

Research Article

Special Issue: In honor of David Jerison

Alexandru D. Ionescu*, Ákos Magyar, Mariusz Mirek, and Tomasz Z. Szarek

Polynomial sequences in discrete nilpotent groups of step 2

<https://doi.org/10.1515/ans-2023-0085>

received December 5, 2022; accepted July 2, 2023

Abstract: We discuss some of our work on averages along polynomial sequences in nilpotent groups of step 2. Our main results include boundedness of associated maximal functions and singular integrals operators, an almost everywhere pointwise convergence theorem for ergodic averages along polynomial sequences, and a nilpotent Waring theorem. Our proofs are based on analytical tools, such as a nilpotent Weyl inequality, and on complex almost-orthogonality arguments that are designed to replace Fourier transform tools, which are not available in the noncommutative nilpotent setting. In particular, we present what we call a *nilpotent circle method* that allows us to adapt some of the ideas of the classical circle method to the setting of nilpotent groups.

Keywords: discrete nilpotent groups, pointwise ergodic theorems, nilpotent circle method, Weyl inequality

MSC 2020: 37A30, 37A45, 42B25

1 Introduction

The goal of this article is twofold. We first review some recent results on averages of functions along polynomial sequences in discrete nilpotent Lie groups of step 2, and the main ideas in the proofs. Then we use one of the main ingredients, a nilpotent Weyl inequality, to prove a new theorem on a nilpotent version of the Waring problem.

The natural general setting for our analysis consists of a *discrete nilpotent group* \mathbb{G} of step d , which by definition is assumed to be a discrete, co-compact subgroup of a connected and simply connected nilpotent Lie group $\mathbb{G}^\#$ of step d , and a *polynomial sequence* $A : \mathbb{Z} \rightarrow \mathbb{G}$, which is a map satisfying $A(0) = 1$ and $D^{k_0}A \equiv 1$ for some $k_0 \geq 1$. Here, D^k is the k -fold differencing operator defined recursively by

$$D^0 A(n) = A(n), \quad D^{k+1} A(n) = D^k A(n)^{-1} D^k A(n+1), \quad n \in \mathbb{Z}.$$

We consider a class of operators defined by taking averages along polynomial sequences in discrete nilpotent groups. As in the continuous case, one can consider discrete maximal operators, which have applications to pointwise ergodic theorems, and discrete Calderón-Zygmund operators.

Dedicated to David Jerison, on the occasion of his 70th birthday.

* **Corresponding author: Alexandru D. Ionescu**, Department of Mathematics, Princeton University, Fine Hall, Washington Rd., Princeton, NJ 08544, United States, e-mail: aionescu@math.princeton.edu

Ákos Magyar: Department of Mathematics, University of Georgia, Athens, GA 30602, United States, e-mail: magyar@math.uga.edu

Mariusz Mirek: Institute for Advanced Study, Princeton, NJ 08540, USA; Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA; Instytut Matematyczny, Uniwersytet Wrocławski, 50-384 Wrocław, Poland, e-mail: mariusz.mirek@rutgers.edu

Tomasz Z. Szarek: Instytut Matematyczny, Uniwersytet Wrocławski, 50-384 Wrocław, Poland; BCAM - Basque Center for Applied Mathematics, 48009 Bilbo, Bizkaia, Spain, e-mail: tzsarek@bcamath.org

1.1 The main theorem

Our main theorem in this article concerns L^p boundedness of maximal averages along polynomial sequences in discrete nilpotent groups of step 2, L^p pointwise ergodic theorems, and L^2 boundedness of singular integrals. More precisely:

Theorem 1.1. (Main result) Assume that \mathbb{G} is a discrete nilpotent group \mathbb{G} of step 2, and $A : \mathbb{Z} \rightarrow \mathbb{G}$ is a polynomial sequence. Then:

(i) (ℓ^p boundedness of maximal averages) Assume $f : \mathbb{G} \rightarrow \mathbb{C}$ is a function, and let

$$\mathcal{M}f(g) = \sup_{N \geq 0} \frac{1}{2N+1} \sum_{|n| \leq N} |f(A^{-1}(n) \cdot g)|, \quad g \in \mathbb{G}.$$

Then, for any $p \in (1, \infty]$,

$$\|\mathcal{M}f\|_{\ell^p(\mathbb{G})} \lesssim_p \|f\|_{\ell^p(\mathbb{G})}.$$

(ii) (L^p pointwise ergodic theorems) Assume \mathbb{G} acts by measure-preserving transformations on a σ -finite measure space X , $f \in L^p(X)$, $p \in (1, \infty)$, and let

$$A_N f(x) = \frac{1}{2N+1} \sum_{|n| \leq N} f(A^{-1}(n) \cdot x), \quad x \in X. \quad (1.1)$$

Then the sequence $A_N f$ converges pointwise almost everywhere and in the L^p norm as $N \rightarrow \infty$.

(iii) (ℓ^2 boundedness of singular averages) Assume $K : \mathbb{R} \rightarrow \mathbb{R}$ is a Calderón-Zygmund kernel, i.e., a C^1 function satisfying

$$\sup_{t \in \mathbb{R}} [(1 + |t|)|K(t)| + (1 + |t|)^2 |K'(t)|] \leq 1, \quad \sup_{N \geq 0} \left| \int_{-N}^N K(t) dt \right| \leq 1. \quad (1.2)$$

Assume that $f : \mathbb{G} \rightarrow \mathbb{C}$ is a (compactly supported) function, and let

$$Hf(g) = \sum_{n \in \mathbb{Z}} K(n) f(A^{-1}(n) \cdot g), \quad g \in \mathbb{G}.$$

Then

$$\|Hf\|_{\ell^2(\mathbb{G})} \lesssim \|f\|_{\ell^2(\mathbb{G})}.$$

The theorem follows by combining the main results in [35] for parts (i) and (ii) and [37] for part (iii). We discuss now some connections between this theorem and other related results in the literature.

1.1.1 Continuous Radon transforms

The discrete maximal averages and the discrete singular averages defined in Theorem 1.1 can be thought of as discrete analog of the continuous Radon transforms, which are averages along suitable curves or surfaces. The theory of continuous Radon transforms has been extensively studied, motivated mainly by problems at the interface of Fourier analysis and geometry of surfaces in Euclidean spaces or nilpotent groups, and is very well understood. This includes L^q estimates for the full range of exponents $q > 1$ and multidimensional averages, see, e.g., [17, 18, 51].

1.1.2 The Furstenberg-Bergelson-Leibman conjecture

Discrete averages, both of the maximal and singular type, have been considered motivated mainly by open problems in ergodic theory. A fundamental problem in the ergodic theory is to establish convergence in norm

and pointwise almost everywhere for the polynomial ergodic averages as in (1.1) as $N \rightarrow \infty$ for functions $f \in L^p(X)$, $1 \leq p \leq \infty$. The problem goes back to at least the early 1930s with von Neumann's mean ergodic theorem [56] and Birkhoff's pointwise ergodic theorem [10] and led to profound extensions such as Bourgain's polynomial pointwise ergodic theorem [11–13] and Furstenberg's ergodic proof [25] of Szemerédi's theorem [53] in particular. Furstenberg's proof was also the starting point of ergodic Ramsey theory, which resulted in many natural generalizations of Szemerédi's theorem, including a polynomial Szemerédi theorem of Bergelson and Leibman [7].

This motivates the following far reaching conjecture known as the Furstenberg-Bergelson-Leibman conjecture [8, Section 5.5, p. 468].

Conjecture 1.2. Assume that $d, k \geq 1$ are integers, $(X, \mathcal{B}(X), \mu)$ is a probability space, and assume that $T_1, \dots, T_d : X \rightarrow X$ is a given family of invertible measure-preserving transformations on the space $(X, \mathcal{B}(X), \mu)$ that generates a nilpotent group of step k . Assume that $m \geq 1$ is an integer and $P_{1,1}, \dots, P_{1,j}, \dots, P_{d,m} : \mathbb{Z} \rightarrow \mathbb{Z}$ are polynomial maps with integer coefficients such that $P_{i,j}(0) = 0$. Then for any $f_1, \dots, f_m \in L^\infty(X)$, the nonconventional multilinear polynomial averages

$$A_{N; X, T_1, \dots, T_d}^{P_{1,1}, \dots, P_{1,j}, \dots, P_{d,m}}(f_1, \dots, f_m)(x) = \frac{1}{2N+1} \sum_{n \in [-N, N] \cap \mathbb{Z}} \prod_{j=1}^m f_j(T_1^{P_{1,j}(n)} \dots T_d^{P_{d,j}(n)} x) \quad (1.3)$$

converge for μ -almost every $x \in X$ as $N \rightarrow \infty$.

Conjecture 1.2 is a major open problem in the ergodic theory that was promoted in person by Furstenberg, see [1, p. 6662], before being published in [8]. Our main result, Theorem 1.1 (ii), proves this conjecture in the linear case $m = 1$, provided that the family of transformations $T_1, \dots, T_d : X \rightarrow X$ generates a nilpotent group of step $k = 2$.

1.1.3 Earlier pointwise ergodic theorems

The basic linear case $m = d = k = 1$ with $P_{1,1}(n) = n$ follows from Birkhoff's original ergodic theorem [10]. On the other hand, the commutative case $m = d = k = 1$ with an arbitrary polynomial $P = P_{1,1}$ with integer coefficients was a famous open problem of Bellow [3] and Furstenberg [26], solved by Bourgain in his breakthrough papers [11–13].

Some particular examples of averages (1.3) with $m = 1$ and polynomial mappings with degree at most two in the step two nilpotent setting were studied in [36, 45].

The multilinear theory $m \geq 2$, in contrast to the linear theory, is widely open even in the commutative case $k = 1$. Only a few results in the bilinear $m = 2$ and commutative $d = k = 1$ setting are known. Bourgain [14] proved pointwise convergence when $P_{1,1}(n) = an$ and $P_{1,2}(n) = bn$, $a, b \in \mathbb{Z}$. More recently, Krause et al. [40] established pointwise convergence for the polynomial Furstenberg and Weiss averages [27, 28] corresponding to $P_{1,1}(n) = n$ and $P_{1,2}(n) = P(n)$, $\deg P \geq 2$.

1.1.4 Norm convergence

Except for these few cases, there are no other results concerning pointwise convergence for the averages (1.3). The situation is completely different, however, for the question of norm convergence, which is much better understood.

A breakthrough article of Walsh [57] (see also [1]) gives a complete picture of $L^2(X)$ norm convergence of the averages (1.3) for any $T_1, \dots, T_d \in \mathbb{G}$, where \mathbb{G} is a nilpotent group of transformations of a probability space. Prior to this, there was an extensive body of research toward establishing $L^2(X)$ norm convergence, including groundbreaking works of Host and Kra [31], Ziegler [59], Bergelson [4], and Leibman [42]. See also [2, 19, 24, 32, 54] and the survey articles [5, 6, 23] for more details and references, including a comprehensive historical background.

1.1.5 Additional remarks

Bergelson and Leibman [8] showed that convergence may fail if the transformations T_1, \dots, T_d generate a solvable group, so the nilpotent setting is probably the appropriate setting for Conjecture 1.2. The restriction $p > 1$ is necessary in the case of nonlinear polynomials as shown in [15, 41].

If $(X, \mathcal{B}(X), \mu)$ is a probability space and the family of measure preserving transformations (T_1, \dots, T_{d_1}) is *totally ergodic*, then Theorem 1.1(ii) implies that

$$\lim_{N \rightarrow \infty} A_{N; X}^{P_1, \dots, P_{d_1}}(f)(x) = \int_X f(y) d\mu(y), \quad (1.4)$$

μ -almost everywhere on X . We recall that a family of measure preserving transformations (T_1, \dots, T_{d_1}) is called ergodic on X if $T_j^{-1}(B) = B$ for all $j \in \{1, \dots, d_1\}$ implies $\mu(B) = 0$ or $\mu(B) = 1$ and is called totally ergodic if the family $(T_1^n, \dots, T_{d_1}^n)$ is ergodic for all $n \in \mathbb{Z}_+$.

1.2 The universal step-two group \mathbb{G}_0

The proof of Theorem 1.1 will follow from our second main result, Theorem 1.3, for averages on universal nilpotent groups of step two. We start with some definitions. For integers $d \geq 1$, we define

$$Y_d := \{(l_1, l_2) \in \mathbb{Z} \times \mathbb{Z} : 0 \leq l_2 < l_1 \leq d\}$$

and the “universal” step-two nilpotent Lie groups $\mathbb{G}_0^\# = \mathbb{G}_0^\#(d)$

$$\mathbb{G}_0^\# := \{(x_{l_1 l_2})_{(l_1, l_2) \in Y_d} : x_{l_1 l_2} \in \mathbb{R}\}, \quad (1.5)$$

with the group multiplication law

$$[x \cdot y]_{l_1 l_2} := \begin{cases} x_{l_1 0} + y_{l_1 0} & \text{if } l_1 \in \{1, \dots, d\} \text{ and } l_2 = 0, \\ x_{l_1 l_2} + y_{l_1 l_2} + x_{l_1 0} y_{l_2 0} & \text{if } l_1 \in \{1, \dots, d\} \text{ and } l_2 \in \{1, \dots, l_1 - 1\}. \end{cases} \quad (1.6)$$

Alternatively, we can also define the group $\mathbb{G}_0^\#$ as the set of elements

$$g = (g^{(1)}, g^{(2)}), \quad g^{(1)} = (g_{l_1 0})_{l_1 \in \{1, \dots, d\}} \in \mathbb{R}^d, \quad g^{(2)} = (g_{l_1 l_2})_{(l_1, l_2) \in Y_d} \in \mathbb{R}^{d'}, \quad (1.7)$$

where $d' = d(d-1)/2$ and $Y_d' := \{(l_1, l_2) \in Y_d : l_2 \geq 1\}$. Letting

$$R_0 : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d'} \quad \text{denote the bilinear form } [R_0(x, y)]_{l_1 l_2} := x_{l_1 0} y_{l_2 0}, \quad (1.8)$$

we notice that the product rule in the group $\mathbb{G}_0^\#$ is given by

$$[g \cdot h]^{(1)} = g^{(1)} + h^{(1)}, \quad [g \cdot h]^{(2)} = g^{(2)} + h^{(2)} + R_0(g^{(1)}, h^{(1)}) \quad (1.9)$$

if $g = (g^{(1)}, g^{(2)})$ and $h = (h^{(1)}, h^{(2)})$. For any $g = (g^{(1)}, g^{(2)}) \in \mathbb{G}_0^\#$, its inverse is given by

$$g^{-1} = (-g^{(1)}, -g^{(2)} + R_0(g^{(1)}, g^{(1)})).$$

The second variable of $g = (g^{(1)}, g^{(2)}) \in \mathbb{G}_0^\#$ is called the central variable. Based on the product structure (1.9) of the group $\mathbb{G}_0^\#$, it is not difficult to see that $g \cdot h = h \cdot g$ for any $g = (g^{(1)}, g^{(2)}) \in \mathbb{G}_0^\#$ and $h = (0, h^{(2)}) \in \mathbb{G}_0^\#$.

