Furstenberg, *A polynomial Szemerédi theorem*, Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), Bolyai Soc. Math. Stud., vol. 2, János Bolyai Math. Soc., Budapest, 1996, pp. 253–269. 8

[13] W. T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geometric and Functional Analysis **8** (1998), 529–551. 1, 2

[14] W. T. Gowers, *Arithmetic progressions in sparse sets*, Current developments in mathematics, 2000, Int. Press, Somerville, MA, 2001, pp. 149–196. 1

[15] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geometric and Functional Analysis **11** (2001), 465–588. 1, 2

[16] B. Green, *100 open problems*, Manuscript, available on request. 3

[17] B. Green, *On arithmetic structures in dense sets of integers*, Duke Mathematical Journal **114** (2002), 215–238. 2, 20

[18] B. Green, *Roth's theorem in the primes*, Annals of Mathematics. Second Series **161** (2005), 1609–1636. 5

[19] B. Green and T. Tao, *An inverse theorem for the Gowers $U^3(G)$ norm*, Proceedings of the Edinburgh Mathematical Society. Series II **51** (2008), 73–153. 7, 8, 10, 11, 22, 23

[20] B. Green and T. Tao, *Quadratic uniformity of the Möbius function*, Université de Grenoble. Annales de l'Institut Fourier **58** (2008), 1863–1935. 36, 38

[21] B. Green and T. Tao, *An arithmetic regularity lemma, an associated counting lemma, and applications*, An irregular mind, Bolyai Soc. Math. Stud., vol. 21, János Bolyai Math. Soc., Budapest, 2010, pp. 261–334. 8, 11, 13, 16

[22] B. Green and T. Tao, *Linear equations in primes*, Annals of Mathematics. Second Series **171** (2010), 1753–1850. 10

[23] B. Green and T. Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Annals of Mathematics. Second Series **175** (2012), 465–540. 8, 10, 12, 13, 14, 15, 16, 18, 19, 20, 23, 27, 35, 36, 38, 42

[24] B. Green, T. Tao, and T. Ziegler, *An inverse theorem for the Gowers U^4-norm*, Glasgow Mathematical Journal **53** (2011), 1–50. 46

[25] T. Kamae and M. Mendès France, *van der Corput's difference theorem*, Israel Journal of Mathematics **31** (1978), 335–342. 1

[26] Z. Kelley and R. Meka, *Strong bounds for 3-progressions*, arXiv:2302.05537. 9

[27] B. Kuca, *Further bounds in the polynomial Szemerédi theorem over finite fields*, Acta Arithmetica **198** (2021), 77–108. 2

[28] B. Kuca, *On several notions of complexity of polynomial progressions*, Ergodic Theory and Dynamical Systems **43** (2023), 1269–1323. 2

[29] M. Lazard, *Sur les groupes nilpotents et les anneaux de Lie*, Annales Scientifiques de l'École Normale Supérieure. Troisième Série **71** (1954), 101–190. 35
[30] A. Leibman, *Polynomial sequences in groups*, Journal of Algebra **201** (1998), 189–206. 35
[31] A. Leibman, *Polynomial mappings of groups*, Israel Journal of Mathematics **129** (2002), 29–60. 35
[32] A. Leibman, *Orbit of the diagonal in the power of a nilmanifold*, Transactions of the American Mathematical Society **362** (2010), 1619–1658. 10
[33] J. Leng, *The partition rank vs. analytic rank problem for cyclic groups I. Equidistribution for periodic nilsequences*, arXiv:2306.13820. 2, 8, 36
[34] J. Leng, *A Quantitative Bound For Szemerédi's Theorem for a Complexity One Polynomial Progression over* $\mathbb{Z}/n\mathbb{Z}$, arXiv:2205.05540. 2, 3, 6, 7, 8, 14
[35] J. Lucier, *Intersective sets given by a polynomial*, Acta Arithmetica **123** (2006), 57–95. 1
[36] F. Manners, *True complexity and iterated Cauchy–Schwarz*, arXiv:2109.05731. 5
[37] M. B. Nathanson, *Additive Number Theory The Classical Bases*, vol. 164, Springer Science & Business Media, 1996. 38
[38] S. Peluse, *Three-term polynomial progressions in subsets of finite fields*, Israel Journal of Mathematics **228** (2018), 379–405. 2
[39] S. Peluse, *On the polynomial Szemerédi theorem in finite fields*, Duke Mathematical Journal **168** (2019), 749–774. 2, 3, 6
[40] S. Peluse, *Bounds for sets with no polynomial progressions*, Forum of Mathematics. Pi **8** (2020), e16, 55. 2, 4, 6, 21, 29, 45
[41] S. Peluse and S. Prendiville, *Quantitative bounds in the nonlinear Roth theorem*, arXiv:1903.02592. 2, 6
[42] S. Peluse and S. Prendiville, *A polylogarithmic bound in the nonlinear Roth theorem*, International Mathematics Research Notices. IMRN (2022), 5658–5684. 2, 6
[43] S. Prendiville, *Quantitative bounds in the polynomial Szemerédi theorem: the homogeneous case*, Discrete Analysis (2017), Paper No. 5, 34. 2
[44] A. Rice, *A maximal extension of the best-known bounds for the Furstenberg-Sárközy theorem*, Acta Arithmetica **187** (2019), 1–41. 1
[45] T. Sanders, *On the Bogolyubov-Ruzsa lemma*, Analysis & PDE **5** (2012), 627–655. 8
[46] A. Sárkőzy, *On difference sets of sequences of integers. I*, Acta Mathematica. Academiae Scientiarum Hungaricae **31** (1978), 125–149. 1, 7
[47] A. Sárközy, *On difference sets of sequences of integers. III*, Acta Math. Acad. Sci. Hungar. **31** (1978), 355–386. 1
[48] S. Slijepčević, *A polynomial Sárközy-Furstenberg theorem with upper bounds*, Acta Mathematica Hungarica **98** (2003), 111–128. 1
[49] T. Tao, *Higher order Fourier analysis*, Graduate Studies in Mathematics, vol. 142, American Mathematical Society, Providence, RI, 2012. 36
[50] T. Tao and J. Teräväinen, *Quantitative bounds for Gowers uniformity of the Möbius and von Mangoldt functions*, arXiv:2107.02158. 3, 8, 10, 13, 14, 34, 38
[51] T. D. Wooley and T. Ziegler, *Multiple recurrence and convergence along the primes*, American Journal of Mathematics **134** (2012), 1705–1732. 5
[52] T. Ziegler, *A non-conventional ergodic theorem for a nilsystem*, Ergodic Theory and Dynamical Systems **25** (2005), 1357–1370. 8

Appendix A. Conventions regarding nilsequences and effective equidistribution

We begin this appendix by giving the precise definition of the complexity of a nilmanifold; this definition is exactly as in [50, Definition 6.1].

Definition A.1. Let $s \geq 1$ be an integer and let $K > 0$. A *filtered nilmanifold* G/Γ *of degree* s *and complexity at most* K consists of the following:

- a nilpotent, connected, and simply connected Lie group G of dimension m, which can be identified with its Lie algebra $\log G$ via the exponential map $\exp\colon \log G \to G$;
- a filtration $G_\bullet = (G_i)_{i\geq 0}$ of closed connected subgroups G_i of G with
$$G = G_0 = G_1 \geqslant G_1 \geqslant \cdots \geqslant G_s \geqslant G_{s+1} = \mathrm{Id}_G$$
such that $[G_i, G_j] \subseteq G_{i+j}$ for all $i, j \geq 0$;
- a discrete cocompact subgroup Γ of G; and

- a linear basis $\mathcal{X} = \{X_1, \ldots, X_m\}$ of $\log G$, known as a *Mal'cev basis*.

