The quantitative behaviour of polynomial orbits on nilmanifolds

By Ben Green and Terence Tao

Abstract

A theorem of Leibman asserts that a polynomial orbit $(g(n)\Gamma)_{n\in\mathbb{Z}}$ on a nilmanifold G/Γ is always equidistributed in a union of closed subnilmanifolds of G/Γ. In this paper we give a quantitative version of Leibman's result, describing the uniform distribution properties of a finite polynomial orbit $(g(n)\Gamma)_{n\in[N]}$ in a nilmanifold. More specifically we show that there is a factorisation $g = \varepsilon g'\gamma$, where $\varepsilon(n)$ is "smooth," $(\gamma(n)\Gamma)_{n\in\mathbb{Z}}$ is periodic and "rational," and $(g'(n)\Gamma)_{n\in P}$ is uniformly distributed (up to a specified error δ) inside some subnilmanifold G'/Γ' of G/Γ for all sufficiently dense arithmetic progressions $P \subseteq [N]$.

Our bounds are uniform in N and are polynomial in the error tolerance δ. In a companion paper we shall use this theorem to establish the Möbius and Nilsequences conjecture from an earlier paper of ours.

1. Introduction

Nilmanifolds. In the last few years it has come to be appreciated that *nilmanifolds*, together with orbits on them, play a fundamental role in combinatorial number theory. Their relevance was certainly apparent in [7] and has been displayed quite dramatically in recent ergodic-theoretic work of Host-Kra [15] and Ziegler [35]. More recently the authors have explored how nilmanifolds arise in additive combinatorics [9] and in the study of linear equations in the primes [11]. The present paper is a part of that programme (and in particular will be used to prove the *Möbius and Nilsequences conjecture* from [11] in the companion [12] to this paper) but, since it concerns only the intrinsic properties of nilmanifolds, may be read independently of any of the other work. The reader interested in the background may consult the surveys [8], [17], [31], or the paper [11].

We begin by setting out our notation for nilmanifolds.

Definition 1.1 (Filtrations and nilmanifolds). Let G be a connected, simply connected Lie group with identity element id_G. For the purposes of this paper we define a *filtration* $G_\bullet$ on G to be a sequence of closed connected subgroups

$$G = G_0 = G_1 \supseteq G_2 \supseteq \cdots \supseteq G_d \supseteq G_{d+1} = \{\mathrm{id}_G\}$$

which has the property that $[G_i, G_j] \subseteq G_{i+j}$ for all integers $i, j \geqslant 0$. The least integer d for which $G_{d+1} = \{\mathrm{id}_G\}$ is called the *degree* of the filtration $G_\bullet$ and here, as usual, the commutator group $[H, K]$ is the group generated by $\{[h, k] : h \in H, k \in K\}$, where $[h, k] := hkh^{-1}k^{-1}$ is the commutator of h and k. If G possesses a filtration, then we say that G is *nilpotent*. Let $\Gamma \subseteq G$ be a uniform subgroup (i.e., a discrete, cocompact subgroup). Then the quotient $G/\Gamma = \{g\Gamma : g \in G\}$ is called a *nilmanifold*. We also write $g(\mathrm{mod}\ \Gamma)$ for $g\Gamma$.

Throughout the paper we will write $m = \dim G$ and $m_i = \dim G_i$, $i = 1, \ldots, d$.

Remark. The assumptions of connectedness and simple-connectedness for G are not completely standard, but are very convenient for us. In any situation in which we apply our theorems, we expect to be able to reduce to this case. If a filtration $G_\bullet$ of degree d exists, then it is easy to see that the *lower central series filtration*[1] defined by $G = G_0 = G_1$, $G_{i+1} = [G, G_i]$ terminates with $G_{s+1} = \{\mathrm{id}_G\}$ for some integer $s \leqslant d$. We call the minimal such integer s the *step* of the nilpotent Lie group G. In this paper the degree d will play a vastly more important role than the step s, since it will be important to work with filtrations more general than the lower central series.

Examples. The simplest examples of nilmanifolds arise when $s = 1$ in which case we may, after a linear transformation, take $G = \mathbb{R}^m$ and $\Gamma = \mathbb{Z}^m$. The lower central series filtration is given by $G = G_0 = G_1$ and $G_2 = \{\mathrm{id}_G\}$. The nilmanifold G/Γ is then referred to as a *torus*. Note that in this example the group operation is written *additively*, as is conventional for abelian groups. When we are working with non-abelian groups we shall write the group operation multiplicatively. The simplest non-abelian example is given by the 3-dimensional *Heisenberg nilmanifold*, in which $s = 2$. We will study this object in some detail later on. Here we take

$$(1.1) \qquad G = \begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \Gamma = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}.$$

The lower central series filtration is given by $G = G_0 = G_1$,

$$G_2 = \begin{pmatrix} 1 & 0 & \mathbb{R} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

---

[1] It is not hard to see that the lower central series filtration *is* a filtration, in that we have $[G_i, G_j] \subseteq G_{i+j}$ for all i, j.

and $G_3 = \{\mathrm{id}_G\}$. Observe that a fundamental domain for the action of Γ on G is

$$(1.2) \qquad \left\{ \begin{pmatrix} 1 & x_1 & x_2 \\ 0 & 1 & x_3 \\ 0 & 0 & 1 \end{pmatrix} : 0 \leqslant x_1, x_2, x_3 < 1 \right\}.$$

Thus one can view G/Γ as a unit cube, with the sides glued together in a twisted fashion.

This paper will be concerned with the qualitative and quantitative equidistribution of various algebraic sequences on nilmanifolds. We first set out our notation for equidistribution.

Definition 1.2 (Equidistribution). Let G/Γ be a nilmanifold. Here and in the sequel we endow G/Γ with the unique normalised Haar measure, we let $[N] := \{n \in \mathbb{Z} : 1 \leqslant n \leqslant N\}$, and we write $\mathbb{E}_{a \in A} f(a) := \frac{1}{|A|} \sum_{a \in A} f(A)$ for the average of f on the set A.

(i) An infinite sequence $(g(n)\Gamma)_{n \in \mathbb{N}}$ in G/Γ is said to be *equidistributed* if we have

$$\lim_{N \to \infty} \mathbb{E}_{n \in [N]} F(g(n)\Gamma) = \int_{G/\Gamma} F$$

for all continuous functions $F : G/\Gamma \to \mathbb{C}$.

(ii) An infinite sequence $(g(n)\Gamma)_{n \in \mathbb{Z}}$ in G/Γ is said to be *totally equidistributed* if the sequences $(g(an + r)\Gamma)_{n \in \mathbb{N}}$ are equidistributed for all $a \in \mathbb{Z} \backslash \{0\}$ and $r \in \mathbb{Z}$.

(iii) Given a length $N > 0$ and an error tolerance $\delta > 0$, a finite sequence $(g(n)\Gamma)_{n \in [N]}$ is said to be *δ-equidistributed* if we have

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

for all Lipschitz functions $F : G/\Gamma \to \mathbb{C}$, where

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

and the metric $d_{G/\Gamma}$ on G/Γ will be defined in Definition 2.2 in the next section (it will involve choosing a *Mal'cev basis* $\mathcal{X}$ for G/Γ).

(iv) A finite sequence $(g(n)\Gamma)_{n \in [N]}$ is said to be *totally δ-equidistributed* if we have

$$\left| \mathbb{E}_{n \in P} F(g(n)\Gamma) - \int_{G/\Gamma} F \right| \leqslant \delta \|F\|_{\mathrm{Lip}}$$

for all Lipschitz functions $F : G/\Gamma \to \mathbb{C}$ and all arithmetic progressions $P \subset [N]$ of length at least δN.

We will be interested in the *qualitative* question of when a sequence $(g(n)\Gamma)_{n\in\mathbb{N}}$ is equidistributed (or totally equidistributed), as well as the more *quantitative* question of when a finite sequence $(g(n)\Gamma)_{n\in[N]}$ is δ-equidistributed (or totally δ-equidistributed). Such questions, and corresponding questions in more general settings (for example when G/Γ is a homogeneous space of a general, not necessarily nilpotent, Lie group), play a fundamental role in number theory; see [34] for a discussion. These questions are also closely related to the celebrated theorem of Ratner [28] on unipotent flows, although as we are restricting attention to nilmanifolds, we will not need the full force of Ratner's theorem (or quantitative versions thereof) here.

Qualitative equidistribution theory of linear sequences. To begin the discussion let us first restrict attention to *linear* sequences.

Definition 1.3 (Linear sequences). A *linear sequence* in a group G is any sequence $g : \mathbb{Z} \to G$ of the form $g(n) := a^n x$ for some $a, x \in G$. A *linear sequence* in a nilmanifold G/Γ is a sequence of the form $(g(n)\Gamma)_{n\in\mathbb{Z}}$, where $g : \mathbb{Z} \to G$ is a linear sequence in G.

In the additive case $G = \mathbb{R}^m$, $\Gamma = \mathbb{Z}^m$, a linear sequence takes the form $(an + x (\mathrm{mod}\ \mathbb{Z}^m))_{n\in\mathbb{Z}}$. In this case one can understand equidistribution satisfactorily using Kronecker's theorem and its variants. For instance, to answer qualitative questions about equidistribution in this case, we have the following classical result.

THEOREM 1.4 (Qualitative Kronecker theorem). *Let* $m \geqslant 1$, *and let* $(g(n)(\mathrm{mod}\ \mathbb{Z}^m))_{n\in\mathbb{N}}$ *be a linear sequence in the torus* $\mathbb{R}^m/\mathbb{Z}^m$. *Then exactly one of the following statements is true*:

(i) $(g(n)(\mathrm{mod}\ \mathbb{Z}^m))_{n\in\mathbb{N}}$ *is equidistributed in* $\mathbb{R}^m/\mathbb{Z}^m$.
(ii) *There exists a nontrivial* character $\eta : \mathbb{R}^m \to \mathbb{R}/\mathbb{Z}$, *i.e., a continuous additive homomorphism which annihilates* $\mathbb{Z}^m$ *but does not vanish entirely, such that* $\eta \circ g$ *is constant.* (*Equivalently, if* $g(n) = an + x$, *there exists a nonzero* $k \in \mathbb{Z}^m$ *such that* $k \cdot a \in \mathbb{Z}$.)

In particular, $(g(n)(\mathrm{mod}\ \mathbb{Z}^m))_{n\in\mathbb{Z}}$ *is equidistributed if and only if it is totally equidistributed.*

Remarks. An equivalent formulation of this theorem is that if the linear sequence

$$(g(n)(\mathrm{mod}\ \mathbb{Z}^m))_{n\in\mathbb{N}}$$

is not equidistributed, then this sequence instead takes values in a finite union of proper subtori of G/Γ. This can be viewed as an extremely simple special

case of the theorems of Ratner [28] and Shah [29]. More quantitative results can be obtained via Fourier analysis;[2] see Proposition 3.1 below.

A remarkable theorem of Leon Green allows one to reduce qualitative questions about the distribution of orbits on nilmanifolds of step $s > 1$ to the abelian case just described.

Definition 1.5 (Horizontal torus). Given a nilmanifold G/Γ, the *horizontal torus* is defined to be $(G/\Gamma)_{\mathrm{ab}} := G/[G,G]\Gamma$. We let $\pi : G \to (G/\Gamma)_{\mathrm{ab}}$ be the canonical projection map. A *horizontal character* is a continuous additive homomorphism $\eta : G \to \mathbb{R}/\mathbb{Z}$ which annihilates Γ; observe that such characters in fact annihilate $[G,G]\Gamma$ and so can be viewed as characters on the horizontal torus. We say that a horizontal character is *nontrivial* if it is not identically zero.

It follows from results of Mal'cev [24], and in particular the existence of so-called *Mal'cev bases*, that $(G/\Gamma)_{\mathrm{ab}}$ really is a torus and in fact is isomorphic to $\mathbb{R}^{m_{\mathrm{ab}}}/\mathbb{Z}^{m_{\mathrm{ab}}}$, where $m_{\mathrm{ab}} := \dim_{\mathbb{R}}(G) - \dim_{\mathbb{R}}([G,G])$. We will not actually need this characterisation, as the properties of horizontal characters $\eta : G \to \mathbb{R}/\mathbb{Z}$ will be our main focus. Readers may find it useful to keep this in mind, however.

THEOREM 1.6 (Leon Green's theorem). *Let $(g(n)\Gamma)_{n \in \mathbb{Z}}$ be a linear sequence in a nilmanifold G/Γ. Then the orbit $(g(n)\Gamma)_{n \in \mathbb{N}}$ is equidistributed in G/Γ if and only if the projected orbit $(\pi(g(n)\Gamma))_{n \in \mathbb{N}}$ is equidistributed in the horizontal torus $(G/\Gamma)_{\mathrm{ab}}$. (In particular, $(g(n)\Gamma)_{n \in \mathbb{Z}}$ is equidistributed if and only if it is totally equidistributed.)*

Proof. See [1], [13]. Leon Green used representation theory to establish his result, but a more elementary proof was subsequently found by Parry [26]. □

Example. Suppose that G/Γ is the Heisenberg example (1.1). Then

$$[G,G] = \begin{pmatrix} 1 & 0 & \mathbb{R} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and $(G/\Gamma)_{\mathrm{ab}}$ may be identified with $\mathbb{R}^2/\mathbb{Z}^2$, the projection π being given by

$$\pi \left[\begin{pmatrix} 1 & x_1 & x_2 \\ 0 & 1 & x_3 \\ 0 & 0 & 1 \end{pmatrix} \right] := (x_1, x_3).$$

Leon Green's theorem implies that the orbit $(a^n \Gamma)_{n \in \mathbb{N}}$, where

$$a = \begin{pmatrix} 1 & \alpha_1 & \alpha_2 \\ 0 & 1 & \alpha_3 \\ 0 & 0 & 1 \end{pmatrix},$$

---

[2]In this simple setting one could also use more classical tools such as Minkowski's geometry of numbers, and in the $m = 1$ case one could even use continued fractions. However, these methods do not seem to extend easily to higher steps.

is equidistributed in G/Γ if and only if $1, \alpha_1$, and α_3 are independent over $\mathbb{Q}$. It is already somewhat nontrivial to establish this result directly.

By Kronecker's theorem, we can then recast Theorem 1.6 in the following equivalent formulation.

THEOREM 1.7 (Leon Green's theorem, again). *Let* $(g(n)\Gamma)_{n \in \mathbb{Z}}$ *be a linear sequence in a nilmanifold* G/Γ. *Then exactly one of the following statements is true*:

   (i) $(g(n)\Gamma)_{n \in \mathbb{N}}$ *is equidistributed in* G/Γ.
   (ii) *There exists a nontrivial horizontal character* $\eta : G \to \mathbb{R}/\mathbb{Z}$ *such that* $\eta \circ g$ *is constant*.

Qualitative equidistribution theory of polynomial sequences. While our primary applications are concerned with linear sequences, it turns out for various technical reasons that it is important to work in the more general class of *polynomial sequences.*

Definition 1.8 (Polynomial sequences in nilpotent groups). Suppose that G is a nilpotent group with a filtration $G_\bullet$. Let $g : \mathbb{Z} \to G$ be a sequence. If $h \in \mathbb{Z}$, we write $\partial_h g := g(n+h)g(n)^{-1}$. We say that g is a polynomial sequence with coefficients in $G_\bullet$, and write $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$, if $\partial_{h_i} \ldots \partial_{h_1} g$ takes values in G_i for all positive integers i and for all choices of $h_1, \ldots, h_i \in \mathbb{Z}$. In this case we say that g has degree d. If g lies in $\mathrm{poly}(G_\bullet)$ for some filtration $G_\bullet$, then we simply say that g is a polynomial sequence.

This definition is a little abstract. However we will show in Section 6 that $g : \mathbb{Z} \to G$ is a polynomial sequence if and only if g has the form $g(n) = a_1^{p_1(n)} \ldots a_k^{p_k(n)}$, where $a_1, \ldots, a_k \in G$ and the $p_i : \mathbb{N} \to \mathbb{N}$ are polynomials. In particular, a linear sequence $g(n) = a^n x$ is a polynomial sequence, and in fact since $\partial_{h_1} g(n) = a^{h_1}$ and $\partial_{h_2} \partial_{h_1} g(n) = \mathrm{id}_G$, it is clear that such a sequence has coefficients in the lower central series filtration $G_\bullet$. Note carefully that the degree of a linear sequence is equal to the step s of the underlying Lie group G and is not equal to one as the name "linear" might suggest.

A remarkable result of Lazard and Leibman [18], [19], [20] asserts that $\mathrm{poly}(\mathbb{Z}, G_\bullet)$ is a group. We will prove this in Section 6, and it will play a key role in several of our arguments.

Theorem 1.6 was extended by Liebman [22] to the case when $g(n)$ is a polynomial sequence rather than a linear one. In particular, he showed the following generalisation of Theorem 1.7.

THEOREM 1.9 (Leibman's theorem [22]). *Suppose that* G/Γ *is a nilmanifold and that* $g : \mathbb{Z} \to G$ *is a polynomial sequence. Then exactly one of the following statements is true*:

(i) $(g(n)\Gamma)_{n\in\mathbb{N}}$ is equidistributed in G/Γ.

(ii) There exists a nontrivial horizontal character $\eta : G \to \mathbb{R}/\mathbb{Z}$ such that $\eta \circ g$ is constant.

Remark. This theorem significantly generalises the classical theorem of Weyl that a polynomial sequence in $\mathbb{R}/\mathbb{Z}$ is equidistributed unless all of its nonconstant coefficients are rational. We will in fact use a quantitative version of Weyl's theorem in our arguments; see Proposition 4.3 below.

We can iterate this theorem to establish a factorisation result. We first need some notation.

Definition 1.10 (Rational subgroup). Let G/Γ be a nilmanifold. A *rational subgroup* of G is a closed connected subgroup G' of G such that $G'\Gamma/\Gamma \cong G'/\Gamma' = G'/(G' \cap \Gamma)$ is a closed submanifold of G/Γ (or equivalently, that Γ' is a cocompact subgroup of G'). We say that G' is *proper* if $G' \neq G$.

Example. If G/Γ is a nilmanifold (that is to say if there *exists* a uniform subgroup $\Gamma \leqslant G$), one can show that each member G_i of the lower central series is a rational subgroup; see, e.g., [4] or [24].

Definition 1.11 (Rational sequence). Let G/Γ be a nilmanifold. A *rational group element* is any $g \in G$ such that $g^r \in \Gamma$ for some integer $r > 0$. A *rational point* is any point in G/Γ of the form $g\Gamma$ for some rational group element g. A sequence $(g(n)\Gamma)_{n\in\mathbb{Z}}$ is *rational* if every element $g(n)\Gamma$ in the sequence is a rational point.

Remark. It is not difficult to show that the rational group elements form a dense subgroup of G that contains Γ; see Lemma A.11. We will show in Lemma A.12 that any polynomial sequence in G/Γ which is rational is automatically periodic.

COROLLARY 1.12 (Factorisation theorem for polynomial sequences). *Let $(g(n)\Gamma)_{n\in\mathbb{Z}}$ be a polynomial sequence in a nilmanifold G/Γ. Then there exists a rational subgroup G' of G and a factorisation $g = \varepsilon g'\gamma$, where $\varepsilon \in G$ is a constant, $g' : \mathbb{Z} \to G'$ is a polynomial sequence such that $(g'(n)\Gamma')_{n\in\mathbb{N}}$ is totally equidistributed in G'/Γ' (where $\Gamma' := G \cap \Gamma$), and $\gamma : \mathbb{Z} \to G$ is a polynomial sequence such that the sequence $(\gamma(n)\Gamma)_{n\in\mathbb{N}}$ is rational (and hence, by Lemma A.12(i), is periodic).*

Proof. We give a sketch of this argument only; we will repeat this argument in more detail when proving Theorem 1.19 below.

We induct on the dimension m of G/Γ, assuming that the claim has already been proven for all nilmanifolds of lesser dimension. By replacing $g(n)$ with $g(0)^{-1}g(n)$ if necessary (absorbing the $g(0)$ factor into the ε term), we may normalise so that $g(0) = \mathrm{id}_G$. If $(g(n)\Gamma)_{n\in\mathbb{Z}}$ is equidistributed on G/Γ,

then it is totally equidistributed by Leibman's theorem, and we are done (with $g' = g$, $G' = G$, and ε, γ trivial). So we may assume that $(g(n)\Gamma)_{n\in\mathbb{Z}}$ is not equidistributed. By Leibman's theorem, there exists a nontrivial horizontal character $\eta : G \to \mathbb{R}/\mathbb{Z}$ such that $\eta \circ g$ is constant. In fact, by our normalisation $g(0) = \mathrm{id}_G$, we must have $\eta \circ g \equiv 0$, thus g takes values in $\ker(\eta)$. It is then not difficult to factorise $g = g_0\gamma_0$, where γ_0 is a polynomial sequence with $(\gamma_0(n)\Gamma)_{n\in\mathbb{Z}}$ rational and periodic, and g_0 is a polynomial sequence taking values in the proper rational subgroup $G' \leqslant G$, defined to be the connected component of $\ker(\eta)$ which contains the origin. The claim then follows by applying the induction hypothesis to the sequence $(g_0(n)\Gamma')_{n\in\mathbb{Z}}$ in the nilmanifold G'/Γ', which has dimension $m - 1$, and using the fact that the product of two rational group elements is again rational, as well as the trivial observation that rational group elements of G' are automatically rational group elements of G also. $\qquad\square$

Remark. In words, this corollary asserts that in the qualitative setting, one can decompose

(arbitrary polynomial sequence)

$$= (\text{constant}) \times (\text{totally equidistributed}) \times (\text{periodic}).$$

An inspection of the proof reveals that one can in fact take the constant ε to be $g(0)$.

As a corollary, we obtain a Ratner-Shah type theorem for polynomial sequences in nilmanifolds, first established by Leibman [22].

COROLLARY 1.13 (Leibman's Ratner-Shah type theorem for nilmanifolds). *Let $(g(n)\Gamma)_{n\in\mathbb{Z}}$ be a polynomial sequence in a nilmanifold G/Γ. Then there exists a rational subgroup G' of G, a group element $\varepsilon \in G$, and a rational periodic sequence $(x_n)_{n\in\mathbb{Z}}$ in G/Γ with some period q such that for every $r \in \mathbb{Z}$, the sequence $(g(qn + r)\Gamma)_{n\in\mathbb{Z}}$ is totally equidistributed in $\varepsilon G' x_r$.*

Remark. Shah [29] obtained a similar result for arbitrary discrete unipotent (but linear) flows on a finite volume homogeneous space; the case of continuous unipotent linear flows was treated earlier by Ratner [28] (see [25] for further discussion). Leibman's proof of Corollary 1.13 does not use these results, but instead proceeds in two stages. Firstly, by iterating Theorem 1.6 (or more precisely a generalisation of this theorem to the case when G is not necessarily connected), a version of Corollary 1.13 for linear sequences is obtained. Secondly, by utilising a lifting trick of Furstenberg [6, p. 31], the polynomial case is deduced from the linear case. As we shall discuss shortly, these arguments do not work well in the quantitative case, and one must instead grapple with polynomial sequences directly.

Quantitative equidistribution results. This paper stems from an attempt to establish quantitative versions of the above theorems for *finite* orbits. Unfortunately, the need for quantitative bounds on all aspects of these results forces us to introduce a substantial amount of new notation.

Definition 1.14 (Asymptotic notation). We use $Y = O(X)$ or $Y \ll X$ to denote the estimate $|Y| \leqslant CX$ for some absolute constant C. When we need to indicate dependence of C on various parameters, we shall indicate this by subscripts; thus for instance $O_{d,m}(X)$ denotes a quantity bounded in magnitude by $C_{d,m}X$ for some $C_{d,m}$ depending only on the quantities d, m.

Definition 1.15 (Circle norm). If $x \in \mathbb{R}/\mathbb{Z}$, we use $\|x\|_{\mathbb{R}/\mathbb{Z}} := \mathrm{dist}(x, \mathbb{Z})$ to denote the distance of x to the origin (thus $\|a(\mathrm{mod}\ \mathbb{Z})\|_{\mathbb{R}/\mathbb{Z}} = |a|$ whenever $-1/2 < a \leqslant 1/2$). If $x \in \mathbb{R}$, we write $\|x\|_{\mathbb{R}/\mathbb{Z}}$ for $\|x(\mathrm{mod}\ \mathbb{Z})\|_{\mathbb{R}/\mathbb{Z}}$.

Our first main result is the following quantitative version of Theorem 1.9. Note that some of the terminology in this theorem will not be formally introduced until the next section, but this should not prevent the reader from gaining a rough appreciation of the statement.

THEOREM 1.16 (Quantitative Leibman theorem). *Let $m, d \geqslant 0$, $0 < \delta < 1/2$, and $N \geqslant 1$. Let G/Γ be an m-dimensional nilmanifold together with a filtration $G_\bullet$ of degree d and a $\frac{1}{\delta}$-rational Mal'cev basis $\mathcal{X}$ adapted to this filtration. Suppose that $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$. Then at least one of the following statements is true:*

  (i) *$(g(n)\Gamma)_{n \in [N]}$ is δ-equidistributed in G/Γ.*
  (ii) *There exists a nontrivial horizontal character $\eta : G \to \mathbb{R}/\mathbb{Z}$ with $|\eta| \ll \delta^{-O_{m,d}(1)}$ such that $\|\eta \circ g(n) - \eta \circ g(n-1)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O_{m,d}(1)}/N$ for all $n \in [N]$.*

Remarks. The notions of a "$\frac{1}{\delta}$-rational Mal'cev basis adapted to $G_\bullet$" of the modulus $|\eta|$ of a horizontal character and of the metric which is implicit in the notion of δ-equidistribution are technical and will be defined precisely in Definitions 2.4, 2.6, and 2.2 respectively.

Theorem 1.16 asserts that the sequence $(g(n)\Gamma)_{n \in [N]}$ is either δ-equidistributed up to time N, or else it is very far from being equidistributed up to time $\delta^{O_{m,d}(1)}N$, being concentrated very close to a union of $\delta^{-O_{m,d}(1)}$ subtori. One should view N as being very large compared to $1/\delta$, otherwise the content of the proposition is trivial. It is not hard to deduce Theorem 1.9 from Theorem 1.16; we leave this to the reader as an exercise.

For technical reasons it will be convenient later to strengthen the statement (ii) slightly, so as to also control higher "derivatives" $\partial^j(\eta \circ g)$; see the next section for more information.

Whereas in the qualitative setting one always works in the limit $N \to \infty$, in the quantitative setting one works with a fixed (but large) N. As N increases, there can be transitions in the behaviour of the finite sequence $(g(n)\Gamma)_{n \in [N]}$, in which the equidistribution (or lack thereof) changes significantly (cf. the "coalescence of progressions" phenomenon [32, Ch. 12]); these transitions are a new feature of the quantitative setting, which are not readily visible in the qualitative one. We illustrate this with a simple example.

Example. Consider the (additive) example $G = \mathbb{R}$, $\Gamma = \mathbb{Z}$, and $g(n) = (\frac{1}{2} + \sigma)n$, where $0 < \sigma \leqslant \frac{\delta}{100}$ is a parameter. In this case we have $m = d = 1$. If N is much larger than $1/\sigma$, we see that $(g(n)(\mathrm{mod}\ \mathbb{Z}))_{n \in [N]}$ is δ-equidistributed. On the other hand, if N is much smaller than $1/\sigma$, we see that $(g(n)(\mathrm{mod}\ \mathbb{Z}))_{n \in [N]}$ fails to be δ-equidistributed; indeed it is highly concentrated around 0 and $1/2$ in this case. However, if we let $\eta : G \to \mathbb{R}/\mathbb{Z}$ be the nontrivial horizontal character $\eta(x) := 2x(\mathrm{mod}\ \mathbb{Z})$, we see that $\eta(g(n))$ is slowly varying in the sense of (ii). The transitional regime when N is comparable to $1/\sigma$ is interesting; there is enough irregularity to prevent δ-equidistribution on the sequence $(g(n)(\mathrm{mod}\ \mathbb{Z}))_{n \in [N]}$, but in order to obtain near-constancy of $\eta(g(n))$, one in fact has to pass to shorter sequences such as $(g(n)(\mathrm{mod}\ \mathbb{Z}))_{n \in [\delta^{100}N]}$. The need to work on a variety of different scales like this is very much a feature of additive combinatorics, particularly those parts of it that have the flavour of "quantitative ergodic theory". The work of Bourgain [3] on Roth's theorem is another example.