Let $\mathbb{G}_0 = \mathbb{G}_0(d)$ denote the discrete subgroup

$$\mathbb{G}_0 := \mathbb{G}_0^\# \cap \mathbb{Z}^{|Y_d|}. \quad (1.10)$$

Let $A_0 : \mathbb{R} \rightarrow \mathbb{G}_0^\#$ denote the canonical polynomial map (or the moment curve on $\mathbb{G}_0^\#$)

$$[A_0(x)]_{l_1 l_2} := \begin{cases} x^{l_1} & \text{if } l_2 = 0, \\ 0 & \text{if } l_2 \neq 0, \end{cases} \quad (1.11)$$

and notice that $A_0(\mathbb{Z}) \subseteq \mathbb{G}_0$. For $x = (x_{l_1 l_2})_{(l_1, l_2) \in Y_d} \in \mathbb{G}_0^\#$ and $\Lambda \in (0, \infty)$, we define

$$\Lambda \circ x := (\Lambda^{l_1 + l_2} x_{l_1 l_2})_{(l_1, l_2) \in Y_d} \in \mathbb{G}_0^\#. \quad (1.12)$$

Notice that the dilations $\Lambda \circ$ are group homomorphisms on the group \mathbb{G}_0 that are compatible with the map A_0 , i.e., $\Lambda \circ A_0(x) = A_0(\Lambda x)$.

Let $\chi : \mathbb{R} \rightarrow [0, 1]$ be a smooth function supported on the interval $[-2, 2]$. Given any real number $N \geq 1$ and a function $f : \mathbb{G}_0 \rightarrow \mathbb{C}$, we can define a smoothed average along the moment curve A_0 by the formula

$$M_N^\chi(f)(x) := \sum_{n \in \mathbb{Z}} N^{-1} \chi(N^{-1}n) f(A_0(n)^{-1} \cdot x), \quad x \in \mathbb{G}_0. \quad (1.13)$$

The main advantage of working on the group \mathbb{G}_0 with the polynomial map A_0 is the presence of the compatible dilations $\Lambda \circ$ defined in (1.12), which lead to a natural family of associated balls. This can be efficiently exploited by noting that M_N^χ is a convolution operator on \mathbb{G}_0 .

The convolution of functions on the group \mathbb{G}_0 is defined by the formula

$$(f * g)(x) := \sum_{y \in \mathbb{G}_0} f(y^{-1} \cdot x) g(y) = \sum_{z \in \mathbb{G}_0} f(z) g(x \cdot z^{-1}). \quad (1.14)$$

Then it is not difficult to see that $M_N^\chi(f)(x) = f * G_N^\chi(x)$, where

$$G_N^\chi(x) := \sum_{n \in \mathbb{Z}} N^{-1} \chi(N^{-1}n) \mathbb{1}_{\{A_0(n)\}}(x), \quad x \in \mathbb{G}_0. \quad (1.15)$$

We are now ready to state our second main result.

Theorem 1.3. (Boundedness on \mathbb{G}_0) *Let $\mathbb{G}_0 = \mathbb{G}_0(d)$, $d \geq 1$, be the discrete nilpotent group defined in (1.10) and A_0 the polynomial sequence defined in (1.11). Then*

(i) (Maximal estimates) *If $1 < p \leq \infty$ and $f \in \ell^p(\mathbb{G}_0)$, then*

$$\left\| \sup_{N \geq 1} |M_N^\chi(f)| \right\|_{\ell^p(\mathbb{G}_0)} \lesssim_p \|f\|_{\ell^p(\mathbb{G}_0)}, \quad (1.16)$$

where M_N^χ is defined as in (1.13).

(ii) (Long variational estimates) *If $1 < p < \infty$ and $\rho > \max\left\{p, \frac{p}{p-1}\right\}$, and $\tau \in (1, 2]$, then*

$$\|V^\rho(M_N^\chi(f) : N \in \mathbb{D}_\tau)\|_{\ell^p(\mathbb{G}_0)} \lesssim_{p, \rho, \tau} \|f\|_{\ell^p(\mathbb{G}_0)}, \quad (1.17)$$

where $\mathbb{D}_\tau := \{\tau^n : n \in \mathbb{N}\}$. See (1.18) for the definition of the ρ -variation seminorms V^ρ .

(iii) (Singular integrals) *If $K : \mathbb{R} \rightarrow \mathbb{R}$ is a Calderón-Zygmund kernel as in (1.2), $f : \mathbb{G}_0 \rightarrow \mathbb{C}$ is a (compactly supported) function, and*

$$H_0 f(g) := \sum_{n \in \mathbb{Z}} K(n) f(A_0^{-1}(n) \cdot g), \quad g \in \mathbb{G}_0,$$

then

$$\|H_0 f\|_{\ell^2(\mathbb{G}_0)} \lesssim \|f\|_{\ell^2(\mathbb{G}_0)}.$$

1.3 Remarks and overview of the proof

We discuss now some of the main ideas in the proofs of Theorems 1.1 and 1.3.

1.3.1 The Calderón transference principle

One can show that Theorem 1.1 is a consequence of Theorem 1.3 upon performing lifting arguments and adapting the Calderón transference principle [16]. Indeed, if $\mathbb{G}^\#$ is a connected and simply connected nilpotent Lie group of step 2, with Lie algebra \mathcal{G} , then one can choose the so-called exponential coordinates of the second kind associated to a Malcev basis of the Lie algebra \mathcal{G} (see [20], Sec. 1.2) in such a way that

$$\mathbb{G}^\# \simeq \{(x, y) \in \mathbb{R}^{b_1} \times \mathbb{R}^{b_2} : (x, y) \cdot (x', y') = (x + x', y + y' + R(x, x'))\},$$

where $b_1, b_2 \in \mathbb{Z}_+$ depend on the Lie algebra \mathcal{G} and $R : \mathbb{R}^{b_1} \times \mathbb{R}^{b_1} \rightarrow \mathbb{R}^{b_2}$ is a bilinear form.

Moreover, if $\mathbb{G} \leq \mathbb{G}^\#$ is a discrete co-compact subgroup, then one can choose the Malcev basis such that the discrete subgroup \mathbb{G} is identified with the integer lattice $\mathbb{Z}^b = \mathbb{Z}^{b_1} \times \mathbb{Z}^{b_2}$ (see [20], Thm. 5.1.6 and Prop. 5.3.2). Recall that $A : \mathbb{Z} \rightarrow \mathbb{G}$ is a polynomial sequence satisfying $A(0) = 1$. The main point is that one can choose d sufficiently large and a group morphism $T : \mathbb{G}_0 \rightarrow \mathbb{G}^\#$ such that

$$A(n) = T(A_0(n)) \quad \text{for any } n \in \mathbb{Z}.$$

Then one can use this group morphism to transfer bounds on operators on the universal group \mathbb{G}_0 to bounds on operators on the group \mathbb{G} . Theorem 1.1 is thus a consequence of Theorem 1.3, and our main goal therefore is to prove Theorem 1.3.

1.3.2 The variation spaces V^ρ

For any family $(a_t : t \in \mathbb{I})$ of elements of \mathbb{C} indexed by a totally ordered set \mathbb{I} , and any exponent $1 \leq \rho < \infty$, the ρ -variation seminorm is defined by

$$V^\rho(a_t : t \in \mathbb{I}) := \sup_{J \in \mathbb{Z}_+} \sup_{\substack{t_0 < \dots < t_J \\ t_j \in \mathbb{I}}} \left(\sum_{j=0}^{J-1} |a(t_{j+1}) - a(t_j)|^\rho \right)^{1/\rho}, \quad (1.18)$$

where the supremum is taken over all finite increasing sequences in \mathbb{I} . It is easy to see that $\rho \mapsto V^\rho$ is nonincreasing, and for every $t_0 \in \mathbb{I}$, one has

$$\sup_{t \in \mathbb{I}} |a_t| \leq |a_{t_0}| + V^\rho(a_t : t \in \mathbb{I}) \leq \sup_{t \in \mathbb{I}} |a_t| + V^\rho(a_t : t \in \mathbb{I}). \quad (1.19)$$

In particular, the maximal estimate (1.16) follows from the variational estimate (1.17). The main point of proving stronger variational estimates such as (1.17), with general parameters $\tau \in (1, 2]$, is that it gives an elegant path to deriving pointwise ergodic theorems (which would not follow directly just from maximal estimates such as (1.16)). At the same time, the analysis of variational inequalities has many similarities with the analysis of maximal inequalities and is not substantially more difficult. This is due in large part to the Rademacher-Menshov inequality (see [47, Lemma 2.5]): for any $2 \leq \rho < \infty$ and $j_0, m \in \mathbb{N}$ so that $j_0 < 2^m$ and any sequence of complex numbers $(a_k : k \in \mathbb{N})$, we have

$$V^\rho(a_j : j_0 \leq j \leq 2^m) \leq \sqrt{2} \sum_{i=0}^m \left(\sum_{j \in [j_0 2^{-i}, 2^{m-i}-1] \cap \mathbb{Z}} |a_{(j+1)2^i} - a_{j2^i}|^2 \right)^{1/2}. \quad (1.20)$$

1.3.3 ℓ^p theory

The problem of passing from ℓ^2 estimates to ℓ^p estimates in the context of discrete polynomial averages has been investigated extensively in recent years (see, for example, [46], and the references therein).

The full $\ell^p(\mathbb{G}_0)$ bounds in Theorem 1.3 rely on first proving $\ell^2(\mathbb{G}_0)$ bounds. In fact, we first establish (1.17) for $p = 2$ and $\rho > 2$. Then we use the positivity of the operators M_N^χ (i.e., $M_N^\chi(f) \geq 0$ if $f \geq 0$) to prove the maximal operator bounds (1.16) for all $p \in (1, \infty]$. Finally, we use vector-valued interpolation between the bounds (1.17) with $p = 2$ and $\rho > 2$ and (1.16) with $p \in (1, \infty]$ to complete the proof of Theorem 1.3.

1.3.4 Some technical remarks

Theorem 1.3 (i) and (ii) extends the results of [46,48] to the noncommutative, nilpotent setting. Its conclusions remain true for rough averages, i.e., when $\chi = \mathbb{1}_{[-1,1]}$ in (1.13), but it is more convenient to work with smooth averages.

The restriction $p > 1$ in Theorem 1.3 (i) and (ii) is sharp due to [15,41]. However, the range of $\rho > \max\left\{p, \frac{p}{p-1}\right\}$ is only sharp when $p = 2$ due to Lépingle's inequality [43]. One could hope to improve this

to the full range $\rho > 2$, but we do not address this here since the limited range $\rho > \max\left\{p, \frac{p}{p-1}\right\}$ is already sufficient for us to establish Theorem 1.1.

The restriction $p = 2$ in the singular integral bounds in part (ii) is probably not necessary. In the commutative case, one can prove boundedness in the full range $p \in (1, \infty)$ [38], but the proof depends on exploiting certain Fourier multipliers, and we do not know at this time if a similar definitive result holds in the nilpotent case.

1.4 The main difficulty and a nilpotent circle method

Bourgain's seminal articles [11–13] generated a large amount of research and progress in the field. Many other discrete operators have been analyzed by many authors motivated by problems in analysis and ergodic theory. See, for example, [15,36,38–41,44–48,49,50,52] for some results of this type and more references. A common feature of all of these results, which plays a crucial role in the proofs, is that one can use Fourier analysis techniques, in particular, the powerful framework of the classical circle method, to perform the analysis.

Our situation in Theorem 1.3 is different. The main conceptual issue is that there is no good Fourier transform on nilpotent groups, compatible with the structure of the underlying convolution operators and at the level of analytical precision of the classical circle method. At a more technical level, there is no good resolution of the delta function compatible with the group multiplication on the group \mathbb{G}_0 . This prevents us from using a naive implementation of the circle method. The classical delta function resolution

$$\mathbb{1}_{\{0\}}(x^{-1} \cdot y) = \int_{\mathbb{T}^d \times \mathbb{T}^{d'}} e((y^{(1)} - x^{(1)}) \cdot \theta^{(1)}) e((y^{(2)} - x^{(2)}) \cdot \theta^{(2)}) d\theta^{(1)} d\theta^{(2)}$$

does not detect the group multiplication correctly. Here, $(y^{(1)} - x^{(1)}) \cdot \theta^{(1)}$ and $(y^{(2)} - x^{(2)}) \cdot \theta^{(2)}$ denote the usual scalar product of vectors in \mathbb{R}^d and $\mathbb{R}^{d'}$, respectively.

These issues lead to very significant difficulties in the proof and require substantial new ideas. Our main new construction in [35] is what we call a *nilpotent circle method*, an iterative procedure, starting from the center of the group and moving down along its central series. At every stage, we identify “minor arcs,” and bound their contributions using Weyl's inequalities (the classical Weyl inequality as well as a nilpotent Weyl inequality which was proved in [37]). The final stage involves “major arcs” analysis, which relies on a combination of continuous harmonic analysis on groups $\mathbb{G}_0^\#$ and arithmetic harmonic analysis over finite integer rings modulo $Q \in \mathbb{Z}_+$. We outline this procedure in Section 3.

At the implementation level, classical Fourier techniques are replaced with almost orthogonality methods based on exploiting high order T^*T arguments for operators defined on the discrete group \mathbb{G}_0 . Investigating high powers of T^*T (i.e., $(T^*T)^r$ for a large $r \in \mathbb{Z}_+$) is consistent with a general heuristic lying behind the proof of Waring-type problems, which says that the more variables that occur in Waring-type equations, the easier it is to find solutions, and we are able to make this heuristic rigorous in our problem. Manipulating the parameter r , by taking r to be very large, we can always decide how many variables we have at our disposal, making our operators “smoother and smoother.”