We, furthermore, require that this data obeys the following conditions:

(1) for $1 \leq i, j \leq m$, one has Lie algebra relations

$$[X_i, X_j] = \sum_{i,j<k\leq m} c_{ijk} X_k$$

for rational numbers c_{ijk} of height at most K;

(2) for each $1 \leq i \leq s$, the Lie algebra $\log G_i$ is spanned by $\{X_j : m - \dim(G_i) < j \leq m\}$; and

(3) the subgroup Γ consists of all elements of the form $\exp(t_1 X_1) \cdots \exp(t_m X_m)$ with $t_i \in \mathbf{Z}$.

We note that the conditions imply $[G, G_s] = \mathrm{Id}_G$, i.e., G_s is contained in the center of G (commutes with every element).

Next, we will define polynomial sequences in filtered nilpotent groups. This concrete definition is equivalent (by [23, Lemma 6.7]) to the one given in [23].

Definition A.2. We adopt the conventions of Definition A.1. Let G be a filtered nilpotent group of degree s. A function $g \colon \mathbf{Z} \to G$ is a *polynomial sequence* if there exist elements $g_i \in G_i$ for $i = 0, \ldots, s$ such that

$$g(n) = g_0 g_1^{\binom{n}{1}} \cdots g_s^{\binom{n}{s}},$$

where $\binom{n}{i} = \frac{1}{i!} \prod_{j=0}^{i-1} (n - j)$, for all $n \in \mathbf{Z}$

We will denote the set of polynomial sequences $g \colon \mathbf{Z} \to G$ relative to the filtration $G_\bullet$ of G by $\mathrm{Poly}(\mathbf{Z}, G_\bullet)$. It turns out that $\mathrm{Poly}(\mathbf{Z}, G_\bullet)$ is a group under the natural multiplication of sequences–this is due to Lazard [29] and Leibman [30, 31].

We will also require the definition of rational points, sequences, and subgroups.

Definition A.3. We adopt the conventions of Definition A.1. We say that $\gamma \in G$ is *Q-rational* if there exists an integer $0 < r \leq Q$ such that $\gamma^r \in \Gamma$. A *Q-rational point in G/Γ* is any point of the form $\gamma\Gamma$ for some $\gamma \in G$ that is Q-rational. A sequence $(\gamma(n))_{n=1}^{\infty}$ in G is *Q-rational* if all elements in the sequence are Q-rational.

Finally, we say a closed connected subgroup G' of G is *Q-rational relative to $\mathcal{X}$* if its Lie algebra $\mathfrak{g}'$ is spanned by linear combinations of the form $\sum_{i\in[m]} a_i X_i$ with $a_1, \ldots, a_m \in \mathbf{Q}$ all of height at most Q.

Now we can define Mal'cev coordinates, the explicit metrics on G and G/Γ used in our work, and the precise definition of the Lipschitz norm of functions on G/Γ. These definitions are exactly as in [23, Appendix A].

Definition A.4. We adopt the conventions of Definition A.1. Given a Mal'cev basis $\mathcal{X}$ and $g \in G$, there exists $(u_1, \ldots, u_m) \in \mathbf{R}^m$ such that

$$g = \exp(u_1 X_1) \cdots \exp(u_m X_m),$$

and we define the *Mal'cev coordinates $\psi = \psi_{\mathcal{X}} \colon G \to \mathbf{R}^m$ for g relative to $\mathcal{X}$* by

$$\psi(g) := (u_1, \ldots, u_m).$$

We then define a metric $d = d_{\mathcal{X}}$ on G by

$$d(x, y) := \left\{ \sum_{i=1}^{n} \min(|\psi(x_i x_{i+1}^{-1})|, |\psi(x_{i+1} x_i^{-1})|) \colon n \in \mathbf{N}, x_1, \ldots, x_{n+1} \in G, x_1 = x, x_{n+1} = y \right\},$$

where $|\cdot|$ denotes the ℓ^∞-norm on $\mathbf{R}^m$, and define a metric on G/Γ by

$$d(x\Gamma, y\Gamma) = \inf_{\gamma,\gamma'\in\Gamma} d(x\gamma, y\gamma').$$

Furthermore, for any function $F\colon G/\Gamma \to \mathbf{C}$, we define

$$\|F\|_{\mathrm{Lip}} := \|F\|_\infty + \sup_{\substack{x,y\in G/\Gamma \\ x\neq y}} \frac{|F(x)-F(y)|}{d(x,y)}.$$

We now define the notion of equidistribution of a sequence on G/Γ which we will require.

Definition A.5. Given a length N, a sequence $(g(n)\Gamma)_{n\in\Gamma}$ is δ-equidistributed if for all Lipschitz functions $F\colon G/\Gamma \to \mathbf{C}$ we have that

$$\left|\mathbf{E}_{n\in[N]} F(g(n)\Gamma) - \int_{G/\Gamma} F\right| \leq \delta \|F\|_{\mathrm{Lip}}.$$

We will require the notion of a horizontal character and the notion of a function F having a vertical frequency; our definitions are exactly as in [23, Definitions 1.5, 3.3, 3.4, 3.5].

Definition A.6. Given a filtered nilmanifold G/Γ, the *horizontal torus* is defined to be

$$(G/\Gamma)_{\mathrm{ab}} := G/[G,G]\Gamma.$$

A *horizontal character* is a continuous homomorphism $\eta\colon G \to \mathbf{T}$ that annihilates Γ; such characters may be equivalently viewed as characters on the horizontal torus. A horizontal character is *nontrivial* if it is not identically zero.

Furthermore, if the nilmanifold G/Γ has degree s, the vertical torus is defined to be

$$G_s/(G_s \cap \Gamma).$$

A *vertical character* is a continuous homomorphism $\xi\colon G_s \to \mathbf{T}$ that annihilates $\Gamma \cap G_s$. Setting $m_s = \dim G_s$, one may use the last m_s coordinates of the Mal'cev coordinate map to identify G_s and $G_s/(G_s\cap\Gamma)$ with $\mathbf{R}^{m_s}$ and $\mathbf{R}^{m_s}/\mathbf{Z}^{m_s}$, respectively. Thus, we may identify any vertical character ξ with a unique $k \in \mathbf{Z}^{m_s}$ such that $\xi(x) = k\cdot x$ under this identification $G_s/(\Gamma\cap G_s) \cong \mathbf{R}^{m_s}/\mathbf{Z}^{m_s}$. We refer to k as the *frequency* of the character ξ, we write $|\xi| := \|k\|_\infty$ to denote the magnitude of the frequency ξ, and say that a function $F\colon G/\Gamma \to \mathbf{C}$ *has a vertical frequency* ξ if

$$F(g_s\cdot x) = e(\xi(g_s))F(x)$$

for all $g_s \in G_s$ and $x \in G/\Gamma$.

Finally, we will require the definition of the smoothness norm of a polynomial sequence $\mathbf{Z}\to\mathbf{T}$.