Of course, by specialising to linear sequences, Theorem 1.16 also implies a quantitative version of Leon Green's theorem. The proof of Theorem 1.16 could be simplified somewhat in this case. Such a theorem is not especially useful, however. The following example may help to illustrate why, in the quantitative setting, the consideration of linear sequences leads naturally to the "polynomial" world.

Example (The skew torus). Let us consider the Heisenberg example (1.1) once more, taking now

$$a := \begin{pmatrix} 1 & 2\alpha & \alpha \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

where $\alpha := N^{-3/2}$. Set

$$g(n) := a^n = \begin{pmatrix} 1 & 2n\alpha & n^2\alpha \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix}.$$

Translating to the fundamental domain, we obtain

$$g(n)\Gamma = \left[\begin{pmatrix} 1 & \{2n\alpha\} & \{-n^2\alpha\} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right].$$

(Here, and for the rest of the paper, we define $\{x\} := x - \lfloor x \rfloor$, where $\lfloor x \rfloor$ is the greatest integer less than or equal to x.) The orbit $(g(n)\Gamma)_{n \in [N]}$ is certainly not

close to equidistributed in G/Γ, and indeed the projected orbit $(\pi(g(n)\Gamma))_{n\in[N]}$ stays very close to the trivial subtorus $T \subseteq \mathbb{R}^2/\mathbb{Z}^2$, which consists simply of the point $\{(0,0)\}$.

Now $\pi^{-1}(T)$ is of course isomorphic to a one-dimensional torus $\mathbb{R}/\mathbb{Z}$. However the orbit $(g(n)\Gamma)_{n\in[N]}$ does not approximate a linear orbit on this torus; rather, it has *quadratic* behaviour. Thus $(g(n)\Gamma)_{n\in[N]}$ is very close to $(g'(n)\Gamma')_{n\in[N]}$ on $G'/\Gamma' \cong \mathbb{R}/\mathbb{Z}$, where

$$G' := \left(\begin{smallmatrix} 1 & 0 & \mathbb{R} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right),$$

$$\Gamma' := \left(\begin{smallmatrix} 1 & 0 & \mathbb{Z} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right),$$

and

$$(1.3) \qquad g'(n) := \left(\begin{smallmatrix} 1 & 0 & -n^2\alpha \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix} \right).$$

Thus, in order to approximate the linear sequence $(g(n)\Gamma)_{n\in[N]}$ by a lower-dimensional sequence, the latter sequence needs to be polynomial. Note however that if one had the luxury of passing from $[N]$ to a much shorter progression, e.g., $[N^{1/100}]$, then the lower-dimensional sequence would remain linear. In the limit $N \to \infty$, N and $N^{1/100}$ both go to infinity, which may help explain why in the qualitative setting one can avoid polynomial sequences entirely and work purely in the category of linear sequences. Unfortunately, for the quantitative applications we have in mind (in particular, the number-theoretic application in [12]), we cannot afford to reduce the scale N in such a drastic manner.[3]

In much the same way that Theorem 1.9 could be iterated in order to establish Corollary 1.12, we can iterate Theorem 1.16 to obtain a quantitative factorisation theorem. To state it we need quantitative versions of the "rationality" concepts of Definition 1.11 and also the new notion of *smooth* sequences, which must be introduced in place of constant sequences in the finitary setting.

Definition 1.17 (Rational sequences, quantitative definitions). Let G/Γ be a nilmanifold, and let $Q > 0$ be a parameter. We say that $\gamma \in G$ is *Q-rational* if $\gamma^r \in \Gamma$ for some integer r, $0 < r \leqslant Q$. A *Q-rational point* is any point in G/Γ of the form $\gamma\Gamma$ for some Q-rational group element γ. A sequence $(\gamma(n))_{n\in\mathbb{Z}}$ is *Q-rational* if every element $\gamma(n)\Gamma$ in the sequence is a Q-rational point.

---

[3]This is ultimately because it is known how to obtain nontrivial control on averages of number-theoretic functions such as the Möbius function μ on intervals such as $[N, N + N \log^{-A} N]$, but not in intervals such as $[N, N + N^{1/100}]$, even if one assumes strong hypotheses such as GRH.

Definition 1.18 (Smooth sequences). Let G/Γ be a nilmanifold with a Mal'cev basis $\mathcal{X}$. Let $(\varepsilon(n))_{n\in\mathbb{Z}}$ be a sequence in G, and let $M, N \geqslant 1$. We say that $(\varepsilon(n))_{n\in\mathbb{Z}}$ is (M,N)-*smooth* if we have $d(\varepsilon(n), \mathrm{id}_G) \leqslant M$ and $d(\varepsilon(n), \varepsilon(n-1)) \leqslant M/N$ for all $n \in [N]$, where the metric $d = d_\mathcal{X}$ on G will be defined in Definition 2.2.

Note that the notion of a (M,N)-smooth sequence collapses to that of a constant sequence in the limit $N \to \infty$ (holding M fixed).

THEOREM 1.19 (Factorisation theorem). *Let $m, d \geqslant 0$, and let $M_0, N \geqslant 1$ and $A > 0$ be real numbers. Suppose that G/Γ is an m-dimensional nilmanifold together with a filtration $G_\bullet$ of degree d. Suppose that $\mathcal{X}$ is an M_0-rational Mal'cev basis $\mathcal{X}$ adapted to $G_\bullet$ and that $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$. Then there is an integer M with $M_0 \leqslant M \ll M_0^{O_{A,m,d}(1)}$, a rational subgroup $G' \subseteq G$, a Mal'cev basis $\mathcal{X}'$ for G'/Γ' in which each element is an M-rational combination of the elements of $\mathcal{X}$, and a decomposition $g = \varepsilon g' \gamma$ into polynomial sequences $\varepsilon, g', \gamma \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ with the following properties*:

(i) $\varepsilon : \mathbb{Z} \to G$ *is (M,N)-smooth*;

(ii) $g' : \mathbb{Z} \to G'$ *takes values in G', and the finite sequence $(g'(n)\Gamma')_{n\in[N]}$ is totally $1/M^A$-equidistributed in G'/Γ', using the metric $d_{\mathcal{X}'}$ on G'/Γ'*;

(iii) $\gamma : \mathbb{Z} \to G$ *is M-rational, and $(\gamma(n)\Gamma)_{n\in\mathbb{Z}}$ is periodic with period at most M*.

Remark. In words, this corollary asserts that in the quantitative setting, one can decompose

(arbitrary polynomial sequence)

$$= (\text{smooth}) \times (\text{totally equidistributed}) \times (\text{periodic}).$$

The notion of a subgroup G' being M-rational relative to a Mal'cev basis $\mathcal{X}$ will be defined in Definition 2.5. This result has some faint resemblance to the Szemerédi regularity lemma [30], although with the key difference that our bounds here are all polynomial in nature.

The derivation of Theorem 1.19 from Theorem 1.16 will be performed in Section 8–10.

We will use Theorem 1.19 in [12] in order to establish the *Möbius and Nilsequences conjecture* $\mathrm{MN}(s)$ from [11] for arbitrary step s. For this application, it is important that all bounds here are only polynomial in M and that the equidistribution is established on progressions of length linear in N (as opposed to N^c for some small $c > 0$).

Just as Corollary 1.12 implies a Ratner-type theorem, namely Corollary 1.13, it is not hard to deduce the following result from Theorem 1.19.

COROLLARY 1.20 (Ratner-type theorem for polynomial nilsequences). *Let $m, d \geqslant 0$, $0 < \delta < 1/2$, and $N \geqslant 1$. Suppose that G/Γ is an m-dimensional nilmanifold, that $G_\bullet$ is a filtration of degree d on G, and that $\mathcal{X}$ is a $1/\delta$-rational Mal'cev basis adapted to $G_\bullet$. Suppose that $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$. Then we may decompose $[N]$ as a union $P_1 \cup \cdots \cup P_k$ of arithmetic progressions with length $\gg \delta^{O_{m,d}(1)} N$ and the same common difference q, $1 \leqslant q \ll \delta^{-O_{m,d}(1)}$, such that each orbit $(g(n)\Gamma)_{n \in P_i}$ lies within δ (using the metric $d_\mathcal{X}$) of $x_i G' y_i \Gamma/\Gamma \subseteq G/\Gamma$, where $x_i \in G$, $y_i \in G$ is $\delta^{-O_{m,d}(1)}$-rational, and G' is a closed subgroup of G which is $\delta^{-O_{m,d}(1)}$-rational relative to $\mathcal{X}$. (This notion will be defined in the next section.)*

Remark. The reader may wish to compare this with [5], another recent result on quantitative variants of Ratner's theorem.

Let us conclude this introduction by remarking that our main theorem actually applies to *multiparameter* polynomial mappings $g : \mathbb{Z}^t \to G$. In the infinitary setting such a generalisation was obtained by Leibman [21], and his result has subsequently been applied in such papers as [2] and [23]. We have taken the trouble to derive multiparameter extensions of our main results with analogous finitary applications in mind; see Theorems 8.6 and 10.2.

2. Precise statements of results

In this section we define various "quantitative" concepts (such as Q-rational Mal'cev bases, subgroups which are Q-rational relative to such a basis, and the metrics $d_\mathcal{X}$ and $d_{G/\Gamma}$) which were needed to properly state the main results from the introduction section. We also give a more precise version of Theorem 1.16, which we will then spend the next several sections proving.

Mal'cev bases and metrics on G/Γ. The notion of *Mal'cev coordinates* play a vital role in the quantitative theory of nilmanifolds. They allow us to put a metric on G/Γ, which in turn allows us to define the notion of equidistribution; they also quantify the "rationality" of various objects associated to the nilmanifold. Mal'cev coordinates were introduced in [24], which contains a nice discussion; they are covered quite extensively in the book [4], particularly Chapters 1 and 5. We will also need several more quantitative statements about Mal'cev coordinates, which we have placed in Appendix A. We recommend that the reader dip into that appendix as and when required.

We will make use of the Lie algebra $\mathfrak{g}$ of G together with the exponential map $\exp : \mathfrak{g} \to G$. When G is a connected, simply-connected nilpotent Lie group, the exponential map is a diffeomorphism; see [4, Th. 1.2.1]. In particular, we have a logarithm map $\log : G \to \mathfrak{g}$. One does not really need to have an understanding of the exponential and logarithm maps beyond some of their formal properties, which we will list as we need them, in order to understand this paper.

Definition 2.1 (Mal'cev bases). Let G/Γ be a m-dimensional nilmanifold and let $G_\bullet$ be a filtration. A basis $\mathcal{X} = \{X_1, \ldots, X_m\}$ for the Lie algebra $\mathfrak{g}$ over $\mathbb{R}$ is called a *Mal'cev basis* for G/Γ adapted to $G_\bullet$ if the following four conditions are satisfied:

(i) For each $j = 0, \ldots, m-1$, the subspace $\mathfrak{h}_j := \mathrm{Span}(X_{j+1}, \ldots, X_m)$ is a Lie algebra ideal in $\mathfrak{g}$, and hence $H_j := \exp \mathfrak{h}_j$ is a normal Lie subgroup of G.

(ii) For every $0 \leqslant i \leqslant s$, we have $G_i = H_{m-m_i}$;

(iii) Each $g \in G$ can be written uniquely as $\exp(t_1 X_1) \exp(t_2 X_2) \ldots \exp(t_m X_m)$ for $t_i \in \mathbb{R}$;

(iv) Γ consists precisely of those elements which, when written in the above form, have all $t_i \in \mathbb{Z}$.

Remarks. Our main results only make sense if the nilmanifold G/Γ is already equipped with a Mal'cev basis $\mathcal{X}$, since they involve quantitative dependencies that can only be described using such a basis. However it is a well-known result of Mal'cev [24] that *any* nilmanifold G/Γ can be equipped with a Mal'cev basis adapted to the lower central series filtration. Indeed the very existence of a discrete and cocompact subgroup Γ guarantees that the lower central series is rational by [4, Th. 5.1.8 (a)] and [4, Cor. 5.2.2]. One may then apply [4, Prop. 5.3.2] to deduce the existence of a Mal'cev basis adapted to the lower central series. More generally there is a Mal'cev basis adapted to any filtration $G_\bullet$ which consists of rational subgroups (cf. Definition 1.10).

We refer to the t_i as the *Mal'cev coordinates* of g, and we define the *Mal'cev coordinate map* $\psi = \psi_{\mathcal{X}} : G \to \mathbb{R}^m$ to be the map

$$(2.1) \qquad\qquad \psi(g) := (t_1, \ldots, t_m),$$

thus for instance $\Gamma = \psi^{-1}(\mathbb{Z}^m)$. If $\mathcal{X}'$ is another Mal'cev basis (relative to some filtration), then we write $\psi' = \psi_{X'}$. Only very occasionally will we need to use the notation $\psi_{\mathcal{Y}}$ to indicate the coordinate map relative to some further basis $\mathcal{Y}$.

Remarks. In the literature, Mal'cev coordinates are invariably discussed in the context of the lower central series filtration and are referred to as *coordinates of the second kind*. Coordinates of the first kind or *exponential coordinates* are derived by writing $\log g \in \mathfrak{g}$ as a linear combination $\log g = s_1 X_1 + \cdots + s_m X_m$ of elements of the basis $\mathcal{X}$, and we write $\psi_{\exp}(g) = \psi_{\mathcal{X},\exp}(g) := (s_1, \ldots, s_m)$ for the coordinates of g obtained in this fashion. However, we shall mostly work using coordinates of the second kind.

We can use a Mal'cev basis $\mathcal{X}$ to put a (slightly artificial) metric structure on G and on G/Γ.

Definition 2.2 (Metrics on G and G/Γ). Let G/Γ be a nilmanifold with Mal'cev basis $\mathcal{X}$. We define $d = d_{\mathcal{X}} : G \times G \to \mathbb{R}_{\geqslant 0}$ to be the largest metric such that $d(x, y) \leqslant |\psi(xy^{-1})|$ for all $x, y \in G$, where $|\cdot|$ denotes the ℓ^{∞}-norm on $\mathbb{R}^m$. More explicitly, we have

$$d(x,y) = \inf\left\{ \sum_{i=0}^{n-1} \min(|\psi(x_{i-1}x_i^{-1})|, |\psi(x_i x_{i-1}^{-1})|) : x_0, \ldots, x_n \in G; x_0 = x; x_n = y \right\}.$$

This descends to a metric on G/Γ by setting

$$d(x\Gamma, y\Gamma) := \inf\{d(x', y') : x', y' \in G; x' \equiv x(\mathrm{mod}\ \Gamma); y' \equiv y(\mathrm{mod}\ \Gamma)\}.$$

It turns out that this is indeed[4] a metric on G/Γ; this essentially follows from the discreteness of Γ in G, and we will prove it in Lemma A.15. Since d is right-invariant, we also have

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

When the letter d is used for a metric, it will always denote the metric $d_{\mathcal{X}}$ relative to some basis $\mathcal{X}$ that is already under discussion. The symbol d' will be used for the metric defined using some other basis $\mathcal{X}'$. On the very rare occasions (for example in the proof of Lemma 7.4) where the metric relative to some further basis is under consideration, we will indicate this explicitly using subscripts.

Quantitative rationality. Now we define the concept of rational nilmanifolds and subgroups.

Definition 2.3 (Height). The *height* of a real number x is defined as $\max(|a|, |b|)$ if $x = a/b$ is rational in reduced form and ∞ if x is irrational.

Definition 2.4 (Rationality of a basis). Let G/Γ be a nilmanifold, and let $Q > 0$. We say that a Mal'cev basis $\mathcal{X}$ for G/Γ is Q-*rational* if all of the structure constants c_{ijk} in the relations

$$[X_i, X_j] = \sum_k c_{ijk} X_k$$

are rational with height at most Q.

Definition 2.5 (Rational subgroups). Suppose that a nilmanifold G/Γ is given together with a Mal'cev basis $\mathcal{X} = \{X_1, \ldots, X_m\}$ and that $Q > 0$.

---

[4]We note that this metric structure is a little more specific than in some of our previous papers, notably in [11, §8]. This will not cause any difficulty, as the metrics in that paper are equivalent to the one given here, up to constants depending on G, Γ, and $\mathcal{X}$. Indeed, at small scales, d agrees with the distance function given by the unique right-invariant Riemannian metric on G whose value at the origin is equal to that of the Euclidean metric at the origin of $\mathbb{R}^m$, pulled back by ψ; see also Lemma A.4.

Suppose that $G' \subseteq G$ is a closed connected subgroup. We say that G' is Q-*rational* relative to $\mathcal{X}$ if the Lie algebra $\mathfrak{g}'$ has a basis $\mathcal{X}' = \{X'_1, \ldots, X'_{m'}\}$ consisting of linear combinations $\sum_{i=1}^{m} a_i X_i$, where a_i are rational numbers with height at most Q for all i.

Definition 2.6 (Modulus of a horizontal character). Suppose that G/Γ is a nilmanifold with a Mal'cev basis $\mathcal{X}$. Suppose that $\eta : G \to \mathbb{R}/\mathbb{Z}$ is a horizontal character, that is to say a homomorphism from G to $\mathbb{R}/\mathbb{Z}$ which annihilates Γ. Then, when written in coordinates relative to $\mathcal{X}$, properties (iii) and (iv) of Proposition 2.1 imply that $\eta(g) = k \cdot \psi(g)$ for some unique $k \in \mathbb{Z}^m$. We write $|\eta| := |k|$.

Smooth polynomial sequences. For technical reasons it will be convenient to quantify the smoothness of sequences, such as the sequence $\varepsilon(n)$ appearing in Theorem 1.19, in a slightly different manner from that used so far.

Definition 2.7 (Smoothness norms). Suppose that $g : \mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is a polynomial sequence of degree d. Then g may be written uniquely as

$$g(n) = \alpha_0 + \alpha_1 \binom{n}{1} + \cdots + \alpha_d \binom{n}{d},$$

where α_i is in fact equal to $\partial^i g(0)$. For any $N > 0$, we define the *smoothness norm*

$$\|g\|_{C^\infty[N]} := \sup_{1 \leqslant j \leqslant d} N^j \|\alpha_j\|_{\mathbb{R}/\mathbb{Z}}.$$

The smoothness norm $\| \cdot \|_{C^\infty[N]}$ is designed to capture the notion of a polynomial sequence which is slowly-varying. Indeed, the following lemma is easily verified.

LEMMA 2.8 (Smooth polynomials vary slowly). *Let $g : \mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be a polynomial sequence of degree d, and let $N > 0$. Then for any $n \in [N]$, we have*

$$\|g(n) - g(n-1)\|_{\mathbb{R}/\mathbb{Z}} \ll_d \frac{1}{N} \|g\|_{C^\infty[N]}.$$

In view of this lemma, we see that Theorem 1.16 will be an immediate consequence of the following more precise statement. This is in fact the main technical result in our paper, and we will use it to derive all our other main results.

THEOREM 2.9 (Quantitative Leibman theorem). *Let $m, d \geqslant 0$, $0 < \delta < 1/2$, and $N \geqslant 1$. Suppose that G/Γ is an m-dimensional nilmanifold together with a filtration $G_\bullet$ and that $\mathcal{X}$ is a $\frac{1}{\delta}$-rational Mal'cev basis adapted to $G_\bullet$. Suppose that $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$. If $(g(n)\Gamma)_{n \in [N]}$ is not δ-equidistributed, then there is a horizontal character η with $0 < |\eta| \ll \delta^{-O_{m,d}(1)}$ such that*

$$\|\eta \circ g\|_{C^\infty[N]} \ll \delta^{-O_{m,d}(1)}.$$

Notes on reading the paper. As with so many papers, some parts of this work are merely technical and other parts represent deeper ideas of greater interest. There are quite a number of computations in this paper in which one has to show, say, that a certain integer is bounded polynomially by another, or that a certain basis is $O(\delta^{-O(1)})$-rational. All such computations are of the technical variety and should certainly be ignored on a first reading. They are all in a sense "clear;" their proofs proceed by algebra of a type which could hardly be expected to introduce nonpolynomial dependencies. It is possible that this could even be encoded in some relatively soft "proof-theoretic" language, but we have chosen not to follow such a path.

We begin with several sections containing motivating examples. In Section 3 we will discuss linear flows on tori $\mathbb{R}^m/\mathbb{Z}^m$, in Section 4 we shall discuss polynomial flows on $\mathbb{R}/\mathbb{Z}$, and in Section 5 we will look at linear flows on the 2-step Heisenberg nilmanifold (1.1). Some lemmas from these sections will be required in the sequel.

We then begin the study of the general case. In Section 6 we study the algebraic properties of polynomial sequences on nilpotent groups following Lazard and Leibman. There is a rich general theory here which is not evident from the study of the abelian and Heisenberg examples.

We then turn to the full proof of Theorem 2.9, the quantitative Leibman theorem. This is the technical heart of the paper and is given in the (rather long) Section 7.

In Section 8 use a straightforward iteration argument to bootstrap Theorem 2.9 to a multiparameter version of itself, namely Theorem 8.6. In Section 9 we then establish a preliminary multiparameter factorisation theorem, Proposition 9.2, which is a fairly short consequence of Theorem 8.6. In Section 10 we then iterate this proposition, obtaining a multiparameter theorem (Theorem 10.2) which then easily implies Theorem 1.19 (and hence Corollary 1.20) as special cases.

The appendix contains basic results on bases and nilmanifolds.

There is unfortunately a large amount of notation in this paper. In Figure 1 the key objects in the argument are briefly described.

3. **A quantitative Kronecker theorem**

In this section we prove Theorem 2.9 for linear sequences on the torus $\mathbb{R}^m/\mathbb{Z}^m$; that is to say we establish a quantitative Kronecker theorem. The methods and the result are very standard.

PROPOSITION 3.1 (Quantitative Kronecker Theorem). *Let $m \geqslant 1$, let $0 < \delta < 1/2$, and let $\alpha \in \mathbb{R}^m$. If the sequence $(\alpha n (\mathrm{mod}\ \mathbb{Z}^m))_{n \in [N]}$ is not δ-equidistributed in the additive torus $\mathbb{R}^m/\mathbb{Z}^m$, then there exists $k \in \mathbb{Z}^m$ with $0 < |k| \ll \delta^{-O_m(1)}$ such that $\|k \cdot \alpha\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O_m(1)}/N$.*

G	nilpotent group	Def. 1.1
$G_\bullet = (G_i)_{i=0}^\infty$	filtration on G	Def. 1.1
G/Γ	nilmanifold	Def. 1.1
$(G/\Gamma)_{\mathrm{ab}} = G/[G,G]\Gamma$	horizontal torus	Def. 1.5
$G_d/(\Gamma \cap G_d) \cong \mathbb{R}^{m_d}/\mathbb{Z}^{m_d}$	vertical torus	Def. 3.3
$d \geqslant 0$	degree of the filtration $G_\bullet$	Def. 1.1
$s \geqslant 0$	step of G	Def. 1.1
$m \geqslant 0$	dimension of G	Def. 1.1
m_i	dimension of G_i	Def. 1.1
m_{ab}	dimension of horizontal torus	Def. 1.5
m_{lin}	$m - m_2$	§7
$m_* = m_{\mathrm{ab}} - m_{\mathrm{lin}}$	nonlinearity degree of $G_\bullet$	§7
$\eta : G \to \mathbb{R}/\mathbb{Z}$	horizontal character	Def. 1.5
$\xi : G_d \to \mathbb{R}/\mathbb{Z}$	vertical character	Def. 3.4
$\mathcal{X} = (X_i)_{i=1}^m,\ \mathcal{X}' = (X_i')_{i=1}^m$	Mal'cev bases	Def. 2.1
ψ, ψ'	coordinate maps relative to $\mathcal{X}, \mathcal{X}'$	(2.1)
d, d'	metrics defined using $\mathcal{X}, \mathcal{X}'$	Def. 2.2
$Q \geqslant 1$	rationality bound for $\mathcal{X}$ (usually $Q = 1/\delta$)	Def. 2.4
$\pi : G \to (G/\Gamma)_{\mathrm{ab}}$	projection onto the horizontal torus	Def. 1.5
$F : G/\Gamma \to \mathbb{C}$	Lipschitz function	Def. 2.2
$0 < \delta < 1/2$	level of equidistribution	Def. 1.2
$N \geqslant 1$	length of sequence	Def. 1.2
$g : \mathbb{Z} \to G$	a polynomial sequence	Def. 1.8
$\mathrm{poly}(\mathbb{Z}, G_\bullet)$	polynomial sequences with coeffs in $G_\bullet$	Def. 1.8
$t \geqslant 1$	number of parameters	§8

Figure 1. A list of key objects in the paper, together with brief descriptions of these objects, and the location where they are first defined or introduced.

Remark. We leave it to the reader to check that this really *is* the specialisation of Theorem 2.9 to the case of linear orbits on the torus $\mathbb{R}^m/\mathbb{Z}^m$. This may be found helpful in understanding some of our notation. Note in particular that in this case the horizontal torus is simply $\mathbb{R}^m/\mathbb{Z}^m$, and we may take π to be the identity map.

Proof. By Definition 1.2, there is a Lipschitz function $F : \mathbb{R}^m/\mathbb{Z}^m \to \mathbb{R}$ such that

$$(3.1) \qquad |\mathbb{E}_{n \in [N]} F(\alpha n (\mathrm{mod}\ \mathbb{Z}^m)) - \int_{\mathbb{R}^m/\mathbb{Z}^m} F\, d\theta| > \delta \|F\|_{\mathrm{Lip}}.$$

At the expense of replacing δ by $\delta/2$ we may translate F, add a constant to it, and rescale in such a way that $\int F = 0$ and $\|F\|_{\mathrm{Lip}} = 1$. By approximating F by smooth functions we may assume that F is smooth (we do this to avoid

any technical issues regarding convergence of Fourier series). We now use a standard manœuvre to approximate F by a function which has finite support in frequency space (cf. [10, Lemma A.9]).

Consider the Fejér kernel $K : \mathbb{R}^m/\mathbb{Z}^m \to \mathbb{R}^+$ defined by

$$K(\theta) := \frac{1}{\operatorname{mes}(Q)} 1_Q * \frac{1}{\operatorname{mes}(Q)} 1_Q(\theta),$$

where $Q := [-\frac{\delta}{16m}, \frac{\delta}{16m}]^m \subset \mathbb{R}^m/\mathbb{Z}^m$ is a small cube and $*$ denotes the usual convolution operation on the torus $\mathbb{R}^m/\mathbb{Z}^m$. It is immediate that K is a non-negative function supported in Q with

$$(3.2) \qquad \int_{\mathbb{R}^m/\mathbb{Z}^m} K = 1.$$

A simple calculation also establishes the estimate

$$(3.3) \qquad \sum_{k \in \mathbb{Z}^m : |k| \geqslant M} |\widehat{K}(k)| \ll_m \delta^{-2m} M^{-1}$$

for all $M > 1$, where the Fourier coefficient is defined by

$$\widehat{K}(k) := \int_{\mathbb{R}^m/\mathbb{Z}^m} K(\theta) e(-\theta \cdot k) \, d\theta$$

and $e(x) := e^{2\pi i x}$ is the standard character on $\mathbb{R}/\mathbb{Z}$. We also have the crude bound

$$(3.4) \qquad |\widehat{F}(k)| \leqslant \|F\|_\infty \leqslant \|F\|_{\operatorname{Lip}} \leqslant 1$$

for all $k \in \mathbb{Z}^m$.