1.5 General discrete nilpotent groups

The primary goal is, of course, to remove the restriction that the discrete nilpotent groups \mathbb{G} in Theorem 1.1 are of step 2 and thus establish the full Conjecture 1.2 in the linear $m = 1$ case for arbitrary invertible measure-preserving transformations T_1, \dots, T_d that generate a nilpotent group of any step $k \geq 2$. The iterative argument we outline in Section 3 below could, in principle, be extended to higher step groups, at least as long as the group and the polynomial sequence have suitable “universal”-type structure, as one could try to go down along the central series of the group and prove minor arcs and transition estimates at every stage.

However, this is only possible if one can prove suitable analog of the nilpotent Weyl’s inequalities in Proposition 2.1 on general nilpotent groups of step $k \geq 3$. The point is to have a small (not necessarily optimal, but nontrivial) gain for bounds on oscillatory sums over many variables, corresponding to the kernels of high power $(T^*T)^r$ operators, whenever frequencies are restricted to the minor arcs. In our case, the formulas are explicit, see the identities (2.10), and we can use ideas of Birch [9] and Davenport [21] for Diophantine forms in many variables to control the induced oscillatory sums, but the analysis seems to be more complicated for the higher step nilpotent groups.

1.6 Waring-type problems

The classical Waring problem, solved by Hilbert [30] in 1909, concerns the possibility of writing any positive integer as a sum of finitely many p powers: for any integer $p \geq 1$, there is $r = r(p)$ such that any integer $y \in \mathbb{Z}_+$ can be written in the form

$$y = \sum_{i=1}^r m_i^p, \quad \text{for some nonnegative integers } m_1, \dots, m_r. \quad (1.21)$$

There is a vast amount of literature on this problem and its many possible extensions. In particular, the symmetric system of equations

$$\sum_{j=1}^r (m_j^s - n_j^s) = 0 \quad (1 \leq s \leq d), \quad (1.22)$$

first studied by Vinogradov [55] in relation to the Waring problem, have been the focus of intense recent research, see [58] for some recent results.

We are interested here in understanding the analogs question on our discrete nilpotent Lie group \mathbb{G}_0 , and for our given polynomial sequence A_0 : Can one represent elements $g \in \mathbb{G}_0$ in the form

$$g = A_0(n_1)^{-1} \cdot A_0(m_1) \cdot \dots \cdot A_0(n_r)^{-1} \cdot A_0(m_r), \quad (1.23)$$

for some integers $n_1, m_1, \dots, n_r, m_r$, provided that r is large enough? We are, in fact, interested in proving a quantitative statement on the number of such representations, for integers $n_1, m_1, \dots, n_r, m_r \in [N] := [-N, N] \cap \mathbb{Z}$.

We notice that many group elements g cannot be written in the form (1.23), due to local obstructions; for instance, if g can be represented in the form (1.23) then necessarily $g_{10} \equiv g_{20} \equiv \dots \equiv g_{d0} \pmod{2}$, $g_{10} = g_{30} = \dots \pmod{3}$ etc. We remark also there is a significant difference between the classical Waring

problem (1.21) and its nilpotent analog (1.23), namely the positivity of the p -powers which imposes size restrictions on the variables x_i in terms of the prescribed output value y .

For integers $r, N \geq 1$ and $g \in \mathbb{G}_0$ let

$$S_{r,N}(g) := |\{(m, n) \in [N]^{2r} : A_0(n_1)^{-1} \cdot A_0(m_1) \cdot \dots \cdot A_0(n_r)^{-1} \cdot A_0(m_r) = g\}|. \quad (1.24)$$

A qualitative variant of the Waring problem on nilpotent groups was recently investigated in [33,34], see also the references given there. Our main result in this direction is the following quantitative version on the nilpotent group \mathbb{G}_0 :

Theorem 1.4. (i) *There is an integer $r_0(d) \geq 1$ such that if $r \geq r_0(d)$ is sufficiently large and $g \in \mathbb{G}_0$, then*

$$S_{r,N}(g) = N^{2r} \left(\prod_{(l_1, l_2) \in Y_d} N^{-|l_1| - |l_2|} \right) \left[\mathfrak{S}(g) \int_{\mathbb{R}^{d+d'}} \Phi(\zeta) \epsilon(-(N^{-1} \circ g) \cdot \zeta) d\zeta + O_r(N^{-1/2}) \right], \quad (1.25)$$

uniformly in $N \in \mathbb{N}$. Here, the singular series \mathfrak{S} is defined by

$$\mathfrak{S}(g) := \sum_{a/q \in \mathcal{R}_\infty^{d+d'} \cap [0,1]^{d+d'}} \overline{G(a/q)} \epsilon(-g \cdot a/q) \quad (1.26)$$

and the singular integral Φ is defined by

$$\Phi(\xi) = \int_{[-1,1]^{2r}} \epsilon(D(z, w) \cdot \xi) dz dw, \quad \xi \in \mathbb{R}^{d+d'}. \quad (1.27)$$

In particular, all elements $g \in \mathbb{G}_0$ cannot be represented in the form (1.23) more than a constant times the expected number of representations, i.e.,

$$S_{r,N}(g) \lesssim_r N^{2r} \left(\prod_{(l_1, l_2) \in Y_d} N^{-|l_1| - |l_2|} \right) \quad \text{for any } g \in \mathbb{G}_0. \quad (1.28)$$

(ii) For $r_0(d)$ as earlier, if $r \geq r_0(d)$ and r is even, then there is a sufficiently large integer $Q = Q(r)$ such that

$$S_{r,N}(g) = N^{2r} \left(\prod_{(l_1, l_2) \in Y_d} N^{-|l_1| - |l_2|} \right) [c_r(g) + O_{r,g}(N^{-1/2})], \quad (1.29)$$

for any $g \in \mathbb{H}_Q$ (see definition (3.32)), where $c_r(g) \approx_r 1$ uniformly in g .

We will provide a complete proof of this theorem in Section 4.

1.7 Organization

The rest of this article is organized as follows: in Section 2, we present several nilpotent Weyl estimates proved in [37], which play a key role in the analysis of minor arcs. In Section 3, we outline our main new method, the nilpotent circle method, developed in [35] to prove maximal and variational estimates on nilpotent groups. In Section 4, we prove Theorem 1.4, the main new result in this article.

2 A nilpotent Weyl inequality on the group \mathbb{G}_0

In this section, we derive explicit formulas used in high order T^*T arguments and discuss a key ingredient in our analysis, namely, Weyl inequalities on the group \mathbb{G}_0 .

2.1 High-order T^*T arguments and product kernels

Many of our $\ell^2(\mathbb{G}_0)$ estimates will be based on high-order T^*T arguments. Assume that

$$S_1, T_1, \dots, S_r, T_r : \ell^2(\mathbb{G}_0) \rightarrow \ell^2(\mathbb{G}_0)$$

are convolution operators defined by some $\ell^1(\mathbb{G}_0)$ kernels $L_1, K_1, \dots, L_r, K_r : \mathbb{G}_0 \rightarrow \mathbb{C}$, i.e., $S_j f = f * L_j$ and $T_j f = f * K_j$ for $j \in \{1, \dots, r\}$. Then the adjoint operators S_1^*, \dots, S_r^* are also convolution operators, defined by the kernels L_1^*, \dots, L_r^* given by

$$L_j^*(g) = \overline{L_j(g^{-1})}.$$

Moreover, by using (1.14), for any $f \in \ell^2(\mathbb{G}_0)$ and $x \in \mathbb{G}_0$, we have

$$(S_1^* T_1 \dots S_r^* T_r f)(x) = \sum_{h_1, g_1, \dots, h_r, g_r \in \mathbb{G}_0} \left\{ \prod_{j=1}^r L_j^*(h_j) K_j(g_j) \right\} f(g_r^{-1} \cdot h_r^{-1} \cdot \dots \cdot g_1^{-1} \cdot h_1^{-1} \cdot x). \quad (2.1)$$

In other words $(S_1^* T_1 \dots S_r^* T_r f)(x) = (f * A^r)(x)$, where the kernel A^r is given by

$$A^r(y) = \sum_{h_1, g_1, \dots, h_r, g_r \in \mathbb{G}_0} \left\{ \prod_{j=1}^r \overline{L_j(h_j)} K_j(g_j) \right\} \mathbb{1}_{\{0\}}(g_r^{-1} \cdot h_r \cdot \dots \cdot g_1^{-1} \cdot h_1 \cdot y). \quad (2.2)$$

To use these formulas, we decompose $h_j = (h_j^{(1)}, h_j^{(2)})$, $g_j = (g_j^{(1)}, g_j^{(2)})$ as in (1.7). Then

$$[h_1^{-1} \cdot g_1 \cdot \dots \cdot h_r^{-1} \cdot g_r]^{(1)} = \sum_{1 \leq j \leq r} (-h_j^{(1)} + g_j^{(1)}), \quad (2.3)$$

$$[h_1^{-1} \cdot g_1 \cdot \dots \cdot h_r^{-1} \cdot g_r]^{(2)} = \sum_{1 \leq j \leq r} \{-(h_j^{(2)} - g_j^{(2)}) + R_0(h_j^{(1)}, h_j^{(1)} - g_j^{(1)})\} + \sum_{1 \leq l < j \leq r} R_0(-h_l^{(1)} + g_l^{(1)}, -h_j^{(1)} + g_j^{(1)}), \quad (2.4)$$

as a consequence of applying (1.9) inductively.

In many of our applications, the operators $S_1, T_1, \dots, S_r, T_r$ are equal and, more importantly, are defined by a kernel K that has product structure, i.e.,

$$\begin{aligned} S_1 f &= T_1 f = \dots = S_r f = T_r f = f * K, \\ K(g) &= K(g^{(1)}, g^{(2)}) = K^{(1)}(g^{(1)}) K^{(2)}(g^{(2)}). \end{aligned} \quad (2.5)$$

In this case, we can derive an additional formula for the kernel A^r . We use the identity

$$\mathbb{1}_{\{0\}}(x^{-1} \cdot y) = \int_{\mathbb{T}^d \times \mathbb{T}^{d'}} \epsilon((y^{(1)} - x^{(1)}), \theta^{(1)}) \epsilon((y^{(2)} - x^{(2)}), \theta^{(2)}) d\theta^{(1)} d\theta^{(2)},$$

where $\epsilon(z) = e^{2\pi i z}$. The formula (2.2) shows that

$$A^r(y) = \int_{\mathbb{T}^d \times \mathbb{T}^{d'}} \epsilon(y^{(1)}, \theta^{(1)}) \epsilon(y^{(2)}, \theta^{(2)}) \Sigma^r(\theta^{(1)}, \theta^{(2)}) d\theta^{(1)} d\theta^{(2)}, \quad (2.6)$$

where

$$\Sigma^r(\theta^{(1)}, \theta^{(2)}) = \sum_{h_j, g_j \in \mathbb{G}_0} \left\{ \prod_{j=1}^r \overline{K(h_j)} K(g_j) \right\} \prod_{i=1}^2 \epsilon(-[h_1^{-1} \cdot g_1 \cdot \dots \cdot h_r^{-1} \cdot g_r]^{(i)}, \theta^{(i)}).$$

Recalling the product formula (2.5), we can write

$$\Sigma^r(\theta^{(1)}, \theta^{(2)}) = \Pi^r(\theta^{(1)}, \theta^{(2)}) \Omega^r(\theta^{(2)}), \quad (2.7)$$

for any $(\theta^{(1)}, \theta^{(2)}) \in \mathbb{T}^d \times \mathbb{T}^{d'}$, where

$$\begin{aligned} \Pi^r(\theta^{(1)}, \theta^{(2)}) := & \sum_{h_j^{(1)}, g_j^{(1)} \in \mathbb{Z}^d} \left\{ \prod_{j=1}^r \overline{K^{(1)}(h_j^{(1)})} K^{(1)}(g_j^{(1)}) \right\} \epsilon \left(\theta^{(1)}, \sum_{1 \leq j \leq r} (h_j^{(1)} - g_j^{(1)}) \right) \\ & \times \epsilon \left(-\theta^{(2)}, \left\{ \sum_{1 \leq j \leq r} R_0(h_j^{(1)}, h_j^{(1)} - g_j^{(1)}) + \sum_{1 \leq l < j \leq r} R_0(-h_l^{(1)} + g_l^{(1)}, -h_j^{(1)} + g_j^{(1)}) \right\} \right) \end{aligned} \quad (2.8)$$

and

$$\begin{aligned} \Omega^r(\theta^{(2)}) := & \sum_{h_j^{(2)}, g_j^{(2)} \in \mathbb{Z}^{d'}} \left\{ \prod_{j=1}^r \overline{K^{(2)}(h_j^{(2)})} K^{(2)}(g_j^{(2)}) \right\} \epsilon \left(\theta^{(2)}, \sum_{1 \leq j \leq r} (h_j^{(2)} - g_j^{(2)}) \right) \\ & = \left| \sum_{g^{(2)} \in \mathbb{Z}^{d'}} K^{(2)}(g^{(2)}) \epsilon(-\theta^{(2)}, g^{(2)}) \right|^{2r}. \end{aligned} \quad (2.9)$$

2.2 Weyl estimates

After applying high-order T^*T arguments, we often need to estimate exponential sums and oscillatory integrals involving polynomial phases. With the notation in Section 1.2, for $r \geq 1$, let $D, \widetilde{D} : \mathbb{R}^r \times \mathbb{R}^r \rightarrow \mathbb{G}_0^\#$ be defined by

$$\begin{aligned} D((n_1, \dots, n_r), (m_1, \dots, m_r)) &= A_0(n_1)^{-1} \cdot A_0(m_1) \cdot \dots \cdot A_0(n_r)^{-1} \cdot A_0(m_r), \\ \widetilde{D}((n_1, \dots, n_r), (m_1, \dots, m_r)) &= A_0(n_1) \cdot A_0(m_1)^{-1} \cdot \dots \cdot A_0(n_r) \cdot A_0(m_r)^{-1}. \end{aligned} \quad (2.10)$$

By definition, we have

$$[A_0(n)]_{l_1 l_2} = \begin{cases} n^{l_1} & \text{if } l_2 = 0, \\ 0 & \text{if } l_2 \geq 1, \end{cases} \quad [A_0(n)^{-1}]_{l_1 l_2} = \begin{cases} -n^{l_1} & \text{if } l_2 = 0, \\ n^{l_1 + l_2} & \text{if } l_2 \geq 1. \end{cases}$$