Definition A.7. Any polynomial sequence $g\colon \mathbf{Z}\to\mathbf{T}$ can be expressed uniquely as

$$g(n) = \sum_{i=0}^{d} \alpha_i \binom{n}{i}$$

with $\alpha_0,\ldots,\alpha_d \in \mathbf{T}$ [49, Exercise 1.6.11]. We then define

$$\|g\|_{C^\infty[N]} := \max_{1\leq j\leq d} N^j \|\alpha_j\|_{\mathbf{T}}.$$

We will need the fact that any Lipschitz function of a nilsequence can be well-approximated by a sum of vertical characters. The statement we require is, essentially, [33, Lemma A.6]; our proof closely follows [23, Lemma 3.7], given a sufficiently explicit estimate for approximating functions on the torus as a sum of characters. We provide a proof below, as the statement in [33, Lemma A.6] has several typos. To give the proof, we require a version of Fourier expansion on the torus, which we will obtain by quantifying the proof of [49, Proposition 1.1.13] (or [20, Lemma A.9]).

Lemma A.8. *Fix $0 < \varepsilon < 1/2$, and let $F\colon \mathbf{T}^d \to \mathbf{C}$ with $\|F\|_{\mathrm{Lip}} \le L$, where, for $x, y \in \mathbf{T}^d$, we have $d(x,y) = \max_{1\le i\le d}\|x_i - y_i\|_{\mathbf{T}}$. There exists an absolute constant $C = C_{A.8} > 0$ such that we can write*

$$F(x) = \sum_{|\xi|\le (CLd\varepsilon^{-1})^2} c_\xi e(\xi\cdot x) + \widetilde{F}(x)$$

for a choice of $\widetilde{F}, c_\xi$ with $\|\widetilde{F}\|_\infty \le \varepsilon$ and $\sum_\xi |c_\xi| \le (3CLd\varepsilon^{-1})^{5d}$.

Proof. Let $R \ge 1$ be a integer cutoff parameter to be chosen later and define

$$F_R(x) := \sum_{k\in\mathbf{Z}^d} R^d \mu_R(k)\widehat{F}(k)e(k\cdot x),$$

recalling the definition (2.1) of μ_R. It is a basic fact from Fourier analysis that

$$F_R(x) = \int_{\mathbf{T}^d} F(y)K_R(x-y)dy,$$

where

$$K_R(y) = \prod_{i=1}^d \frac{1}{R}\bigg(\frac{\sin(\pi R y_i)}{\sin(\pi y_i)}\bigg)^2 = \prod_{i=1}^d\left(\sum_{|h|\le R}\bigg(1-\frac{|h|}{R}\bigg)e(hy_i)\right).$$

Noting that $\int_{\mathbf{T}^d} K_R(y)dy = 1$,

$$\frac{1}{R}\left(\frac{\sin(\pi R y_i)}{\sin(\pi y_i)}\right)^2 \le C_0\frac{R}{(1+R\|y_i\|_{\mathbf{T}})^2}$$

for some absolute constant $C_0 > 0$, and F is L-Lipschitz, we get

$$\begin{aligned}\|F - F_R\|_\infty &\le \sup_{x\in\mathbf{T}^d}\int_{\mathbf{T}^d}|F(x)-F(y)|\cdot K_R(x-y)\,dy\\ &\le L\sup_{x\in\mathbf{T}^d}\int_{\mathbf{T}^d}\max_{1\le i\le d}\|x_i-y_i\|_{\mathbf{T}}\cdot K_R(x-y)\,dy\\ &\le L\sup_{x\in\mathbf{T}^d}\int_{\mathbf{T}^d}\sum_{1\le i\le d}\|x_i-y_i\|_{\mathbf{T}}\cdot K_R(x-y)\,dy\\ &\le Ld\int_{\mathbf{T}}\|y\|_{\mathbf{T}}\cdot\frac{1}{R}\bigg(\frac{\sin(\pi R y)}{\sin(\pi y)}\bigg)^2\,dy \le C\frac{Ld}{\sqrt{R}}\end{aligned}$$

for some absolute constant $C > 0$. The result follows by taking $R = \lceil (CLd\varepsilon^{-1})^2\rceil$, noting that $\|\widehat{F}\|_\infty \le L$, $\mu_R(k) \le R^{-d}$, and that $\mu_R(k)$ is supported on $k\in\mathbf{Z}^d$ such that $\|k\|_\infty \le R$. □

We now extend this result to general filtered nilmanifolds by using Fourier analysis on the final nontrivial group of the filtration, G_s.

Lemma A.9. *Fix $0 < \varepsilon < 1/2$, let G/Γ be a filtered nilmanifold of dimension m, degree s, and complexity at most K, and let $F\colon G/\Gamma \to \mathbf{C}$ satisfy $\|F\|_{\mathrm{Lip}} \le L$. Then one may represent*

$$F(x) = \sum_{|\xi|\le \mathrm{poly}_m(LK\varepsilon^{-1})} F_\xi(x) + G(x),$$

with

- *(1) $\|G\|_\infty \le \varepsilon$;*
- *(2) F_ξ has vertical frequency ξ; and*
- *(3) F_ξ has Lipschitz norm bounded by $\mathrm{poly}_m(LK\varepsilon^{-1})$.*

Remark. The bounds in this specific lemma could likely be substantially improved with a more careful treatment.

Proof. The proof follows exactly as in [23, Lemma 3.7] (with the quantification as suggested by [50]) so we will be brief with details. Let $R \geq 1$ be an integer cutoff and let K_R denote the same kernel as in the proof of Lemma A.8. Define

$$F_1(y) := \int_{\mathbf{R}^{m_s}/\mathbf{Z}^{m_s}} F(\Theta y) K_R(\Theta)\ d\Theta$$

where we have identified the last group in the filtration with $\mathbf{R}^{m_s}/\mathbf{Z}^{m_s}$ for the appropriate integer $m_s \leq m$ (and therefore Θy makes sense for $y \in G/\Gamma$, explicitly defined as $\psi_{\mathcal{X}}^{-1}(\Theta^*)y$ where Θ^* is 0 in the first $m - m_s$ coordinates and Θ in the final m_s). Fourier expansion in $\mathbf{R}^{m_s}/\mathbf{Z}^{m_s}$ gives that

$$F_1(y) := \sum_{k \in \mathbf{Z}^{m_s}} F^{\wedge}(y;k)(R^{m_s}\mu_R(k))$$

where

$$F^{\wedge}(y;k) := \int_{\mathbf{R}^{m_s}/\mathbf{Z}^{m_s}} F(\Theta y) e(-k \cdot \Theta)\ d\Theta.$$

The estimates from Lemma A.8 now complete the proof, noting that metric on G/Γ when descended to the torus is $\mathrm{poly}_m(K)$-equivalent to the standard metric on $\mathbf{R}^{m_s}/\mathbf{Z}^{m_s}$. ☐

Appendix B. Circle method estimates

The material within this appendix consists of standard circle method computations, aside from proving an L^∞-comparison estimate between certain W-tricked quadratic Gauss sums and the Fourier transform of an interval. This comparison is essentially contained within the work of Browning and Prendiville [9].

B.1. L^6-bound on the Fourier transform.

We first require a log-free variant of Weyl's inequality (see [20, Lemma A.11]).

Lemma B.1 (Weyl's inequality)**.** *There exists an absolute constant $C = C_{B.1} > 0$ such that the following holds. Let $\alpha, \beta \in \mathbf{T}$, $\delta \in (0,1)$, and let I be an interval in $\mathbf{Z}$. If*

$$\left|\sum_{y \in I} e(\alpha y^2 + \beta y)\right| \geq \delta |I|,$$

then either $|I| \leq C\delta^{-C}$ or there is a positive integer $q \leq C\delta^{-C}$ such that

$$\|q\alpha\| \leq C\delta^{-C}|I|^{-2}.$$

We next require the following basic estimate regarding exponential sum estimates, which is based on [37, Chapter 4].