Set $F_1 := F * K$. Since $\|F\|_{\operatorname{Lip}} = 1$, and K is supported in Q and satisfies (3.2), a standard computation shows that

$$\|F - F_1\|_\infty \leqslant \delta/8.$$

Choose $M := C_m \delta^{-2m-1}$ for some suitably large C_m, and set

$$F_2(\theta) := \sum_{k \in \mathbb{Z}^m : 0 < |k| \leqslant M} \widehat{F_1}(k) e(k \cdot \theta).$$

Noting that $\widehat{F_1}(0) = 0$, facts (3.3), (3.4), and the Fourier inversion formula imply that

$$\|F_1 - F_2\|_\infty \leqslant \delta/8.$$

It follows that $\|F - F_2\|_\infty \leqslant \delta/4$, which means, in view of the failure of (3.1), that

$$|\mathbb{E}_{n \in [N]} F_2(n\alpha \mathbb{Z}^m)| \geqslant \delta/4.$$

Applying (3.4) once more we see that there is some k, $0 < |k| \leqslant M$, such that

$$|\mathbb{E}_{n \in [N]} e(nk \cdot \alpha)| \gg_m \delta M^m \gg \delta^{O_m(1)}.$$

The result now follows immediately from the standard estimate

$$|\mathbb{E}_{n\in[N]}e(nt)| \ll \min\left(1, \frac{1}{N\|t\|_{\mathbb{R}/\mathbb{Z}}}\right),$$

which follows from summing the geometric progression. □

Let us now record a corollary of the $m = 1$ version of this result which will be used several times in the sequel. This gives stronger information in the case that $(n\alpha(\mathrm{mod}\ \mathbb{Z}))_{n\in[N]}$ is *very* far from being equidistributed.

LEMMA 3.2 (Strongly recurrent linear functions are highly non-diophantine). *Let $\alpha \in \mathbb{R}$, $0 < \delta < 1/2$, and $0 < \epsilon \leqslant \delta/2$, and let $I \subseteq \mathbb{R}/\mathbb{Z}$ be an interval of length ϵ such that $\alpha n \in I$ for at least δN values of $n \in [N]$. Then there is some $k \in \mathbb{Z}$ with $0 < |k| \ll \delta^{-O(1)}$ such that $\|k\alpha\|_{\mathbb{R}/\mathbb{Z}} \ll \epsilon\delta^{-O(1)}/N$.*

Proof. Taking F to be a Lipschitz approximation to the interval I, we see immediately that our assumption precludes $(\alpha n(\mathrm{mod}\ \mathbb{Z}))_{n\in[N]}$ from being δ^{10}-equidistributed. It follows from the case $m = 1$ of Proposition 3.1 that there is some $k \in \mathbb{Z}$, $|k| \ll \delta^{-C}$, such that $\|k\alpha\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-C}/N$, where $C = O(1)$. Write $\beta := \|q\alpha\|_{\mathbb{R}/\mathbb{Z}}$. Let $n_0 \in \mathbb{Z}$ be arbitrary, and suppose that n' ranges over any interval of integers J of length at most $1/\beta$. The number of n' for which $\alpha(n_0 + qn')\mathbb{Z} \in I$ is then at most $1 + \epsilon/\beta$. Since $[N]$ may be divided into $\leqslant 2q + \beta N$ progressions of the form $\{n_0 + qn' : n' \in J\}$, we obtain from our assumption the inequality

$$(3.5) \quad \delta N \leqslant \#\{n \in [N] : \alpha n \mathbb{Z} \in I\} \leqslant (1 + \frac{\epsilon}{\beta})(2q + \beta N) \ll q + \frac{\epsilon q}{\beta} + \beta N + \epsilon N.$$

Now the lemma is trivial if $N \ll \delta^{-10C}$ and follows immediately from Proposition 3.1 when $\epsilon \gg \delta^{10C}$, so suppose that neither of these is the case. Then all of the terms except the second on the right-hand side of (3.5) are negligible, and we deduce that

$$\delta N \ll q\epsilon/\beta.$$

This immediately implies the result. □

The main idea in the proof of Proposition 3.1, of course, was that the space of Lipschitz functions is essentially spanned by the space of pure phase functions $e(k \cdot \theta)$. Thus we were able to assert that if the condition (3.1) fails for some F, then it also fails (albeit with a smaller value of δ) for a pure phase function with not-too-large frequency.

A similar observation turns out to be essential in the analysis of polynomial sequences on general nilmanifolds G/Γ (cf. the proof of [22, Th. 2.17]). Though we will not be discussing general sequences for quite a while, this does seem to be an appropriate place to state and prove a lemma which generalises the

observations just made. For this, we will be working primarily on the vertical torus.

Definition 3.3 (Vertical torus). Suppose that G/Γ is a nilmanifold and that $G_\bullet$ is a filtration of degree d. Note that G_d then lies in the centre of G. We define the *vertical torus* to be $G_d/(\Gamma \cap G_d)$ and the *vertical dimension* m_d to be $m_d := \dim G_d$; the last m_d coordinates of the Mal'cev coordinate map ψ may be used to canonically identify G_d and $G_d/(\Gamma \cap G_d)$ with $\mathbb{R}^{m_d}$ and $\mathbb{R}^{m_d}/\mathbb{Z}^{m_d}$ respectively. Also observe that the vertical torus acts canonically on the nilmanifold G/Γ, thus we can define[5] $\theta y \in G/\Gamma$ for all $\theta \in \mathbb{R}^{m_d}/\mathbb{Z}^{m_d}$ and $y \in G/\Gamma$.

Definition 3.4 (Vertical characters). A *vertical character* is a continuous homomorphism $\xi : G_d \to \mathbb{R}/\mathbb{Z}$ such that $\Gamma \cap G_d \subseteq \ker \xi$ (in particular, ξ can also be meaningfully defined on $G_d/\Gamma_d \cong \mathbb{R}^{m_d}/\mathbb{Z}^{m_d}$). Any such character has the form $\xi(x) = k \cdot x$ for a unique $k \in \mathbb{Z}^{m_d}$, where we identify G_d with $\mathbb{R}^{m_d}$. We refer to k as the *frequency* of the character ξ and $|\xi| := |k|$ as the *frequency magnitude*. For instance the trivial character $\xi \equiv 0$ has frequency 0.

Definition 3.5 (Vertical oscillation). Let $F : G/\Gamma \to \mathbb{C}$ be a Lipschitz function and suppose that ξ is a vertical character. We say that F has *vertical oscillation* ξ if we have $F(g_d \cdot x) = e(\xi(g_d))F(x)$ for all $g_d \in G_d$ and $x \in G/\Gamma$.

The next definition is a repetition of Definition 1.2, except that we specialise to functions with a fixed vertical oscillation ξ.

Definition 3.6 (Equidistribution along a vertical character). Let $g : \mathbb{Z} \to G$ be a polynomial sequence. We say that $(g(n)\Gamma)_{n \in [N]}$ is δ-*equidistributed* along a vertical character ξ if

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

for all Lipschitz functions $F : G/\Gamma \to \mathbb{C}$ with vertical oscillation ξ.

The next lemma states that in order to check whether a sequence is equidistributed, it suffices to test that sequence against functions possessing a vertical oscillation.

LEMMA 3.7 (Vertical oscillation reduction). *Let G/Γ be a nilmanifold together with a filtration $G_\bullet$ of degree d. Let m_d be as above, and let $0 < \delta < 1/2$. Suppose that $g : \mathbb{Z} \to G$ is a polynomial sequence and that $(g(n)\Gamma)_{n \in [N]}$ is not*

---

[5]Here we have a slight clash between the additive notation for the torus $\mathbb{R}^{m_d}/\mathbb{Z}^{m_d}$ and the multiplicative notation for the group G. We hope this will not confuse the reader.

δ-*equidistributed. Then there is a vertical character ξ with $|\xi| \ll \delta^{-O_{m_d}(1)}$ such that $(g(n)\Gamma)_{n \in [N]}$ is not $\delta^{O_{m_d}(1)}$-equidistributed along the vertical oscillation ξ.*

Proof. We merely sketch this, for the argument is little more than a repetition of that used to prove Proposition 3.1. We begin with the same reductions. That is, assuming the existence of an $F : G/\Gamma \to \mathbb{C}$ such that

$$(3.6) \qquad \left| \mathbb{E}_{n \in [N]} F(g(n)\Gamma) - \int_{G/\Gamma} F \right| \geqslant \delta \|F\|_{\mathrm{Lip}},$$

we weaken δ to $\delta/2$ and assume that $\int_{G/\Gamma} F = 0$, that $\|F\|_{\mathrm{Lip}} = 1$, and that F is smooth.

Let K be the same Fejér-type kernel as before, and now take $F_1 : G \to \mathbb{C}$ to be the function obtained by convolving with K in each $G_d/(\Gamma \cap G_d) \cong \mathbb{R}^{m_d}/\mathbb{Z}^{m_d}$-fibre; that is to say,

$$F_1(y) := \int_{\mathbb{R}^{m_d}/\mathbb{Z}^{m_d}} F(\theta y)K(\theta)d\theta.$$

Fourier expansion on $\mathbb{R}^{m_d}/\mathbb{Z}^{m_d}$ gives

$$F_1(y) = \sum_{k \in \mathbb{Z}^{m_d}} F^{\wedge}(y; k)\widehat{K}(k),$$

where

$$F^{\wedge}(y; k) := \int_{\mathbb{R}^{m_d}/\mathbb{Z}^{m_d}} F(\theta y)e(-k \cdot \theta)d\theta.$$

Now for $g_d \in G_d \cong \mathbb{R}^{m_d}$, we have

$$F^{\wedge}(g_d y; k) = \int F((\theta + g_d)y)e(-k \cdot \theta)\, d\theta = e(k \cdot g_d)F^{\wedge}(y; f);$$

thus each function $F^{\wedge}(y; k)$ has vertical oscillation ξ, where $\xi(x) := k \cdot x$ is the vertical character with frequency k.

Using exactly the same estimates as in the proof of Proposition 3.1, we have $\|F - F_2\|_{\infty} \leqslant \delta/4$, where

$$F_2(y) := \sum_{k \in \mathbb{Z}^{m_d} : |k| \leqslant Q} F^{\wedge}(y; k)\widehat{K}(k)$$

for some $Q = C_{m_d}\delta^{-2m_d-1}$. The rest of the argument proceeds exactly as before, and we see that if we take $\tilde{F}(y) := F^{\wedge}(y; k)$ for suitable $k \in \mathbb{Z}^{m_d}$, $|k| \ll \delta^{-O_{m_d}(1)}$, we have

$$\left| \mathbb{E}_{n \in [N]} \tilde{F}(g(n)\Gamma) - \int_{G/\Gamma} \tilde{F} \right| \gg \delta^{O_{m_d}(1)} \|\tilde{F}\|_{\mathrm{Lip}}.$$

Thus $(g(n)\Gamma)_{n \in [N]}$ is not $\delta^{O_{m_d}(1)}$-equidistributed along the vertical character ξ, as desired. $\qquad \square$

4. **The van der Corput trick and polynomial flows on tori**

In the last section we introduced one important trick — the idea of decomposing a Lipschitz function into phases using Fourier analysis. In this section we introduce a second trick - namely, the use of van der Corput's inequality — and use this trick to study polynomial sequences on tori $\mathbb{R}^m/\mathbb{Z}^m$. Although our language is somewhat different, this is really just a *reprise* of the standard theory of Weyl sums as used, for instance, in the study of Waring's problem (see, for example, [33]).

LEMMA 4.1 (van der Corput inequality). *Let N, H be positive integers and suppose that $(a_n)_{n\in[N]}$ is a sequence of complex numbers. Extend (a_n) to all of $\mathbb{Z}$ by defining $a_n := 0$ when $n \notin [N]$. Then*

$$|\mathbb{E}_{n\in[N]}a_n|^2 \leqslant \frac{N+H}{HN} \sum_{|h|\leqslant H} \left(1 - \frac{|h|}{H}\right) \mathbb{E}_{n\in[N]}a_n\overline{a_{n+h}}.$$

Proof. We have

$$\sum_n a_n = \frac{1}{H} \sum_{-H<n\leqslant N} \sum_{h=0}^{H-1} a_{n+h}.$$

Thus, applying the Cauchy-Schwarz inequality, we have

$$\left|\sum_n a_n\right|^2 = \frac{1}{H^2}\left|\sum_{-H<n\leqslant N} \sum_{h=0}^{H-1} a_{n+h}\right|^2$$

$$\leqslant \frac{N+H}{H^2} \sum_{-H<n\leqslant N} \left|\sum_{h=0}^{H-1} a_{n+h}\right|^2$$

$$= \frac{N+H}{H^2} \sum_{-H<n\leqslant N} \sum_{h=0}^{H-1} \sum_{h'=0}^{H-1} a_{n+h}\overline{a}_{n+h'},$$

which is equivalent to the right-hand side of the claimed inequality. $\square$

We will use the following simple (and rather crude) corollary of this, which we phrase in the contrapositive.

COROLLARY 4.2 (van der Corput). *Let N be a positive integer, and suppose that $(a_n)_{n\in[N]}$ is a sequence of complex numbers with $|a_n| \leqslant 1$. Extend (a_n) to all of $\mathbb{Z}$ by defining $a_n := 0$ when $n \notin [N]$. Suppose that $0 < \delta < 1$ and that*

$$|\mathbb{E}_{n\in[N]}a_n| \geqslant \delta.$$

Then for at least $\delta^2 N/8$ values of $h \in [N]$, we have

$$|\mathbb{E}_{n\in[N]}a_{n+h}\overline{a_n}| \geqslant \delta^2/8.$$

Proof. The result is vacuous if $N \leqslant 4/\delta^2$, so assume this is not the case. Suppose for a contradiction that the result is false. Apply Lemma 4.1 with $H = N$. Then it is easy to see that we have

$$\delta^2 \leqslant |\mathbb{E}_{n \in [N]} a_n| \leqslant \frac{2}{N} \sum_{|h| \leqslant N} |\mathbb{E}_{n \in [N]} a_n \overline{a_{n+h}}| \leqslant \frac{2}{N} \left(1 + 2 \left(\frac{\delta^2 N}{8} + \frac{\delta^2 N}{8} \right) \right),$$

where we have used the trivial estimate $|\mathbb{E}_{n \in [N]} a_n \overline{a_{n+h}}| \leqslant 1$ for those $h \in [N]$ such that $|\mathbb{E}_{n \in [N]} a_n \overline{a_{n+h}}| \geqslant \delta^2/8$, of which there are no more than $\delta^2 N/8$. Rearranging and using the fact that $N > 4/\delta^2$ we see that this is a contradiction. $\qquad \square$

The next proposition is the main result of this section and is Theorem 2.9 in the case $G = \mathbb{R}$, $\Gamma = \mathbb{Z}$, and with $g : \mathbb{Z} \to G$ an arbitrary polynomial.

PROPOSITION 4.3 (Weyl). *Suppose that $g : \mathbb{Z} \to \mathbb{R}$ is a polynomial of degree d, and let $0 < \delta < 1/2$. Then either $(g(n) (\mathrm{mod}\ \mathbb{Z}))_{n \in [N]}$ is δ-equidistributed, or else there is an integer k, $1 \leqslant k \ll \delta^{-O_d(1)}$, such that $\|kg (\mathrm{mod}\ \mathbb{Z})\|_{C^\infty[N]} \ll \delta^{-O_d(1)}$.*

We will deduce this from the following, which is nothing but a reformulation of Weyl's exponential sum estimate (see, e.g., [33]).

LEMMA 4.4 (Weyl's exponential sum estimate). *Suppose that $g : \mathbb{Z} \to \mathbb{R}$ is a polynomial of degree d with leading coefficient α_d and that*

$$|\mathbb{E}_{n \in [N]} e(g(n))| \geqslant \delta$$

for some $0 < \delta < 1/2$. Then there is $k \in \mathbb{Z}$, $|k| \ll \delta^{-O_d(1)}$, such that

$$\|k\alpha_d\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O_d(1)}/N^d.$$

Proof. We proceed by induction on d, the result having been established in Section 3 in the case $d = 1$. We may assume that $N > \delta^{-C'_d}$ for some large C'_d since the result is trivial otherwise. Applying van der Corput's estimate in the form of Corollary 4.2, we deduce that there are $\gg \delta^2 N$ values of $h \in [N]$ such that

$$|\mathbb{E}_{n \in [N]} e(g(n+h) - g(n))| \gg \delta^2.$$

For each such h, $g(n+h) - g(n)$ is a polynomial with degree $d-1$ and leading coefficient $hd\alpha_d$. Thus by the induction hypothesis there is, for $\gg \delta^2$ values of $h \in [N]$, some $1 \leqslant q_h \ll \delta^{-O_d(1)}$ such that we have

$$\|hq_h d\alpha_d\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O_d(1)}/N^{d-1}$$

for each of these values of h. Pigeonholing in the q_h, this implies that there is q, $1 \leqslant q \ll \delta^{-O_d(1)}$, such that

$$\|h\alpha_d\|_{\mathbb{R}/\mathbb{Z}, \delta^{-O_d(1)}} \ll \delta^{-O_d(1)}/N^{d-1}$$

for $\gg \delta^{O_d(1)}N$ values of $h \in [N]$. Since N is so large, Lemma 3.2 may applied to conclude that there is $q' \ll \delta^{-O_d(1)}$ such that

$$\|qq'\alpha_d\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O_d(1)}/N^d.$$

Redefining $q := qq'$, the result follows. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Proposition 4.3. In this proof we allow all implied constants to depend on d. Suppose that $g : \mathbb{Z} \to \mathbb{R}$ is a polynomial sequence of degree d such that the orbit $(g(n)\mathbb{Z})_{n \in [N]}$ on $\mathbb{R}/\mathbb{Z}$ is not δ-equidistributed. Expand g as a Taylor series

$$(4.1) \qquad\qquad g(n) = \binom{n}{d}\alpha_d + \cdots + \binom{n}{1}\alpha_1 + \alpha_0,$$

and suppose as a hypothesis for induction on r, $0 \leqslant r < d$, that we have shown that each of the coefficients $\alpha_d, \alpha_{d-1}, \ldots, \alpha_{d-r}$ is nearly rational in the sense that $\|q\alpha_{d-i}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N^{d-i}$ for some $q \ll \delta^{-O(1)}$ for $i = 0, \ldots, r$. (The implied constants in the $O()$ notation may increase with each induction step, but there are only d such steps, and we are allowing these constants to depend on d, so this is harmless.) The statement we are trying to prove, Proposition 4.3, is the case $r = d - 1$.

Now by the argument used in proving Proposition 3.1 (or indeed by simply quoting Lemma 3.7), there is $k \in \mathbb{Z}$, $0 < |k| \ll \delta^{-O(1)}$, such that

$$(4.2) \qquad\qquad |\mathbb{E}_{n \in [N]}e(kg(n))| \gg \delta^{O(1)}.$$

The base case $r = 0$ of the induction follows immediately from Lemma 4.4. Suppose now that we have established the result for some r and wish to establish it for $r + 1$. Set

$$g'(n) := g(n) - \binom{n}{d}\alpha_d - \cdots - \binom{n}{d-r}\alpha_{d-r} = \binom{n}{d-r-1}\alpha_{d-r-1} + \cdots + \alpha_0.$$

Set $Q := qd!$, and write $\alpha_{d-i} = a_{d-i}/q + O(\delta^{-O(1)}/N^{d-i})$, $i = 0, \ldots, r$ for some integers a_{d-i}. For any $n_0 \in \mathbb{Z}$ for any $n' \in \mathbb{Z}$, we have

$$g'(n_0 + Qn') - g'(n_0)$$
$$= g(n_0 + Qn') - g(n_0) - \frac{1}{q}\sum_{i=0}^{r} a_{d-i}\left[\binom{n_0 + Qn'}{d-i} - \binom{n_0}{d-i}\right]$$
$$+ O(\delta^{-O(1)})\sum_{i=0}^{r} \frac{1}{N^{d-i}}\left[\binom{n_0 + Qn'}{d-i} - \binom{n_0}{d-i}\right].$$

Set $N' := \lfloor \delta^{C'_d}N \rfloor$ for some suitably large C'_d and suppose that $n' \in [N']$ and also that $|n_0| \leqslant 2N$. Then the last term here is $O(\delta^{C'_d - O(1)})$. The first term is

an integer, since

$$\binom{n_0+Q}{j} - \binom{n_0}{j} = \sum_{i=1}^{j}\binom{Q}{i}\binom{n_0}{j-i} \equiv 0 (\mathrm{mod}\ q)$$

for all $j \leqslant d$. Thus we see that if $n' \in [N']$ and $|n_0| \leqslant 2N$, then

$$(4.3) \quad g'(n_0+Qn') - g'(n_0) = g(n_0+Qn') - g(n_0) + O(\delta^{C'_d-O(1)})(\mathrm{mod}\ \mathbb{Z}).$$

Splitting $[N]$ into progressions of common difference Q and length $[N']$ plus a negligible error, we see from (4.2) that there is n_0, $|n_0| \leqslant 2N$, such that

$$|\mathbb{E}_{n'\in[N']} e(kg(n_0+Qn'))| \gg \delta^{O(1)}.$$

It follows from (4.3) that

$$|\mathbb{E}_{n'\in[N']} e(kg'(n_0+Qn'))| \gg \delta^{O(1)}.$$

By Lemma 4.4 we see that the leading coefficient

$$\alpha' := kQ^{d-r-1}\alpha_{d-r-1}/(d-r-1)!$$

of this polynomial is nearly rational in the sense that there is $1 \leqslant q' \ll \delta^{-O(1)}$ such that $\|q'\alpha'\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N^{d-r-1}$. It follows that there is $1 \leqslant q'' \ll \delta^{-O(1)}$ such that $\|q''\alpha_{d-r-1}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N^{d-r-1}$. Setting $\tilde{q} := qq''$ we now clearly have $1 \leqslant \tilde{q} \ll \delta^{-O(1)}$ and also $\|\tilde{q}\alpha_{d-i}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N^{d-i}$ for $i = 0, \ldots, r+1$.

This concludes the proof of the inductive step and hence of the proposition.

We will also need a "strong recurrence" result for polynomials $g : \mathbb{Z} \to \mathbb{R}$, generalizing the linear result, Lemma 3.2, that we obtained in the last section. This is in fact an easy deduction from Proposition 4.3 and Lemma 3.2.

LEMMA 4.5 (Strongly recurrent polynomials are highly non-diophantine). *Let $d \geqslant 0$, and suppose that $g : \mathbb{Z} \to \mathbb{R}$ is a polynomial sequence of degree d. Suppose that $0 < \delta < 1/2$ and $\epsilon \leqslant \delta/2$, that $I \subseteq \mathbb{R}/\mathbb{Z}$ is an interval of length ϵ, and that $g(n)(\mathrm{mod}\ \mathbb{Z}) \in I$ for at least δN values of $n \in [N]$. Then there is a $k \in \mathbb{Z}$, $0 < |k| \ll \delta^{-O_d(1)}$, such that $\|kg(\mathrm{mod}\ \mathbb{Z})\|_{C^\infty[N]} \ll \epsilon\delta^{-O_d(1)}$.*

Proof. In this proof we allow all implied constants to depend on d. If $\epsilon \gg \delta^{C_d}$ for some large C_d depending only on d, then the result follows immediately from Proposition 4.3, so assume this is not the case. Expand g in a Taylor series as in (4.1), with coefficients $\alpha_0, \ldots, \alpha_d$. It follows from the assumption that none of the polynomials λg, $\lambda \leqslant \delta/2\epsilon$, is $\delta^{O(1)}$-equidistributed on $[N]$. Thus by Proposition 4.3 we have that for each $\lambda \leqslant \delta/2\epsilon$, there is $q_\lambda \ll \delta^{-O(1)}$ such that $\|q_\lambda\lambda\alpha_i\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N^i$ for $i = 0, \ldots, d$. Pigeonholing in the possible values of q_λ we see that there is $q \ll \delta^{-O(1)}$ such that for $\gg \delta^{O(1)}/\epsilon$ values of $\lambda \leqslant \delta/2\epsilon$ we have $\|\lambda q\alpha_i\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N^i$ for each $i = 0, \ldots, d$. It follows from Lemma 3.2 that for each i there is $q_i \ll \delta^{-O(1)}$ such that $\|q_i\alpha_i\|_{\mathbb{R}/\mathbb{Z}} \ll \epsilon\delta^{-C_d}/N^i$. Writing $\tilde{q} := q_1 \ldots q_d$ we see that $\tilde{q} \ll \delta^{-O(1)}$ and that $\|q\alpha_i\|_{\mathbb{R}/\mathbb{Z}} \ll \epsilon\delta^{-O(1)}/N^i$ for all i. This concludes the proof of the proposition. $\square$

5. **The Heisenberg example**

In this section we discuss the first example which is not just a rephrasing of classical work on equidistribution, establishing Theorem 2.9 for a linear sequence on the Heisenberg nilmanifold (1.1); thus $s = d = 2$ and $m = 3$. Strictly speaking, this section is not necessary in order to prove Theorem 2.9 in the general case, however we present this "worked example" here in order to illustrate the key ideas of the main argument in a simplified model setting. (Also, a key computation in this setting, namely Proposition 5.3, will be reused in the main argument.) As in the preceding section, the idea is to use van der Corput's inequality to reduce the problem to a simpler problem and, in particular, to reduce to a "1-step" or "abelian" problem that can be treated by the tools of the previous section. This turns out to work, but it will take a certain amount of algebraic manipulation to see the 1-step structure emerge from van der Corput's inequality applied to the 2-step Heisenberg situation.

Let us begin with a brief tour of the Heisenberg example (1.1). We have $\mathfrak{g} = \left(\begin{smallmatrix} 0 & \mathbb{R} & \mathbb{R} \\ 0 & 0 & \mathbb{R} \\ 0 & 0 & 0 \end{smallmatrix} \right)$, with the exponential map being given by

$$\exp \left(\begin{smallmatrix} 0 & x & y \\ 0 & 0 & z \\ 0 & 0 & 0 \end{smallmatrix} \right) = \left(\begin{smallmatrix} 1 & x & y + \frac{1}{2}xz \\ 0 & 1 & z \\ 0 & 0 & 1 \end{smallmatrix} \right)$$

and the logarithm map by

$$\log \left(\begin{smallmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{smallmatrix} \right) = \left(\begin{smallmatrix} 0 & x & y - \frac{1}{2}xz \\ 0 & 0 & z \\ 0 & 0 & 0 \end{smallmatrix} \right).$$

Observe that $\log \Gamma$ is not quite a lattice in $\mathbb{R}^3$, although it is a finite union of lattices.

Consider the elements $X_1, X_2, X_3 \in \mathfrak{g}$, defined by $X_1 := \left(\begin{smallmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix} \right)$, $X_2 := \left(\begin{smallmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{smallmatrix} \right)$, and $X_3 := \left(\begin{smallmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{smallmatrix} \right)$. It is easy to see that $\mathcal{X} = \{X_1, X_2, X_3\}$ is a Mal'cev basis adapted to the lower central series filtration $G_\bullet$. A simple computation confirms that

$$\exp(t_1 X_1) \exp(t_2 X_2) \exp(t_3 X_3) = \left(\begin{smallmatrix} 1 & t_1 & t_1 t_2 + t_3 \\ 0 & 1 & t_2 \\ 0 & 0 & 1 \end{smallmatrix} \right),$$

and so the Mal'cev coordinate map $\psi_{\mathcal{X}} : G \to \mathbb{R}^3$ is given by

$$\psi_{\mathcal{X}} \left(\begin{smallmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{smallmatrix} \right) = (x, z, y - xz).$$

The horizontal torus is isomorphic to $(\mathbb{R}/\mathbb{Z})^2$, and the projection $\pi : G \to (\mathbb{R}/\mathbb{Z})^2$ is given by $\pi \left(\begin{smallmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{smallmatrix} \right) = (x, z)$.