Thus, by using (2.3) and (2.4), for $x = (x_1, \dots, x_r) \in \mathbb{R}^r$ and $y = (y_1, \dots, y_r) \in \mathbb{R}^r$, one has

$$[D(x, y)]_{l_1 l_2} = \begin{cases} \sum_{j=1}^r (y_j^{l_1} - x_j^{l_1}) & \text{if } l_2 = 0, \\ \sum_{1 \leq j_1 < j_2 \leq r} (y_{j_1}^{l_1} - x_{j_1}^{l_1})(y_{j_2}^{l_2} - x_{j_2}^{l_2}) + \sum_{j=1}^r (x_j^{l_1 + l_2} - x_j^{l_1} y_j^{l_2}) & \text{if } l_2 \geq 1, \end{cases} \quad (2.11)$$

and

$$[\widetilde{D}(x, y)]_{l_1 l_2} = \begin{cases} \sum_{j=1}^r (x_j^{l_1} - y_j^{l_1}) & \text{if } l_2 = 0, \\ \sum_{1 \leq j_1 < j_2 \leq r} (x_{j_1}^{l_1} - y_{j_1}^{l_1})(x_{j_2}^{l_2} - y_{j_2}^{l_2}) + \sum_{j=1}^r (y_j^{l_1 + l_2} - x_j^{l_1} y_j^{l_2}) & \text{if } l_2 \geq 1. \end{cases} \quad (2.12)$$

For $P \in \mathbb{Z}_+$ assume $\phi_p^{(j)}, \psi_p^{(j)} : \mathbb{R} \rightarrow \mathbb{R}$, $j \in \{1, \dots, r\}$, are $C^1(\mathbb{R})$ functions with the properties

$$\sup_{1 \leq j \leq r} [|\phi_p^{(j)}| + |\psi_p^{(j)}|] \leq \mathbb{1}_{[-P, P]}, \quad \sup_{1 \leq j \leq r} \int_{\mathbb{R}} |[\phi_p^{(j)}]'(x)| + |[\psi_p^{(j)}]'(x)| dx \leq 1. \quad (2.13)$$

For $\theta = (\theta_{l_1 l_2})_{(l_1, l_2) \in Y_d} \in \mathbb{R}^{|Y_d|}$, $r \in \mathbb{Z}_+$, and $P \in \mathbb{Z}_+$ let

$$S_{P, r}(\theta) = \sum_{n, m \in \mathbb{Z}^r} \epsilon(-D(n, m), \theta) \left\{ \prod_{j=1}^r \phi_p^{(j)}(n_j) \psi_p^{(j)}(m_j) \right\}$$

and

$$\tilde{S}_{P,r}(\theta) = \sum_{n,m \in \mathbb{Z}^r} \epsilon(-\tilde{D}(n, m) \cdot \theta) \left\{ \prod_{j=1}^r \phi_p^{(j)}(n_j) \psi_p^{(j)}(m_j) \right\},$$

where D and \tilde{D} are defined as in (2.11)–(2.12).

The following key estimates are proved in [37, Proposition 5.1 and Lemma 3.1]:

Proposition 2.1. (i) (Nilpotent Weyl estimate) For any $\varepsilon > 0$, there is $r = r(\varepsilon, d) \in \mathbb{Z}_+$ sufficiently large such that for all $P \in \mathbb{Z}_+$ we have

$$|S_{P,r}(\theta)| + |\tilde{S}_{P,r}(\theta)| \lesssim_\varepsilon P^{2r} P^{-1/\varepsilon}, \quad (2.14)$$

provided that there is $(l_1, l_2) \in Y_d$ and an irreducible fraction $a/q \in \mathbb{Q}$, $q \in \mathbb{Z}_+$, such that

$$|\theta_{l_1 l_2} - a/q| \leq 1/q^2 \quad \text{and} \quad q \in [P^\varepsilon, P^{l_1 + l_2 - \varepsilon}]. \quad (2.15)$$

(ii) (Nilpotent Gauss sums) For any irreducible fraction $a/q \in \mathbb{Q}^{|Y_d|}$, $a = (a_{l_1 l_2})_{(l_1, l_2) \in Y_d} \in \mathbb{Z}^{|Y_d|}$, $q \in \mathbb{Z}_+$, we define the arithmetic coefficients

$$G(a/q) = q^{-2r} \sum_{v, w \in \mathbb{Z}_q^r} \epsilon(-D(v, w) \cdot (a/q)), \quad \tilde{G}(a/q) = q^{-2r} \sum_{v, w \in \mathbb{Z}_q^r} \epsilon(-\tilde{D}(v, w) \cdot (a/q)). \quad (2.16)$$

Then for any $\varepsilon > 0$, there is $r = r(\varepsilon, d) \in \mathbb{Z}_+$ sufficiently large such that

$$|G(a/q)| + |\tilde{G}(a/q)| \lesssim_\varepsilon q^{-1/\varepsilon}. \quad (2.17)$$

We also need a related integral estimate, see Lemma 5.4 in [37]:

Proposition 2.2. Given $\varepsilon > 0$ there is $r = r(\varepsilon, d)$ sufficiently large as in Proposition 2.1, such that

$$\left| \int_{\mathbb{R}^r \times \mathbb{R}^r} \left\{ \prod_{j=1}^r \phi_j(x_j) \psi_j(y_j) \right\} \epsilon(-D(x, y) \cdot \beta) dx dy \right| \lesssim \langle \beta \rangle^{-1/\varepsilon},$$

$$\left| \int_{\mathbb{R}^r \times \mathbb{R}^r} \left\{ \prod_{j=1}^r \phi_j(x_j) \psi_j(y_j) \right\} \epsilon(-\tilde{D}(x, y) \cdot \beta) dx dy \right| \lesssim \langle \beta \rangle^{-1/\varepsilon}, \quad (2.18)$$

for any $\beta \in \mathbb{R}^{|Y_d|}$ (here and later on, we use the Japanese bracket notation $\langle \beta \rangle = (1 + |\beta|^2)^{1/2}$) and for any $C^1(\mathbb{R})$ functions $\phi_1, \psi_1, \dots, \phi_r, \psi_r : \mathbb{R} \rightarrow \mathbb{C}$ satisfying, for any $j \in \{1, \dots, r\}$, the bounds

$$|\phi_j| + |\psi_j| \leq \mathbb{1}_{[-1,1]}(x), \quad \int_{\mathbb{R}} [|\partial_x \phi_j(x)| + |\partial_x \psi_j(x)|] dx \leq 1.$$

These statements should be compared with classical Weyl-type estimates, which are proved, for example, in [52, Proposition 1]:

Proposition 2.3. (i) Assume that $P \geq 1$ is an integer and $\phi_P : \mathbb{R} \rightarrow \mathbb{R}$ is a $C^1(\mathbb{R})$ function satisfying

$$|\phi_P| \leq \mathbb{1}_{[-P, P]}, \quad \int_{\mathbb{R}} |\phi_P'(x)| dx \leq 1. \quad (2.19)$$

Assume that $\varepsilon > 0$ and $\theta = (\theta_1, \dots, \theta_d) \in \mathbb{R}^d$ has the property that there is $l \in \{1, \dots, d\}$ and an irreducible fraction $a/q \in \mathbb{Q}$ with $q \in \mathbb{Z}_+$, such that

$$|\theta_l - a/q| \leq 1/q^2 \quad \text{and} \quad q \in [P^\varepsilon, P^{l-\varepsilon}]. \quad (2.20)$$

Then there is a constant $\bar{C} = \bar{C}_d \geq 1$ such that

$$\left| \sum_{n \in \mathbb{Z}} \phi_p(n) \epsilon(-(\theta_1 n + \dots + \theta_d n^d)) \right| \lesssim_\epsilon P^{1-\epsilon/\bar{C}}. \quad (2.21)$$

(ii) For any irreducible fraction $\theta = a/q \in (\mathbb{Z}/q)^d$, $a = (a_1, \dots, a_d) \in \mathbb{Z}^d$, $q \in \mathbb{Z}_+$, we have

$$\left| q^{-1} \sum_{n \in \mathbb{Z}_q} \epsilon(-(\theta_1 n + \dots + \theta_d n^d)) \right| \lesssim q^{-1/\bar{C}}. \quad (2.22)$$

Notice a formal similarity between Propositions 2.1 and 2.3. They both involve a small but nontrivial gain of a power of P as soon as one of the coefficients of the relevant polynomials is far from rational numbers with small denominators. These estimates can therefore be used efficiently to estimate minor arcs contributions.

We note, however, that the proof of the nilpotent Weyl estimates in Proposition 2.1 is much more involved than the proof of Proposition 2.3. It relies on some classical ideas of Birch [9] and Davenport [21,22] on treating polynomials in many variables, but one has to identify and exploit suitable nondegeneracy properties of the explicit (but complicated) polynomials D and \tilde{D} in (2.11)–(2.12) to make the proof work. All the details of the proof are provided in [37, Section 5].

3 A nilpotent circle method

To illustrate our main method, we focus on a particular case of Theorem 1.3, namely, on proving boundedness of the maximal function M_N^χ on $\ell^2(\mathbb{G}_0)$. For simplicity of notation, for $k \in \mathbb{N}$ and $x \in \mathbb{G}_0$, let

$$\begin{aligned} \mathcal{M}_k f(x) &= M_{2^k}^\chi f(x) = \sum_{n \in \mathbb{Z}} 2^{-k} \chi(2^{-k} n) f(A_0(n)^{-1} \cdot x) = (f * K_k)(x), \\ K_k(x) &= G_{2^k}^\chi(x) = \sum_{n \in \mathbb{Z}} 2^{-k} \chi(2^{-k} n) \mathbb{1}_{\{A_0(n)\}}(x), \end{aligned} \quad (3.1)$$

see (1.13) and (1.15) for the definitions M_N^χ and G_N^χ , respectively. With this new notation, our main goal is to prove the following:

Theorem 3.1. *For any $f \in \ell^2(\mathbb{G}_0)$, we have*

$$\left\| \sup_{k \geq 0} |\mathcal{M}_k f| \right\|_{\ell^2(\mathbb{G}_0)} \lesssim \|f\|_{\ell^2(\mathbb{G}_0)}. \quad (3.2)$$

In the rest of this section, we outline the proof of this theorem. Our main new construction is an iterative procedure, starting from the center of the group and moving down along its central series, which allows us to use some of the ideas of the classical circle method recursively at every stage. In our case of nilpotent groups of step two, the procedure consists of two basic stages and one additional step corresponding to “major arcs.”

Notice that the kernels K_k have a product structure

$$K_k(g) = L_k(g^{(1)}) \mathbb{1}_{\{0\}}(g^{(2)}), \quad L_k(g^{(1)}) = \sum_{n \in \mathbb{Z}} 2^{-k} \chi(2^{-k} n) \mathbb{1}_{\{0\}}(g^{(1)} - A_0^{(1)}(n)), \quad (3.3)$$

where $A_0^{(1)}(n) = (n, \dots, n^d) \in \mathbb{Z}^d$ and $g = (g^{(1)}, g^{(2)}) \in \mathbb{G}_0$ as in (1.7).

3.1 First stage reduction

We first decompose the singular kernel $\mathbb{1}_{\{0\}}(g^{(2)})$ in the central variable $g^{(2)}$ into smoother kernels. For any $s \in \mathbb{N}$ and $m \in \mathbb{Z}_+$, we define the set of rational fractions

$$\mathcal{R}_s^m = \{a/q : a = (a_1, \dots, a_m) \in \mathbb{Z}^m, q \in [2^s, 2^{s+1}) \cap \mathbb{Z}, \gcd(a_1, \dots, a_m, q) = 1\}. \quad (3.4)$$

We define also $\mathcal{R}_{\leq a}^m := \cup_{0 \leq s \leq a} \mathcal{R}_s^m$. For $x^{(1)} = (x_{h_0}^{(1)})_{h_1 \in \{1, \dots, d\}} \in \mathbb{R}^d$, $x^{(2)} = (x_{h_1 h_2}^{(2)})_{(h_1, h_2) \in Y_d'} \in \mathbb{R}^{d'}$ and $\Lambda \in (0, \infty)$, we define the partial dilations:

$$\Lambda \circ x^{(1)} = (\Lambda^{l_1} x_{h_0}^{(1)})_{h_1 \in \{1, \dots, d\}} \in \mathbb{R}^d, \quad \Lambda \circ x^{(2)} = (\Lambda^{l_1 + l_2} x_{h_1 h_2}^{(2)})_{(h_1, h_2) \in Y_d'} \in \mathbb{R}^{d'}, \quad (3.5)$$

which are induced by the group-dilations defined in (1.12).