Lemma B.2. *Let W and P be as in (2.2) with $W \leq N$ and, for $r \in [W]$, define*

$$P_r(y) = \frac{P(Wy + r) - P(r)}{W}.$$

We have that

$$\sup_{r \in [W]} \int_0^1 \left|\sum_{x \in [\pm N]} e(\Theta P_r(x))\right|^4 d\Theta \ll N^2 \cdot \exp\left(O\left(\frac{\log N}{\log\log N}\right)\right).$$

Proof. Fix $r \in [W]$; the proof will trivially give a bound uniform in r. Let $s(d) = \{(x,y) \in J \times J \colon P_r(x) - P_r(y) = d\}$. Since $|P_r(x) - P_r(y)| \le N^4$ (say) for $x, y \in [\pm N]$, we have $s(d) = 0$ for $|d| \ge N^4$. Furthermore, note that $P_r(x) - P_r(y) = (x-y)(W^2(x+y) + (2Wr+1))$. Therefore, $s(0) \le 2(2N+1)$ and, for $d \neq 0$, the divisor bound implies

$$|s(d)| \le \exp\left(O\left(\frac{\log N}{\log\log N}\right)\right).$$

By definition,

$$\sum_{d \in \mathbf{Z}} s(d) = (2N+1)^2.$$

Therefore,

$$\int_0^1 \left|\sum_{x \in [\pm N]} e(\Theta P_r(x))\right|^4 d\Theta = \int_0^1 \left(\sum_{y \in \mathbf{Z}} s(y) e(y\Theta)\right)^2 d\Theta = \sum_{d \in \mathbf{Z}} s(d) s(-d)$$
$$\le s(0)^2 + \left(\max_{d \in \mathbf{Z}\setminus\{0\}} s(d)\right) \sum_{d \in \mathbf{Z}} s(d) \ll N^2 \cdot \exp\left(O\left(\frac{\log N}{\log\log N}\right)\right). \qquad \square$$

We next record various basic properties of (generalized composite) Gauss sums. Several of these properties are recorded in [6, Exercise 12,23].

Lemma B.3. *Define*

$$G(a,b,c) = \sum_{n=0}^{c-1} e\left(\frac{an^2 + bn}{c}\right).$$

We have the following set of properties:

- *If* $\gcd(c,d) = 1$ *then*

$$G(a,b,cd) = G(ac,b,d)G(ad,b,c);$$

- *If* $\gcd(a,c) > 1$ *then* $G(a,b,c) = 0$ *unless* $\gcd(a,c) | b$. *In this case, it follows that*

$$G(a,b,c) = G\left(\frac{a}{\gcd(a,c)}, \frac{b}{\gcd(a,c)}, \frac{c}{\gcd(a,c)}\right);$$

- *If* $\gcd(a,c) = 1$ *and* $\gcd(c,2) = 1$ *then*

$$|G(a,b,c)| = \sqrt{c};$$

- *If* $c = 2^k$ *for* $k \ge 1$, $\gcd(a,c) = 1$, *and* b *is even then*

$$|G(a,b,c)| \le 2\sqrt{c}.$$

Note that these relations can be used to determine a bound on the magnitude of any composite Gauss sum. We can use the first relation to decompose into prime power moduli c and the second relation to reduce to $\gcd(a,c) = 1$; the third deals with odd prime powers and the final one with even prime powers.

Now for the remainder of this appendix, we say Θ is in the major arcs $\mathfrak{M}$ if there exists $0 \le q_1 < q_2 \le N^{\varepsilon}$ such that

$$\left|\Theta - \frac{q_1}{q_2}\right| \le N^{-2+\varepsilon}$$

for a small constant $\varepsilon > 0$ to be chosen later. Set $\mathfrak{m} = \mathbf{T} \setminus \mathfrak{M}$ to be the minor arcs.

Lemma B.4. *There exists* $\varepsilon_{B.4} > 0$ *such that the following holds. Let* $0 < \varepsilon < \varepsilon_{B.4}$ *and* $0 < \delta \le \delta_{B.4}(\varepsilon)$*. Furthermore, suppose that* W *and* $P_r(\cdot)$ *are as in Lemma B.2 and with* $W \le N^{\delta}$*. Then, we have*

$$\sup_{r\in[W]} \int_{\mathfrak{m}} \left| \sum_{x\in[\pm N]} e(\Theta P_r(x)) \right|^6 d\Theta \ll_{\varepsilon,\delta} N^{4-\delta}$$

and

$$\sup_{\substack{\Theta\in\mathfrak{m}\\ r\in[W]}} \left| \sum_{x\in[\pm T]} e(\Theta P_r(x)) \right| \ll_{\varepsilon,\delta} N^{1-\delta}.$$

Proof. We take $0 < \delta \ll \varepsilon$ to be chosen later. Let $f(\Theta)$ be the expression inside the supremum. By Lemma B.1, for $\Theta \in \mathfrak{m}$ we have $|f(\Theta)| \ll_{\varepsilon,\delta} N^{1-\delta}$. Indeed, if not then we must have $\|qW^2\alpha\| \ll N^{C\delta}N^{-2}$ for some $q \ll N^{C\delta}$ (and appropriate C), noting that the first coefficient of P_r is W^2. If δ is small enough, this violates the definition of the minor arcs. This proves the second desired inequality.

For the first, applying Lemma B.2 and using the above bound we find

$$\begin{aligned}\int_{\mathfrak{m}} |f(\Theta)|^6 \, d\Theta &\le \int_{\mathbf{T}} |f(\Theta)|^4 \, d\Theta \cdot \sup_{\Theta\in\mathfrak{m}} |f(\Theta)|^2 \\ &\ll_{\varepsilon,\delta} N^2 \cdot \exp\left(O\left(\frac{\log N}{\log\log N} \right) \right) \cdot (N^{1-\delta})^2 \\ &\ll N^{4-\delta/2}. \end{aligned}$$

□

We now handle the major arcs. Note that, without loss of generality, either $(q_1, q_2) = (0, 1)$ or $\gcd(q_1, q_2) = 1$ and $1 \le q_1 < q_2 \le N^{\varepsilon}$. Furthermore, given that $\varepsilon < 1/4$ (say), the arcs

$$\mathfrak{M}_{q_1,q_2} := \left\{ \Theta \colon \left| \Theta - \frac{q_1}{q_2} \right| \le N^{-2+\varepsilon} \right\}$$

for such (q_1, q_2) are disjoint. We now state the major arc asymptotic for exponential sums of $P_r(x)$; as this material is completely standard, we omit the proof.

Lemma B.5. *There exists* $\varepsilon = \varepsilon_{B.5}$ *such that the following holds. If* $0 < \varepsilon < \varepsilon_{B.5}$*,* $\Theta \in \mathfrak{M}_{q_1,q_2}$*, and* W*,* $P_r(\cdot)$ *are as in Lemma B.2 with* $W \le N^{\varepsilon}$*, and* $\Theta^* = \Theta - \frac{q_1}{q_2}$*, then*

$$\sum_{x\in[\pm N]} e(\Theta P_r(x)) = q_2^{-1} G(W^2 q_1, (2Wr+1)q_1, q_2) \int_{-N}^{N} e(\Theta^* \cdot W^2 x^2) \, dx + O\left(N^{1/2}\right).$$

We also need the following elementary fact proven via integration by parts.