We shall be working through the special case of Theorem 2.9 in the case when $g : \mathbb{Z} \to G$ is a linear sequence. To simplify the exposition very slightly we will assume that this sequence has no constant term, thus $g(n) = a^n$ for

some $a \in G$. Note that $g \in \text{poly}(\mathbb{Z}, G_\bullet)$, where $G_\bullet$ is the lower central series filtration. Thus the sequence g has degree 2.

PROPOSITION 5.1 (Main theorem, Heisenberg case). *Let G/Γ be the 2-step Heisenberg nilmanifold with the Mal'cev basis $\mathcal{X}$ described above, and let g : $\mathbb{Z} \to G$ be a linear sequence of the form $g(n) = a^n$. Let $\delta > 0$ be a parameter, and let $N \geqslant 1$ be an integer. Then either $(g(n)\Gamma)_{n \in [N]}$ is δ-equidistributed, or else there is a horizontal character η with $0 < |\eta| \ll \delta^{-O(1)}$ such that $\|\eta(a)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N$.*

Remark. Note that, since $g(n)$ is linear, the last condition here is equivalent to the statement that $\|\eta \circ g\|_{C^\infty[N]} \ll \delta^{-O(1)}$.

Proof. By Lemma 3.7 we may assume that there is a function $F : G/\Gamma \to \mathbb{C}$ with a vertical oscillation ξ with $\|\xi\| \ll \delta^{-O(1)}$, and $\|F\|_{\text{Lip}} = 1$, such that

$$(5.1) \qquad \left| \mathbb{E}_{n \in [N]} F(a^n \Gamma) - \int_{G/\Gamma} F \right| \gg \delta^{O(1)}.$$

We split into two cases: $\xi \equiv 0$ and $\xi \not\equiv 0$.

If $\xi \equiv 0$, then F is G_2-invariant, which means we may factor through π to get a function $\tilde{F} : \mathbb{R}^2/\mathbb{Z}^2 \to \mathbb{C}$ defined by

$$F(x) = \tilde{F}(\pi(x)).$$

It is clear that $\|\tilde{F}\|_{\text{Lip}} \leqslant 1$. Equation (5.1) implies that

$$|\mathbb{E}_{n \in [N]} \tilde{F}(n\pi(a)) - \int_{\mathbb{R}^2/\mathbb{Z}^2} \tilde{F}| \gg \delta^{O(1)} \|\tilde{F}\|_{\text{Lip}}.$$

Proposition 5.1, in this case, now follows immediately from Proposition 3.1. Note how the G_2-invariance allowed us to reduce a 2-step problem into a 1-step one.

Suppose then that $\xi \not\equiv 0$. The integral of F over every translate of $G_2/(\Gamma \cap G_2)$ is then zero, and hence $\int_{G/\Gamma} F = 0$. Thus (5.1) becomes

$$|\mathbb{E}_{n \in [N]} F(a^n \Gamma)| \geqslant \delta^{O(1)}.$$

We now come to one of the key ideas of the proof, which is to apply the van der Corput lemma, Corollary 4.2. This tells us that there are $\gg \delta^{O(1)} N$ values of $h \in [N]$ such that

$$(5.2) \qquad |\mathbb{E}_{n \in [N]} F(a^{n+h} \Gamma) \overline{F(a^n \Gamma)}| \gg \delta^{O(1)}.$$

It is very natural to try and interpret this in terms of a nilsequence on the product nilmanifold G^2/Γ^2. To do this we first observe by direct computation that any $x \in G$ may be factored uniquely as $\{x\}[x]$, where $\psi(\{x\}) \in [0,1)^3$ and $[x] \in \Gamma$.

Let us, then, factor $a^h = \{a^h\}[a^h]$. The inequality (5.2) implies that

$$|\mathbb{E}_{n\in[N]} F(a^n\{a^h\}\Gamma)\overline{F(a^n\Gamma)}| \gg \delta^{O(1)}$$

for $\gg \delta^{O(1)}N$ values of h. This can be rewritten as

(5.3) $$|\mathbb{E}_{n\in[N]} \tilde{F}_h(\tilde{a}_h^n \Gamma^2)| \gg \delta^{O(1)}$$

for $\gg \delta^{O(1)}N$ values of h, where $\tilde{F}_h : G^2/\Gamma^2 \to \mathbb{C}$ is given by

$$\tilde{F}_h(x,y) := F(\{a^h\}x)\overline{F(y)}$$

and the element $\tilde{a}_h$ is given by

$$\tilde{a}_h := (\{a^h\}^{-1}a\{a^h\}, a).$$

At first sight, the estimates (5.3) do not appear much better than our original estimate (5.1); indeed, it seems "worse" since we are now working on a 6-dimensional 2-step nilmanifold rather than a 3-dimensional 2-step one.

The crucial observation, however, is that all the elements $\tilde{a}_h$ in fact lie not just in G^2, but in the smaller group

$$G^\square = G \times_{G_2} G := \{(g,g') : g^{-1}g' \in G_2\}.$$

This is also a 2-step nilpotent, connected, simply connected Lie group (of dimension 4). It is not hard to check that $[G^\square, G^\square]$ is the *diagonal group* $G_2^\Delta := \{(g_2, g_2) : g_2 \in G_2\}$ and that one can take for a Mal'cev basis of $G^\square/\Gamma^\square$ the collection $\mathcal{X}^\square = \{X_1^\square, X_2^\square, X_3^\square, X_4^\square\}$ given by

$$X_1^\square = \begin{pmatrix} 0 & 1 & \{0,0\} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, X_2^\square = \begin{pmatrix} 0 & 0 & \{0,0\} \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, X_3^\square = \begin{pmatrix} 0 & 0 & \{1,0\} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad X_4^\square = \begin{pmatrix} 0 & 0 & \{1,1\} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where we have written

$$\begin{pmatrix} 0 & x & \{y,y'\} \\ 0 & 0 & z \\ 0 & 0 & 0 \end{pmatrix} := \left(\begin{pmatrix} 0 & x & y \\ 0 & 0 & z \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & x & y' \\ 0 & 0 & z \\ 0 & 0 & 0 \end{pmatrix} \right).$$

This allows us to identify the horizontal torus of $G^\square/\Gamma^\square$ with $\mathbb{R}^3/\mathbb{Z}^3$ by projecting onto the first three coordinates.

Now (5.3) implies that for $\gg \delta^{O(1)}N$ values of h, we have

(5.4) $$|\mathbb{E}_{n\in[N]} F_h^\square((a_h^\square)^n \Gamma^\square)| \gg \delta^{O(1)},$$

where $F_h^\square$ and $a_h^\square$ are the restrictions of $\widetilde{F}_h$ and $\tilde{a}_h$ to $G^\square$, and $\Gamma^\square := \Gamma \times_{\Gamma \cap G_2} \Gamma$. By inspecting the action of G_2^2 on $F_h^\square$ (and the hypothesis $\xi \not\equiv 0$) we also conclude that $\int_{G^\square/\Gamma^\square} F_h^\square = 0$.

Now, the group $G^\square$ is still 2-step nilpotent, so we do not appear to have reduced to a 1-step situation yet. However, recall that F has vertical oscillation ξ. Using this and the fact that g_2 is central in G, we obtain

$$F_h^\square((g_2, g_2) \cdot (g, g')) = F(\{a^h\}g_2 g)\overline{F(g_2 g')}$$
$$= \xi(g_2)\overline{\xi(g_2)}F(\{a^h\}g)\overline{F(g')} = F_h^\square((g, g')).$$

Thus $F_h^\square$ is $[G^\square, G^\square]$-invariant. In (5.4) we may therefore factor through the projection $\pi^\square$ to obtain

$$|\mathbb{E}_{n \in [N]} \tilde{F}_h(n\pi^\square(\tilde{a}_h))| \gg \delta^{O(1)}$$

for $\gg \delta^{O(1)} N$ values of h, where the function $\tilde{F}_h : \mathbb{R}^3/\mathbb{Z}^3 \to \mathbb{C}$ is defined by

$$\tilde{F}_h(\pi^\square(x)) = F_h^\square(x\Gamma^\square).$$

We leave it to the reader to check that $\|\tilde{F}_h\|_{\mathrm{Lip}} = O(1)$ (in the general case to follow this computation is given in more detail). Since $F_h^\square$ has mean zero, we see that $\tilde{F}_h$ has mean zero also.

We are now finally in a situation in which we may apply "1-step" tools. Indeed, from Proposition 3.1 we see that for each h, there is some $k_h^\square \in \mathbb{Z}^3$, $|k_h^\square| \ll \delta^{-O(1)}$, such that

$$\|k_h^\square \cdot \pi^\square(\tilde{a}_h)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N.$$

Pigeonholing in h, we may assume that $k_h^\square = k^\square$ is independent of h. Define $\eta : G^\square \to \mathbb{R}/\mathbb{Z}$ by

$$\eta(x) := k^\square \cdot \pi^\square(x).$$

Then η is an additive homomorphism which annihilates $[G^\square, G^\square]$ and $\Gamma^\square$, and we have

(5.5) $$\|\eta(\tilde{a}_h)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N$$

for $\gg \delta^{O(1)} N$ values of $h \in [N]$.

Our task now is to "piece together" these pieces of information for many different h to deduce Proposition 5.1. We begin by factoring the character η on $G^\square$ into two simpler components, which originate from G (or G_2) rather than $G^\square$.

LEMMA 5.2 (Decomposition of η). *There exist horizontal characters $\eta_1 : G \to \mathbb{R}/\mathbb{Z}$ and $\eta_2 : G_2 \to \mathbb{R}/\mathbb{Z}$ on on G and G_2 respectively (thus η_1 annihilates Γ and η_2 annihilates $\Gamma \cap G_2$) such that*

(5.6) $$\eta(g', g) = \eta_1(g) + \eta_2(g'g^{-1})$$

for all $(g, g') \in G^\square$. Furthermore, we have $|\eta_1|, |\eta_2| \ll \delta^{-O(1)}$.

Proof. Since η is an additive homomorphism we have $\eta(g', g) = \eta((g'g^{-1}, 1) \cdot (g, g)) = \eta(g, g) + \eta(g'g^{-1}, 1)$. Thus if we define $\eta_1(g) := \eta(g, g)$ and $\eta_2(g_2) := \eta(g_2, \mathrm{id}_G)$, then (5.6) is immediately seen to hold. Now η_1 is a horizontal character because η annihilates $\Gamma^\square$, which contains Γ^Δ. Furthermore $\Gamma^\square$ also contains $(\Gamma \cap G_2) \times \mathrm{id}_G$, and hence η_2 annihilates $\Gamma \cap G_2$ as claimed. The bounds on $|\eta_1|$ and $|\eta_2|$ are left as an exercise to the reader; one may compute explicitly with the Mal'cev bases $\mathcal{X}^\square$ and $\mathcal{X}$ on $G^\square/\Gamma^\square$ and G/Γ respectively. $\square$

Using this decomposition and the fact that, in the Heisenberg group, we have the identity $x^{-1}yxy^{-1} = [x, y]$ since $[x, y]$ is central, we see that

$$\eta(\tilde{a}_h) = \eta_1(a) + \eta_2([a, \{a^h\}]).$$

Now a straightforward computation with matrices confirms that if $\psi(x) = (t_1, t_2, t_3)$ and $\psi(y) = (u_1, u_2, u_3)$, then $\psi([x, y]) = (0, 0, t_1 u_2 - t_2 u_1)$, and also that if $\psi(a) = (\gamma_1, \gamma_2, *)$, then $\psi(\{a^h\}) = (\{\gamma_1 h\}, \{\gamma_2 h\}, *)$, where we do not care about the values of the coordinates marked with an asterisk $*$. Thus if we write $\gamma := (\gamma_1, \gamma_2) = \pi(a)$ and $\zeta := (-\gamma_2, \gamma_1)$, then

$$\eta(\tilde{a}_h) = k_1 \cdot \gamma + k_2 \zeta \cdot \{\gamma h\},$$

where $k_1, k_2 = O(\delta^{-O(1)})$ are the frequencies of η_1, η_2 respectively. Thus if (5.5) holds, then

$$(5.7) \qquad \|k_1 \cdot \gamma + k_2 \zeta \cdot \{\gamma h\}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N$$

for $\gg \delta^{O(1)} N$ values of h.

The next proposition derives diophantine information concerning γ and ζ from a hypothesis such as this. In fact we handle a slightly more general situation, since this will be useful when we come to handle the general case of Theorem 2.9. In the following proposition we shall take $\alpha = 0$ and $m = 2$; the proof when $\alpha = 0$ is actually considerably shorter and the reader may care to work through that case to better understand the argument.

PROPOSITION 5.3 (Bracket polynomial lemma). *Let $\delta \in (0, 1)$ and let $N \geqslant 1$ be an integer. Suppose that $\alpha, \beta \in \mathbb{R}$ and that $|\alpha| \leqslant 1/\delta N$. Suppose that $\gamma \in \mathbb{R}^m/\mathbb{Z}^m$ and that $\zeta \in \mathbb{R}^m$ satisfies $|\zeta| \leqslant 1/\delta$. Suppose that for at least δN values of $h \in [N]$, we have*

$$(5.8) \qquad \|\beta + \alpha h + \zeta \cdot \{\gamma h\}\|_{\mathbb{R}/\mathbb{Z}} \leqslant 1/\delta N.$$

Then either $|\zeta_i| \ll_m \delta^{-O_m(1)}/N$ for all $1 \leqslant i \leqslant m$, or else there is some $k \in \mathbb{Z}^m$, $|k| \ll_m \delta^{-O_m(1)}$, such that $\|k \cdot \gamma\|_{\mathbb{R}/\mathbb{Z}} \ll_m \delta^{-O_m(1)}/N$.

Proof. If $\sup_i |\zeta_i| \leqslant 1/\delta N$, then we are done, so assume this is not the case. Then the assumption implies that $\|\beta + \alpha h\|_{\mathbb{R}/\mathbb{Z}} \leqslant (1 + m) \sup_i |\zeta_i|$ for $\geqslant \delta N$ values of $h \in [N]$. Then Lemma 3.2 implies that there is $q \ll \delta^{-C}$ such that $\|q\alpha\|_{\mathbb{R}/\mathbb{Z}} \ll_m \sup_i |\zeta_i| \delta^{-C}/N$ for some absolute constant $C > 0$. Since we are assuming that $|\alpha| \leqslant 1/\delta N$ this forces us to conclude that in fact $|\alpha| \ll_m \sup_i |\zeta_i| \delta^{-C}/N$ unless $N \ll_m \delta^{-O(1)}$, in which case the result is trivial in any case.

Split $[N]$ into intervals of length between N' and $2N'$, where $N' := c_m \delta^{C+1} N$ and $c_m > 0$ is a small number to be chosen later. By the pigeonhole principle, we can find one of these intervals I in which there are $\geqslant \delta|I|$ values of h such that (5.8) holds. If c_m is chosen sufficiently small, then αh does not

vary by more than $\frac{\delta}{20}\sup_i|\zeta_i|$ on such an interval, and we conclude that there is θ such that

$$\|\theta + \zeta \cdot \{\gamma h\}\|_{\mathbb{R}/\mathbb{Z}} \leqslant \frac{\delta}{20}\sup_i|\zeta_i| + \frac{1}{\delta N}$$

for at least $\delta|I|$ values of $h \in I$. Now if $\sup_i|\zeta_i| \leqslant \frac{20}{\delta^2 N}$, then the proposition holds, so we may assume that this is not the case, in which eventuality we have

$$(5.9) \qquad\qquad \|\theta + \zeta \cdot \{\gamma h\}\|_{\mathbb{R}/\mathbb{Z}} \leqslant \frac{\delta}{10}|\zeta_i|$$

for some $i \in [m]$ and for at least $\delta|I|$ values of $h \in I$. We then set

$$\Omega := \left\{ t \in \mathbb{R}^m/\mathbb{Z}^m : \|\theta + \zeta \cdot \{t\}\|_{\mathbb{R}/\mathbb{Z}} \leqslant \frac{\delta}{10}|\zeta_i| \right\}$$

and

$$\tilde{\Omega} := \{x \in \mathbb{R}^m/\mathbb{Z}^m : \operatorname{dist}(x, \Omega) < \delta/10\}.$$

For fixed $u \in \mathbb{R}^m/\mathbb{Z}^m$, the slice

$$\{t \in \tilde{\Omega} : t_j = u_j \text{ for } j \neq i\}$$

is a union of intervals of length less than $\delta/2$, and so $\operatorname{vol}(\tilde{\Omega}) \leqslant \delta/2$. Let $F : \mathbb{R}^m/\mathbb{Z}^m \to \mathbb{R}^+$ be the function

$$F(x) := \max\left(1 - \frac{10\operatorname{dist}(x, \Omega)}{\delta}, 0 \right).$$

Then $F = 1$ on Ω, and so our assumption implies that

$$(5.10) \qquad\qquad \mathbb{E}_{n \in I} F(\gamma n) \geqslant \delta.$$

On the other hand F is supported on $\widetilde{\Omega}$, and so

$$(5.11) \qquad\qquad \int_{\mathbb{R}^m/\mathbb{Z}^m} F(x)\,dx \leqslant \operatorname{vol}(\tilde{\Omega}) \leqslant \frac{\delta}{2}.$$

Thus of course

$$\left| \mathbb{E}_{n \in I} F(\gamma n) - \int_{\mathbb{R}^m/\mathbb{Z}^m} F(x)\,dx \right| \leqslant \frac{\delta}{2}.$$

However F has been constructed so that $\|F\|_{\mathrm{Lip}} \ll 1/\delta$ (we leave this as an exercise), and so we conclude that $(\gamma n)_{n \in I}$ is not $c\delta^2$-equidistributed. Applying Proposition 3.1 we conclude that there is $1 \leqslant k \ll \delta^{-O_m(1)}$ such that $\|k \cdot \gamma\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O_m(1)}/N' \ll \delta^{-O_m(1)}/N$, and the claim follows. $\qquad\square$

Recall that our efforts to prove Proposition 5.1 established condition (5.7). Applying Proposition 5.3 and recalling that $\gamma = (\gamma_1, \gamma_2)$ and $\zeta = (-\gamma_2, \gamma_1)$ we see that in all cases there is some nonzero $k' \in \mathbb{Z}^2$ with $|k'| \ll \delta^{-O(1)}$ such that $\|k' \cdot \gamma\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N$, that is to say $\|k' \cdot \pi(a)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N$. This concludes the proof of Proposition 5.1. $\qquad\square$

Let us pause for a moment to consider the form of the argument just presented. There were two places where we reduced matters to a simpler situation. First of all, in the case $\xi \equiv 0$ we were able to consider F as a function on a 1-step nilmanifold. Secondly, when we applied the van der Corput trick we found ourselves with a function $F_h^{\square}$ which had 0 as a vertical frequency, and so we were again able to reduce to the 1-step case, although we had to restrict the ambient nilmanifold (from G^2/Γ^2 to $G^{\square}/\Gamma^{\square}$) and also quotient out by a commutator group $[G^{\square}, G^{\square}]$ before the 1-step structure became manifest. This already makes it clear that some kind of induction is going on, and in the general case we will see this quite clearly.

6. Polynomial sequences in nilpotent groups

Our analysis of linear sequences on the Heisenberg example captured much of the essence of the proof of Theorem 2.9 in general. What it did not reveal, however, was the rather subtle structure of the space of polynomial sequences $g : \mathbb{Z} \to G$. In this section we begin by establishing a remarkable result of Lazard [18], which asserts that $\mathrm{poly}(\mathbb{Z}, G_{\bullet})$ is a group for any filtration $G_{\bullet}$. Lazard's proof uses the Lie algebra $\mathfrak{g}$ and it works if G is a connected and simply-connected Lie group (as in the present paper). However it turns out that the result is true with no topological assumptions on G, and indeed in the greater generality of so-called polynomial mappings from H to G, where H is an arbitrary group. This result is due to Leibman [20] (see also [19] for a proof of the special case $H = \mathbb{Z}$).

We will then use the Lazard-Leibman results to derive sundry further results concerning the representation of elements of $\mathrm{poly}(\mathbb{Z}, G_{\bullet})$ in coordinates. In fact, keeping in mind our intention to prove multiparameter results in Section 8, we develop the theory of polynomial maps $\mathrm{poly}(\mathbb{Z}^t, G_{\bullet})$.

Definition 6.1 (Polynomial maps). Let H be a group and let G be a nilpotent group with a filtration $G_{\bullet}$. If $g : H \to G$ is a map and if $h \in H$, we write $\partial_h g$ for the map defined by $\partial_h g(x) = g(xh)g(x)^{-1}$. We say that g is a polynomial map with coefficients in $G_{\bullet}$ if we have $\partial_{h_i} \ldots \partial_{h_1} g(x) \in G_i$ for all choices of i and for all $h_1, \ldots, h_i \in H$ and $x \in G$. We write $\mathrm{poly}(H, G_{\bullet})$ for the collection of all such mappings. If $g : H \to G$ is a map, we say that g is a *polynomial sequence of degree at most d* if there exists a filtration $G_{\bullet}$ of degree at most d such that g has coefficients in $G_{\bullet}$.

PROPOSITION 6.2 (Lazard-Leibman theorem [20]). *Let H be a group, let G be a nilpotent group, and let $G_{\bullet}$ be a filtration. Then $\mathrm{poly}(H, G_{\bullet})$, the space of polynomial maps $g : H \to G$ having coefficients in $G_{\bullet}$, is a group.*

Remarks. This result is contained in [20]. (Although the result is only stated in the case that $G_{\bullet}$ is the lower central series filtration, the proof does

not use this fact.) Our proof is a little different, relying on the machinery of Host-Kra cube groups. These were featured for the first time in [15, §§5, 11] and were discussed subsequently in [11, App. E]. See also the recent preprint [16]. We thank Sasha Leibman for helpful conversations concerning these methods.

One should mention at this point the Hall-Petresco theorem [14], [27], which established a special case of the Lazard-Leibman theorem. This theorem states that if $G_\bullet$ is the lower central series filtration, then the sequence $n \mapsto a^n b^n$ lies in $\mathrm{poly}(\mathbb{Z}, G_\bullet)$ for any $a, b \in G$.

In this section it is convenient to generalise the notion of a filtration somewhat. By a *prefiltration* $G_\bullet$ on a nilpotent group G we mean a sequence

$$G \supseteq G_0 \supseteq G_1 \supseteq \cdots \supseteq G_d \supseteq \{\mathrm{id}_G\}$$

of subgroups with the property that $[G_i, G_j] \subseteq G_{i+j}$ for all $i, j \geqslant 0$. The only difference between a prefiltration and a filtration (cf. Definition 1.1) is that we no longer require that $G = G_0 = G_1$. The definition of $\mathrm{poly}(H, G_\bullet)$ extends in a completely obvious way to prefiltrations.

For each integer $k \geqslant 0$, we are going to define the *Host-Kra cube group* $\mathrm{HK}^k(G_\bullet)$ associated to the prefiltration $G_\bullet$. This will be a subgroup of $G^{\{0,1\}^k}$, the product of 2^k copies of G indexed by the cube $\{0,1\}^k$. Before giving the definition, we need to set up some nomenclature concerning these cubes.

Each element $\omega \in \{0,1\}^k$ corresponds in an obvious way to a subset of $[k]$, and we write $\omega \subseteq \omega'$ when the corresponding sets are nested. An *upper face* F is a subset of $\{0,1\}^k$ of the form $F(\omega_0) := \{\omega \in \{0,1\}^k : \omega \supseteq \omega_0\}$. There are, of course, 2^k upper faces, one for each $\omega_0 \in \{0,1\}^k$. The *codimension* $\mathrm{codim}(F)$ of F is simply the number of ones in ω_0. Note that if F, F' are two upper faces, then $F \cap F'$ is also an upper face, and $\mathrm{codim}(F \cap F') \leqslant \mathrm{codim}(F) + \mathrm{codim}(F')$.

Given an upper face F and an element $x \in G$, we write x^F for the element of $G^{\{0,1\}^k}$ defined by

$$(x^F)_\omega = \begin{cases} x & \text{if } \omega \in F, \\ \mathrm{id}_G & \text{otherwise.} \end{cases}$$

Write Γ_F for the subgroup of $G^{\{0,1\}^k}$ consisting of all elements x^F with $x \in G_{\mathrm{codim}(F)}$, where G_i is the ith group in the prefiltration $G_\bullet$; we call such a group an *upper face group*.

Definition 6.3 (Host-Kra cube group). Let $G_\bullet$ be a prefiltration on a nilpotent group G, and let $k \geqslant 0$ be an integer. Then the *Host-Kra cube group* $\mathrm{HK}^k(G_\bullet)$ is the subgroup of $G^{\{0,1\}^k}$ generated by the upper face groups Γ_F.

The Host-Kra cube group can, it turns out, be described in a rather explicit way. Write $\prec$ for the reverse lexicographic ordering on $\{0,1\}^k$; thus $\omega \prec \omega'$ if an only if there is some j such that $\omega_j < \omega'_j$ and $\omega_i = \omega'_i$ for

$i = j + 1, \ldots, k$. This induces an ordering on the upper faces F. We write $F(\omega) \succ F(\omega')$ if and only if $\omega \prec \omega'$. Let $F_0 \prec F_1 \prec \cdots \prec F_{2^k-1}$ be the complete list of upper faces in this order; thus $F_0 = \{1^k\}$ and $F_{2^k-1} = \{0,1\}^k$.

LEMMA 6.4 (Description of Host-Kra cube group). *We have*

$$\mathrm{HK}^k(G_\bullet) = \Gamma_{F_0} \cdot \Gamma_{F_1} \cdot \ldots \Gamma_{F_{2^k-1}}.$$

That is, every element of $\mathrm{HK}^k(G)$ *may be written as* $\gamma_0^{F_0} \ldots \gamma_{2^k-1}^{F_{2^k-1}}$, *where* $\gamma_i \in G_{\mathrm{codim}(F_i)}$. *The representation is in fact unique.*

Proof. The key point here is the inclusion

(6.1) $$[\Gamma_F, \Gamma_{F'}] \subseteq \Gamma_{F \cap F'}.$$

This follows immediately from the fact that

$$[G_{\mathrm{codim}(F)}, G_{\mathrm{codim}(F')}] \subseteq G_{\mathrm{codim}(F)+\mathrm{codim}(F')} \subseteq G_{\mathrm{codim}(F \cap F')}.$$

Using this fact repeatedly, we shift all elements coming from Γ_{F_0} to the left. We then shift all elements coming from Γ_{F_1} to the left, and so on. We leave the routine details and the proof that the representation is unique (which we do not actually need) to the reader. □

Host-Kra cube groups and polynomial maps. It is now time to develop the link between Host-Kra cube groups $\mathrm{HK}^k(G_\bullet)$ and polynomial maps $g \in \mathrm{poly}(H, G_\bullet)$. To do this we introduce a the notion of a *parallelepiped* on H. This is an element in $H^{\{0,1\}^k}$ of the form $(xh^\omega)_{\omega \in \{0,1\}^k}$, where $x \in H$, $h = (h_1, \ldots, h_k)$ is a k-tuple of elements of H, and $h^\omega := h_1^{\omega_1} \ldots h_k^{\omega_k}$. For example the tuple (x, xh_1, xh_2, xh_1h_2) is a parallelepiped in $H^{\{0,1\}^2}$, and $(x, xh_1, xh_2, xh_1h_2, xh_3, xh_1h_2, xh_2h_3, xh_1h_2h_3)$ is a parallelepiped in $H^{\{0,1\}^3}$. Write $H^{[k]}$ for the set of parallelepipeds in $H^{\{0,1\}^k}$. (If H is abelian, $H^{[k]}$ is actually a group, but this need not be the case in general and in any case is not important here.)