We fix $\eta_0 : \mathbb{R} \rightarrow [0, 1]$ a smooth even function such that $1_{[-1, 1]} \leq \eta_0 \leq 1_{[-2, 2]}$. For $t \in \mathbb{R}$ and integers $j \geq 1$, we define

$$\eta_j(t) := \eta_0(2^{-j}t) - \eta_0(2^{-j+1}t), \quad 1 = \sum_{j=0}^{\infty} \eta_j. \quad (3.6)$$

For any $A \in [0, \infty)$, we define

$$\eta_{\leq A} := \sum_{j \in [0, A] \cap \mathbb{Z}} \eta_j. \quad (3.7)$$

By a slight abuse of notation, we also let η_j and $\eta_{\leq A}$ denote the smooth radial functions on \mathbb{R}^m , $m \geq 1$, defined by $\eta_j(x) = \eta_j(|x|)$ and $\eta_{\leq A}(x) = \eta_{\leq A}(|x|)$. We fix also two small constants $\delta = \delta(d) \ll \delta' = \delta'(d)$ such that $\delta' \in (0, (10d)^{-10}]$ and $\delta \in (0, (\delta')^4]$, and a large constant $D = D(d) \gg \delta^{-8}$, which depend on arithmetic properties of the polynomial sequence A_0 (more precisely on the structural constants in Propositions 2.1–2.2) such that

$$1 \ll 1/\delta' \ll 1/\delta \ll r = r(\delta, \delta', d) \ll D. \quad (3.8)$$

For $k \geq D^2$, we fix two cutoff functions $\phi_k^{(1)} : \mathbb{R}^d \rightarrow [0, 1]$, $\phi_k^{(2)} : \mathbb{R}^{d'} \rightarrow [0, 1]$, such that

$$\phi_k^{(1)}(g^{(1)}) := \eta_{\leq \delta k}(2^{-k} \circ g^{(1)}), \quad \phi_k^{(2)}(g^{(2)}) := \eta_{\leq \delta k}(2^{-k} \circ g^{(2)}). \quad (3.9)$$

For $k \in \mathbb{N}$ so that $k \geq D^2$ and for any 1-periodic sets of rationals $\mathcal{A} \subseteq \mathbb{Q}^d$, $\mathcal{B} \subseteq \mathbb{Q}^{d'}$, we define the periodic Fourier multipliers by

$$\begin{aligned} \Psi_{k, \mathcal{A}}(\xi^{(1)}) &:= \sum_{a/q \in \mathcal{A}} \eta_{\leq \delta' k}(2^k \circ (\xi^{(1)} - a/q)), \quad \xi^{(1)} \in \mathbb{T}^d, \\ \Xi_{k, \mathcal{B}}(\xi^{(2)}) &:= \sum_{b/q \in \mathcal{B}} \eta_{\leq \delta k}(2^k \circ (\xi^{(2)} - b/q)), \quad \xi^{(2)} \in \mathbb{T}^{d'}. \end{aligned} \quad (3.10)$$

For $k \geq D^2$ and $s \in [0, \delta k] \cap \mathbb{Z}$, we define the periodic Fourier multipliers $\Xi_{k, s} : \mathbb{R}^{d'} \rightarrow [0, 1]$,

$$\Xi_{k, s}(\xi^{(2)}) := \Xi_{k, \mathcal{R}_s^{d'}}(\xi^{(2)}) = \sum_{a/q \in \mathcal{R}_s^{d'}} \eta_{\leq \delta k}(2^k \circ (\xi^{(2)} - a/q)). \quad (3.11)$$

For $k \geq D^2$, we write

$$\begin{aligned} \mathbf{1}_{\{0\}}(g^{(2)}) &= \int_{\mathbb{T}^{d'}} e(g^{(2)} \cdot \xi^{(2)}) d\xi^{(2)} \\ &= \sum_{s \in [0, \delta k] \cap \mathbb{Z}} \int_{\mathbb{T}^{d'}} e(g^{(2)} \cdot \xi^{(2)}) \Xi_{k, s}(\xi^{(2)}) d\xi^{(2)} + \int_{\mathbb{T}^{d'}} e(g^{(2)} \cdot \xi^{(2)}) \Xi_k^c(\xi^{(2)}) d\xi^{(2)}, \end{aligned} \quad (3.12)$$

where

$$\Xi_k^c := 1 - \sum_{s \in [0, \delta k] \cap \mathbb{Z}} \Xi_{k, s}. \quad (3.13)$$

Then we decompose $K_k = K_k^c + \sum_{s \in [0, \delta k] \cap \mathbb{Z}} K_{k, s}$, where, with the notation in (3.3), we have

$$K_{k, s}(g) := L_k(g^{(1)}) N_{k, s}(g^{(2)}), \quad K_k^c(g) := L_k(g^{(1)}) N_k^c(g^{(2)}), \quad (3.14)$$

and

$$\begin{aligned} N_{k, s}(g^{(2)}) &:= \phi_k^{(2)}(g^{(2)}) \int_{\mathbb{T}^{d'}} e(g^{(2)} \cdot \xi^{(2)}) \Xi_{k, s}(\xi^{(2)}) d\xi^{(2)}, \\ N_k^c(g^{(2)}) &:= \phi_k^{(2)}(g^{(2)}) \int_{\mathbb{T}^{d'}} e(g^{(2)} \cdot \xi^{(2)}) \Xi_k^c(\xi^{(2)}) d\xi^{(2)}. \end{aligned} \quad (3.15)$$

We first show that we can bound the contributions of the minor arcs in the central variables.

Lemma 3.2. *For any integer $k \geq D^2$ and $f \in \ell^2(\mathbb{G}_0)$, we have*

$$\|f * K_k^c\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-k/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \quad (3.16)$$

Then we prove our first transition estimate, i.e., we show that we can bound the contributions of the kernels $K_{k,s}$ corresponding to scales $k \geq 0$ not very large. More precisely, for any $s \geq 0$, we define

$$\kappa_s = 2^{2D(s+1)^2}. \quad (3.17)$$

Lemma 3.3. *For any integer $s \geq 0$ and $f \in \ell^2(\mathbb{G}_0)$, we have*

$$\left\| \sup_{\max(D^2, s/\delta) \leq k < 2\kappa_s} |f * K_{k,s}| \right\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-s/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \quad (3.18)$$

In the commutative setting, minor arcs estimates such as (3.16) follow using Weyl estimates and the Plancherel theorem. As we do not have a useful Fourier transform on the group \mathbb{G}_0 , our main tool to prove the bounds (3.16) is a high-order T^*T argument. More precisely, we analyze the kernel of the convolution operator $\{(\mathcal{K}_k^c)^* \mathcal{K}_k^c\}^r$, where $\mathcal{K}_k^c f = f * K_k^c$ and r is sufficiently large, and show that its $\ell^1(\mathbb{G}_0)$ norm is $\lesssim 2^{-k}$. The main ingredient in this proof is the noncommutative Weyl estimate in Proposition 2.1 (i).

To prove the transition estimates (3.18), we use the Rademacher-Menshov inequality and Khintchine's inequality (leading to logarithmic losses) to reduce to proving the bounds

$$\left\| \sum_{k \in [J, 2J]} \mathfrak{u}_k(f * H_{k,s}) \right\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-4s/D^2} \|f\|_{\ell^2(\mathbb{G}_0)} \quad (3.19)$$

for any $J \geq \max(D^2, s/\delta)$ and any coefficients $\mathfrak{u}_k \in [-1, 1]$, where $H_{k,s} = K_{k+1,s} - K_{k,s}$. For this, we use a high order version of the Cotlar-Stein lemma, which relies again on precise analysis of the kernel of the convolution operator $\{(\mathcal{H}_{k,s})^* \mathcal{H}_{k,s}\}^r$, where $\mathcal{H}_{k,s} f = f * H_{k,s}$ and r is sufficiently large. The key exponential gain of $2^{-4s/D^2}$ in (3.19) is due to the noncommutative Gauss sums estimate, see Proposition 2.1 (ii).

3.2 Second stage reduction

In view of Lemmas 3.2–3.3, it remains to prove that

$$\left\| \sup_{k \geq \kappa_s} |f * K_{k,s}| \right\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-s/D^2} \|f\|_{\ell^2(\mathbb{G}_0)} \quad (3.20)$$

for any fixed integer $s \geq 0$. The kernels $K_{k,s}$ are now reasonably well adapted to a natural family of nonisotropic balls in the central variables, at least when $2^s \approx 1$, and we need to start decomposing in the noncentral variables.

We examine the kernels $L_k(g^{(1)})$ defined in (3.3) and rewrite them in the form

$$\begin{aligned} L_k(g^{(1)}) &= \sum_{n \in \mathbb{Z}} 2^{-k} \chi(2^{-k}n) \mathbb{1}_{\{0\}}(-A_0^{(1)}(n) + g^{(1)}) \\ &= \phi_k^{(1)}(g^{(1)}) \int_{\mathbb{T}^d} \epsilon(g^{(1)} \cdot \xi^{(1)}) S_k(\xi^{(1)}) d\xi^{(1)}, \end{aligned} \quad (3.21)$$

where $g^{(1)} \cdot \xi^{(1)}$ denotes the usual scalar product of vectors in \mathbb{R}^d , and

$$S_k(\xi^{(1)}) = \sum_{n \in \mathbb{Z}} 2^{-k} \chi(2^{-k}n) \epsilon(-A_0^{(1)}(n) \cdot \xi^{(1)}). \quad (3.22)$$

For any integers $Q \in \mathbb{Z}_+$ and $m \in \mathbb{Z}_+$, we define the set of fractions

$$\widetilde{\mathcal{R}}_Q^m = \{a/Q : a = (a_1, \dots, a_m) \in \mathbb{Z}^m\}. \quad (3.23)$$

For any integer $s \geq 0$, we fix a large denominator

$$Q_s := (\lfloor 2^{D(s+1)} \rfloor)! = 1 \cdot 2 \cdot \dots \cdot \lfloor 2^{D(s+1)} \rfloor, \quad (3.24)$$

and using (3.10) define the periodic multipliers

$$\begin{aligned} \Psi_{k,s}^{\text{low}}(\xi^{(1)}) &= \Psi_{k, \widetilde{\mathcal{R}}_{Q_s}^d}(\xi^{(1)}) = \sum_{a/q \in \widetilde{\mathcal{R}}_{Q_s}^d} \eta_{\leq \delta'k}(2^k \circ (\xi^{(1)} - a/q)), \\ \Psi_{k,s,t}(\xi^{(1)}) &= \Psi_{k, \mathcal{R}_t^d \setminus \widetilde{\mathcal{R}}_{Q_s}^d}(\xi^{(1)}) = \sum_{a/q \in \mathcal{R}_t^d \setminus \widetilde{\mathcal{R}}_{Q_s}^d} \eta_{\leq \delta'k}(2^k \circ (\xi^{(1)} - a/q)), \\ \Psi_k^c(\xi^{(1)}) &= 1 - \Psi_{k,s}^{\text{low}}(\xi^{(1)}) - \sum_{t \in [0, \delta'k] \cap \mathbb{Z}} \Psi_{k,s,t}(\xi^{(1)}) = 1 - \sum_{a/q \in \mathcal{R}_{\leq \delta'k}^d} \eta_{\leq \delta'k}(\tau^k \circ (\xi^{(1)} - a/q)). \end{aligned} \quad (3.25)$$

Since $k \geq \kappa_s = 2^{2D(s+1)^2}$, we see that $Q_s \leq 2^{\delta^2 k}$. Therefore, the supports of the cutoff functions $\eta_{\leq \delta'k}(2^k \circ (\xi^{(1)} - a/q))$ are all disjoint, and the multipliers $\Psi_{k,s}^{\text{low}}$, $\Psi_{k,s,t}$, Ψ_k^c take values in the interval $[0, 1]$. Notice also that $\Psi_{k,s,t} \equiv 0$ unless $t \geq D(s+1)$ and that the cutoffs used in these definitions depend on $\delta'k$ not on δk as in the case of the central variables.

We examine formula (3.21) and define the kernels $L_{k,s}^{\text{low}}, L_{k,s,t}, L_k^c : \mathbb{Z}^d \rightarrow \mathbb{C}$ by

$$L_*(g^{(1)}) = \phi_k^{(1)}(g^{(1)}) \int_{\mathbb{T}^d} \mathfrak{e}(g^{(1)} \cdot \xi^{(1)}) S_k(\xi^{(1)}) \Psi_*(\xi^{(1)}) d\xi^{(1)}, \quad (3.26)$$

where $(L_*, \Psi_*) \in \{(L_{k,s}^{\text{low}}, \Psi_{k,s}^{\text{low}}), (L_{k,s,t}, \Psi_{k,s,t}), (L_k^c, \Psi_k^c)\}$. For any $k \geq \kappa_s$, we obtain $K_{k,s} = G_{k,s}^{\text{low}} + \sum_{t \leq \delta'k} G_{k,s,t} + G_{k,s}^c$, where the kernels $G_{k,s}^{\text{low}}, G_{k,s,t}, G_{k,s}^c : \mathbb{Z}^{|Y_d|} \rightarrow \mathbb{C}$ are defined by

$$\begin{aligned} G_{k,s}^{\text{low}}(g) &= L_{k,s}^{\text{low}}(g^{(1)}) N_{k,s}(g^{(2)}), \\ G_{k,s,t}(g) &= L_{k,s,t}(g^{(1)}) N_{k,s}(g^{(2)}), \\ G_{k,s}^c(g) &= L_k^c(g^{(1)}) N_{k,s}(g^{(2)}). \end{aligned} \quad (3.27)$$

Our next step is to show that the contributions of the minor arcs corresponding to the kernels $G_{k,s}^c$ can be suitably bounded:

Lemma 3.4. *For any integers $s \geq 0$ and $k \geq \kappa_s$, and for any $f \in \ell^2(\mathbb{G}_0)$, we have*

$$\|f * G_{k,s}^c\|_{\ell^2(\mathbb{G}_0)} \leq 2^{-k/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \quad (3.28)$$

Then we prove our second transition estimate, bounding the contributions of the operators defined by the kernels $G_{k,s,t}$ for intermediate values of k .

Lemma 3.5. *For any integers $s \geq 0$, and $t \geq D(s+1)$, and $f \in \ell^2(\mathbb{G}_0)$, we have*

$$\left\| \sup_{\max(\kappa_s, t/\delta') \leq k < 2\kappa_t} f * G_{k,s,t} \right\|_{\ell^2(\mathbb{G}_0)} \leq 2^{-t/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}, \quad (3.29)$$

where $\kappa_t = 2^{2D(t+1)^2}$ as in (3.17).

The proofs of these estimates are similar to the proofs of the corresponding first stage estimates (3.16) and (3.18), using high-order T^*T arguments. However, instead of using the nilpotent oscillatory sums estimates in Proposition 2.1, we use the classical estimates from Proposition 2.3 here. We emphasize, however, that the underlying nilpotent structure is very important and that these estimates are only possible after performing the two reductions in the first stage, namely, the restriction to major arcs corresponding to denominators $\approx 2^s$ and the restriction to parameters $k \geq \kappa_s$. We finally remark that the circle method could not have been applied simultaneously to both central and noncentral variables, as we would not have been able to control efficiently the phase functions arising in the corresponding exponential sums and oscillatory integrals, especially on major arcs.

3.3 Final stage: major arcs contributions

After these reductions, it remains to bound the contributions of the “major arcs” in both the central and the noncentral variables. More precisely, we prove the following bounds:

Lemma 3.6. (i) For any integer $s \geq 0$ and $f \in \ell^2(\mathbb{G}_0)$, we have

$$\left\| \sup_{k \geq \kappa_s} |f * G_{k,s}^{\text{low}}| \right\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-s/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \quad (3.30)$$

(ii) For any integers $s \geq 0$, $t \geq D(s+1)$, and $f \in \ell^2(\mathbb{G}_0)$, we have

$$\left\| \sup_{k \geq \kappa_t} |f * G_{k,s,t}| \right\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-t/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \quad (3.31)$$

The main idea here is different: we write the kernels $G_{k,s}^{\text{low}}$ and $G_{k,s,t}$ as tensor products of two components up to acceptable errors. One of these components is essentially a maximal average operator on a continuous group, which can be analyzed using the classical method of Christ [17]. The other component is an arithmetic operator-valued analog of the classical Gauss sums, which leads to the key factors $2^{-s/D^2}$ and $2^{-t/D^2}$ in (3.30) and (3.31).