Lemma B.6. *We have*

$$\left| \int_{-\gamma}^{\gamma} e(x^2) \, dx \right| \ll \min(|\gamma|, 1).$$

Proof. By negation symmetry and the triangle inequality it suffices to assume that $\gamma \ge 2$. Now

$$\begin{aligned}\left| \int_{-\gamma}^{\gamma} e(x^2) \, dx \right| &\le \left| \int_{-1}^{1} e(x^2) \, dx \right| + 2 \left| \int_{1}^{\gamma} e(x^2) \, dx \right| \\ &\le 2 + 4 \sup_{t\ge 1} \left| \int_{x\ge t} e(x^2) \, dx \right| = 2 + 4 \sup_{t\ge 1} \left| \int_{t^2}^{\infty} \frac{e(x)}{2\sqrt{x}} \, dx \right| \\ &\le 2 + 4 \sup_{t\ge 1} \left| \frac{e(t^2)}{4\pi i t} \right| + 4 \left| \int_{t^2}^{\infty} \frac{e(x)}{8\pi i x^{3/2}} \, dx \right| \ll 1. \end{aligned}$$

□

We now in position to derive the necessary L^6-bound; this is essentially an exercise in bounding certain integrals and quadratic Gauss sums.

Lemma B.7. *There exists $\varepsilon_{B.7} > 0$ such that the following holds. Let $0 < \varepsilon < \varepsilon_{B.7}$ and $0 < \delta \leq \delta_{B.7}(\varepsilon)$. Furthermore, suppose that W and $P_r(\cdot)$ are as in Lemma B.2 and with $W \leq N^\delta$. We have*

$$\sup_{r\in[W]} \int_{\mathbf{T}} \Big| \sum_{x\in[\pm N]} e(\Theta P_r(x)) \Big|^6 \, d\Theta \ll N^4 W^{-2},$$

where the implied constant is absolute.

Proof. Choosing $\varepsilon_{B.7}$ sufficiently small and using Lemmas B.4 and B.5, it suffices to prove that

$$\Bigg(G(0,0,1)^6 + \sum_{\substack{1\leq q_1<q_2\leq n^\varepsilon \\ \gcd(q_1,q_2)=1}} (q_2^{-1} G(W^2 q_1, (2Wr+1)q_1, q_2))^6\Bigg) \cdot \int_{-N^{-2+\varepsilon}}^{N^{-2+\varepsilon}} \Big| \int_{-N}^{N} e(\Theta^* \cdot W^2 x^2)\, dx \Big|^6 d\Theta^*$$

is bounded as in the statement of the lemma. Note that, as $\gcd(W^2, 2Wr+1) = 1$ and $\gcd(q_1, q_2) = 1$, we have $|G(W^2 q_1, (2Wr+1)q_1, q_2)| \leq 4\sqrt{q_2}$ using Lemma B.3, and thus the first term is seen to be bounded by a constant. The required bound on the other term, the integral, follows from Lemma B.6: we obtain after change of variables the bound

$$\Big| \int_{-N}^{N} e(\Theta^* W^2 x^2)\, dx \Big| \ll \min\{N, (\Theta^*)^{-1/2} W^{-1}\}$$

whose 6-th power integrates to $O(N^4 W^{-2})$, as desired. □

B.2. L^∞-comparison estimate.

Lemma B.8. *There exists $\varepsilon = \varepsilon_{B.8} > 0$ such that the following holds. Let $N \geq 1$ and W, w be as in (2.2) with $|W| \leq N^\varepsilon$. Furthermore, define*

$$\nu^*(d) = (NW)^{1/2} \cdot \mathbb{1}[d \in \{P(k) \colon k \in \mathbf{Z}\} \cap [N]]$$

and

$$\nu(d) = \sqrt{\frac{N}{d}} \mathbb{1}_{1\leq d\leq N}.$$

Then, we have

$$\sup_{\substack{\Theta\in\mathbf{T} \\ k\in[W]}} \Big| \sum_{d\in\mathbf{Z}} e(d\Theta)(\nu^*(Wd+k) - \nu(Wd+k)) \Big| \ll \frac{N}{W\sqrt{w}}.$$

Proof. Unwinding the definitions and noting that W is sufficiently small, it suffices to prove that

$$\sup_{\substack{\Theta\in\mathbf{T} \\ k\in[W]}} \Bigg| \sum_{|d|\leq N^{1/2}W^{-3/2}} (NW)^{1/2} e(P(Wd+k^*)\Theta) - \sum_{1\leq d\leq NW^{-1}} \sqrt{\frac{N}{Wd}} e((Wd+k)\Theta) \Bigg| \ll \frac{N}{W\sqrt{w}}$$

with $k^* \in [W]$ the unique choice (due to Proposition 6.5) such that $P(k^*) \equiv k \mod W$. Replacing N by NW, it suffices to prove that

$$\sup_{\substack{\Theta\in\mathbf{Z} \\ k\in[W]}} \Bigg| \sum_{|d|\leq N^{1/2}W^{-1}} N^{1/2} W e(P(Wd+k^*)\Theta) - \sum_{1\leq d\leq N} \sqrt{\frac{N}{d}} e((Wd+k)\Theta) \Bigg| \ll \frac{N}{\sqrt{w}}.$$

Consider potential Θ, k which do not satisfy this. We first consider the second summation. By summing a geometric series, we find that

$$\begin{aligned}\Big|\sum_{1\le d\le N}\sqrt{\frac{N}{d}}\exp((Wd+k)\Theta)\Big| &= \Big|\sum_{1\le d\le N}\sqrt{\frac{N}{d}}\exp((Wd)\Theta)\Big| \\ &\ll \sum_{1\le t\le N-1}\left(\sqrt{\frac{N}{t}}-\sqrt{\frac{N}{t+1}}\right)\Big|\sum_{1\le d\le t}e(dW\Theta)\Big| + \Big|\sum_{1\le d\le N}\exp(dW\Theta)\Big| \\ &\ll \frac{1}{\|W\Theta\|_{\mathbf{T}}} + \sum_{1\le t\le N} N^{1/2}t^{-3/2}\min\left(t, \frac{1}{\|W\Theta\|_{\mathbf{T}}}\right) \\ &\ll \frac{1}{\|W\Theta\|_{\mathbf{T}}} + \frac{N^{1/2}}{\|W\Theta\|_{\mathbf{T}}^{1/2}}.\end{aligned}$$

This is smaller than $\frac{N}{\sqrt{w}}$ unless $\|W\Theta\|_{\mathbf{T}} \le w/N$. Now we consider the first summation. Using a version of Weyl's inequality that accounts for the linear coefficient (see, e.g., [23, Proposition 4.3]), we have that if

$$\sum_{|d|\le N^{1/2}W^{-1}} N^{1/2}We(P(Wd+k^*)\Theta) \gg \frac{N}{\sqrt{w}},$$

then there exists $q \le w^{O(1)}$ such that

$$\|q\cdot W^3\Theta\|_{\mathbf{T}} \le \frac{w^{O(1)}W^2}{N}, \qquad \|q\cdot(2W^2k^*+W)\Theta\|_{\mathbf{T}} \le \frac{w^{O(1)}W}{N^{1/2}}.$$

As W is sufficiently small and $\gcd(2Wk^*+1, W) = 1$, we can see that it suffices to handle Θ such that there exists $q' \le w^{O(1)}$ for which

$$\|q'\cdot W\Theta\|_{\mathbf{T}} \le \frac{w^{O(1)}}{N}.$$

Indeed, one can use argumentation similar to that appearing in the proof of Proposition 5.6. (Note that both the first and second sums being large implies this condition.)