Suppose that $g : H \to G$ is a map. Then for any $k \geqslant 0$, there is an obvious induced map $g^{\{0,1\}^k} : H^{\{0,1\}^k} \to G^{\{0,1\}^k}$.

PROPOSITION 6.5 (Characterisation of polynomial maps). *Suppose that H is a group, that G is a nilpotent group together with a prefiltration $G_\bullet$, and that $g : H \to G$. Then g lies in* $\mathrm{poly}(H, G_\bullet)$ *if and only if $g^{\{0,1\}^k}$ maps $H^{[k]}$ to $\mathrm{HK}^k(G_\bullet)$ for all $k \geqslant 0$.*

Remark. The reader might find it useful, as an exercise to get to grips with the notation, to verify this in the case $H = G$ and g being the identity mapping.

We note that Proposition 6.2 is an immediate consequence of Proposition 6.5. Indeed if $g^{\{0,1\}^k}$ and $\tilde{g}^{\{0,1\}^k}$ both map $H^{[k]}$ to $\mathrm{HK}^k(G_\bullet)$, then so does $(g\tilde{g})^{\{0,1\}^k}$, since $\mathrm{HK}^k(G_\bullet)$ is a group.

Proof of Proposition 6.5. We start by establishing the *only if* direction of the proposition, proving by induction on k that $g^{\{0,1\}^k}$ does indeed map $H^{[k]}$ to $\mathrm{HK}^k(G_\bullet)$ when $g \in \mathrm{poly}(H, G_\bullet)$. This is clear when $k = 0$. Suppose it is known for a given value of $k \geqslant 0$. If X is a set, we may regard $X^{\{0,1\}^{k+1}}$ as a product of two copies of $X^{\{0,1\}^k}$, the first factor corresponding to those ω with $\omega_{k+1} = 0$ and the second to those ω with $\omega_{k+1} = 1$. With this notation, every $\tilde{z} \in H^{[k+1]}$ may be written $\tilde{z} = (z, zh_{k+1})$, where $z := (xh^\omega)_{\omega \in \{0,1\}^k}$. We may factor $g^{\{0,1\}^{k+1}}(\tilde{z})$ as a product of two elements, namely

$$(6.2) \qquad g^{\{0,1\}^{k+1}}(\tilde{z}) = (\mathrm{id}_G^{\{0,1\}^k}, (\partial_{h_{k+1}} g)^{\{0,1\}^k}(z)) \cdot (g^{\{0,1\}^k}(z), g^{\{0,1\}^k}(z)).$$

By the inductive hypothesis we have $g^{\{0,1\}^k}(z) \in \mathrm{HK}^k(G_\bullet)$. The derivative $\partial_{h_{k+1}} g : H \to G$ is a polynomial map with coefficients in the prefiltration $\overleftarrow{G}_\bullet$ defined by $\overleftarrow{G}_i := G_{i+1}$; note that this *is* a prefiltration, since

$$[\overleftarrow{G}_i, \overleftarrow{G}_j] = [G_{i+1}, G_{j+1}] \subseteq G_{i+j+2} \subseteq G_{i+j+1} = \overleftarrow{G}_{i+j}.$$

By a second application of the inductive hypothesis we therefore have $(\partial_{h_{k+1}} g)^{\{0,1\}^k}(z) \in \mathrm{HK}^k(\overleftarrow{G}_\bullet)$. In view of (6.2) it therefore suffices to show the inclusions

$$\mathrm{HK}^k(G_\bullet)^\Delta \subseteq \mathrm{HK}^{k+1}(G_\bullet)$$

(where $\mathrm{HK}^k(G_\bullet)^\Delta$ is the diagonal subgroup $\{(t,t) : t \in \mathrm{HK}^k(G_\bullet)\}$) and

$$\mathrm{id}_G^{\{0,1\}^k} \times \mathrm{HK}^k(\overleftarrow{G}_\bullet) \subseteq \mathrm{HK}^{k+1}(G_\bullet).$$

To check the first inclusion it suffices to check elements (γ^F, γ^F) where $\gamma \in G_{\mathrm{codim}(F)}$. But it is easy to see that $(\gamma^F, \gamma^F) = \gamma^{\tilde{F}}$ inside $G^{\{0,1\}^{k+1}}$, where the codimension of the face $\tilde{F}$ inside $\{0,1\}^{k+1}$ equals $\mathrm{codim}(F)$, and the inclusion follows. To check the second inclusion it suffices to check elements $(\mathrm{id}_G^{\{0,1\}^k}, \gamma_F)$ where $\gamma \in \overleftarrow{G}_{\mathrm{codim}(F)} = G_{\mathrm{codim}(F)+1}$. But it is again easy to see that $(\mathrm{id}_G^{\{0,1\}^k}, \gamma^F) = \gamma^{\tilde{F}}$, where now the codimension of $\tilde{F}$ inside $\{0,1\}^{k+1}$ is $\mathrm{codim}(F) + 1$. This concludes the proof of the *only if* part of Proposition 6.5; the perceptive reader will have noticed that we have not yet made any essential use of the main property of prefiltrations, namely the nesting property that $[G_i, G_j] \subseteq G_{i+j}$.

We turn now to the proof of the *if* direction of the proposition. We are to show that if $g^{\{0,1\}^k}$ maps $H^{[k]}$ to $\mathrm{HK}^k(G_\bullet)$ for all k, then $g \in \mathrm{poly}(H, G_\bullet)$. Pick an element $z = (xh^\omega)_{\omega \in \{0,1\}^k}$ in $H^{[k]}$. By Lemma 6.4 (which does use the nesting property of $G_\bullet$), we may write

$$g^{\{0,1\}^k}(z) = \gamma_0^{F_0} \dots \gamma_{2^k-1}^{F_{2^k-1}},$$

where $\gamma_i \in G_{\mathrm{codim}(F_i)}$. Write $\eta_j := \gamma_{2^{j-1}}^{F_{2^{j-1}}} \ldots \gamma_{2^{j-1}}^{F_{2^j -1}}$, $j = 1, \ldots, k$, so that

$$(6.3) \qquad\qquad g^{\{0,1\}^k}(z) = \gamma_0^{1^k} \eta_1 \ldots \eta_k.$$

One may check that the η_i enjoy the following support properties: $(\eta_i)_\omega = \mathrm{id}_G$ unless $\omega_{i+1}, \ldots, \omega_k$ are all nonzero, and $(\eta_i)_\omega = (\eta_i)_{\omega'}$ if ω, ω' differ only in the ω_i coordinate. One may now examine (6.3) coordinatewise, peeling off $\eta_k, \eta_{k-1}, \ldots$ in turn, eventually to conclude that

$$\gamma_0 = \partial_{h_1} \ldots \partial_{h_k} g(x).$$

Now we know that $\gamma_0 \in G_{\mathrm{codim}(F_0)} = G_k$, and thus we have proved that $\partial_{h_1} \ldots \partial_{h_k} g$ takes values in G_k, as required. $\qquad\square$

Polynomial maps in coordinates. From now on we specialise to the case of polynomial maps from $\mathbb{Z}^t$ to G and revert to dealing with filtrations as opposed to prefiltrations. Our aim in this section is to describe the elements of $\mathrm{poly}(\mathbb{Z}^t, G_\bullet)$ using the Mal'cev coordinate map $\psi : G \to \mathbb{R}^m$ relative to some Mal'cev basis $\mathcal{X}$ for G/Γ adapted to the filtration $G_\bullet$.

Definition 6.6 (Multi-binomial coefficients). Let $t \geqslant 1$ be an integer. Suppose that $\vec{n} = (n_1, \ldots, n_t)$ and that $\vec{j} = (j_1, \ldots, j_t) \in \mathbb{Z}_{\geqslant 0}^t$ is a set of indices. Then we write

$$\binom{\vec{n}}{\vec{j}} := \prod_{i=1}^{t} \binom{n_i}{j_i}.$$

A version of the following lemma may be found in [23, §4].

LEMMA 6.7 (Description of $\mathrm{poly}(\mathbb{Z}^t, G_\bullet)$ in bases). *Suppose that G/Γ is a nilmanifold of dimension m and that $\mathcal{X}$ is a Mal'cev basis for G/Γ adapted to some filtration $G_\bullet$. Then $g \in \mathrm{poly}(\mathbb{Z}^t, G_\bullet)$ if and only if the coordinates $\psi(g(n))$ have the form*

$$\psi(g(n)) = \sum_{\vec{j}} t_{\vec{j}} \binom{\vec{n}}{\vec{j}},$$

where each $t_{\vec{j}}$ lies in $\mathbb{R}^m$ and is such that $(t_{\vec{j}})_i = 0$ if $i \leqslant m - m_{|\vec{j}|}$, where $|\vec{j}| := j_1 + \cdots + j_t$.

Remark. The presence of the discrete subgroup Γ is not at all relevant to this lemma; however we have only defined Mal'cev bases in this context.

Proof. We start with the *if* direction. If $g(n)$ has the form stated, then it is a product of sequences of the form $\vec{n} \mapsto a^{\binom{\vec{n}}{\vec{j}}}$, where $a \in G_{|\vec{j}|}$. By the group property of $\mathrm{poly}(\mathbb{Z}^t, G_\bullet)$ it therefore suffices to establish the result in the case that $g(\vec{n})$ is actually *equal* to such a sequence. By induction one sees that the derivative $\partial_{h_1} \ldots \partial_{h_k} g(\vec{n})$ equals $a^{p(h_1, \ldots, h_k; \vec{n})}$, where the maximal degree

$\alpha_1 + \cdots + \alpha_t$ of a monomial $n_1^{\alpha_1} \ldots n_t^{\alpha_t}$ appearing in p is at most $\max(|\vec{j}| - k, 0)$. Thus we see that this derivative lies in $G_{|\vec{j}|}$ if $k \leqslant |\vec{j}|$ and is zero otherwise. It follows that $g \in \mathrm{poly}(\mathbb{Z}^t, G_\bullet)$.

To prove the *only if* direction, let $\mathfrak{h}_j \subset \mathfrak{g}$ be the subspace

$$\mathfrak{h}_j := \mathrm{Span}(X_{j+1}, \ldots, X_m)$$

and set $H_j := \exp(\mathfrak{h}_j)$. By the nesting property of the Mal'cev basis $\mathcal{X}$ (see (A.1)), we see that $H_j \lhd G$.

Suppose as a hypothesis for downward induction on k that the statement has been proved for all $g \in \mathrm{poly}(\mathbb{Z}^t, G_\bullet)$ with $g(\vec{n}) \in H_k$ for all $\vec{n}$, for a certain value of k. This is trivial for $k = m$, in which case $g(\vec{n}) = \mathrm{id}_G$. Suppose that $g(\vec{n}) \in H_{k-1}$ for all $\vec{n}$. Let $\pi : H_{k-1} \to H_{k-1}/H_k \cong \mathbb{R}$ be the natural projection. Then $p_{k-1}(\vec{n}) := \pi(g(\vec{n})\Gamma)$ is a polynomial map from $\mathbb{R}^t$ to $\mathbb{R}$. Suppose that $k - 1 < m - m_i$ and that i is minimal subject to this property. Then for any $h_1, \ldots, h_i \in \mathbb{Z}^t$, we have $\partial_{h_1} \ldots \partial_{h_i} g \in G_i = H_{m-m_i}$, and therefore $\partial_{h_1} \ldots \partial_{h_i} p_{k-1}(\vec{n}) = 0$. Thus the total degree of any monomial in p_{k-1} is at most $i - 1$. Therefore we may write the sequence $h(\vec{n})$ defined by

$$h(\vec{n}) := \exp(X_{k-1})^{p_{k-1}(\vec{n})}$$

as a product of sequences $\exp(X_{k-1})^{t_{\vec{j}} \binom{\vec{n}}{\vec{j}}}$ with $|\vec{j}| \leqslant i - 1$. By the minimality of i we have $X_{k-1} \in \mathfrak{g}_{i-1}$, and so each of these sequences lies in $\mathrm{poly}(\mathbb{Z}^t, G_\bullet)$, and hence so does h. It follows that the sequence $\tilde{g}(n) := g(n)h(n)^{-1}$ lies in $\mathrm{poly}(\mathbb{Z}^t, G_\bullet)$. But this new sequence $\tilde{g}$ has $\tilde{g}(n) \in H_k$, and hence we may proceed by induction. $\qquad\square$

A useful and easily-derived corollary of Lemma 6.7 is that $\mathrm{poly}(\mathbb{Z}^t, G_\bullet)$ is closed under dilations.

COROLLARY 6.8 (Dilation of polynomial sequences). *Suppose that $g \in \mathrm{poly}(\mathbb{Z}^t, G_\bullet)$ and that $a_1, \ldots, a_t, b_1, \ldots, b_t \in \mathbb{Z}$. Then the sequence $\vec{n} \mapsto g(a_1 + b_1 n_1, \ldots, a_t + b_t n_t)$ also lies in $G_\bullet$.*

We remarked in the introduction that a sequence $g : \mathbb{Z} \to G$ is polynomial with coefficients in some filtration $G_\bullet$ if and only if g has the form

$$(6.4) \qquad\qquad g(n) = a_1^{p_1(n)} \ldots a_k^{p_k(n)}$$

for polynomials $p_1, \ldots, p_k$ with integer coefficients. Although this result is not required in the paper, it is certainly conceivable that one might wish to apply the main theorems of the paper to a sequence that is presented in an explicit form such as (6.4) and does not obviously satisfy the more abstract condition of Definition 1.8.

The fact that every polynomial sequence has the form (6.4) is an easy consequence of Lemma 6.7. To establish the converse, consider first the lower

central series filtration $G_\bullet$ which has degree s, the step of the nilpotent Lie group G. Let d be the maximum degree occurring amongst the polynomials p_i and define a finer filtration $G'_\bullet$ of degree sd by setting $G'_i := G_{\lceil i/d \rceil}$. This *is* a filtration since

$$[G'_i, G'_j] = [G_{\lceil i/d \rceil}, G_{\lceil j/d \rceil}] \subseteq G_{\lceil i/d \rceil + \lceil j/d \rceil} \subseteq G_{\lceil (i+j)/d \rceil} = G'_{i+j}.$$

Any sequence of the form $n \mapsto a^{\binom{n}{j}}$, $j \leqslant d$ has coefficients in $G'_\bullet$ since $G'_i = G$ for $i = 0, 1, \ldots, d$ and the $(d+1)$st derivative of such a sequence is trivial. Since g is a product of such sequences and $\mathrm{poly}(\mathbb{Z}, G'_\bullet)$ is a group we see that $g \in \mathrm{poly}(\mathbb{Z}, G'_\bullet)$.

We note that if G/Γ has a Q-rational Mal'cev basis adapted to the lower central series then, by the results of the appendix, there is a $Q^{O_{d,s}(1)}$-rational Mal'cev basis for G/Γ adapted to $G'_\bullet$.

We leave it to the reader to formulate and prove an analogous result for polynomial mappings from $\mathbb{Z}^t$ to G.

7. The general case of the main theorem

We are now in a position to attack the general case of Theorem 2.9. Our analysis of the Heisenberg example in Section 5 suggested that the argument will involve an induction on the degree d of $G_\bullet$. In that case there were two different scenarios in which we reduced from the case $d = 2$ to the case $d = 1$. Whilst the same is true in general, the introduction of genuinely polynomial sequences (rather than just linear ones) necessitates a further inductive loop on the quantity $m_* := m_{\mathrm{ab}} - m_{\mathrm{lin}}$, which we call the *nonlinearity degree*. To see why, consider the following slightly informal example.

Example. Let G/Γ be the Heisenberg example, and let $g(n) = \begin{pmatrix} 1 & \alpha_1 & \alpha_3 \\ 0 & 1 & \alpha_2 \\ 0 & 0 & 1 \end{pmatrix}^n$, where α_1, α_2, and α_3 are highly independent over $\mathbb{Q}$. Then there is no horizontal character η of low frequency such that $\|\eta \circ g\|_{C^\infty[N]}$ is small.

Now we have $\partial g = g$ and $\partial^i g = \mathrm{id}_G$ for $i \geqslant 2$, and so g has coefficients in the subgroup sequence $G_\bullet$ defined by $G_{(0)} := G_{(1)} := G_{(2)} := G$, $G_{(3)} := G_{(4)} := G_2$, and $G_{(i)} := \{\mathrm{id}_G\}$ for $i \geqslant 5$. With this choice we have $G^\square = G \times G$. However $g_h^\square$ takes values in $G \times_{G_2} G$, and hence $\eta^\square \circ g_h^\square = 0$ for any horizontal character $\eta^\square$ with frequency of the form $(a, b, -a, -b) \in \mathbb{Z}^4$. Thus, a lack of uniform distribution for $g_h^\square$ does not imply lack of uniform distribution for g.

The problem in the above example is that the filtration $G_\bullet$ was far too "coarse" to accurately capture the differential structure of the sequence g. Indeed g also takes values in the minimal (lower central series) filtration, as we saw in Section 5.

In the light of the above example, we can expect that it will sometimes be necessary to pass to a "finer" filtration of the same degree d in order to

properly capture the differential structure of g. This finer filtration will have a smaller value of the nonlinearity degree m_*, and thus we introduce an extra inductive loop to incorporate this parameter. To be precise we shall prove, by induction on d and m_*, the following slight variant of Theorem 2.9.

THEOREM 7.1 (Variant of Main Theorem). *Let $m, d \geqslant 0$ be integers with $m_* \leqslant m$. Let $0 < \delta < 1/2$, and suppose that $N \geqslant 1$. Suppose that G/Γ is a nilmanifold and that $G_\bullet$ is a filtration of degree d and with nonlinearity degree m_*. Suppose that $\mathcal{X}$ is a $1/\delta$-rational Mal'cev basis adapted to $G_\bullet$, and suppose that $g \in \operatorname{poly}(\mathbb{Z}, G_\bullet)$. If $(g(n)\Gamma)_{n \in [N]}$ is not δ-equidistributed, then there is a horizontal character η with $0 < |\eta| \ll \delta^{-O_{m,m*,d}(1)}$ such that*

$$\|\eta \circ g\|_{C^\infty[N]} \ll \delta^{-O_{m,m*,d}(1)}.$$

It is clear that this does imply Theorem 2.9, since the dependence of the $O(1)$ exponents on m_* may be suppressed once Theorem 7.1 has been proven by induction. In our proof there will be an outer inductive loop over d and an inner one over m_*. In other words we shall assume that Theorem 7.1 holds for all pairs (d', m'_*) in which either $d' < d$ or for which $d' = d$ and $m'_* < m_*$, and deduce the case (d, m_*).

Henceforth we allow all constants implicit in the $\ll$ or O-notation to depend on d, m and m_*.

We begin with some simple reductions. By Lemma 3.7 we may assume that the orbit $(g(n)\Gamma)_{n \in [N]}$ is not $\delta^{O(1)}$-equidistributed along some vertical frequency $\xi \in \mathbb{Z}^{m_d}$ with $|\xi| \ll \delta^{-O(1)}$. Thus there is some function $F : G/\Gamma \to \mathbb{C}$ with $\|F\|_{\mathrm{Lip}} \leqslant 1$ and vertical frequency ξ such that

$$(7.1) \qquad |\mathbb{E}_{n \in [N]} F(g(n)\Gamma) - \int_{G/\Gamma} F| \gg \delta^{O(1)}.$$

If $\xi = 0$ then F is G_d-invariant and we may descend to G/G_d, together with the filtration $G_\bullet/G_d$ which has length $d - 1$, and invoke our inductive hypothesis. We pause to give the rather straightforward details.

Write $\overline{G} := G/G_d$ and $\overline{\Gamma} := \Gamma/(\Gamma \cap G_d)$. Then $\overline{G}/\overline{\Gamma}$ is a nilmanifold togther with a filtration $\overline{G}_\bullet$ of length $d - 1$, where $\overline{G}_i := G_i/G_d$. The Mal'cev basis $\mathcal{X} = \{X_1, \ldots, X_m\}$ may be reduced to give a $\frac{1}{\delta}$-rational Mal'cev basis $\overline{\mathcal{X}} = \{\overline{X}_1, \ldots, \overline{X}_{\overline{m}}\}$ for $\overline{G}/\overline{\Gamma}$ adapted to $\overline{G}_\bullet$, where $\overline{m} := m - m_d$.

Write $\overline{g} : \mathbb{Z} \to \overline{G}$ for the reduction of $g \pmod{G_d}$. By the G_d-invariance the function F descends to a Lipschitz function $\overline{F} : \overline{G}/\overline{\Gamma} \to \mathbb{C}$ with $\|\overline{F}\|_{\mathrm{Lip}} \leqslant \|F\|_{\mathrm{Lip}}$, and so (7.1) implies that

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

(Here we have used the fact that normalised Haar measure on $\overline{G}/\overline{\Gamma}$ is be obtained by quotienting that on G/Γ by G_d.)

We may now apply the inductive hypothesis to obtain a horizontal character $\overline{\eta} : \overline{G} \to \mathbb{C}$ on $\overline{G}$ of frequency magnitude $0 < |\overline{\eta}| \ll \delta^{-O(1)}$ such that

$$\|\overline{\eta} \circ \overline{g}\|_{C^\infty[N]} \ll \delta^{-O(1)}.$$

If we let $\eta : G \to \mathbb{C}$ be the horizontal character on G defined by $\eta(x) = \overline{\eta}(\overline{x})$, then we have $\overline{\eta} \circ \overline{g} = \eta \circ g$ and $|\eta| = |\overline{\eta}|$. This concludes the proof in the case $\xi = 0$.

Suppose henceforth that $\xi \neq 0$. Since F has ξ as a vertical frequency, (7.1) becomes

(7.2) $$|\mathbb{E}_{n \in [N]} F(g(n)\Gamma)| \gg \delta^{O(1)}.$$

We proceed initially with two additional reductions. The first is to the case $g(0) = \mathrm{id}_G$. Factorise $g(0) = \{g(0)\}[g(0)]$ as in Lemma A.14. Set $\tilde{g}(n) := \{g(0)\}^{-1} g(n) g(0)^{-1} \{g(0)\}$. Then we have $|\mathbb{E}_{n \in [N]} \tilde{F}(\tilde{g}(n)\Gamma)| \geqslant \delta$, where $\tilde{F}(x) := F(\{g(0)\}x)$. But $\tilde{F}$ still has vertical oscillation ξ and, by Lemma A.5, it has Lipschitz constant $O(1)$. Noting that $\|\eta \circ g\|_{C^\infty[N]} = \|\eta \circ \tilde{g}\|_{C^\infty[N]}$ we see that if we have Theorem 7.1 for $\tilde{g}$ then we also have it for g.

The second reduction is to the case when $|\psi(g(1))| \leqslant 1$. (This is needed in the lead up to (7.16).) To do this, factorise $g(1) = \{g(1)\}[g(1)]$ as in Lemma A.14. Set $\tilde{g}(n) := g(n)[g(1)]^{-n}$. Then $\tilde{g}(n)\Gamma = g(n)\Gamma$, $\tilde{g}(0) = \mathrm{id}_G$, $\tilde{g} \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$, and $\pi(\tilde{g}(n)\Gamma) = \pi(g(n)\Gamma)$, so proving Theorem 7.1 for g is equivalent to proving it for $\tilde{g}$.

Henceforth we assume $g(0) = \mathrm{id}_G$ and $|\psi(g(1))| \leqslant 1$.

As in Section 5 we apply Van der Corput's Lemma (Corollary 4.2) to (7.2) to deduce that for $\gg \delta^{O(1)} N$ values of h, we have

(7.3) $$|\mathbb{E}_{n \in [N]} F(g(n+h)\Gamma)\overline{F(g(n)\Gamma)}| \gg \delta^{O(1)}.$$

For each fixed h, this may be interpreted as a statement about the polynomial sequence $(g(n+h), g(n))$ on the product group G^2. However, guided by our experience with the Heisenberg group, it is natural to try and interpret it as a sequence on a somewhat smaller group. To this end, we define the *nonlinear part* g_2 of g by

(7.4) $$g_2(n) := g(n)g(1)^{-n}.$$

Motivated by what we did in Section 5, we may then rewrite (7.3) in the form

(7.5) $$|\mathbb{E}_{n \in [N]} \tilde{F}_h(\tilde{g}_h(n)\Gamma^2)| \gg \delta^{O(1)},$$

where

$$\tilde{F}_h(x, y) := F(\{g(1)^h\}x)\overline{F(y)}$$

and

$$\tilde{g}_h(n) := (\{g(1)^h\}^{-1} g_2(n+h) g(1)^n \{g(1)^h\}, g_2(n) g(1)^n). \tag{7.6}$$

It turns out that g_h takes values in $G^\square := G \times_{G_2} G$, just as we found in our analysis of the Heisenberg case. To prove this, note that $G_2 \supseteq [G, G]$ and so G becomes abelian after quotienting out by the normal subgroup G_2. Thus we need only prove that $g_2(n) \in G_2$ for all n. We have $\partial^2 g(n) = \mathrm{id}_G$ modulo G_2. Since $g(0) = \mathrm{id}_G$, this implies by an easy induction that $g(n) = g(1)^n$ modulo G_2, and so g_2 does indeed take values in G_2.

We may therefore replace (7.5) by

$$|\mathbb{E}_{n \in [N]} F_h^\square(g_h^\square(n) \Gamma^\square)| \gg \delta^{O(1)} \tag{7.7}$$

by restricting everything in that equation to an object on $G^\square$.

Note that, exactly as in the Heisenberg case, $F_h^\square$ is invariant under $G_d^\Delta = \{(g_d, g_d) : g_d \in G_d\}$. Indeed, since G_d is central in G, we have

$$F_h^\square((g_d, g_d) \cdot x^\square) = F(\{g(1)^h\} g_d x) \overline{F(g_d x')}$$
$$= e(\xi(g_d)) e(-\xi(g_d')) F(\{g(1)^h\} x) \overline{F(x')}$$
$$= F_h^\square(x^\square).$$

Thus $F_h^\square$ descends to a function $\overline{F_h^\square}$ on $\overline{G^\square} := G^\square / G_d^\Delta$ and we may write (7.7) as

$$|\mathbb{E}_{n \in [N]} \overline{F_h^\square}(\overline{g_h^\square}(n) \overline{\Gamma^\square})| \gg \delta^{O(1)}, \tag{7.8}$$

where $\overline{\Gamma^\square} := \Gamma^\square / (\Gamma \cap G_d^\Delta)$.

The next proposition is central to our whole argument in that it clarifies the sense in which $\overline{G^\square}$ is "less complex" than G.

PROPOSITION 7.2 (Reduction in degree). *Define* $(G^\square)_i := G_i \times_{G_{i+1}} G_i$ *for* $i = 1, \ldots, d$. *Then* $(G^\square)_\bullet$ *is a filtration on* $G^\square$ *of degree* d. *Since* $(G^\square)_d = G_d^\Delta$, *it descends under quotienting by* G_d^Δ *to a filtration* $\overline{(G^\square)}_\bullet$ *of degree* $d-1$ *on* $\overline{G^\square}$. *Each polynomial sequence* $g_h^\square$ *lies in* $\mathrm{poly}(\mathbb{Z}, (G^\square)_\bullet)$, *and hence each reduced polynomial sequence* $\overline{g_h^\square}$ *lies in* $\mathrm{poly}(\mathbb{Z}, \overline{(G^\square)}_\bullet)$.

Proof. We start with a lemma.