More precisely, for any integer $Q \geq 1$, we define the subgroup

$$\mathbb{H}_Q = \{h = (Qh_{l_1 l_2})_{(l_1, l_2) \in Y_d} \in \mathbb{G}_0 : h_{l_1, l_2} \in \mathbb{Z}\}. \quad (3.32)$$

Clearly, $\mathbb{H}_Q \subseteq \mathbb{G}_0$ is a normal subgroup. Let \mathbb{J}_Q denote the coset

$$\mathbb{J}_Q = \{b = (b_{l_1 l_2})_{(l_1, l_2) \in Y_d} \in \mathbb{G}_0 : b_{l_1, l_2} \in \mathbb{Z} \cap [0, Q-1]\}, \quad (3.33)$$

with the natural induced group structure. Notice that

$$\text{the map } (b, h) \mapsto b \cdot h \text{ defines a bijection from } \mathbb{J}_Q \times \mathbb{H}_Q \text{ to } \mathbb{G}_0. \quad (3.34)$$

Assume that $Q \geq 1$ and $2^k \geq Q$. For any $a \in \mathbb{Z}^d$ and $\xi \in \mathbb{R}^d$, let

$$\begin{aligned} J_k(\xi) &= 2^{-k} \int_{\mathbb{R}} \chi(2^{-k}x) \epsilon[-A_0^{(1)}(x) \cdot \xi] dx = \int_{\mathbb{R}} \chi(y) \epsilon[-A_0^{(1)}(y) \cdot (2^k \circ \xi)] dy, \\ S(a/Q) &= Q^{-1} \sum_{n \in \mathbb{Z}_Q} \epsilon[-A_0^{(1)}(n) \cdot a/Q]. \end{aligned} \quad (3.35)$$

The point is that the kernels $G_{k,s}^{\text{low}}$ and $G_{k,s,t}$ can be decomposed as tensor products. Indeed, to decompose $G_{k,s,t}$ (the harder case), we set $Q = Q_t = (\lfloor 2^{D(t+1)} \rfloor)!$ as in (3.24). Then we show that if $k \geq \kappa_t$ (so $2^k \gg Q_t^4$), $h \in \mathbb{H}_{Q_t}$ and $b_1, b_2 \in \mathbb{G}_0$ satisfy $|b_1| + |b_2| \leq Q^4$, then

$$G_{k,s,t}(b_1 \cdot h \cdot b_2) \approx W_{k,Q_t}(h) V_{\mathcal{R}_t^d \setminus \overline{\mathcal{R}}_{Q_t^d}^d, \mathcal{R}_s^{d'}, Q_t}(b_1 \cdot b_2), \quad (3.36)$$

up to acceptable summable errors. Here,

$$\begin{aligned} W_{k,Q}(h) &= Q^{d+d'} \phi_k(h) \int_{\mathbb{R}^d \times \mathbb{R}^{d'}} \eta_{\leq \delta' k}(2^k \circ \xi) \eta_{\leq \delta k}(2^k \circ \theta) \epsilon(h \cdot (\xi, \theta)) J_k(\xi) d\xi d\theta, \\ V_{\mathcal{A}, \mathcal{B}, Q}(b) &= Q^{-d-d'} \left[\sum_{\sigma^{(1)} \in \mathcal{A} \cap [0,1]^d} S(\sigma^{(1)}) \epsilon[b^{(1)} \cdot (\sigma^{(1)})] \right] \left[\sum_{\sigma^{(2)} \in \mathcal{B} \cap [0,1]^{d'}} \epsilon[b^{(2)} \cdot (\sigma^{(2)})] \right], \end{aligned}$$

and $\phi_k(h) = \phi_k^{(1)}(h^{(1)}) \phi_k^{(2)}(h^{(2)})$, $h = (h^{(1)}, h^{(2)}) \in \mathbb{H}_Q$, is defined in (3.9), $b = (b^{(1)}, b^{(2)}) \in \mathbb{G}_0$, and the functions J_k and S are defined in (3.35).

Finally, we show that the kernels $V_{s,t} = V_{\mathcal{R}_t^d \setminus \overline{\mathcal{R}}_{Q_t^d}^d, \mathcal{R}_s^{d'}, Q_t}$ (which can be interpreted as an operator-valued Gauss sums) define bounded operators on $\ell^2(\mathbb{J}_{Q_t})$,

$$\|f * \mathbb{J}_{Q_t} V_{s,t}\|_{\ell^2(\mathbb{J}_{Q_t})} \lesssim 2^{-t/D} \|f\|_{\ell^2(\mathbb{J}_{Q_t})}.$$

Moreover, the kernels W_{k,Q_t} are close to classical maximal operators and one can show that

$$\left\| \sup_{k \geq \kappa_t} |f * \mathbb{H}_{Q_t} W_{k,Q_t}| \right\|_{\ell^2(\mathbb{H}_{Q_t})} \lesssim \|f\|_{\ell^2(\mathbb{H}_{Q_t})}.$$

The desired bounds (3.31) follow using the approximation formula (3.36).

4 A nilpotent Waring theorem on the group \mathbb{G}_0 : proof of Theorem 1.4

We prove now Theorem 1.4. Observe that $D(n, m) = A_0(n_1)^{-1} \cdot A_0(m_1) \cdot \dots \cdot A_0(n_r)^{-1} \cdot A_0(m_r)$. By using the classical delta function, we can write

$$S_{r,N}(g) = \sum_{m,n \in [N]^r} \int_{\mathbb{T}^{d+d'}} \epsilon(D(n, m) \cdot \xi) \epsilon(-g \cdot \xi) d\xi. \quad (4.1)$$

Step 1. We start by decomposing the integration in ξ into major and minor arcs. For any integer $m \geq 1$ and any positive number $M > 0$, we define the set of rational fractions

$$\mathcal{R}_{\leq M}^m = \{a/q : a = (a_1, \dots, a_m) \in \mathbb{Z}^m, q \in [1, M] \cap \mathbb{Z}, \gcd(a_1, \dots, a_m, q) = 1\}. \quad (4.2)$$

Notice that we use a bit different definition of $\mathcal{R}_{\leq M}^m$ than in (3.4). We fix a small constant $\delta = \delta(d) \ll 1$ and a smooth radial function $\eta_0 : \mathbb{R}^{|\mathbb{Y}_d|} \rightarrow [0, 1]$ such that $\mathbb{1}_{|x| \leq 1} \leq \eta_0(x) \leq \mathbb{1}_{|x| \leq 2}$, $x \in \mathbb{R}^{|\mathbb{Y}_d|}$. For $A > 0$, let $\eta_{\leq A}(x) = \eta_0(A^{-1}x)$, $x \in \mathbb{R}^{|\mathbb{Y}_d|}$; here, we use a bit different definition of $\eta_{\leq A}$ than in (3.7). Then, we introduce the projections

$$\Xi_N(\xi) = \sum_{\substack{a/q \in \mathcal{R}_{\leq N^\delta}^{d+d'} \\ \leq N^\delta}} \eta_{\leq N^\delta}(N \circ (\xi - a/q)), \quad \xi \in \mathbb{T}^{d+d'}, \quad N \in \mathbb{N}, \quad (4.3)$$

and decompose the integration in (4.1) into major and minor arcs, i.e., we define

$$S_{r,N,\text{maj}}(g) = \sum_{m,n \in [N]^r} \int_{\mathbb{T}^{d+d'}} \epsilon(D(n, m) \cdot \xi) \epsilon(-g \cdot \xi) \Xi_N(\xi) d\xi, \quad (4.4)$$

$$S_{r,N,\text{min}}(g) = \sum_{m,n \in [N]^r} \int_{\mathbb{T}^{d+d'}} \epsilon(D(n, m) \cdot \xi) \epsilon(-g \cdot \xi) (1 - \Xi_N(\xi)) d\xi. \quad (4.5)$$

Notice that $S_{r,N}(g) = S_{r,N,\text{min}}(g) + S_{r,N,\text{maj}}(g)$. Moreover,

$$|S_{r,N,\text{min}}(g)| \lesssim_r N^{2r-1} \left(\prod_{(l_1, l_2) \in \mathbb{Y}_d} N^{-|l_1| - |l_2|} \right), \quad N \in \mathbb{N}, \quad g \in \mathbb{G}_0, \quad (4.6)$$

provided that r is sufficiently large, as a consequence of Proposition 2.1 (i) and the Dirichlet principle; in fact, we use Proposition 2.1(i) with $\phi_N^{(j)} = \psi_N^{(j)} = \mathbb{1}_{[N]}$, $1 \leq j \leq r$, which is still valid as can be seen by a careful reading of the proof of this result contained in [37]. Therefore, the contribution of the minor arcs $S_{r,N,\text{min}}(g)$ can be absorbed by the error term in (1.25).

Step 2. Next, we deal with the major arcs contributions. Notice that

$$S_{r,N,\text{maj}}(g) = \sum_{\substack{a/q \in \mathcal{R}_{\leq N^\delta}^{d+d'} \\ \leq N^\delta}} \epsilon(-g \cdot a/q) \int_{\mathbb{R}^{d+d'}} \eta_{\leq N^\delta}(N \circ \xi) I_{r,N,a/q}(\xi) \epsilon(-g \cdot \xi) d\xi, \quad (4.7)$$

where

$$I_{r,N,a/q}(\xi) = \sum_{m,n \in [N]^r} \epsilon(D(n, m) \cdot (a/q)) \epsilon(D(n, m) \cdot \xi). \quad (4.8)$$

Observe that for $a/q \in \mathcal{R}_{\leq N^\delta}^{d+d'} \cap [0, 1)^{d+d'}$ and $|N \circ \xi| \leq N^\delta$, we have

$$\begin{aligned} I_{r,N,a/q}(\xi) &= \sum_{m,n \in [N/q]^r} \sum_{u,v \in \mathbb{Z}_q^r} \epsilon(D(v, w) \cdot (a/q)) \epsilon(D(qn, qm) \cdot \xi) + O(qN^{2r-1+\delta}) \\ &= N^{2r} \overline{G(a/q)} \Phi(N \circ \xi) + O(qN^{2r-1+\delta}), \end{aligned}$$

where $G(a/q)$ is defined in (2.16) and Φ is defined in (1.27).

Therefore, if $\delta \leq (10d)^{-4}$, then we have

$$\begin{aligned} S_{r,N,\text{maj}}(g) &= N^{2r} \left(\prod_{(l_1, l_2) \in Y_d} N^{-|l_1| - |l_2|} \right) \\ &\quad \times \left[\sum_{a/q \in \mathcal{R}_{\leq N^\delta}^{d+d'} \cap [0, 1)^{d+d'}} \overline{G(a/q)} \epsilon(-g \cdot a/q) \int_{\mathbb{R}^{d+d'}} \eta_{\leq N^\delta}(\xi) \Phi(\xi) \epsilon(-g \cdot (N^{-1} \circ \xi)) d\xi + O_r(N^{-1/2}) \right]. \end{aligned} \quad (4.9)$$

It follows from Proposition 2.1(ii) and Proposition 2.2 that

$$|G(a/q)| \lesssim_r q^{-1/\delta^2}, \quad (a, q) = 1, \quad (4.10)$$

and

$$|\Phi(\zeta)| \lesssim_r \langle \zeta \rangle^{-1/\delta^2}, \quad \zeta \in \mathbb{R}^{d+d'}, \quad (4.11)$$

provided that r is sufficiently large. Therefore, recalling the definition (1.26),

$$\begin{aligned} |\mathfrak{S}(g)| &\lesssim_r 1, \\ \left| \mathfrak{S}(g) - \sum_{a/q \in \mathcal{R}_{\leq N^\delta}^{d+d'} \cap [0, 1)^{d+d'}} \overline{G(a/q)} \epsilon(-g \cdot a/q) \right| &\lesssim_r \sum_{q \geq N^\delta} q^{d+d'-1/\delta^2} \lesssim_r N^{-1/(2\delta)}. \end{aligned} \quad (4.12)$$

Moreover, we have

$$\begin{aligned} \left| \int_{\mathbb{R}^{d+d'}} \eta_{\leq N^\delta}(\xi) \Phi(\xi) \epsilon(-g \cdot (N^{-1} \circ \xi)) d\xi \right| &\lesssim_r 1, \\ \left| \int_{\mathbb{R}^{d+d'}} \eta_{\leq N^\delta}(\xi) \Phi(\xi) \epsilon(-g \cdot (N^{-1} \circ \xi)) d\xi - \int_{\mathbb{R}^{d+d'}} \Phi(\xi) \epsilon(-g \cdot (N^{-1} \circ \xi)) d\xi \right| &\lesssim_r N^{-1/(2\delta)}. \end{aligned} \quad (4.13)$$

It follows from (4.9), (4.12), and (4.13) that

$$S_{r,N,\text{maj}}(g) = N^{2r} \left(\prod_{(l_1, l_2) \in Y_d} N^{-|l_1| - |l_2|} \right) \times \left[\mathfrak{S}(g) \int_{\mathbb{R}^{d+d'}} \Phi(\xi) \epsilon(-g \cdot (N^{-1} \circ \xi)) d\xi + O_r(N^{-1/2}) \right]. \quad (4.14)$$

The desired conclusion (1.25) follows using also (4.6). This completes the proof of part (i) of the theorem.