For such Θ, we may write $W\Theta = \frac{q_1}{q_2} + \Theta^*$ with $|\Theta^*| \le \frac{w^{O(1)}}{N}$ where $q_2 \le w^{O(1)}$ and $\gcd(q_1, q_2) = 1$. We now have

$$\begin{aligned}&\sum_{|d|\le N^{1/2}W^{-1}} N^{1/2}We(P(Wd+k^*)\Theta) \\ &= \exp(k\Theta)\cdot\sum_{|d|\le N^{1/2}W^{-1}} N^{1/2}We\left((W\Theta)\left(\frac{P(Wd+k^*)-P(k^*)}{W}\right) + \frac{P(k^*)-k}{W}\cdot W\Theta\right) \\ &= \exp\left(k\Theta + \frac{q_1\cdot(P(k^*)-P(k))}{q_2W}\right)\sum_{|d|\le N^{1/2}W^{-1}} N^{1/2}We((W\Theta)(W^2d^2+(2Wk^*+1)d)) + O\left(N^{1/2}\right).\end{aligned}$$

Now using the major arc bounds in Lemma B.5, and noting that $q_2^{-1}G(W^2q_1, (2Wk^*+1)q_1, q_2) = 0$ if $\gcd(q_2, W) > 1$, the above sum is bounded by $N^{1-\Omega(1)}$ in this case. Furthermore, if $q_2 \ne 1$ and $\gcd(q_2, W) = 1$, then q_2 has a prime factor larger than w, and therefore $q_2^{-1}G(W^2q_1, (2Wk^*+1)q_1, q_2) \ll w^{-1/2}$ and the sum becomes bounded by $O((N^{1/2}W)(N^{1/2}W^{-1})w^{-1/2}) = O(Nw^{-1/2})$.

Thus, it suffices to focus on the case where $q_2 = 1$ and thus $q_1 = 0$. Note that

$$\sup_{|\Theta^*|\le\frac{w^{O(1)}}{N}}\Big|\sum_{|d|\le N^{1/2}W^{-1}} N^{1/2}W\Big(e(\Theta^*(W^2d^2+(2Wk^*+1)d)) - e(\Theta^*(W^2d^2))\Big)\Big| \le N^{3/4}.$$

Now that we have significantly reduced our initial situation of general Θ, k. To prove the lemma, it now simply suffices to prove that

$$\sup_{|\Theta^*|\le \frac{w^{O(1)}}{N}} \Big| \sum_{|d|\le N^{1/2}W^{-1}} N^{1/2}We(\Theta^*(W^2d^2)) - \sum_{1\le d\le N} \sqrt{\frac{N}{d}}e(d\Theta^*)\Big| \ll \frac{N}{\sqrt{w}}.$$

For $|\Theta^*| \le w^{O(1)}/N$, we have

$$\begin{aligned}
&\Big| \sum_{|d|\le N^{1/2}W^{-1}} N^{1/2}We(\Theta^*(W^2d^2)) - \sum_{1\le d\le N} \sqrt{\frac{N}{d}}e(d\Theta^*)\Big| \\
&\le \Big| \sum_{1\le d\le N^{1/2}W^{-1}} 2N^{1/2}We(\Theta^*(W^2d^2)) - \sum_{1\le d\le N} \sqrt{\frac{N}{d}}e(d\Theta^*)\Big| + N^{3/4} \\
&\le \sum_{1\le d\le N^{1/2}W^{-1}} \Big|2N^{1/2}We(\Theta^*(W^2d^2)) - \sum_{W^2d^2\le t\le W^2d^2+2W^2d} \sqrt{\frac{N}{t}}e(t\Theta^*)\Big| + O\left(N^{4/5}\right) \\
&\le \sum_{1\le d\le N^{1/2}W^{-1}} \Big|2N^{1/2}We(\Theta^*(W^2d^2)) - \sum_{W^2d^2\le t\le W^2d^2+2W^2d} \sqrt{\frac{N}{W^2d^2}}e(W^2d^2\Theta^*)\Big| + O\left(N^{4/5}\right) \\
&\le \sum_{1\le d\le N^{1/2}W^{-1}} \Big|2N^{1/2}W - (2W^2d+1)\sqrt{\frac{N}{W^2d^2}}\Big| + O\left(N^{4/5}\right) \ll N^{4/5},
\end{aligned}$$

as desired. □

Appendix C. Miscellaneous estimates

In this appendix, we prove a variety of miscellaneous estimates largely concerning changing the parameters of the U^k-norms. We first require the following elementary inequality.

Fact C.1. *For all positive integers $k \ge 1$, we have that*

$$|\sin(kx)| \le k|\sin x|.$$

Proof. We proceed by induction; the result is trivial for $k = 1$. For the inductive step, note that

$$|\sin((j+1)x)| \le |\sin(jx)\cos x + \sin x \cos jx| \le |\sin(jx)| + |\sin(x)| \le (j+1)|\sin x|. \qquad \square$$

We now prove a lemma saying that the U^k-norms behave well with respect to rescaling the width.

Lemma C.2. *Given a function $f\colon \mathbf{Z} \to \mathbf{C}$ with finite support, subsets of integers $Q_1, \ldots, Q_{k-1}$, and positive integers L_1, L_2 such that $L_2 \mid L_1$, we have*

$$\|f\|^{2^k}_{\square^k_{Q_1,\ldots,Q_{k-1},[L_1]}} \le \|f\|^{2^k}_{\square^k_{Q_1,\ldots,Q_{k-1},[L_2]}}.$$

Proof. We first reduce to the case $k = 1$. Consider expanding the box-norm; note that

$$\|f\|^{2^k}_{\square^k_{\square_{Q_1,\ldots,Q_{k-1},[L]}}} = \mathbf{E}_{h_i,h_i'\in Q_i}\|\Delta'_{(h_i,h_i')_{i\in[k-1]}}f\|^2_{U^1_{[L]}}$$

for $L \in \{L_1, L_2\}$. Therefore it suffices to prove that

$$\sum_{x\in\mathbf{Z}} \mathbf{E}_{h,h'\in[L_1]}\Delta'_{h,h'}f(x) \le \sum_{x\in\mathbf{Z}} \mathbf{E}_{h,h'\in[L_2]}\Delta'_{h,h'}f(x).$$

Set $W_{L_i}(x) = \mathbb{1}_{0<x\le L_i}/L_i$, and note that

$$\sum_{x\in\mathbf{Z}} \mathbf{E}_{h,h'\in[L_i]}\Delta_{h,h'}f(x) = \sum_{x,h,h'\in\mathbf{Z}} f(x+h)\overline{f(x+h')}W_{L_i}(h)W_{L_i}(h')$$
$$= \int_0^1 |\widehat{f}(\Theta)|^2|\widehat{W_{L_i}}(\Theta)|^2\ d\Theta = \int_{-1/2}^{1/2} |\widehat{f}(\Theta)|^2\left(\frac{\sin(L_i\pi\Theta)}{L_i\sin(\pi\Theta)}\right)^2\ d\Theta.$$

Applying Fact C.1 with $k = L_1/L_2$ yields

$$\int_{-1/2}^{1/2} |\widehat{f}(\Theta)|^2\left(\frac{\sin(L_1\pi\Theta)}{L_1\sin(\pi\Theta)}\right)^2\ d\Theta \le \int_{-1/2}^{1/2} |\widehat{f}(\Theta)|^2\left(\frac{\sin(L_2\pi\Theta)}{L_2\sin(\pi\Theta)}\right)^2\ d\Theta,$$

as desired. □

We next prove the analogous inequality with respect to rescaling the difference parameters within the U^k-norm.