LEMMA 7.3. *Suppose that* H_1, H_2 *and* K_1, K_2 *are normal subgroups of a group* G, *that* H_1, H_2 *generate a group* H, *and that* K_1, K_2 *generate a group* K. *Then* $[H, K]$ *is generated by the groups* $[H_i, K_j]$, $1 \leqslant i, j \leqslant 2$.

Proof. The groups $[H_i, K_j]$ are all normal, and thus the group they generate is also normal. If we quotient by that group, then H_1, H_2 commute with K_1, K_2, and thus H commutes with K. The claim follows. $\square$

Now observe that $(G^\square)_i$ is generated by G_{i+1}^2 and $G_i^\triangle$. In view of the lemma, it therefore suffices to establish that all four of the quantities

$$[G_i^\triangle, G_j^\triangle], [G_i^\triangle, G_{j+1}^2], [G_{i+1}^2, G_j^\triangle], [G_{i+1}^2, G_{j+1}^2]$$

lie in $G_{i+j}^\square$. Using the fact that $G_\bullet$ is a filtration, the first quantity is manifestly contained in $G_{i+j}^\triangle$ and the last three lie in G_{i+j+1}^2. It follows immediately that $(G^\square)_\bullet$ is indeed a filtration.

Next we show that $g_h^\square \in \text{poly}(\mathbb{Z}, (G^\square)_\bullet)$. Here we make serious use of the fact that $\text{poly}(\mathbb{Z}, G_\bullet)$ is a group for the first time. Recall that

$$(7.9) \qquad g_h^\square(n) := \left(\{g(1)^h\}^{-1} g_2(n+h) g(1)^n \{g(1)^h\}, g_2(n) g(1)^n \right).$$

Now $\text{poly}(\mathbb{Z}, (G^\square)_\bullet)$ is a group, and it is also closed under conjugation by elements of G^2. Since $(g(1)^n, g(1)^n)$ is obviously in $\text{poly}(\mathbb{Z}, (G^\square)_\bullet)$, it suffices to check that $(g_2(n+h), g_2(n)) \in \text{poly}(\mathbb{Z}, (G^\square)_\bullet)$. Of course, $g_2 \in \text{poly}(\mathbb{Z}, G_\bullet)$ and hence, by Lemma 6.7, it is a product of elements $g_i^{\binom{n}{i}}$ with $g_i \in G_i$. It therefore suffices to show that $(g_i^{\binom{n+h}{i}}, g_i^{\binom{n}{i}}) \in (G^\square)_\bullet$. Taking jth derivatives, it suffices to check that $g_i^{\binom{n+h}{i-j}} \equiv g_i^{\binom{n}{i-j}} (\text{mod } G_{j+1})$. For $j < i$, this follows from the fact that $g_i \in G_i$, whilst for $j \geqslant i$ it is trivial. $\qquad \square$

In order to apply the inductive hypothesis, we must specify a Mal'cev basis $\overline{\mathcal{X}^\square}$ for $\overline{G^\square / \Gamma^\square}$ adapted to the sequence $\overline{(G^\square)}_\bullet$, and it must then be checked that $\overline{F_h^\square}$ is Lipschitz with respect to the metric $d_{\overline{\mathcal{X}^\square}}$. These are rather tedious matters, and we recommend that the reader take the following lemma on trust on a first reading of the paper.

LEMMA 7.4 (Rationality bounds for the relative square). *There is an $O(\delta^{-O(1)})$-rational Mal'cev basis $\mathcal{X}^\square = \{X_1^\square, \ldots, X_{m^\square}^\square\}$ for $G^\square / \Gamma^\square$ adapted to the filtration $(G^\square)_\bullet$ with the property that $\psi_{\mathcal{X}^\square}(x, x')$ is a polynomial of degree $O(1)$ with rational coefficients of height $\delta^{-O(1)}$ in the coordinates $\psi(x), \psi(x')$. With respect to the metric $d_{\mathcal{X}^\square}$, we have $\|F_h^\square\|_{\text{Lip}} \ll \delta^{-O(1)}$ uniformly in h.*

Proof. We consider $G^\square$ as a subgroup of $G \times G$. Recall (cf. Definition A.7) the definition of a *weak basis*. It is clear that

$$\mathcal{X} \times \mathcal{X} = \{(X_1, 0), (0, X_1), \ldots, (X_m, 0), (0, X_m)\}$$

is a $\delta^{-O(1)}$-rational weak basis for $G/\Gamma \times G/\Gamma$ and that each of the groups $(G^\square)_i := G_i \times_{G_{i+1}} G_i$ is $\delta^{-O(1)}$-rational with respect to this basis. By Proposition A.10 it follows that there is a Mal'cev basis $\mathcal{X}^\square = \{X_1^\square, \ldots, X_{m^\square}^\square\}$ for $G^\square / \Gamma^\square$, adapted to the filtration $(G^\square)_\bullet$, with the property that each $X_i^\square$ is a $\delta^{-O(1)}$-rational combination of the elements of $\mathcal{X} \times \mathcal{X}$. By adding the elements $(X_1, 0), \ldots, (X_{m_{\text{lin}}}, 0)$ to $\mathcal{X}^\square$, we obtain a weak basis $\mathcal{Y}$ for $G/\Gamma \times G/\Gamma$ which

enjoys the nesting property (A.1). From Lemma A.2 it follows that each coordinate of $\psi_{\mathcal{Y}}(x, x')$ is a polynomial of degree $O(1)$ and with coefficients $\delta^{-O(1)}$ in the coordinates $\psi_{\mathcal{X} \times \mathcal{X}}(x, x')$. Restricting to those pairs (x, x') which lie in $G^{\square}$, we obtain the stated property.

Recall that $F_h^{\square}(x^{\square}) = F(\{g(1)^h\}x)\overline{F(x')}$. Now by definition we have $|\psi_{\mathcal{X}}(\{g(1)^h\})| \leqslant 1$. By Lemma A.5 (and Lemma A.14, which guarantees that every $x \in G/\Gamma$ has a representative with coordinates bounded by $O(1)$), we see that $(x, x') \mapsto F(\{g(1)^h\}x)\overline{F(x')}$ defines a function on $G \times G$ whose Lipschitz constant with respect to the product metric $d \times d$ is $\ll \delta^{-O(1)}$. Now by Lemma A.6 and the construction of $\mathcal{X}^{\square}$ we therefore have $\|F_h^{\square}\|_{\mathrm{Lip}} \ll \delta^{-O(1)}$ where, remember, the Lipschitz constant is being computed with respect to the metric $d_{\mathcal{X}^{\square}}$. $\qquad\square$

Let us now resume the discussion starting from (7.8). We begin by reprising some of the straightforward arguments at the start of the section (where we dealt with the case $\xi = 0$). By reducing the first $\overline{m^{\square}} := m^{\square} - m_d$ elements of $\mathcal{X}^{\square}$, we obtain an $O(\delta^{-O(1)})$-rational Mal'cev basis $\overline{\mathcal{X}^{\square}} = \{\overline{X_1^{\square}}, \ldots, \overline{X_{\overline{m^{\square}}}^{\square}}\}$ for $\overline{G^{\square}}/\overline{\Gamma^{\square}}$ adapted to the filtration $\overline{(G^{\square})}_\bullet$. With respect to the metric $d_{\overline{\mathcal{X}^{\square}}}$ we have $\|\overline{F_h^{\square}}\|_{\mathrm{Lip}} \ll \delta^{-O(1)}$.

Since $\overline{(G^{\square})}_\bullet$ has degree $d-1$, our inductive hypothesis is applicable and we conclude that for $\gg \delta^{O(1)}$ values of $h \in [N]$, there is some horizontal character $\overline{\eta}_h : \overline{G^{\square}} \to \mathbb{R}/\mathbb{Z}$ with $0 < |\overline{\eta}_h| \ll \delta^{-O_m(1)}$ and

$$\|\overline{\eta}_h \circ \overline{g_h^{\square}}\|_{C^\infty[N]} \ll \delta^{-O_m(1)}.$$

By pigeonholing in h we may assume that $\overline{\eta} = \overline{\eta}_h$ is independent of h. Writing $\eta : G^{\square} \to \mathbb{R}/\mathbb{Z}$ for the horizontal character defined by $\eta(x) = \overline{\eta}(\overline{x})$, we see that $0 < |\eta| \ll \delta^{-O_m(1)}$ and that

$$(7.10) \qquad \|\eta \circ g_h^{\square}\|_{C^\infty[N]} \ll \delta^{-O_m(1)}.$$

The next lemma, which is almost identical to Lemma 5.2, allows us to write η in terms of maps defined on G rather than $G^{\square}$.

LEMMA 7.5. *We have a decomposition* $\eta(g', g) = \eta_1(g) + \eta_2(g'g^{-1})$ *for all* $(g', g) \in G^{\square}$, *where* $\eta_1 : G \to \mathbb{R}/\mathbb{Z}$ *is a horizontal character on* G, *and* $\eta_2 : G_2 \to \mathbb{R}/\mathbb{Z}$ *is a horizontal character on* G_2 *which also annihilates* $[G, G_2]$. *Furthermore, we have* $|\eta_1|, |\eta_2| \ll \delta^{-O(1)}$.

Proof. If we define $\eta_1(g) := \eta(g, g)$ and $\eta_2(g_2) := \eta(g_2, \mathrm{id}_G)$ for $g \in G$ and $g_2 \in G_2$, then the decomposition follows since η is an additive homomorphism. Since η annihilates $[G^{\square}, G^{\square}]$, which contains $[G^{\Delta}, G_2 \times \mathrm{id}_G] = [G, G_2] \times \mathrm{id}_G$, we see that η_2 annihilates $[G, G_2]$; since η annihilates $\Gamma^{\square}$, which contains both Γ^{Δ} and $(\Gamma \cap G_2) \times \mathrm{id}_G$, we see that η_1 and η_2 annihilate Γ and $\Gamma \cap G_2$ respectively.

It remains to check the boundedness properties. Writing

$$\eta(x, x') = k^{\square} \cdot \psi_{\mathcal{X}^{\square}}(x, x'),$$

where $k^{\square} \in \mathbb{Z}^{m_{\square}}$, we have by definition that $|k^{\square}| \ll \delta^{-O(1)}$. The integer vectors k_1 and k_2 used to define $|\eta_1|$ and $|\eta_2|$ are then given by

$$k_1 \cdot \psi(x) = \eta_1(x) = \eta(x, x) = k^{\square} \cdot \psi_{\mathcal{X}^{\square}}(x, x)$$

and

$$k_2 \cdot \psi(x) = \eta_2(x) = \eta(x, \mathrm{id}_G) = k^{\square} \cdot \psi_{\mathcal{X}^{\square}}(x, \mathrm{id}_G).$$

The fact that $|k_1|, |k_2| \ll \delta^{-O(1)}$ now follows immediately from the fact, established in Lemma 7.4, that $\psi_{\mathcal{X}^{\square}}(x, x')$ is a polynomial of degree $O(1)$ with rational coefficients of height $O(\delta^{-O(1)})$ in the coordinates $\psi(x), \psi(x')$. $\square$

Now let us return to (7.10) and reinterpret this in terms of the decomposition of η just given. Recalling the formula (7.9) for $g_h^{\square}(n)$, we therefore have

$$\eta(g_h^{\square}(n)) = \eta_1(g(n)) + \eta_2(\{g(1)^h\}^{-1}g_2(n+h)g(1)^n\{g(1)^h\}g(1)^{-n}g_2(n)^{-1})$$

which, since η_2 vanishes on $[G, G_{(2)}]$, is equal to

$$\eta_1(g(n)) + \eta_2(g_2(n+h)\{g(1)^h\}^{-1}g(1)^n\{g(1)^h\}g(1)^{-n}g_2(n)^{-1})$$
$$= \eta_1(g(n)) + \eta_2(g_2(n+h)) - \eta_2(g_2(n)) + \eta_2(\{g(1)^h\}^{-1}g(1)^n\{g(1)^h\}g(1)^{-n}).$$

One easily verifies by induction on n that $y^{-1}x^n y x^{-n} \equiv [x, y]^n (\mathrm{mod}\,[G, [G, G]])$. Since η_2 annihilates $[G, G_2]$, which contains $[G, [G, G]]$, we can therefore simplify the above a little further to

(7.11)
$$\eta(g_h^{\square}(n)) = \eta_1(g(n)) + \eta_2(g_2(n+h)) - \eta_2(g_2(n)) + n\eta_2([g(1), \{g(1)^h\}])$$
$$:= P(n) + Q(n+h) - Q(n) + \sigma(h)n,$$

where $P, Q : \mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ are polynomial sequences of degree at most d.

The next lemma is specifically designed to handle the situation that has arisen here. In this lemma it is convenient to reprise a notation from earlier papers of ours (such as [10]): if $\alpha \in \mathbb{R}/\mathbb{Z}$ and $Q > 1$, we write $\|\alpha\|_{\mathbb{R}/\mathbb{Z}, Q} := \inf_{1 \leqslant q \leqslant Q} \|q\alpha\|_{\mathbb{R}/\mathbb{Z}}$. In a similar spirit, for any $f : \mathbb{Z} \to \mathbb{R}/\mathbb{Z}$, define

$$\|f\|_{C^{\infty}[N], Q} := \inf_{1 \leqslant q \leqslant Q} \|qf\|_{C^{\infty}[N]}.$$

LEMMA 7.6 (Polynomials lemma). *Suppose that $P, Q : \mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ are polynomial sequences of degree at most d with $P(0) = 0$ and $Q(0) = \partial Q(0) = 0$ and that $\sigma : [N] \to \mathbb{R}/\mathbb{Z}$ is an arbitrary map. Suppose that there are $\gg \delta^{O(1)}N$ values of $h \in [N]$ such that*

$$\|P(n) + Q(n+h) - Q(n) + \sigma(h)n\|_{C^{\infty}[N]} \ll \delta^{-O(1)}.$$

Then $\|\partial^i Q\|_{\mathbb{R}/\mathbb{Z}, \delta^{-O(1)}} \ll \delta^{-O(1)}/N^i$ *for* $i \geqslant 3$, *and*

$$\|P(1) + \alpha h + \sigma(h)\|_{\mathbb{R}/\mathbb{Z}, \delta^{-O(1)}} \ll \delta^{-O(1)}/N$$

for $\gg \delta^{O(1)} N$ *values of* $h \in [N]$, *where*

(7.12) $$\alpha := \partial^2 Q(0).$$

Proof. The assumption implies, looking at the second derivative at $n = 0$, that

$$\|\partial^2(P - Q)(0) + \partial^2 Q(h)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N^2$$

for $\gg \delta^{O(1)} N$ values of $h \in [N]$. Applying Lemma 4.5 then implies that

$$\|\partial^2(P - Q)(0) + \partial^2 Q\|_{C^\infty[N], \delta^{-O(1)}} \ll \delta^{-O(1)}/N^2.$$

Thus, as stated, we have

$$\|\partial^i Q\|_{\mathbb{R}/\mathbb{Z}, \delta^{-O(1)}} \ll \delta^{-O(1)}/N^i$$

for $i \geqslant 3$, which means in view of the Taylor expansion of Q that we can write

$$Q(n) = \alpha \binom{n}{2} + R(n),$$

where $R(0) = R(1) = R(2) = 0$ and $\|R\|_{C^\infty[N], \delta^{-O(1)}} \ll \delta^{-O(1)}$. Substituting back into our assumption yields that

$$\left\| P(n) + (\alpha h + \sigma(h))n + R(n + h) - R(h) + \alpha \binom{h}{2} \right\|_{C^\infty[N]} \ll \delta^{-O(1)}$$

for $\gg \delta^{O(1)} N$ values of $h \in [N]$. Differentiating at zero and recalling that $P(0) = 0$, we obtain

$$\|P(1) + \sigma(h) + \alpha h + \partial R(h)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N,$$

which implies, in view of the properties of R, that

$$\|P(1) + \sigma(h) + \alpha h\|_{\mathbb{R}/\mathbb{Z}, \delta^{-O(1)}} \ll \delta^{-O(1)}/N.$$

This completes the proof. $\qquad \square$

Now let us recall (7.11). We know that $\|\eta \circ g_h^\square\|_{C^\infty[N]} \ll \delta^{-O(1)}$ for $\gg \delta^{O(1)} N$ values of h, so let us apply the lemma with $P := \eta_1 \circ g$,

(7.13) $$Q := \eta_2 \circ g_2,$$

and $\sigma(h) := \eta_2([g(1), \{g(1)\}^h])$. By pigeonholing in h we see that there is some $q \leqslant \delta^{-O(1)}$ for which

$$\|q\eta_1(g(1)) + q\eta_2([g(1), \{g(1)^h\}]) + q\alpha h\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N.$$

By redefining η_1 and η_2 (none of the boundedness properties of Lemma 7.11 are lost by doing this), we may write this as

$$(7.14) \qquad \|\eta_1(g(1)) + \eta_2([g(1), \{g(1)^h\}]) + q\alpha h\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N.$$

We now proceed as in Section 5, using Mal'cev bases to work with explicit bracket polynomials.

Since η_2 annihilates $[G, [G, G]] \subseteq [G, G_2]$, we see that the map $x \mapsto \eta_2([g(1), x])$ is a homomorphism. Thus there exists $\zeta \in \mathbb{R}^m$ such that

$$(7.15) \qquad\qquad \eta_2([g(1), x]) = \zeta \cdot \psi(x) (\text{mod } \mathbb{Z})$$

for all $x \in G$. Since η_2 annihilates $[G, G_2]$, all but the first m_{lin} coordinates of ζ are zero. Since we have reduced to the case $|\psi(g(1))| \leqslant 1$ and the basis $\mathcal{X}$ is $\frac{1}{\delta}$-rational, it follows that $|\zeta| \ll \delta^{-O(1)}$.

We now define $\beta := \eta_1(g(1))$ and $\gamma := \psi(g(1))$. Now since $[G, G] \subseteq G_2$, the map $\psi_{\mathrm{lin}} : G \to \mathbb{R}^{m_{\mathrm{lin}}}$ which picks out the first m_{lin} Mal'cev coordinates is a homomorphism, and therefore the first m_{lin} coordinates of $\psi(g(1)^h)$ are just γh. We may now rewrite (7.14) as

$$(7.16) \qquad\qquad \|\beta + q\alpha h + \zeta \cdot \{\gamma h\}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N$$

for $\gg \delta^{O(1)}N$ values of $h \in [N]$.

This assumption is the same as in Proposition 5.3, except that we do not have a bound on $|q\alpha|$. However, we have

CLAIM 7.7. *At least one of the following statements holds*:

(i) *There is $r \ll \delta^{-O(1)}$ such that $\|r\zeta_i(\text{mod } \mathbb{Z})\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N$ for $i = 1, \ldots, m_{\mathrm{lin}}$;*

(ii) *There exists $k \in \mathbb{Z}^{m_{\mathrm{lin}}}$, $0 < |k| \ll \delta^{-O(1)}$ such that $\|k \cdot \gamma\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N$.*

Proof. We apply Proposition 5.3 with $\zeta' := (\zeta, 1) \in \mathbb{R}^{m_{\mathrm{lin}}} \times \mathbb{R}$, $\gamma' := (\gamma, q\alpha) \in \mathbb{R}^{m_{\mathrm{lin}}} \times \mathbb{R}$ and $\alpha' := 0$, deducing that either $|\zeta_i'| \ll \delta^{-O(1)}/N$ for all $i = 1, \ldots, m_{\mathrm{lin}}$ (in which case (i) holds) or else there exist $k \in \mathbb{Z}^{m_{\mathrm{lin}}}$ and $r \in \mathbb{Z}$, not both zero and with $|k|, |r| \ll \delta^{-O(1)}$, such that $\|k \cdot \gamma + qr\alpha\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N$. If $r = 0$, then (ii) holds, so assume that $r \neq 0$. Multiplying (7.16) through by r we see that for $\geqslant \delta N$ values of $h \in [N]$, we have

$$\|\tilde{\beta} + \tilde{\alpha} h + \tilde{\zeta} \cdot \{\gamma h\}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N,$$

where $\tilde{\beta} := r\beta$, $\tilde{\alpha} := \{k \cdot \gamma + qr\alpha\}$ satisfies $|\tilde{\alpha}| \leqslant \delta^{-O(1)}/N$ and $\tilde{\zeta} := r\zeta - k$. Thus we may apply Proposition 5.3 once more to conclude that either $|\tilde{\zeta}_i| \ll \delta^{-O(1)}/N$ for $i = 1, \ldots, m_{\mathrm{lin}}$, which implies (i), or else there is a nonzero $\tilde{k} \in \mathbb{Z}^{m_{\mathrm{lin}}}$ such that $\|\tilde{k} \cdot \gamma\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N$, which implies (ii). This establishes the claim. $\square$

If Claim 7.7(ii) holds, then consider the map $\eta : G \to \mathbb{R}/\mathbb{Z}$ defined by

$$\eta(x) := k \cdot \psi(x) (\mathrm{mod}\ \mathbb{Z}).$$

Since $k \in \mathbb{Z}^{m_{\mathrm{lin}}}$, η is a horizontal character and we have $|\eta| = |k| \ll \delta^{-O(1)}$. Finally, we have

$$\eta \circ g(n) = \eta(g(1)^n) = nk \cdot \gamma (\mathrm{mod}\ \mathbb{Z}),$$

and so $\|\eta \circ g\|_{C^{\infty}[N]} \ll \delta^{-O(1)}$. This completes the proof of Theorem 7.1 in this case.

Suppose then Claim 7.7(i) of the claim holds. For each $i = 1, \ldots, m$, consider the map $\tau_i : G \to \mathbb{R}/\mathbb{Z}$ defined by

$$\tau_i(x) := r\eta_2([x, \exp(X_i)]).$$

Since $[\Gamma, \Gamma] \subseteq \Gamma$ and $[G, G] \subseteq G_2$, we see from the properties established in Lemma 7.5 that τ_i is a horizontal character which annihilates G_2. It is not hard to establish that $|\tau_i| \ll \delta^{-O(1)}$. To do this we write (as usual)

$$\tau_i(x) = k_i \cdot \psi(x) (\mathrm{mod}\ \mathbb{Z}),$$

where $k_i \in \mathbb{Z}^m$ (and in fact $k_i \in \mathbb{Z}^{m_{\mathrm{lin}}}$ since τ_i annihilates G_2). From the definition of τ_i, the bound $r \ll \delta^{-O(1)}$, the $\frac{1}{\delta}$-rationality of the basis $\mathcal{X}$ and Lemma A.3, we have

$$(k_i)_j = \tau_i(\exp(X_j)) = r\eta_2([\exp(X_j), \exp(X_i)]) \ll \delta^{-O(1)},$$

and so indeed $|\tau_i| = |k_i| \ll \delta^{-O(1)}$. Now we have

$$\tau_i \circ g(n) = n\eta_i(g(1)) = rn\zeta_i (\mathrm{mod}\ \mathbb{Z}),$$

where the last equality follows from (7.15). By property (i), this implies that

$$\|\tau_i \circ g\|_{C^{\infty}[N]} \ll \delta^{-O(1)},$$

and so once again we have proved Theorem 7.1 unless $\tau_i = 0$ for all $i = 1, \ldots, m$.

So far we have been successful in deducing Theorem 7.1 by induction on the degree d, but we know from the example at the start of this section that it is not always possible to make such a deduction as $G_{\bullet}$ may be "reducible" for g. It turns out that the case we have not yet covered corresponds to this situation.

Suppose then that $\tau_i = 0$ for all i, so that $\eta_2([x, \exp(X_i)]) = 0$ for all $x \in G$ and all $i \in [m]$. Since the homomorphism η_2 annihilates $[G, [G, G]] \subseteq [G, G_2]$, we see using the identity $[x, yz] = [x, z][z^{-1}, [x, y]][x, y]$ that the map $y \mapsto \eta_2([x, y])$ is a homomorphism for any fixed x. It follows that $\eta_2([x, y]) = 0$ for all $x, y \in G$, or in other words that η annihilates $[G, G]$. Thus $\zeta = 0$ (cf. (7.15)) and (7.16) degenerates to

$$\|\beta + q\alpha h\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N$$

for $\gg \delta^{O(1)} N$ values of $h \in [N]$. By Lemma 3.2 this implies that

$$\|\alpha\|_{\mathbb{R}/\mathbb{Z}, \delta^{-O(1)}} \ll \delta^{-O(1)}/N^2,$$

and thus by (7.12),

$$\|\partial^2 Q\|_{\mathbb{R}/\mathbb{Z}, \delta^{-O(1)}} \ll \delta^{-O(1)}/N^2,$$

where Q was defined in (7.13). We have $Q(0) = Q(1) = 0$ and, by Lemma 7.6, $\|\partial^i Q\|_{\mathbb{R}/\mathbb{Z}, \delta^{-O(1)}} \ll \delta^{-O(1)}/N^i$ for $i \geqslant 3$. Thus

$$\|\eta_2 \circ g_2\|_{C^\infty[N], \delta^{-O(1)}} \ll \delta^{-O(1)}.$$

Thus there exists q, $1 \leqslant q \leqslant \delta^{-O(1)}$, such that

$$\|q\eta_2 \circ g_2\|_{C^\infty[N]} \ll \delta^{-O(1)}.$$

For notational simplicity we rename $q\eta_2$ as η_2, thus

(7.17) $$\|\eta_2 \circ g_2\|_{C^\infty[N]} \ll \delta^{-O(1)}.$$

Roughly speaking, this statement means that g exhibits some essentially *linear* behaviour (in the "direction" orthogonal to η_2) inside G_2. For our purposes this means that G_2 was too large to accurately capture the quadratic and higher order terms of g, and we must pass to a finer filtration $G'_\bullet$ which does not have this drawback. This is the point in the proof where we induct on the nonlinearity degree m_*.

Now $\eta_2 : G_2 \to \mathbb{R}/\mathbb{Z}$ has the form

$$\eta_2(x) = k \cdot \psi(x) (\mathrm{mod} \ \mathbb{Z}),$$

where $k \in \mathbb{Z}^{m_2} \subseteq \mathbb{Z}^m$ satisfies $|k| \ll \delta^{-O(1)}$. In the ensuing discussion we will also need the lift $\tilde{\eta}_2 : G_2 \to \mathbb{R}$ defined by

$$\tilde{\eta}_2(x) := k \cdot \psi(x).$$

Now the map $\theta : G_2 \times G_2 \to \mathbb{R}$ defined by $\theta(x, y) := \tilde{\eta}_2(xy) - \tilde{\eta}_2(x) - \tilde{\eta}_2(y)$ is continuous, $\mathbb{Z}$-valued, and vanishes when $x = y = \mathrm{id}_G$. Since $G_2 \times G_2$ is connected, it follows that $\theta = 0$ identically, and hence the lift $\tilde{\eta}_2$ is a homomorphism.

LEMMA 7.8 (A finer subgroup sequence). *Define* $G'_0 = G'_1 = G$ *and* $G'_i = G_i \cap \ker \tilde{\eta}_2$ *for* $i \geqslant 2$. *Then* $G'_\bullet = (G'_i)_{i=0}^\infty$ *is a filtration with degree at most* d *and nonlinearity degree* $m'_* \leqslant m_* - 1$. *Each* G'_i *is closed, connected and* $\delta^{-O(1)}$-*rational (with respect to our Mal'cev basis* $\mathcal{X}$ *on* G/Γ *adapted to* $G_\bullet$).