Step 3. We analyze now the singular series \mathfrak{S} defined in (1.26). Observe that

$$\mathfrak{S}(h) = \sum_{q \geq 1} A(q, h), \quad A(q, h) := \sum_{(a, q)=1} \overline{G(a/q)} \epsilon(-h \cdot a/q), \quad (4.15)$$

for any $h \in \mathbb{G}_0$. Notice that the sequence $A(q, h)$ is multiplicative in the sense that $A(q_1 q_2, h) = A(q_1, h) A(q_2, h)$ provided that $(q_1, q_2) = 1$ and $h \in \mathbb{G}_0$. Therefore, letting \mathbb{P} denote the set of primes,

$$\mathfrak{S}(h) = \prod_{p \in \mathbb{P}} B(p, h), \quad B(p, h) := 1 + \sum_{n \geq 1} A(p^n, h). \quad (4.16)$$

For $h \in \mathbb{G}_0$ and $q \geq 1$, let

$$M(q, h) := |\{(m, n) \in \mathbb{Z}_q^{2r} : D(n, m) = h \pmod{q}\}|. \quad (4.17)$$

We prove that for any $h \in \mathbb{G}_0$, $p \in \mathbb{P}$ and integer $n \geq 1$, we have

$$1 + \sum_{v=1}^n A(p^v, h) = \frac{M(p^n, h)}{p^{n(2r-d-d')}}. \quad (4.18)$$

Indeed, for any integer $q \geq 1$, we have

$$\begin{aligned} M(q, h) &= q^{-d-d'} \sum_{t \in \mathbb{Z}_q^{d+d'}} \sum_{m, n \in \mathbb{Z}_q^r} \epsilon((D(n, m) - h) \cdot (t/q)) \\ &= q^{-d-d'} \sum_{q_1 | q} \sum_{w \in \mathbb{Z}_{q/q_1}^{d+d'}, (w, q/q_1)=1} \sum_{m, n \in \mathbb{Z}_q^r} \epsilon((D(n, m) - h) \cdot (wq_1/q)) \\ &= q^{-d-d'} \sum_{q_2 | q} \sum_{w \in \mathbb{Z}_{q/q_2}^{d+d'}, (w, q_2)=1} \sum_{m, n \in \mathbb{Z}_q^r} \epsilon((D(n, m) - h) \cdot (w/q_2)) \\ &= q^{-d-d'} \sum_{q_2 | q} \sum_{w \in \mathbb{Z}_{q_2}^{d+d'}, (w, q_2)=1} q^{2r} \overline{G(w/q_2)} \epsilon(-h \cdot (w/q_2)) \\ &= q^{2r-d-d'} \sum_{q_2 | q} A(q_2, h). \end{aligned}$$

The identity (4.18) follows by applying this with $q = p^n$, $p \in \mathbb{P}$. In particular, $\mathfrak{S}(h)$ and $B(p, h)$ are real nonnegative numbers,

$$\mathfrak{S}(h), B(p, h) \in [0, \infty) \quad \text{for any } h \in \mathbb{G}_0, p \in \mathbb{P}. \quad (4.19)$$

Moreover, by using formulas (4.16) and (4.18), we obtain

$$B(p, h) = \frac{M(p^n, h)}{p^{n(2r-d-d')}} + O_r(2^{-n/\delta})$$

for any $n \geq 1$, $h \in \mathbb{G}_0$ and $p \in \mathbb{P}$. We would like to show now that $\mathfrak{S}(h) \gtrsim_r 1$ for all elements $h \in \mathbb{H}_Q$, in order to be able to exploit the expansion (1.25); Q will be defined below. We notice first that for any integer r sufficiently large, there is $p_0(r) \in \mathbb{P}$ such that

$$1/2 \leq \prod_{p \in \mathbb{P}, p \geq p_0(r)} B(p, h) \leq 3/2, \quad (4.20)$$

for any $h \in \mathbb{G}_0$, due to the rapid decay of the coefficients $G(a/q)$ in (4.10).

By Lemma 4.1, (and a comment after its statement) if r is even, then there is a point $a_0 = (z_0, w_0)$ such that $D(a_0) = 0$, and there is a $(d + d') \times (d + d')$ minor $J_D(a_0) \neq 0$. By re-indexing the variables, we may assume that this minor is $J_D(a_0) = \det \left(\frac{\partial D}{\partial x_j} (a_0) \right)_{i,j=K+1}^N$, writing $N = 2r$, $K = 2r - d - d'$. In other words, we may assume that the minor corresponding to the last $d + d'$ columns of the Jacobian matrix of D is nonsingular. For a given prime $p \leq p_0(r)$, let $\gamma_p \in \mathbb{N}$ be such that $J_D(a_0) = p^{\gamma_p} u$ with $u \in \mathbb{Z}_+$ and $p \nmid u$. Define $Q = Q(r) := \prod_{p \in \mathbb{P}, p \leq p_0(r)} p^{2\gamma_p+1}$.

For $h \in \mathbb{H}_Q$, we have $D(a_0) = h \pmod{p^{2\gamma_p+1}}$, but $J_D(a_0) \neq 0 \pmod{p^{\gamma_p+1}}$; thus, we are in the position to apply Hensel's lemma (Theorem 4.2) with $N = 2r$ and $K = 2r - d - d'$. Then by Corollary 4.3, we have that $M(p^n, h) \geq p^{(n-2\gamma_p-1)K}$ for all $n > 2\gamma_p$ and hence by (4.18) we have $B(p, h) \geq p^{-(2\gamma_p+1)K-1}$. This proves that

$$\mathfrak{S}(h) \gtrsim_r 1 \quad \text{uniformly for } h \in \mathbb{H}_Q. \quad (4.21)$$

Step 4. Finally, we analyze the contribution of the singular integral. Since

$$\int_{\mathbb{R}^{d+d'}} \Phi(\zeta) \epsilon(-(N^{-1} \circ g) \cdot \zeta) d\zeta = \int_{\mathbb{R}^{d+d'}} \Phi(\zeta) d\zeta + O_{r,g}(N^{-1}),$$

due to (4.11), to prove the approximate identity (1.29), it suffices to prove that

$$\int_{\mathbb{R}^{d+d'}} \Phi(\zeta) d\zeta \gtrsim_r 1. \quad (4.22)$$

We fix a smooth function $\chi : \mathbb{R}^{d+d'} \rightarrow [0, 1]$, satisfying $\chi(x) = 1$ if $|x| \leq 1/2$, $\chi(x) = 0$ if $|x| \geq 2$, and $\int_{\mathbb{R}^{d+d'}} \chi(x) dx = 1$. For $\varepsilon \leq \varepsilon(r)$ sufficiently small, we write

$$\int_{\mathbb{R}^{d+d'}} \Phi(\zeta) \widehat{\chi}(\varepsilon \zeta) d\zeta = \int_{[-1,1]^{2r}} \varepsilon^{-(d+d')} \chi(D(z, w)/\varepsilon) dz dw, \quad (4.23)$$

using the definition (1.27). In particular, by letting $\varepsilon \rightarrow 0$, $\int_{\mathbb{R}^{d+d'}} \Phi(\zeta) d\zeta$ is a real nonnegative number. Moreover, the lower bound (4.22) follows from (4.23) provided that we can show that there is a point $(z_0, w_0) \in [-1, 1]^{2r}$ such that

$$D(z_0, w_0) = 0 \quad \text{and} \quad \text{rank}[\nabla_{z,w} D(z_0, w_0)] = d + d'. \quad (4.24)$$

We notice that this with even r follows easily from Lemma 4.1.

Lemma 4.1. *Let $r \geq r_0(d)$. Then there exists $(n, m) \in \mathbb{Z}^{2r}$ such that*

$$\text{rank}[\nabla_{x,y} D(n, m)] = d + d'. \quad (4.25)$$

Indeed, writing $D_r(x, y) = D(x, y) : \mathbb{R}^{2r} \rightarrow \mathbb{G}_0^\#$, we have that $D_r(x, y) \cdot D_r(n, m)^{-1} = D_{2r}((x, m'), (y, n'))$ with $n' = (n_r, \dots, n_1)$, $m' = (m_r, \dots, m_1)$. Assuming (4.25), it is clear that the map $((x, m'), (y, n')) \mapsto D_{2r}((x, m'), (y, n'))$ has maximal rank at $z_0 = (n, m')$, $w_0 = (m, n')$ and (4.24) follows.

The proof of Lemma 4.1 is based on counting points $(n, m) \in [N]^{2r}$ at which the rank of the map $\nabla_{x,y} D$ drops. This was also crucial in obtaining the nilpotent Weyl estimate (2.14).

Proof of Lemma 4.1. Let N be sufficiently large with respect to r, d . It is enough to show that there is a constant $C_d > 0$ ($C_d = 2d(d + d')^2$ works here) such that

$$|\{n \in [N]^r : \text{rank}[\nabla_x D(n, m)] < d + d'\}| \lesssim_{d,r} N^{(r+1)/2 + C_d}, \quad (4.26)$$

holds uniformly for $m \in [N]^r$. Fix $m \in [N]^r$. If $\text{rank}[\nabla_x D(n, m)] < d + d'$, then by Cramer's rule, there exists $b_{l_1 l_2} \in \mathbb{Z}$, $|b_{l_1 l_2}| \leq N^{(2d-1)(d+d')}$ with $b_{l_1 l_2} \neq 0$ for at least one $0 \leq l_2 < l_1 \leq d$, such that

$$\sum_{0 \leq l_2 < l_1 \leq d} b_{l_1 l_2} [\partial_j D(n, m)]_{l_1 l_2} = 0 \quad \text{for all } 1 \leq j \leq r. \quad (4.27)$$

From (2.11), we have that $[\partial_j D(n, m)]_{l_1 0} = -l_1 n_j^{l_1-1}$, and for $1 \leq l_2$,

$$[\partial_j D(n, m)]_{l_1 l_2} = l_1 n_j^{l_1-1} \sum_{k>j} (n_k^{l_2} - m_k^{l_2}) + l_2 n_j^{l_2-1} \sum_{k<j} (n_k^{l_1} - m_k^{l_1}) - l_1 n_j^{l_1-1} m_j^{l_2} + (l_1 + l_2) n_j^{l_1+l_2-1}. \quad (4.28)$$

We want to only include terms $k \leq j$ and to achieve that we introduce the parameters

$$T_l = T_l(n, m) = \sum_{k=1}^r (n_k^l - m_k^l), \quad \text{for } 1 \leq l < d.$$

Note that $T_l \in [-2rN^{d-1}, 2rN^{d-1}]$. For fixed $T = (T_l)_{1 \leq l < d}$, write

$$\sum_{k>j} (n_k^{l_2} - m_k^{l_2}) = T_{l_2}(n, m) - \sum_{k \leq j} (n_k^{l_2} - m_k^{l_2}),$$

Substituting into (4.28), we obtain, up to lower degree terms in the variables $n = (n_1, \dots, n_r)$,

$$[\partial_j D(n, m)]_{l_1 l_2} = -l_1 \sum_{k \leq j} n_j^{l_1-1} n_k^{l_2} + l_2 \sum_{k < j} n_j^{l_2-1} n_k^{l_1} + (l_1 + l_2) n_j^{l_1+l_2-1}, \quad (4.29)$$

for $1 \leq l_2 < d$. Thus, the system in (4.27) takes the form

$$\sum_{0 \leq l_2 < l_1 \leq d} b_{l_1 l_2} P_{l_1 l_2}^{j,T,m}(n_1, \dots, n_j) = 0, \quad 1 \leq j \leq r. \quad (4.30)$$

Notice that for fixed n_1, \dots, n_{2j-2} with $j \leq r/2$, the left-hand side of (4.30) with j replaced by $2j$ contains the monomials $-b_{l_1 0} l_1 n_{2j}^{l_1-1}$ and $b_{l_1 l_2} l_2 n_{2j}^{l_2-1} n_{2j-1}^{l_1}$ and hence is nonvanishing in the variables n_{2j-1}, n_{2j} . This, thanks to [37, Lemma 5.3], implies that number of solutions to (4.30) is at most $2(d + d')(N + 1)$ in the variables n_{2j-1}, n_{2j} .

As the number of choices for parameters $b = (b_{l_1 l_2})_{0 \leq l_2 < l_1 \leq d}$ and $T = (T_l)_{1 \leq l < d}$ is $\lesssim_{r,d} N^{C_d}$ (with, say $C_d = 2d(d + d^2)$), (4.25) follows. \square

We remark that (4.25) together with the argument proving (4.24) also implies that the map $D_{2r} : \mathbb{R}^{4r} \rightarrow \mathbb{G}_0^\#$ is surjective. Indeed, the image of the map D_r must contain an open ball $B(g, \delta)$; thus, the image of D_{2r} must contain an open ball $B(0, \delta') \subseteq B(g, \delta)B(g, \delta)^{-1}$ centered at the origin, then by homogeneity the whole space $\mathbb{G}_0^\#$.

Lemma 4.1 together with Hensel's lemma is also crucial to show the nonvanishing of the singular series $\mathfrak{S}(h)$ for $h \in \mathbb{H}_Q$. Recall that, given a prime p , the ring of p -adic integers $\widehat{\mathbb{Z}}_p$ is defined as the completion of \mathbb{Z} with respect to the p -adic metric $|m|_p = e^{-k}$, if $m = p^k u$ with $u \in \mathbb{Z}$, $p \nmid u$. Then $\widehat{\mathbb{Z}}_p$ is a so-called complete valuation ring with a unique maximal ideal $I_p = p \widehat{\mathbb{Z}}_p$ and we will write $x = 0 \pmod{p^k}$ if $x \in p^k \widehat{\mathbb{Z}}_p$. We have $|x - y|_p \leq \max\{|x - z|_p, |y - z|_p\}$, and hence, a sequence $(x_j)_{j \in \mathbb{N}}$ is Cauchy if $|x_{j+1} - x_j|_p \rightarrow 0$ as $j \rightarrow \infty$.

It follows that any formal power series $g(x)$ converges at x whenever $x = 0 \pmod{p}$. For a vector $x = (x_1, \dots, x_N) \in \widehat{\mathbb{Z}}_p^N$, we say that $x = 0 \pmod{p^k}$ if $x_j = 0 \pmod{p^k}$, for all $1 \leq j \leq N$. Then any power series $g(x) = g(x_1, \dots, x_N)$ in N variables also converges whenever $x = 0 \pmod{p}$. Moreover, one has the inverse and implicit function theorems for power series maps $g(x) = (g_1(x), \dots, g_N(x)) : \widehat{\mathbb{Z}}_p^N \rightarrow \widehat{\mathbb{Z}}_p^N$ without constant terms. Namely, if the Jacobian of the system at origin $J_g(0) \notin I_p$, i.e., is a unit, then g has an inverse power series map $h(x) = (h_1(x), \dots, h_N(x))$, in the sense that $h(g(x)) = g(h(x)) = x$, see [29, Proposition 5.19]. One also has a corresponding version of the implicit function theorem; for a map $g(x) = (g_{K+1}(x), \dots, g_N(x)) : \widehat{\mathbb{Z}}_p^N \rightarrow \widehat{\mathbb{Z}}_p^{N-K}$ such that $\det \left(\frac{\partial g_i}{\partial x_j}(0) \right)_{i,j=K+1}^N \notin I_p$, the inverse image $V_g := g^{-1}(0)$ can be parameterized as $V_g = (t_1, \dots, t_K, h_{K+1}(t), \dots, h_N(t))$ with $t = (t_1, \dots, t_K)$, which can be seen from the inverse function theorem by extending the map with $g_i(x) = x_i$ for $i = 1, \dots, K$. The following extension of the implicit function is often used to show the nonvanishing of the singular series associated to diophantine systems.