Lemma C.3. *Given an integer $k \ge 1$, there exists $C_k = C_{C.3}(k) > 0$ such that the following holds. Given a 1-bounded function $f\colon \mathbf{Z}\to\mathbf{C}$ such that $\operatorname{supp}(f)\subseteq[\pm N]$, subsets of integers $Q_1,\ldots,Q_{k-1}$ each contained in $[\pm N]$, and positive integers L_1, L_2 such that $N \ge L_1 \ge 2L_2$, we have*

$$\|f(x)\|^{2^k}_{\square^k_{Q_1,\ldots,Q_{k-1},[L_1]}} \le \|f(x)\|^{2^k}_{\square^k_{Q_1,\ldots,Q_{k-1},L_2\cdot[L_1/L_2]}} + O\left(\frac{C_kN\cdot L_2}{L_1}\right)$$

Proof. One can reduce to the case $k=1$ as in Lemma C.2. It, therefore, suffices to prove that

$$\sum_{x\in\mathbf{Z}} \mathbf{E}_{h,h'\in[L_1]}\Delta_{h,h'}f(x) \le \sum_{x\in\mathbf{Z}} \mathbf{E}_{h,h'\in L_2\cdot[L_1/L_2]}\Delta_{h,h'}f(x) + O\left(\frac{N\cdot L_2}{L_1}\right).$$

Via a direct Fourier-analytic computation, we have

$$\sum_{x\in\mathbf{Z}} \mathbf{E}_{h,h'\in[L_1]}\Delta'_{h,h'}f(x) = \int_{-1/2}^{1/2} |\widehat{f}(\Theta)|^2\left(\frac{\sin(L_1\pi\Theta)}{L_1\sin(\pi\Theta)}\right)^2\ d\Theta$$

and

$$\sum_{x\in\mathbf{Z}} \mathbf{E}_{h,h'\in L_2\cdot[L_1/L_2]}\Delta'_{h,h'}f(x) = \int_{-1/2}^{1/2} |\widehat{f}(\Theta)|^2\left(\frac{\sin(\lfloor L_1/L_2\rfloor\pi(L_2\Theta))}{\lfloor L_1/L_2\rfloor\sin(L_2\pi\Theta)}\right)^2\ d\Theta.$$

Using Fact C.1 in the denominator, we have

$$\left(\frac{\sin(L_1\pi\Theta)}{L_1\sin(\pi\Theta)}\right)^2 \le \left(\frac{L_2\sin(L_1\pi\Theta)}{L_1\sin(L_2\pi\Theta)}\right)^2.$$

Next, note that

$$\left|\left(\frac{L_2\sin(L_1\pi\Theta)}{L_1\sin(L_2\pi\Theta)}\right)^2 - \left(\frac{\sin(\lfloor L_1/L_2\rfloor\pi(L_2\Theta))}{\lfloor L_1/L_2\rfloor\sin(L_2\pi\Theta)}\right)^2\right|$$
$$\le \frac{2L_2|\pi\Theta|}{|\sin(L_2\pi\Theta)|^2}\cdot\left|\frac{L_2\sin(L_1\pi\Theta)}{L_1} - \frac{\sin(\lfloor L_1/L_2\rfloor\pi(L_2\Theta))}{\lfloor L_1/L_2\rfloor}\right|$$
$$\ll \frac{L_2|\pi\Theta|}{|\sin(L_2\pi\Theta)|^2}\cdot\left(\frac{L_2^2|\Theta|}{L_1}\right) \ll \frac{L_2^3|\Theta|^2}{L_1|\sin(L_2\pi\Theta)|^2}.$$

Therefore, it follows that

$$\sum_{x\in\mathbf{Z}} \mathbf{E}_{h,h'\in[L_1]}\Delta'_{h,h'}f(x) = \int_{-1/2}^{1/2} |\widehat{f}(\Theta)|^2\left(\frac{\sin(L_1\pi\Theta)}{L_1\sin(\pi\Theta)}\right)^2\ d\Theta$$

$$= \int_{-1/(4L_2)}^{1/(4L_2)} |\widehat{f}(\Theta)|^2 \left(\frac{\sin(L_1\pi\Theta)}{L_1\sin(\pi\Theta)}\right)^2 \, d\Theta + O\left(\frac{N\cdot L_2^2}{L_1^2}\right)$$
$$\leq \int_{-1/(4L_2)}^{1/(4L_2)} |\widehat{f}(\Theta)|^2 \left(\left(\frac{\sin(\lfloor L_1/L_2\rfloor\pi(L_2\Theta))}{\lfloor L_1/L_2\rfloor\sin(L_2\pi\Theta)}\right)^2 + O\left(\frac{L_2^3|\Theta|^2}{L_1|\sin(L_2\pi\Theta)|^2}\right)\right) \, d\Theta + O\left(\frac{N\cdot L_2^2}{L_1^2}\right)$$
$$\leq \sum_{x\in\mathbf{Z}} \mathbf{E}_{h,h'\in L_2\cdot[L_1/L_2]} \Delta'_{h,h'} f(x) + O\left(\frac{N\cdot L_2}{L_1}\right),$$
as desired. □

We will also require the following version of U^2-inverse theorem, which appears as [40, Lemma 2.4].

Lemma C.4 ([40, Lemma 2.4])**.** *Let* $N \geq 1$ *and* $f\colon \mathbf{Z}\to\mathbf{C}$ *be* 1*-bounded such that* $\operatorname{supp}(f)\subseteq[N]$*. If*
$$\|f\|_{U^2_{[\delta' N]}}^4 \geq \delta N$$
then
$$\sup_{\beta\in\mathbf{T}} \left|\sum_{x\in\mathbf{Z}} e(\beta x) f(x)\right| \gg (\delta\delta')^{O(1)} N.$$

We also have the following well-known converse to the U^2-inverse theorem. We include the proof, as our definition of the U^2-norm is slightly nonstandard.