Proof. Let $\pi : G_2 \to G_2/[G_2, G_2]$ be the natural projection. It follows from the Baker-Campbell-Hausdorff formula $\exp(X)\exp(Y) = \exp(X + Y + \frac{1}{2}[X, Y] + \dots)$ that $\pi \circ \exp : \mathfrak{g}_2 \to G_2/[G_2, G_2]$ is a linear map. Since $\tilde{\eta}_2 : G_2 \to \mathbb{R}$ factors through $G_2/[G_2, G_2]$, it follows that $\tilde{\eta}_2 \circ \exp : \mathfrak{g}_2 \to \mathbb{R}$ is also a linear map. For $i = m_{\mathrm{lin}} + 1, \dots, m$, we have $\tilde{\eta}_2 \circ \exp(X_i) = k_i$, an integer

of magnitude $O(\delta^{-O(1)})$. Thus by simple linear algebra we see that each Lie algebra $\mathfrak{g}'_i = \mathfrak{g}_i \cap \ker(\tilde{\eta}_2 \circ \exp)$ is spanned by $O(\delta^{-O(1)})$-rational combinations of the X_i. Thus the G'_i are $O(\delta^{-O(1)})$-rational closed connected subgroups as claimed.

If $i, j \geqslant 2$, then it is clear that $[G'_i, G'_j] \subseteq G'_{i+j}$ since $\eta_2 : G_2 \to \mathbb{R}$ is a homomorphism. We must also check that $[G, G'_i] \subseteq G'_{i+1}$ for $i \geqslant 2$, which follows from the fact that $[G, G_i] \subseteq [G, G_2] \subseteq \ker \eta_2$. The statement about m'_* is immediate from the fact that η is nontrivial, and it is obvious that the degree of $G'_\bullet$ is at most d. $\square$

We now come to the main result of this section, which allows us to pass to a new sequence $g' \in \mathrm{poly}(\mathbb{Z}, G'_\bullet)$ with smaller nonlinearity degree than g.

LEMMA 7.9 (Factorisation lemma). *Suppose that* (7.17) *holds. Then we may factor* $g = \varepsilon g' \gamma$, *where*

(i) $\varepsilon \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$, $\varepsilon(0) = \mathrm{id}_G$, ε *is* $(\delta^{-O(1)}, N)$-*smooth (cf. Definition 1.18) and* $\|\eta \circ \varepsilon\|_{C^\infty[N]} \ll \delta^{-O(1)}$ *for all horizontal characters* $\eta : G \to \mathbb{R}/\mathbb{Z}$ *with* $0 < \|\eta\| \ll \delta^{-O(1)}$;
(ii) $g' \in \mathrm{poly}(\mathbb{Z}, G'_\bullet)$;
(iii) $\gamma \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ *and* $\gamma(n)\Gamma$ *is periodic with period* $Q \ll \delta^{-O(1)}$.

We remark that this lemma is strikingly similar in form to Proposition 9.2 below. The proof of the latter result will, in fact, be closely modelled on the proof of this one, but will be rather easier.

Proof. By Lemma 6.7 and the fact that $g_2(0) = g_2(1) = \mathrm{id}_G$, we have

$$\psi(g_2(n)) = \binom{n}{2}t_2 + \binom{n}{3}t_3 + \cdots + \binom{n}{d}t_d,$$

where $t_i \in \mathbb{R}^m$ and the coordinate $(t_i)_j$ is equal to 0 if $j \leqslant m - m_i$. Thus

$$\tilde{\eta}_2 \circ g_2(n) = \sum_{i=2}^{d} k \cdot t_i \binom{n}{i}.$$

From (7.17) we thus have

$$\|k \cdot t_i\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N^i,$$

$i = 2, \ldots, d$. Since $|k| \ll \delta^{-O(1)}$, we may choose vectors $u_i \in \mathbb{R}^m$ with $(u_i)_j = 0$ if $j \leqslant m - m_i$ such that $|t_i - u_i| \ll \delta^{-O(1)}/N^i$ and $k \cdot u_i \in \mathbb{Z}$ for $i = 2, \ldots, d$.

We may now pick vectors v_i in $\mathbb{R}^m$ with $(v_i)_j = 0$ if $j \leqslant m - m_i$, all of whose coordinates are rationals over some denominator $q \ll \delta^{-O(1)}$, such that $k \cdot u_i = k \cdot v_i$ for $i = 2, \ldots, d$.

Define sequences $\varepsilon, \gamma : \mathbb{Z} \to G$ by

(7.18)
$$\psi(\varepsilon(n)) := \sum_{i=2}^{d} \binom{n}{i} (t_i - u_i) \quad \text{and} \quad \psi(\gamma(n)) := \sum_{i=2}^{d} \binom{n}{i} v_i,$$

and set
$$g'(n) := \varepsilon(n)^{-1} g(n) \gamma(n)^{-1}.$$
Observe from Lemma 6.7 that ε, γ lie in $\mathrm{poly}(\mathbb{Z}, G_\bullet)$ and take values in $G_{(2)}$. We verify the properties of ε, g', and γ in turn.

That $\varepsilon(0) = \mathrm{id}_G$ is obvious. To see that ε is $(\delta^{-O(1)}, N)$-smooth we must confirm that $d(\varepsilon(n), \varepsilon(n-1)) \ll \delta^{-O(1)}/N$ for all $n \in [N]$. Now, as a fairly immediate consequence of the definition of ε, we have that
$$|\psi(\varepsilon(n)) - \psi(\varepsilon(n-1))| \ll \delta^{-O(1)}/N$$
and
$$|\psi(\varepsilon(n))| \ll \delta^{-O(1)}$$
for all $n \in [N]$. The smoothness therefore follows from Lemma A.4. Finally we must establish the statement about $\eta \circ \varepsilon$, where $\eta : G \to \mathbb{R}/\mathbb{Z}$ is a horizontal character. It is clear that any horizontal character $\eta : G \to \mathbb{R}/\mathbb{Z}$ is represented in coordinates as
$$\eta(g) = k \cdot \psi(g) (\mathrm{mod}\ \mathbb{Z}),$$
where $k_i = \eta(\exp(X_i'))$ and so in particular $|k| \ll \delta^{-O(1)}$ if $\|\eta\| \ll \delta^{-O(1)}$. It follows immediately from the definition of ε that $\|\eta \circ \varepsilon\|_{C^\infty[N]} \ll \delta^{-O(1)}$, as required.

Next we show that $g' \in \mathrm{poly}(\mathbb{Z}, G_\bullet')$. Now we have
$$g'(n) = \varepsilon^{-1}(n) g(n) \gamma(n)^{-1} = \varepsilon(n)^{-1} g_2(n) \gamma(n)^{-1} \cdot g(1)^n \cdot [g(1)^{-n}, \gamma(n)].$$
The first derivative of the sequence $n \mapsto g(1)^n$ is $g(1)$ and all higher derivatives are just id_G, so this sequence has coefficients in *any* subgroup sequence. Also the sequence $[g(1)^{-n}, \gamma(n)]$ lies in $\mathrm{poly}(\mathbb{Z}, G_\bullet')$ since it is in $\mathrm{poly}(\mathbb{Z}, G_\bullet)$ and takes values in $[G, G_2]$, which is annihilated by η.

By the group property of $\mathrm{poly}(\mathbb{Z}, G_\bullet')$ it therefore suffices to check that $\varepsilon^{-1} g_2 \gamma^{-1} \in \mathrm{poly}(\mathbb{Z}, G_\bullet')$. Since this sequence lies in $\mathrm{poly}(\mathbb{Z}, G_\bullet)$, we need only check that it is annihilated by η, that is to say that
$$-\eta(\gamma(n)) - \eta(\varepsilon(n)) + \eta(g_2(n)) = 0.$$
Computing using coordinates we see that the left-hand side here is
$$\sum_{i=2}^{d} k \cdot (-v_i + u_i - t_i + t_i) \binom{n}{i},$$
which does indeed vanish by our construction of u_i and v_i.

Finally we must check that $\gamma(n)\Gamma$ is periodic. By Lemma A.11 we see that γ is $\delta^{-O(1)}$-rational (cf. Definition 1.17), and then the result follows instantly from Lemma A.12(ii). $\qquad\square$

We will shortly be completing the proof of Theorem 7.1 in the case that (7.17) holds, which is the only case left to handle. We isolate a technical lemma which allows us to deduce $C^\infty[N]$-properties of polynomials $p(n)$ from properties of $p(an + b)$.

LEMMA 7.10 (Single-parameter extrapolation). *Suppose that $Q, N \geqslant 1$ are integers and a, b are rationals with height at most Q such that $b \neq 0$. Let $p : \mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be a polynomial sequence of degree d, and write $\tilde{p}(n) := p(a+bn)$. Then there is some $q \in \mathbb{Z}$, $1 \leqslant |q| \ll_d Q^{O_d(1)}$, such that*

$$\|qp\|_{C^\infty[N]} \ll_d Q^{O_d(1)} \|\tilde{p}\|_{C^\infty[N]}.$$

We will defer the proof of this lemma to the next section, in which we prove a more general multiparameter version of it (see Lemma 8.4).

Recall now that in our efforts to prove Theorem 7.1 by induction we had reduced to the following situation: $g : \mathbb{Z} \to G$ is a polynomial sequence with $g(0) = \mathrm{id}_G$ and $|\psi(g(1))| \leqslant 1$, and there is a function $F : G/\Gamma \to \mathbb{C}$ with nontrivial vertical oscillation ξ and $\|F\|_{\mathrm{Lip}} \leqslant 1$ such that

$$|\mathbb{E}_{n \in [N]} F(g(n)\Gamma)| \geqslant \delta.$$

Furthermore we reduced to the case when g is "reducible" in the sense that (7.17) holds. This allows us to factor g as in Lemma 7.9, obtaining

$$|\mathbb{E}_{n \in [N]} F(\varepsilon(n) g'(n) \gamma(n) \Gamma)| \geqslant \delta.$$

Choose a $Q \ll \delta^{-O(1)}$ such that $\gamma(n)\Gamma$ is periodic with period Q, and split $[N]$ up into progressions of length between N' and $2N'$, where $N' := \lfloor \delta^C N \rfloor$, and common difference Q. By the pigeonhole principle, there is some such progression $\{n_0 + nQ : n \in [N']\}$ such that

$$\left| \mathbb{E}_{n \in [N']} F\left(\varepsilon(n_0 + nQ) g'(n_0 + nQ) \{\gamma(n_0)\}\Gamma\right) \right| \geqslant \delta/2.$$

Now since ε is $(\delta^{-O(1)}, N)$-smooth we see, using the right-invariance of d, that if C is sufficiently small, then

(7.19) $$\left| \mathbb{E}_{n \in [N']} F\left(\varepsilon(n_0) g'(n_0 + nQ) \{\gamma(n_0)\}\Gamma\right) \right| \geqslant \delta/4.$$

Now $g' \in \mathrm{poly}(\mathbb{Z}, G'_\bullet)$ and hence, by Lemma 6.8, the sequence

$$\tilde{g}(n) := \{g(n_0)\}^{-1} \varepsilon(n_0) g'(n_0 + nQ) \{\gamma(n_0)\}$$

is also in $\mathrm{poly}(\mathbb{Z}, G'_\bullet)$. The inequality (7.19) may be rewritten as

(7.20) $$|\mathbb{E}_{n \in [N']} \tilde{F}(\tilde{g}(n)\Gamma)| \geqslant \delta/4,$$

where $\tilde{F}(x) := F(\{g(n_0)\}x)$. By Lemma A.5, we have $\|\tilde{F}\|_{\mathrm{Lip}} \ll \delta^{-O(1)}$. Noting that $\tilde{g}(0) = \mathrm{id}_G$, we may thus apply the inductive hypothesis that

Theorem 7.1 holds with parameters $(d, m_* - 1)$, deducing that there is some horizontal character $\tilde{\eta}$ with $0 < \|\tilde{\eta}\| \ll \delta^{-O(1)}$ such that

$$\|\tilde{\eta} \circ \tilde{g}\|_{C^\infty[N]} \ll \delta^{-O(1)}.$$

From Lemma 7.10 and the definition of $\tilde{g}$ it follows that there is a horizontal character η with $0 < \|\eta\| \ll \delta^{-O(1)}$ such that

$$\|\eta \circ g''\|_{C^\infty[N]} \ll \delta^{-O(1)},$$

where

$$g''(n) := \{g(n_0)\}^{-1} \varepsilon(n_0) g'(n) \{\gamma(n_0)\}.$$

Since $g'(0) = \mathrm{id}_G$, it follows that

$$\|\eta \circ g'\|_{C^\infty[N]} \ll \delta^{-O(1)}.$$

To complete the proof of the result we must, of course, replace g' by $g := \varepsilon g' \gamma$. To do this, note first that by multiplying η by an integer of size $O(\delta^{-O(1)})$ if necessary, we in fact have

$$\|\eta \circ \gamma\|_{C^\infty[N]} = 0,$$

since the Mal'cev coordinates $\psi(\gamma(n)\Gamma)$ are always rationals over some denominator $\ll \delta^{-O(1)}$. From the property (i) of Lemma 7.9, we have that $\|\eta \circ \varepsilon\|_{C^\infty[N]} \ll \delta^{-O(1)}$. Putting all this together, we obtain

$$\|\eta \circ g\|_{C^\infty[N]} \leqslant \|\eta \circ \varepsilon\|_{C^\infty[N]} + \|\eta \circ g'\|_{C^\infty[N]} + \|\eta \circ \gamma\|_{C^\infty[N]} \ll \delta^{-O(1)},$$

completing (at last!) the proof of Theorem 7.1. $\qquad\square$

Let us remind the reader that, by remarks immediately following the statement of Theorem 7.1, we have also completed the proof of Theorem 2.9.

8. The multiparameter Leibman theorem

We have proved one of our main results, Theorem 2.9. In this section we bootstrap this result into a multiparameter version of itself. Strictly speaking, this step is not necessary in order to establish any of the results stated in the introduction, however the arguments here are not terribly difficult and will be needed in order to obtain multiparameter analogues of the those results.

Recall from Section 6 the definition of $\mathrm{poly}(\mathbb{Z}^t, G_\bullet)$, the group of polynomial sequences $g : \mathbb{Z}^t \to G$ with coefficients in $G_\bullet$. Recall also the definition of, and notation for, multibinomial coefficients $\binom{\vec{n}}{\vec{j}}$.

We need an analogue of the smoothness norms $C^\infty[N]$ in the multiparameter setting. To set these up, we introduce the Taylor coefficients of a polynomial map $g : \mathbb{Z}^t \to \mathbb{R}/\mathbb{Z}$.

Definition 8.1 (Taylor expansion). Suppose that $g : \mathbb{Z}^t \to \mathbb{R}/\mathbb{Z}$ is a polynomial map. Then we define the *Taylor coefficients* $\alpha_{\vec{j}} \in \mathbb{R}/\mathbb{Z}$ for $\vec{j} \in \mathbb{Z}^t$ to be the unique elements of $\mathbb{R}/\mathbb{Z}$ such that

$$g(\vec{n}) = \sum_{\vec{j}} \binom{\vec{n}}{\vec{j}} \alpha_{\vec{j}}$$

for all $\vec{n}$; it is not difficult to verify the existence and uniqueness of these coefficients and to check that if g has degree at most d, then $\alpha_{\vec{j}} = 0$ unless $|\vec{j}| \leqslant d$, where $|\vec{j}| := j_1 + \cdots + j_t$.

Definition 8.2 (Smoothness norms). Suppose that $g : \mathbb{Z}^t \to \mathbb{R}/\mathbb{Z}$ is a polynomial map with Taylor expansion

$$g(\vec{n}) = \sum_{\vec{j}} \alpha_{\vec{j}} \binom{\vec{n}}{\vec{j}}.$$

Then for any t-tuple $\vec{N} = (N_1, \ldots, N_t)$ for $N_1, \ldots, N_t \geqslant 1$, we write $[\vec{N}] := [N_1] \times \ldots \times [N_t]$ and

$$\|g\|_{C^\infty[\vec{N}]} := \sup_{\vec{j} \neq 0} \vec{N}^{\vec{j}} \|\alpha_{\vec{j}}\|_{\mathbb{R}/\mathbb{Z}},$$

where $\vec{N}^{\vec{j}} := N_1^{j_1} \ldots N_t^{j_t}$.

We have the following generalisation of Lemma 2.8.

LEMMA 8.3 (Smooth polynomials vary slowly). *Let* $g : \mathbb{Z}^t \to \mathbb{R}/\mathbb{Z}$ *be a polynomial sequence of degree at most* d, *and suppose that* $\vec{n} \in [\vec{N}]$. *Then for any* $i \in [t]$, *we have*

$$|g(\vec{n}) - g(\vec{n} - \vec{e}_i)| \ll_{t,d} \frac{1}{N_i} \|g\|_{C^\infty[\vec{N}]},$$

where $\vec{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ *is the* i^{th} *basis vector of* $\mathbb{Z}^t$.

Proof. From the Taylor expansion and binomial identities, we have

$$g(\vec{n}) - g(\vec{n} - \vec{e}_i) = \sum_{|\vec{j}| \leqslant d} \binom{\vec{n} - \vec{e}_i}{\vec{j} - \vec{e}_i} \alpha_{\vec{j}}.$$

Thus

$$|g(\vec{n}) - g(\vec{n} - e_i)| \leqslant \frac{1}{N_i} \|g\|_{C^\infty[\vec{N}]} \sum_{\substack{|\vec{j}| \leqslant d \\ \vec{j} \neq 0}} \frac{1}{\vec{N}^{\vec{j} - \vec{e}_i}} \binom{\vec{n} - \vec{e}_i}{\vec{j} - \vec{e}_i} \ll_{t,d} \frac{1}{N_i} \|g\|_{C^\infty[\vec{N}]},$$

as required. □

We now give a multiparameter version of Lemma 7.10, which implies that lemma as the $t = 1$ special case.

LEMMA 8.4 (Multiparameter extrapolation). *Suppose that $t, Q, N_1, \ldots, N_t$, $d \geqslant 1$ are integer parameters and that $a_i, b_i \in \mathbb{Q}$, $i = 1, \ldots, t$ are rationals of height at most Q with $b_i \neq 0$. Let $p : \mathbb{Z}^t \to \mathbb{R}/\mathbb{Z}$ be a polynomial map of degree at most d, and write $\tilde{p}(\vec{n}) := p(a_1 + b_1 n_1, \ldots, a_t + b_t n_t)$. Then there is some $q \in \mathbb{Z}$, $|q| \ll_{d,t} Q^{O_{d,t}(1)}$, such that*

$$\|qp\|_{C^\infty[\vec{N}]} \ll_{d,t} Q^{O_{d,t}(1)} \|\tilde{p}\|_{C^\infty[\vec{N}]}.$$

Proof. First of all observe that, if $a, b \in \mathbb{Q}$ are rationals with height at most Q and $b \neq 0$, we may expand

$$\binom{(n-a)/b}{j} = \sum_{j' \leqslant j} c(a, b, j', j)\binom{n}{j'},$$

where $c(a, b, j', j)$ is a rational number with height $O_j(Q^{O_j(1)})$. Indeed we clearly have $c(a, b, j, j) = b^{-j}$, and we may then compute

$$c(a, b, j-1, j), c(a, b, j-2, j), \ldots$$

in turn.

Multiplying such relations together we obtain a multiparameter version, viz.

$$\prod_{i=1}^t \binom{(n_i - a_i)/b_i}{j_i} = \sum_{\vec{j}' \leqslant \vec{j}} c(\vec{a}, \vec{b}, \vec{j}', \vec{j})\binom{\vec{n}}{\vec{j}},$$

where $\vec{j}' \leqslant \vec{j}$ means that each component of $\vec{j}'$ is at most the corresponding component of $\vec{j}$.

Applying this allows us to give the Taylor coefficients $\alpha_{\vec{j}}$ of p in terms of those of $\tilde{p}$. Indeed we have

$$p(\vec{n}) = \tilde{p}(\frac{n_1 - a_1}{b_1}, \ldots, \frac{n_t - a_t}{b_t})$$

$$= \sum_{\vec{j}} \prod_{i=1}^t \binom{(n_i - a_i)/b_i}{\vec{j}} \tilde{\alpha}_{\vec{j}} = \sum_{\vec{j}} \sum_{\vec{j}' \leqslant \vec{j}} \binom{\vec{n}}{\vec{j}'} c(\vec{a}, \vec{b}, \vec{j}', \vec{j}) \tilde{\alpha}_{\vec{j}},$$

and so

$$\alpha_{\vec{j}} = \sum_{\vec{j}' \geqslant \vec{j}} c(\vec{a}, \vec{b}, \vec{j}, \vec{j}') \tilde{\alpha}_{\vec{j}'}.$$

To obtain the lemma, we simply need to take q to be the product of all the denominators of the rationals $c(\vec{a}, \vec{b}, \vec{j}, \vec{j}')$, which is clearly $\ll_{\vec{d}, t} Q^{O_{d,t}(1)}$. $\square$

Definition 8.5 (Multiparameter equidistribution). Let G/Γ be a nilmanifold and let $\delta > 0$. An finite sequence $(g(\vec{n})\Gamma)_{\vec{n} \in P}$ in G/Γ indexed by a finite

nonempty set P is δ-*equidistributed* if we have

$$\left| \sum_{\vec{n} \in P} F(g(\vec{n})\Gamma) - \int_{G/\Gamma} F \right| \leqslant \delta \|F\|_{\text{Lip}}$$

for all Lipschitz functions $F : G/\Gamma \to \mathbb{C}$. If $\vec{N} = (N_1, \ldots, N_t)$, we say that a sequence $(g(\vec{n})\Gamma)_{\vec{n} \in [N]}$ is *totally δ-equidistributed* if we have

$$\left| \sum_{\vec{n} \in P_1 \times \ldots \times P_t} F(g(\vec{n})\Gamma) - \int_{G/\Gamma} F \right| \leqslant \delta \|F\|_{\text{Lip}}$$

whenever P_i are arithmetic progressions in $[N_i]$ of length at least δN_i for each $1 \leqslant i \leqslant t$.

We can now give the multiparameter version of Theorem 2.9.

THEOREM 8.6 (Multiparameter quantitative Leibman theorem). *Let $s, m, t \geqslant 1$, and $0 < \delta < 1/2$, and let $N_1, \ldots, N_t \geqslant 1$ and $d \geqslant 1$ be integers. Suppose that G/Γ is an m-dimensional nilmanifold equipped with a $\frac{1}{\delta}$-rational Mal'cev basis $\mathcal{X}$ adapted to some filtration $G_\bullet$ of degree d, and suppose that $g \in \text{poly}(\mathbb{Z}^t, G_\bullet)$. Then either $(g(\vec{n})\Gamma)_{\vec{n} \in [\vec{N}]}$ is δ-equidistributed, or else there is some horizontal character η with $0 < \|\eta\| \ll \delta^{-O_{d,m,t}(1)}$ such that*

$$\|\eta \circ g\|_{C^\infty[\vec{N}]} \ll \delta^{-O_{d,m,t}(1)}.$$

Proof. We allow all implied constants to depend on d, m, and t. Suppose that $(g(\vec{n})\Gamma)_{\vec{n} \in [\vec{N}]}$ is not δ-equidistributed. Suppose to begin with that $N_1 \geqslant \delta^{-C}$.

A simple averaging argument confirms that, for $\gg \delta^{O(1)} N_2 \ldots N_t$ values of $(n_2, \ldots, n_t) \in [N_2 \times \cdots \times N_t]$, the polynomial sequence $(g_{n_2, \ldots, n_t}(n))\Gamma)_{n \in [N_1]}$ is not $\delta^{O(1)}$-equidistributed, where $g_{n_2, \ldots, n_t}(n) := g(n, n_2, \ldots, n_t)$.

For each such tuple $(n_2, \ldots, n_t)$, Theorem 2.9 implies that there is some horizontal character $\eta_{n_2, \ldots, n_t}$ with $0 < \|\eta\| \ll \delta^{-O(1)}$ such that

$$\|\eta \circ g_{n_2, \ldots, n_t}\|_{C^\infty[N_1]} \ll \delta^{-O(1)}.$$

By pigeonholing in η and passing to a thinner set of tuples $(n_2, \ldots, n_t)$, we may assume that $\eta_{n_2, \ldots, n_t}$ does not depend on $(n_2, \ldots, n_t)$. Writing $p := \eta \circ g$ and expanding

$$p := \sum_{i_1=0}^{d} p_{i_1}(n_2, \ldots, n_t) \binom{n_1}{i_i},$$

where the p_{i_1} are polynomials, we therefore see that

(8.1) $$\|p_{i_1}(n_2, \ldots, n_t)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N_1^{i_1},$$

for $\gg \delta^{O(1)} N_2 \dots N_d$ values of $(n_2, \dots, n_t)$, for each $i_1 = 0, \dots, d$. In particular (for each i_1) there are $\gg \delta^{O(1)} N_3 \dots N_t$ values of $(n_3, \dots, n_t)$ for which (8.1) holds for $\gg \delta^{O(1)} N_2$ values of n_2.

Suppose that $i_1 > 0$. Writing

$$p_{i_1}(n_2, \dots, n_t) = \sum_{i_2=0}^{t} p_{i_1,i_2}(n_3, \dots, n_t) \binom{n_2}{i_2}$$

and applying Lemma 4.5, we see that for $\gg \delta^{O(1)} N_3 \dots N_t$ tuples $(n_3, \dots, n_t)$, there is $q_{i_1}(n_3, \dots, n_t) \ll \delta^{-O(1)}$ such that

$$\|q_{i_1}(n_3, \dots, n_t) p_{i_1,i_2}(n_3, \dots, n_t)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N_1^{i_1} N_2^{i_2}.$$

Note that the application of Lemma 4.5 is valid because $i_1 > 0$ and $N_1 \geqslant \delta^{-C}$; this guarantees that the parameter ϵ in that lemma is small enough. Pigeonholing in $(n_3, \dots, n_t)$ and passing to a somewhat smaller set of these tuples, we may suppose that $q_{i_1} = q_{i_1}(n_3, \dots, n_t)$ is constant.

We now continue in this vein, obtaining successively quantities $q_{i_1,i_2,\dots,i_r} \ll \delta^{-O(1)}$. At the final stage we obtain

$$\|q_{i_1,\dots,i_t} p_{i_1,\dots,i_t}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/N_1^{i_1} \dots N_t^{i_t}$$

or, in our earlier notation,

(8.2) $$\|q_{\vec{i}} p_{\vec{i}}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/\vec{N}^{\vec{i}}.$$

This has been obtained for all $\vec{i}$ with $i_1 > 0$ on the assumption that $N_1 \geqslant \delta^{-C}$. By switching the indices $i_1, \dots, i_t$ if necessary, we may in fact obtain such a $q_{\vec{i}}$ whenever there is some r with $N_r^{i_r} > \delta^{-C}$. If this is not the case for any r, then (8.2) holds anyway for trivial reasons (for any $q_{\vec{i}} \ll \delta^{-O(1)}$).

Note that by construction the $p_{\vec{i}}$ are simply the Taylor coefficients of p.

Taking $q := \prod_{\vec{i}} q_{\vec{i}}$ we see that $q \ll \delta^{-O(1)}$ and that

$$\|q p_{\vec{i}}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/\vec{N}^{\vec{i}}$$

for each index $\vec{i}$, and thus

$$\|q \eta \circ g\|_{C^\infty[\vec{N}]} \ll \delta^{-O(1)}.$$

The theorem follows. □

9. A multiparameter initial factorisation theorem

Having just established Theorem 8.6, we now use it to obtain an initial factorisation theorem for multiparameter polynomial sequences. We first give a multiparameter version of Definition 1.18, the definition of a smooth sequence. (The multiparameter version of a rational sequence is obvious.)

Definition 9.1 (Multiparameter smooth sequences). Let G/Γ be a nilmanifold with a Mal'cev basis $\mathcal{X}$. Let $(\varepsilon(n))_{n \in \mathbb{Z}^t}$ be a multiparameter sequence in G, let $M \geqslant 1$ be an integer, and let $\vec{N} = (N_1, \dots, N_t)$ with $N_i \geqslant 1$ for all i. We say that $(\varepsilon(n))_{n \in \mathbb{Z}^t}$ is $(M, \vec{N})$-*smooth* if we have $d(\varepsilon(n), \mathrm{id}_G) \leqslant M$ and $d(\varepsilon(\vec{n}), \varepsilon(\vec{n} - \vec{e}_i)) \leqslant M/N_i$ for all $\vec{n} \in [\vec{N}]$.