Theorem 4.2. (Hensel's lemma) *Let $f = (f_{K+1}, \dots, f_N) : \widehat{\mathbb{Z}}_p^N \rightarrow \widehat{\mathbb{Z}}_p^{N-K}$ be a family of polynomials. Assume there exists an $a \in \widehat{\mathbb{Z}}_p^N$ and an integer $\gamma > 0$, such that*

$$f(a) = 0 \pmod{p^{2\gamma+1}}, \quad (4.31)$$

moreover

$$J_f(a) = p^\gamma u, \quad u \neq 0 \pmod{p}, \quad (4.32)$$

where $J_f(a)$ is the Jacobian,

$$J_f(a) = \det \left(\frac{\partial f_i}{\partial x_j}(a) \right)_{i,j=K+1}^N. \quad (4.33)$$

Then there exist power series $h = (h_{K+1}, \dots, h_N)$ such that for all $t = (t_1, \dots, t_K) = 0 \pmod{p}$, one has that

$$f(a + (p^{2\gamma}t, p^\gamma h(t))) = 0. \quad (4.34)$$

This means that for each $j = K + 1, \dots, N$, one has

$$f_j(a_1 + p^{2\gamma}t_1, \dots, a_K + p^{2\gamma}t_K, a_{K+1} + p^\gamma h_{K+1}(t), \dots, a_N + p^\gamma h_N(t)) = 0.$$

This is proved in [29, Lemma 5.21 and Note 5.22]. In fact, it is shown that all $b \in \widehat{\mathbb{Z}}_p^N$ such that $b = a \pmod{p^{2\gamma+1}}$ and $f(b) = 0$ can be parameterized this way. We will use it to obtain the following lower bound, assuming conditions (4.31)–(4.32) hold.

Corollary 4.3. *Let $n > 2\gamma$. Then*

$$|\{b \in \mathbb{Z}_p^N : f(b) = 0 \pmod{p^n}\}| \geq p^{(n-2\gamma-1)K}. \quad (4.35)$$

Proof. Notice that if $t_1 \neq t_2 \pmod{p^{n-2\gamma}}$, then $c_1 \neq c_2 \pmod{p^n}$, where $c_i = a + (p^{2\gamma}t_i, p^\gamma h(t_i))$ for $i = 1, 2$. There are $p^{(n-2\gamma-1)K}$ values of $t \in \mathbb{Z}^K$ such that $t = 0 \pmod{p}$, which fall into different residue classes mod $p^{n-2\gamma}$, thus by (4.34), we have at least this many solutions to $f(c) = 0$ in $\widehat{\mathbb{Z}}_p^N$, which fall into different residue classes mod p^n . For each such c , let $b \in \mathbb{Z}^N$ such that $b = c \pmod{p^n}$. Then clearly $f(b) = 0 \pmod{p^n}$, and all such b 's are distinct mod p^n . \square

Funding information: The first, second and third authors were supported in part by NSF grants DMS-2007008 and DMS-1600840 and DMS-2154712 respectively. The third author was also partially supported by the Department of Mathematics at Rutgers University and by the National Science Centre in Poland, grant Opus 2018/31/B/ST1/00204. The fourth author was partially supported by the National Science Centre of Poland, grant Opus 2017/27/B/ST1/01623, the Juan de la Cierva Incorporación 2019, grant number IJC2019-039661-I, the Agencia Estatal de Investigación, grant PID2020-113156GB-I00/AEI/10.13039/501100011033, the Basque Government through the BERC 2022-2025 program, and by the Spanish Ministry of Sciences, Innovation and Universities: BCAM Severo Ochoa accreditation CEX2021-001142-S.

Conflict of interest: Authors state no conflict of interest.

References

- [1] T. Austin, *A proof of Walsh's convergence theorem using couplings*, Int. Math. Res. Not. IMRN **15** (2015), 6661–6674.
- [2] T. Austin, *On the norm convergence of non-conventional ergodic averages*, Ergodic Theory Dynam. Systems **30** (2010), 321–338.
- [3] A. Bellow, *Measure Theory Oberwolfach 1981*. in: Proceedings of the Conference held at Oberwolfach, June 21–27, 1981. Lecture Notes in Mathematics 945, editors D. Kölzow and D. Maharam-Stone. Springer-Verlag Berlin Heidelberg (1982), Section: Two problems submitted by A. Bellow, pp. 429–431.
- [4] V. Bergelson, *Weakly mixing PET*, Ergodic Theory Dynam. Systems **7** (1987), no. 3, 337–349.
- [5] V. Bergelson, *Ergodic Ramsey theory - an update, ergodic theory of \mathbb{Z}^d -actions*, in: M. Pollicott and K. Schmidt (Eds.), *London Mathematical Society Lecture Note Series*, vol. 228, 1996, pp. 1–61.
- [6] V. Bergelson, *Combinatorial and diophantine applications of ergodic theory (with appendices by A. Leibman and by A. Quas and M. Wierdl)*, in: *Handbook of Dynamical Systems* B. Hasselblatt and A. Katok, (Eds.), Vol. 1B, Elsevier, 2006, pp. 745–841.
- [7] V. Bergelson and A. Leibman, *Polynomial extensions of van der Waerden and Szemerédi's theorems*, J. Amer. Math. Soc. **9** (1996), 725–753.
- [8] V. Bergelson and A. Leibman, *A nilpotent Roth theorem*, Invent. Math. **147** (2002), 429–470.
- [9] B. J. Birch, *Forms in many variables*, Proc. R. Soc. Lond. A **265** (1962), 245–263.
- [10] G. Birkhoff, *Proof of the ergodic theorem*, Proc. Natl. Acad. Sci. USA **17** (1931), no. 12, 656–660.
- [11] J. Bourgain, *On the maximal ergodic theorem for certain subsets of the integers*, Israel J. Math. **61** (1988), 39–72.
- [12] J. Bourgain, *On the pointwise ergodic theorem on L^p for arithmetic sets*, Israel J. Math. **61** (1988), 73–84.
- [13] J. Bourgain, *Pointwise ergodic theorems for arithmetic sets*, with an appendix by the author, H. Furstenberg, Y. Katznelson and D.S. Ornstein, Inst. Hautes Études Sci. Publ. Math. **69** (1989), 5–45.
- [14] J. Bourgain, *Double recurrence and almost sure convergence*, J. Reine Angew. Math. **404** (1990), 140–161.
- [15] Z. Buczolich and R. D. Mauldin, *Divergent square averages*, Ann. Math. **171** (2010), no. 3, 1479–1530.
- [16] A. Calderón, *Ergodic theory and translation invariant operators*, Proc. Natl. Acad. Sci. USA **59** (1968), 349–353.
- [17] M. Christ, *Hilbert transforms along curves: I. nilpotent groups*, Ann. Math. **122/3** (1985), 575–596.
- [18] M. Christ, A. Nagel, E. M. Stein, and S. Wainger, *Singular and maximal Radon transforms: analysis and geometry*, Ann. Math. **150** (1999), no. 2, 489–577.
- [19] Q. Chu, N. Frantzikinakis, and B. Host, *Ergodic averages of commuting transformations with distinct degree polynomial iterates*, Proc. London. Math. Soc. **102** (2011), no. 5, 801–842.
- [20] L. J. Corwin and F. P. Greenleaf, *Representations of nilpotent Lie groups and their applications. Part I. Basic theory and examples*, Cambridge Studies in Advanced Mathematics, vol. 18, Cambridge University Press, Cambridge, 1990.

- [21] H. Davenport, *Cubic Forms in Thirty-two Variables*, Phil. Trans. R. Soc. Lond. A **251** (1959), 193–232.
- [22] H. Davenport, *Analytic Methods for Diophantine Equations and Diophantine Inequalities*, Cambridge University Press, Cambridge, 1959.
- [23] N. Frantzikinakis, *Some open problems on multiple ergodic averages*, Bull. Hellenic Math. Soc. **60** (2016), 41–90.
- [24] N. Frantzikinakis and B. Kra, *Polynomial averages converge to the product of integrals*, Israel J. Math. **148** (2005), 267–276.
- [25] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Anal. Math. **31** (1977), 204–256.
- [26] H. Furstenberg, *Problems Session, Conference on Ergodic Theory and Applications*, University of New Hampshire, Durham, NH, June 1982, .
- [27] H. Furstenberg, *Nonconventional ergodic averages, The legacy of John von Neumann (Hempstead, NY, 1988)*, Proceedings of Symposia in Pure Mathematics, vol. 50, Amer. Math. Soc. Providence, RI, 1990, pp. 43–56.
- [28] H. Furstenberg and B. Weiss, *A mean ergodic theorem for $\frac{1}{N} \sum_{n=1}^N f(T^n x)g(T^{n^2} x)$* , *Convergence in Ergodic Theory and Probability (Columbus, OH, 1993)*, Ohio State Univ. Math. Res. Inst. Publ., vol. 5, De Gruyter, Berlin, 1996, pp. 193–227.
- [29] M. J. Greenberg, *Lectures on Forms in Many Variables*, W.A. Benjamin, Inc., New York, New York, 1969.
- [30] D. Hilbert, *Beweis für die Darstellbarkeit der ganzen Zahlen durch eine feste Anzahl n -ter Potenzen (Waring'sches Problem)*, Math. Ann. **67** (1909), 281–300.
- [31] B. Host and B. Kra, *Non-conventional ergodic averages and nilmanifolds*, Ann. Math. **161** (2005), 397–488.
- [32] B. Host and B. Kra, *Convergence of polynomial ergodic averages*, Israel J. Math. **149** (2005), 1–19.
- [33] Y.-Q. Hu, *Polynomial maps and polynomial sequences in groups*, Available at arXiv:2105.08000.
- [34] Y.-Q. Hu, *Waring's problem for locally nilpotent groups: the case of discrete Heisenberg groups*, Available at arXiv:2011.06683.
- [35] A. Ionescu, A. Magyar, M. Mirek, and T. Z. Szarek, *Polynomial averages and pointwise ergodic theorems on nilpotent groups*, Invent. Math. **231** (2023), 1023–1140.
- [36] A. Ionescu, A. Magyar, E. M. Stein, and S. Wainger, *Discrete Radon transforms and applications to ergodic theory*, Acta Math. **198** (2007), 231–298.
- [37] A. Ionescu, A. Magyar, and S. Wainger, *Averages along polynomial sequences in discrete nilpotent Lie groups: Singular Radon transforms*, In: *Advances in Analysis: the Legacy of Elias M. Stein*, Princeton Mathematical Series, vol. 50, Princeton University Press, Princeton, NJ, 2014, pp. 146–188.
- [38] A. D. Ionescu and S. Wainger, *L^p boundedness of discrete singular Radon transforms*, J. Amer. Math. Soc. **19** (2005), no. 2, 357–383.
- [39] B. Krause, *Discrete analogues in harmonic analysis: maximally monomially modulated singular integrals related to Carleson's theorem*, Available at arXiv:1803.09431.
- [40] B. Krause, M. Mirek, and T. Tao, *Pointwise ergodic theorems for non-conventional bilinear polynomial averages*, Ann. Math. **195** (2022), no. 3, 997–1109.
- [41] P. LaVictoire, *Universally L^1 -bad arithmetic sequences*, J. Anal. Math. **113** (2011), no. 1, 241–263.
- [42] A. Leibman, *Convergence of multiple ergodic averages along polynomials of several variables*, Israel J. Math. **146** (2005), 303–315.
- [43] D. Lépingle, *La variation d'ordre p des semi-martingales*, Z. Wahrscheinlichkeitstheorie Verw. Gebiete **36** (1976), 295–316.
- [44] A. Magyar, E. M. Stein, and S. Wainger, *Discrete analog in harmonic analysis: spherical averages*, Ann. Math. **155** (2002), 189–208.
- [45] A. Magyar, E. M. Stein, and S. Wainger, *Maximal operators associated to discrete subgroups of nilpotent Lie groups*, J. Anal. Math. **101** (2007), 257–312.
- [46] M. Mirek, E. M. Stein, and B. Trojan, *$\ell^p(\mathbb{Z}^d)$ -estimates for discrete operators of Radon type: Variational estimates*, Invent. Math. **209** (2017), no. 3, 665–748.
- [47] M. Mirek, E. M. Stein, and P. Zorin-Kranich, *A bootstrapping approach to jump inequalities and their applications*, Anal. PDE **13** (2020), no. 2, 527–558.
- [48] M. Mirek, E. M. Stein, and P. Zorin-Kranich, *Jump inequalities for translation-invariant operators of Radon type on \mathbb{Z}^d* , Adv. Math. **365** (2020), Article ID 107065, pp. 57.
- [49] L. Pierce, *Discrete fractional Radon transforms and quadratic forms*, Duke Math. J. **161** (2012), 69–106.
- [50] L. Pierce and P.-L. Yung, *A polynomial Carleson operator along the paraboloid*, Rev. Mat. Iberoam. **35** (2019), 339–422.
- [51] F. Ricci and E. M. Stein, *Harmonic analysis on nilpotent groups and singular integrals I. Oscillatory integrals*, J. Funct. Anal. **73** (1987), 179–194.
- [52] E. M. Stein and S. Wainger, *Discrete analog in harmonic analysis, I: ℓ^2 estimates for singular Radon transforms*, Amer. J. Math. **121** (1999), 1291–1336.
- [53] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.
- [54] T. Tao, *Norm convergence of multiple ergodic averages for commuting transformations*, Ergodic Theory Dynam. Syst. **28** (2008), 657–688.
- [55] I. M. Vinogradov, *The method of trigonometrical sums in the theory of numbers. (Russian)*, Trav. Inst. Math. Stekloff **23** (1947).
- [56] J. von Neumann, *Proof of the quasi-ergodic hypothesis*, Proc. Natl. Acad. Sci. USA **18** (1932), 70–82.
- [57] M. Walsh, *Norm convergence of nilpotent ergodic averages*, Ann. Math. **175** (2012), no. 3, 1667–1688.
- [58] T. D. Wooley, *Nested efficient congruencing and relatives of Vinogradov's mean value theorem*, Proc. London Math. Soc. **118** (2019), no. 4, 942–1016.
- [59] T. Ziegler, *Universal characteristic factors and Furstenberg averages*, J. Amer. Math. Soc. **20** (2007), 53–97.