Lemma C.5. *Let* $f\colon \mathbf{Z}\to\mathbf{C}$ *be a* 1*-bounded function with* $\operatorname{supp}(f)\subseteq[\delta^{-1}N]$*. If* $N \gg \delta^{-O(1)}$ *and*
$$\sup_{\beta\in\mathbf{T}} \left|\sum_{x\in\mathbf{Z}} e(\beta x) f(x)\right| \geq \delta N$$
then
$$\|f\|_{U^2_{[N]}}^4 \gg \delta^{O(1)} N.$$

Proof. By adjusting implicit constants, we may assume that δ is smaller than an absolute constant throughout. Let β be such that
$$\left|\sum_{x\in\mathbf{Z}} e(\beta x) f(x)\right| \geq \delta N$$
and define $f^{(1)}(x) = f(x)e(\beta x)$. Note that
$$\|f\|_{U^2_{[N]}}^4 = \sum_{x\in\mathbf{Z}} \mathbf{E}_{\substack{h_1,h_1'\in[N]\\ h_2,h_2'\in[N]}} f(x+h_1+h_2)\overline{f(x+h_1+h_2')f(x+h_1'+h_2)}f(x+h_1'+h_2')$$
$$= \sum_{x\in\mathbf{Z}} \mathbf{E}_{\substack{h_1,h_1'\in[N]\\ h_2,h_2'\in[N]}} f^{(1)}(x+h_1+h_2)\overline{f^{(1)}(x+h_1+h_2')f^{(1)}(x+h_1'+h_2)}f^{(1)}(x+h_1'+h_2')$$
$$= \sum_{x\in\mathbf{Z}} \mathbb{1}_{|x|\leq 5\delta^{-1}N} \mathbf{E}_{h_1,h_1'\in[N]} \left|\mathbf{E}_{h_2\in[N]} f^{(1)}(x+h_1+h_2)\overline{f^{(1)}(x+h_1'+h_2)}\right|^2$$
$$\geq \left(\sum_{x\in\mathbf{Z}} \mathbb{1}_{|x|\leq 5\delta^{-1}N} \mathbf{E}_{h_1,h_1'\in[N]} \left|\mathbf{E}_{h_2\in[N]} f^{(1)}(x+h_1+h_2)\overline{f^{(1)}(x+h_1'+h_2)}\right|\right)^2$$
$$\cdot \left(\sum_{x\in\mathbf{Z}} \mathbb{1}_{|x|\leq 5\delta^{-1}N} \mathbf{E}_{h_1,h_1'\in[N]} 1\right)^{-1}$$

$$\gg \delta N^{-1}\Big|\sum_{x\in\mathbf{Z}} \mathbb{1}_{|x|\le 5\delta^{-1}N}\mathbf{E}_{h_1,h_1'\in[N]}\mathbf{E}_{h_2\in[N]} f^{(1)}(x+h_1+h_2)\overline{f^{(1)}(x+h_1'+h_2)}\Big|^2$$

$$= \delta N^{-1}\Big|\sum_{x\in\mathbf{Z}} \mathbf{E}_{h_1,h_1'\in[N]} f^{(1)}(x+h_1)\overline{f^{(1)}(x+h_1')}\Big|^2$$

$$= \delta N^{-1}\bigg(\int_{\mathbf{T}} |\widehat{f^{(1)}}(\Theta)|^2\cdot\Big(\frac{\sin(N\Theta/2)}{N\sin(\Theta/2)}\Big)^2\, d\Theta\bigg)^2 \gg \delta^{O(1)}N,$$

where, by construction, $|\widehat{f^{(1)}}(0)|\ge \delta N$, and therefore $|\widehat{f^{(1)}}(\Theta)|\ge \delta N/2$ for $|\Theta|\le \delta^4N^{-1}$. □

By writing the 2^k-th power of the U^k-norm for $k\ge 2$ as the sum of the 4-th powers of U^2-norms of differenced functions and applying Lemmas C.2, C.4, and C.5, and then iterating, we thus deduce the following rescaling inequality for the U^k-norm.

Corollary C.6. *Fix an integer $k\ge 2$. Let $f\colon \mathbf{Z}\to\mathbf{C}$ is 1-bounded such that $\operatorname{supp}(f)\subseteq[N]$, $N\ge\delta^{-O_k(1)}$, and*

$$\|f\|_{U^k_{[\delta N]}}^{2^k}\ge \delta N.$$

Then if $\delta'\in[\delta,\delta^{-1}]$, we have

$$\|f\|_{U^k_{[\delta' N]}}^{2^k}\gg \delta^{O_k(1)} N.$$

We next require the elementary fact that if a 1-bounded function correlates with an exponential phase on a arithmetic progression of a positive density, this may be extended to the full interval with only polynomial loss. This is essentially [24, Lemma 3.5(ii)] or [2, Proposition A.4]; we provide a proof for completeness.

Lemma C.7. *Suppose that $f\colon\mathbf{Z}\to\mathbf{C}$ is a 1-bounded function such that $\operatorname{supp}(f)\subseteq[\pm N]$, $N\ge\delta^{-O(1)}$, and there exists an arithmetic progression P contained in $[N]$ such that*

$$\sup_{\beta\in\mathbf{T}}\Big|\sum_{x\in P} e(\beta x)f(x)\Big|\ge\delta N.$$

Then,

$$\sup_{\beta\in\mathbf{T}}\Big|\sum_{x\in \mathbf{Z}} e(\beta x)f(x)\Big|\gg\delta^{O(1)} N.$$

Proof. Since f is 1-bounded, P must have length at least δN. Therefore, $\mathbb{1}_P(x)=\mathbb{1}_{q|(x-a)}\mathbb{1}_I(x)$ for an interval I of length at least δN and $0\le a<q\le\delta^{-1}$. Let

$$P_2=\mathbb{1}_P*\Big(\frac{\mathbb{1}_{q|x}\mathbb{1}_{[\delta^3N]}(x)}{q^{-1}\cdot\delta^3 N}\Big).$$

By construction, there exists $\beta\in\mathbf{T}$ such that

$$\Big|\sum_{x\in\mathbf{Z}} P_2(x)e(\beta x)f(x)\Big|\ge 2^{-1}\delta N.$$

Therefore, letting $f_\beta(x)=e(\beta x)f(x)$ and taking the Fourier transform, we have

$$2^{-1}\delta N\le\sup_{\beta\in\mathbf{T}}\Big|\sum_{x\in\mathbf{Z}}P_2(x)e(\beta x)f(x)\Big| = \sup_{\beta\in\mathbf{T}}\Big|\int_{\mathbf{T}}\widehat{P_2}(\Theta)\cdot\widehat{f_\beta}(\Theta)\,d\Theta\Big|$$

$$\le \sup_{\substack{\beta\in\mathbf{T}\\ \Theta\in\mathbf{T}}}|\widehat{f_\beta}(\Theta)|\cdot\int_{\mathbf{T}}|\widehat{P_2}(\Theta)|\,d\Theta \ll \delta^{-O(1)}\sup_{\substack{\beta\in\mathbf{T}\\ \Theta\in\mathbf{T}}}|\widehat{f_\beta}(\Theta)|$$

$$\ll \delta^{-O(1)} \sup_{\beta \in \mathbf{T}} \Big| \sum_{x \in \mathbf{Z}} e(\beta x) f(x) \Big|,$$

where we bound the L^1-norm of $\widehat{P_2}(\Theta)$ by using the Cauchy–Schwarz inequality. □

Analogously to Corollary C.6, by writing the 2^k-th power of the U^k-norm for $k \geq 2$ as the sum of 4-th powers of the U^2-norms of differenced functions, and applying Lemmas C.2, C.4, C.5, and C.7 and then iterating, we thus deduce another rescaling inequality for the U^k-norm.

Corollary C.8. *Fix an integer $k \geq 2$. Let $L \leq \delta^{-1}$, $f\colon \mathbf{Z} \to \mathbf{C}$ be 1-bounded such that* $\operatorname{supp}(f) \subseteq [\pm N]$*, and $N \geq \delta^{-O_k(1)}$. If*

$$\|f\|_{U^k_{L\cdot[\delta N/L]}}^{2^k} \geq \delta N,$$

then

$$\|f\|_{U^k_{[\delta N]}}^{2^k} \gg \delta^{O_k(1)} N.$$

Department of Mathematics, University of Michigan, East Hall, 530 Church Street, Ann Arbor, MI 48109, USA
Email address: speluse@umich.edu

Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
Email address: {asah,msawhney}@mit.edu