Here, then, is the main result of this section.

PROPOSITION 9.2 (The factorisation of poorly-distributed polynomial sequences). *Let $s, m, \ t \geqslant 1$, let $0 < \delta < 1/2$, and let $N_1, \dots, N_t \geqslant 1$ and $d \geqslant 0$ be integers. Write $\vec{N} := (N_1, \dots, N_t)$. Let G/Γ be an m-dimensional nilmanifold with a $\frac{1}{\delta}$-rational Mal'cev basis $\mathcal{X}$ adapted to a filtration $G_\bullet$ of degree d, and suppose that $g \in \mathrm{poly}(\mathbb{Z}^t, G)$. Suppose that $(g(\vec{n})\Gamma)_{\vec{n} \in [\vec{N}]}$ is not totally δ-equidistributed. Then there is a factorisation $g = \varepsilon g' \gamma$, where $\varepsilon, g', \gamma \in \mathrm{poly}(\mathbb{Z}^t, G_\bullet)$ are polynomial sequences with the following properties*:

(i) *$\varepsilon : \mathbb{Z}^t \to G$ is $(O(\delta^{-O_{d,m,t}(1)}), \vec{N})$-smooth,*

(ii) *$g' : \mathbb{Z}^t \to G'$ takes values in a connected proper subgroup G' of G which is $O(\delta^{-O_{d,m,t}(1)})$-rational relative to $\mathcal{X}$,*

(iii) *$\gamma : \mathbb{Z}^t \to G$ is $\delta^{-O_{d,m,t}(1)}$-rational.*

Proof. We will allow all implied constants to depend on d, m and t.

We first reduce to the case $g(0) = \mathrm{id}_G$, by factorizing $g = \{g(0)\}\tilde{g}[g(0)]$ where $\tilde{g}$ is the polynomial sequence $\tilde{g} := \{g(0)\}^{-1}\tilde{g}[g(0)]^{-1}$, for which $\tilde{g}(0) = \mathrm{id}_G$. If $(g(\vec{n})\Gamma)_{\vec{n} \in [\vec{N}]}$ is not totally δ-equidistributed, then one easily verifies using Lemma A.5 that $(\tilde{g}(\vec{n})\Gamma)_{\vec{n} \in [\vec{N}]}$ is not totally $\tilde{\delta}$-equidistributed for some $\tilde{\delta} \gg \delta^{O(1)}$. Applying the proposition to $\tilde{g}$, we obtain a factorisation $\tilde{g} = \tilde{\varepsilon} g' \tilde{\gamma}$. Setting $\varepsilon := \{g(0)\}\tilde{\varepsilon}$ and $\gamma := \tilde{\gamma}[g(0)]$, we certainly have $g = \varepsilon g' \gamma$. The sequence γ is $\delta^{-O(1)}$-rational by Lemma A.11 and (the multiparameter version of) Lemma A.12. The sequence ε is $(\delta^{-O(1)}, \vec{N})$-smooth by Lemma A.5.

Henceforth, then, we assume that $g(0) = \mathrm{id}_G$. By hypothesis, we can find progressions $P_i := \{a_i + b_i n_i : n_i \in [N_i']\}$ in $[N_i]$ with $N_i' \geqslant \delta N_i$ such that the polynomial sequence $\tilde{g} : \mathbb{Z}^t \to G$ defined by $\tilde{g}(\vec{n}) = g(a_1 + b_1 n_1, \dots, a_t + b_t n_t)$ is such that $(\tilde{g}(\vec{n})\Gamma)_{\vec{n} \in [\vec{N}']}$ fails to be δ-equidistributed, where $\vec{N}' := (N_1', \dots, N_t')$. By Lemma 6.8 we have $\tilde{g} \in \mathrm{poly}(\mathbb{Z}^t, G_\bullet)$. Applying Theorem 2.9 we conclude the existence of a horizontal character $\tilde{\eta} : G \to \mathbb{R}/\mathbb{Z}$ with $0 < \|\tilde{\eta}\| \ll \delta^{-O(1)}$ such that

$$\|\tilde{\eta} \circ \tilde{g}\|_{C^\infty[\vec{N}']} \ll \delta^{-O(1)}.$$

At the expense of worsening the exponent of the $\delta^{-O(1)}$, we may replace $[\vec{N}']$ here by $[\vec{N}]$. Applying Lemma 8.4, we deduce that there is a horizontal character $\eta : G \to \mathbb{R}/\mathbb{Z}$ with $0 < \|\eta\| \ll \delta^{-O(1)}$ such that

(9.1) $$\|\eta \circ g\|_{C^\infty[\vec{N}]} \ll \delta^{-O(1)}.$$

Take G' to be the connected component of $\ker(\eta)$. Then G' is rather clearly a subgroup of G which is $O(\delta^{-O(1)})$-rational relative to $\mathcal{X}$.

Write

$$\psi(g(n)) = \sum_{\vec{j}} t_{\vec{j}} \binom{\vec{n}}{\vec{j}},$$

where $t_{\vec{j}} \in \mathbb{R}^m$. By Lemma 6.7 we know that the coordinate $(t_{\vec{j}})_i$ is equal to 0 if $i \leqslant m - m_{|\vec{j}|}$. The horizontal character η is given in coordinates by

$$\eta \circ g(\vec{n}) = \sum_{\vec{j}} k \cdot t_{\vec{j}} \binom{\vec{n}}{\vec{j}},$$

where $|k| \ll \delta^{-O(1)}$, and (9.1) tells us that $\|k \cdot t_{\vec{j}}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}/\vec{N}^{\vec{j}}$ for all $\vec{j} \neq 0$. Since $|k| \ll \delta^{-O(1)}$ we may choose vectors $u_{\vec{j}} \in \mathbb{R}^{m_1}$ such that $|t_{\vec{j}} - u_{\vec{j}}| \ll \delta^{-O(1)}/\vec{N}^{\vec{j}}$ and $k \cdot u_{\vec{j}} \in \mathbb{Z}$ for all $\vec{j} \neq 0$. We then choose vectors $v_{\vec{j}} \in \mathbb{R}^{m_1}$, all of whose coordinates are rationals with complexity at most $O(\delta^{-O(1)})$, such that $k \cdot u_{\vec{j}} = k \cdot v_{\vec{j}}$ for all $\vec{j} \neq 0$. We may insist that the $u_{\vec{j}}$ and $v_{\vec{j}}$ have the same support properties as the $t_{\vec{j}}$, namely that $(u_{\vec{j}})_i = (v_{\vec{j}})_i = 0$ if $i \leqslant m - m_{|\vec{j}|}$.

Define polynomial sequences $\varepsilon, \gamma : \mathbb{Z}^t \to G$ in terms of their Mal'cev coordinates by

$$\psi(\varepsilon(\vec{n})) = \sum_{\vec{j} \neq 0}(t_{\vec{j}} - u_{\vec{j}})\binom{\vec{n}}{\vec{j}} \qquad \text{and} \qquad \psi(\gamma(\vec{n})) = \sum_{\vec{j} \neq 0} v_{\vec{j}}\binom{\vec{n}}{\vec{j}},$$

and

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

By Lemma 6.7 and the fact that $\mathrm{poly}(\mathbb{Z}^t, G_\bullet)$ is a group, we see that all three of ε, g', and γ lie in $\mathrm{poly}(\mathbb{Z}^t, G_\bullet)$. We must check the claims (i), (ii), and (iii). The claim (ii) is clear. To prove (i), that is to say that ε is $(\delta^{-O(1)}, \vec{N})$-smooth, we need to show that

$$d(\varepsilon(\vec{n}), \varepsilon(\vec{n} - \vec{e}_i)) \ll \delta^{-O(1)}/N_i$$

for $\vec{n} \in \vec{N}$. But as a fairly immediate consequence of the definition of ε, we have the bound

$$|\psi(\varepsilon(\vec{n})) - \psi(\varepsilon(\vec{n} - \vec{e}_i))| \ll \delta^{-O(1)}/N_i,$$

and so the desired bound follows from Lemma A.4. Finally we note that (iii) follows immediately from the definition of γ and the properties of rational points described in Lemma A.11. $\qquad\square$

10. A multiparameter complete factorisation theorem

The last major task of the paper is to iterate Proposition 9.2 to deduce our a multiparameter version of our main result, Theorem 1.19. We first need a technical lemma.

LEMMA 10.1 (Product of smooth sequences is smooth). *Let G/Γ be a nilmanifold of dimension m, and let $M \geqslant 2$ and $N_1, \ldots, N_t \geqslant 1$ be parameters. Suppose that $\mathcal{X}$ is an M-rational Mal'cev basis for G/Γ adapted to some filtration $G_\bullet$ of degree d, and suppose that the maps $\varepsilon_1, \varepsilon_2 : \mathbb{Z}^t \to G$ are $(M, \vec{N})$-smooth in the sense of Definition 9.1. Then the product $\varepsilon_1 \varepsilon_2$ is $(M^{O_{d,m,t}(1)}, \vec{N})$-smooth.*

Proof. By the triangle inequality, we have, for all $\vec{n} \in \vec{N}$,

$$d(\varepsilon_1 \varepsilon_2(\vec{n} - \vec{e_i}), \varepsilon_1 \varepsilon_2(\vec{n})) \leqslant d(\varepsilon_1(\vec{n} - \vec{e_i})\varepsilon_2(\vec{n} - \vec{e_i}), \varepsilon_1(\vec{n})\varepsilon_2(\vec{n} - \vec{e_i}))$$
$$+ d(\varepsilon_1(\vec{n})\varepsilon_2(\vec{n} - \vec{e_i}), \varepsilon_1(\vec{n})\varepsilon_2(\vec{n})).$$

Using the fact that $d(\varepsilon_1(\vec{n}), \mathrm{id}_G), d(\varepsilon_2(\vec{n}), \mathrm{id}_G) \leqslant Q$ for all $\vec{n} \in [\vec{N}]$, the result now follows immediately from the right-invariance of d, Lemma A.5, and Lemma A.4. $\qquad\square$

We can now state and prove the multiparameter version of Theorem 1.19 that we need.

THEOREM 10.2 (Multiparameter factorisation theorem). *Let $s, m, t \geqslant 0$, let $M_0 \geqslant 2$ and $A > 0$, and let $N_1, \ldots, N_t \geqslant 1$ and $d \geqslant 0$. Suppose that G/Γ is an m-dimensional nilmanifold with a M_0-rational Mal'cev basis $\mathcal{X}$ adapted to some filtration $G_\bullet$ of degree d and that $g \in \mathrm{poly}(\mathbb{Z}^t, G_\bullet)$. Then there is a some M, $M_0 \leqslant M \ll M_0^{O_{A,m,d}(1)}$, a subgroup $G' \subseteq G$ which is M-rational with respect to $\mathcal{X}$ and a decomposition $g = \varepsilon g' \gamma$ into sequences $\varepsilon, g', \gamma \in \mathrm{poly}(\mathbb{Z}^t, G_\bullet)$ with the following properties:*

- (i) *ε is $(M, \vec{N})$-smooth;*
- (ii) *g' takes values in G' and, with respect to the restriction of the metric d, the orbit $(g'(\vec{n})\Gamma')_{\vec{n} \in P_1 \times \cdots \times P_t}$ is $1/M^A$-equidistributed in G'/Γ' for any subprogressions $P_i \subseteq [N_i]$ with $|P_i| \geqslant N_i/M^A$;*
- (iii) *γ is a M-rational.*

Proof. Let $1/M_0^A = \delta_1 > \delta_2 > \ldots$ be a sequence of parameters to be specified as the proof unfolds. For each $i = 1, \ldots, t$, let $P_i \subseteq [N_i]$ be a progression of size at least $\delta_1 N_i$. From Proposition 9.2 we know that either $(g(\vec{n}))_{\vec{n} \in P_1 \times \cdots \times P_t}$ is δ_1-equidistributed on G/Γ, or else there is a factorisation

$$g = \varepsilon_1 g_1 \gamma_1,$$

where $\varepsilon_1, g_1, \gamma_1 \in \mathrm{poly}(\mathbb{Z}^t, G_\bullet)$, g_1 takes values in some $O(\delta_1^{-O(1)})$-rational proper subgroup $G' \subseteq G$, ε_1 is $(O(\delta_1^{-O(1)}), \vec{N})$-smooth, and γ_1 is $O(\delta_1^{-O(1)})$-rational. Set $\Gamma' := G' \cap \Gamma$; we are now going to look at the distribution properties of $(g(\vec{n}))$ inside G'/Γ' by applying Proposition 9.2 once more.

To do this we choose an $M_0^{O_{A,d,m}(1)}$-rational Mal'cev basis $\mathcal{X}'$ for G'/Γ' adapted to the filtration $G'_\bullet := G_\bullet \cap G'$. This is possible by Lemma A.10, and we may furthermore ensure that each of the basis elements $\mathcal{X}'_i$ is an $M_0^{O_{A,d,m}(1)}$-rational combination of the X_i. In view of Lemma A.6, we have

$$(10.1) \qquad d'(x,y) \ll M_0^{O_{A,d,m}(1)} d(x,y)$$

for all $x, y \in G'/\Gamma'$.

Take $\delta_2 := cM_0^{-C}$ for some constants c, C depending on m, d, and A. If these are chosen suitably, and if $(g_1(\vec{n}))_{\vec{n} \in P_1 \times \cdots \times P_t}$ is δ_2-equidistributed on G'/Γ' with respect to the metric d' for all progressions P_i with $|P_i| \geqslant \delta_2 N_i$, then by (10.1) the conclusion of the theorem holds. If this is not the case, then we apply Proposition 9.2 once again, obtaining a factorisation $g_1 = \varepsilon_2 g_2 \gamma_2$ where g_2 takes values in some $O(\delta_2^{-O(1)})$-rational proper subgroup $G'' \subseteq G'$, $\varepsilon_2 : \mathbb{Z}^t \to G'$ is $(O(\delta_2^{-O(1)}), \vec{N})$-smooth and $\gamma_2 : \mathbb{Z}^t \to G'$ is $O(\delta_2^{-O(1)})$-rational.

This allows us to write

$$g = \varepsilon_2 \varepsilon_1 g_2 \gamma_1 \gamma_2.$$

Now it follows from Lemma A.6 that $\varepsilon_2 : \mathbb{Z}^t \to G'$ is in fact $(M_0^{O(1)}, \vec{N})$-smooth when regarded as a map into G (smoothness now being measured with respect to the metric d). By Lemma 10.1, $\varepsilon_2 \varepsilon_1 : \mathbb{Z}^t \to G$ is also $(M_0^{O(1)}, \vec{N})$-smooth. By Lemma A.11(v), $\gamma_1 \gamma_2 : \mathbb{Z}^t \to G$ is $O(\delta_2^{-O(1)})$-rational. Thus, taking $\varepsilon := \varepsilon_2 \varepsilon_1$, $\gamma := \gamma_1 \gamma_2$ and $g' := g_2$, the conclusion of the theorem holds unless $(g_2(\vec{n}))_{\vec{n} \in P_1 \times \cdots \times P_t}$ fails to be equidistributed on G''/Γ''. We now proceed as before, introducing a Mal'cev basis $\mathcal{X}''$ and encoding this lack of equidistribution as the failure of $(g_2(\vec{n}))_{\vec{n} \in P_1 \times \cdots \times P_t}$ to be δ_3-equidistributed relative to the metric $d'' = d_{\mathcal{X}''}$ for some $\delta_3 = cM_0^{-C}$. (The constants c, C are, of course, not the same as before.) We may then apply Proposition 9.2 once more, and so on.

It is clear that the total number of iterations is bounded by $m = \dim G$. The implied constants in the $O()$ notation increase with each iteration, but since the total number of iterations is at most $m = O(1)$, this does not cause a difficulty. Thus we obtain a proof of our main theorem. $\square$

It follows from Lemma A.12 (or rather the multidimensional version of it) that $(\gamma(\vec{n})\Gamma)_{\vec{n} \in \mathbb{Z}^t}$ is periodic in each direction in the sense that $\gamma(\vec{n} + Q\vec{e_i})\Gamma = \gamma(\vec{n})\Gamma$ for some $Q \ll M^{O_{s,m,\vec{d}}(1)}$. Setting $t = 1$, we recover Theorem 1.19.

We leave the straightforward deduction of Theorem 1.20 to the reader.

Appendix A. Facts about coordinates and Mal'cev bases

Let us begin this appendix by discussing coordinate systems on a connected, simply-connected nilpotent Lie group G of dimension m. A discrete

and cocompact subgroup Γ, leading to a nilmanifold G/Γ, will be introduced in a little while. Let $\mathfrak{g}$ be the Lie algebra of G, and let $\exp : \mathfrak{g} \to G$ and $\log : G \to \mathfrak{g}$ be the exponential and logarithm maps, which are both diffeomorphisms. In this appendix all implied constants are allowed to depend on m and s, and for notational brevity this dependence will usually be suppressed. The rationality parameter Q will always be assumed to be at least 2.

Let us begin by recalling from Section 2 the notion of coordinates of the first and second kinds.

Definition A.1 (Coordinates). Let $\mathcal{X} = \{X_1, \ldots, X_m\}$ be a basis for $\mathfrak{g}$. If

$$g = \exp(t_1 X_1 + \cdots + t_m X_m),$$

then we say that $(t_1, \ldots, t_m)$ are the coordinates of the *first kind* or *exponential coordinates* for g relative to the basis $\mathcal{X}$. We write $(t_1, \ldots, t_m) = \psi_{\mathcal{X},\exp}(g)$. If

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

then we say that $(u_1, \ldots, u_m)$ are the coordinates of the *second kind* for g relative to $\mathcal{X}$, and we write $(u_1, \ldots, u_m) = \psi_{\mathcal{X}}(g)$.

From now on, in this appendix (as in the main text) we will write $\psi := \psi_{\mathcal{X}}$ and $\psi_{\exp} := \psi_{\mathcal{X},\exp}$. When another basis $\mathcal{X}'$ for some Lie algebra $\mathfrak{g}'$ is present we shall write $\psi' := \psi_{\mathcal{X}'}$ and $\psi'_{\exp} := \psi_{\mathcal{X},\exp}$.

Recall that $\mathcal{X}$ is said to be Q-rational if all the structure constants c_{ijk} in the relations

$$[X_i, X_j] = \sum_k c_{ijk} X_k$$

are rationals of height at most Q.

The effect of a change of basis is easily understood in coordinates of the first kind. (Indeed, it merely effects a linear transformation of coordinates.) Nilmanifolds, however, are best studied using coordinates of the second kind. It is, therefore, no surprise that the following lemma describing the passage between the two types of coordinate system is very useful.

LEMMA A.2 (Coordinates of the first and second type). (i) *Let $\mathcal{X}$ be a basis for $\mathfrak{g}$ with the nesting property that*

(A.1) $$[\mathfrak{g}, X_i] \subseteq \mathrm{Span}(X_{i+1}, \ldots, X_m)$$

for $i = 1, \ldots, m - 1$. Then the compositions $\psi_{\exp} \circ \psi^{-1}$ and $\psi \circ \psi_{\exp}^{-1}$ are both polynomial maps on $\mathbb{R}^m$ with degree $O(1)$. If $\mathcal{X}$ is Q-rational, then all the coefficients of these polynomials are rational of height at most $Q^{O(1)}$.

(ii) Suppose that $G' \subseteq G$ is a closed, connected subgroup of dimension m' with associated Lie algebra $\mathfrak{g}' \subseteq \mathfrak{g}$. Suppose $\mathcal{X}'$ is a basis for $\mathfrak{g}'$ with the nesting property. Then $\psi \circ \psi'^{-1}$ is a polynomial map from $\mathbb{R}^{m'}$ to $\mathbb{R}^m$ and $\psi' \circ \psi^{-1}$ is a polynomial map from $\psi(G') \subseteq \mathbb{R}^m$ to $\mathbb{R}^{m'}$. Both of these maps have degree

$O(1)$. *If $\mathcal{X}$ and $\mathcal{X}'$ are Q-rational and if each element X'_i of $\mathcal{X}'$ is a Q-linear combination of the X_i, then all coefficients of these polynomials are rationals of height $Q^{O(1)}$.*

Proof. (i) Recall the Baker-Campbell-Hausdorff formula, which states that

$$\log(\exp(X)\exp(Y)) = X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] + \ldots,$$

this expression being a sum of $O_s(1)$ terms, each of which is a rational number of height $O_s(1)$ times a commutator of order at most s involving Xs and Ys. Repeated use of this allows us to write $\exp(u_1 X_1) \ldots \exp(u_m X_m)$ in the form $\exp(t_1 X_1 + \cdots + t_m X_m)$. Property (A.1) is easily seen to imply that the t_i are polynomials in the u_i with the specific form

(A.2)
$$t_1 = u_1$$
$$t_2 = u_2 + P_2(u_1)$$
$$t_2 = u_3 + P_3(u_1, u_2)$$
$$\cdots$$
$$t_m = u_m + P_m(u_1, \ldots, u_{m-1}).$$

This establishes the claim for $\psi_{\exp} \circ \psi^{-1}$. To prove the result for $\psi \circ \psi_{\exp}^{-1}$ we simply note that the relations (A.2) are of an "upper triangular" form which is easy to invert. Thus the u_i are given in terms of the t_i by polynomial relations of a similar upper triangular form. The quantitative statements follow by the same arguments, keeping track of the heights of the rational numbers involved. We leave the details to the reader.

(ii) Note the decomposition

$$\psi \circ \psi'^{-1} = (\psi \circ \psi_{\exp}^{-1}) \circ (\psi_{\exp} \circ \psi_{\exp}'^{-1}) \circ (\psi'_{\exp} \circ \psi'^{-1}).$$

Of the three maps here, the first one is a polynomial map from $\mathbb{R}^m$ to $\mathbb{R}^m$ by (i) and the third is a polynomial map from $\mathbb{R}^{m'}$ to $\mathbb{R}^{m'}$. The middle map is simply a linear transformation from $\mathbb{R}^{m'}$ to $\mathbb{R}^m$.

The composition $\psi' \circ \psi^{-1}$ may be dealt with in a very similar manner.

Once again the quantitative claims follow by the same arguments, keeping track of heights. We leave the details to the reader. $\qquad\square$

The upper-triangular form of the relations (A.2) allows us to prove the following key result, which describes group multiplication and inversion in coordinates.

LEMMA A.3 (Multiplication and inversion in coordinates). *Let $\mathcal{X}$ be a basis for $\mathfrak{g}$ with the nesting property (A.1). Let $x, y \in G$, and suppose that $\psi(x) = t$ and $\psi(y) = u$. Then*

$$\psi(xy) = (t_1 + u_1, t_2 + u_2 + P_1(t_1, u_1), \ldots, t_m + u_m$$
$$+ P_{m-1}(t_1, \ldots, t_{m-1}, u_1, \ldots, u_{m-1})),$$

where, for each $i = 1, \ldots, m - 1$, $P_i : \mathbb{R}^i \times \mathbb{R}^i \to \mathbb{R}$ is a polynomial of degree $O(1)$. Furthermore

$$\psi(x^{-1}) = (-t_1, -t_2 + \tilde{P}_1(t_1), \ldots, -t_m + \tilde{P}_{m-1}(t_1, \ldots, t_{m-1})),$$

where $\tilde{P}_i : \mathbb{R}^i \to \mathbb{R}$ is a polynomial of degree $O(1)$. Let $Q \geqslant 2$. If $\mathcal{X}$ is Q-rational, then all the coefficients of the polynomials $P_i, \tilde{P}_i$ are rationals of height $Q^{O(1)}$.

Proof. By (A.2) we know that

$$\psi_{\exp}(x) = (t_1, t_2 + R_1(t_1), \ldots, t_m + R_{m-1}(t_1, \ldots, t_{m-1}))$$

and similarly for $\psi_{\exp}(y)$, where $R_i : \mathbb{R}^i \to \mathbb{R}$ is a polynomial for $i = 1, \ldots, m - 1$. It follows from the Baker-Campbell-Hausdorff formula and the nesting property (A.1) that

$$\psi_{\exp}(xy) = (t_1 + u_1, t_2 + u_2 + S_1(t_1, u_1), \ldots, t_m + u_m$$
$$+ S_{m-1}(t_1, \ldots, t_{m-1}, u_1, \ldots, u_{m-1})),$$

where each $S_i : \mathbb{R}^i \times \mathbb{R}^i \to \mathbb{R}$ is again polynomial. The statement about the form of $\psi(xy)$ now follows from a further application of the relations (A.2), and the statement about $\psi(x^{-1})$ is an immediate corollary of it.

To obtain the quantitative versions of these statements we use the same arguments, keeping track of the heights of the rational numbers involved. We leave the details to the reader. $\square$

Recall at this point Definition 2.2, in which a basis $\mathcal{X}$ is used to define metric $d = d_{\mathcal{X}}$ on G. We defined d to be the largest metric such that $d(x, y) \leqslant |\psi(xy^{-1})|$ for all $x, y \in G$, where $|\cdot|$ denotes the ℓ^∞-norm on $\mathbb{R}^m$. For practical purposes it is important to have an understanding of such metrics in terms of the coordinates $\psi(x)$ and $\psi(y)$, or even in terms of coordinates $\psi'(x), \psi'(y)$ relative to some other basis $\mathcal{X}'$. The following lemma provides some information in this regard. Here, and in the rest of this appendix, we write $d := d_{\mathcal{X}}$ and $d' := d_{\mathcal{X}'}$.

LEMMA A.4 (Bounds for d in terms of coordinates). *Suppose that $Q \geqslant 2$. Suppose that $\mathcal{X}, \mathcal{X}'$ are two Q-rational bases for $\mathfrak{g}$, both satisfying the nesting condition (A.1). Suppose that each X_i' is given by a Q-rational combination of the X_i and vice versa. Then for all $x, y \in G$ with $|\psi'(x)|, |\psi'(y)| \leqslant Q$, we have the bound*

$$\text{(A.3)} \qquad\qquad d(x, y) \ll Q^{O(1)} |\psi'(x) - \psi'(y)|,$$

and for all $x, y \in G$ with $d(x, \mathrm{id}_G), d(y, \mathrm{id}_G) \leqslant Q$, we have the bound

$$\text{(A.4)} \qquad\qquad |\psi'(x) - \psi'(y)| \ll Q^{O(1)} d(x, y).$$

Proof. Inequality (A.3) is by far the easier of the two inequalities claimed here and we prove it first. By definition we have $d(x,y) \leqslant |\psi(xy^{-1})|$. Write $\psi'(x) = t$ and $\psi'(y) = u$; by Lemmas A.2 and A.3 we see that the coordinates $\psi(xy^{-1})$ are

$$(P_1(t,u), \ldots, P_m(t,u)),$$

where each $P_i : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is a polynomial of degree $O(1)$ whose coefficients are rationals of height $Q^{O(1)}$. Each of these polynomials of course vanishes when $t = u$, and so we can write (e.g.)

$$P_1(t,u) = P_1(t,u) - P_1(t,t) = \sum_{i=1}^{m}(t_i - u_i)R_{1,i}(t,u),$$

where each $R_{1,i} : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is a polynomial of degree $O(1)$ whose coefficients are rationals of height $Q^{O(1)}$. (One way to see this is to expand P_1 as a sum of monomials $t^{\vec{\alpha}}u^{\vec{\beta}}$.) The bound (A.3) follows immediately.

The second bound, (A.4), is significantly more difficult. We begin by proving the special case in which $\mathcal{X} = \mathcal{X}'$ and $y = \mathrm{id}_G$, or in other words the following claim:

(A.5) $|\psi(x)| \ll Q^{O(1)}d(x, \mathrm{id}_G)$ uniformly for all x with $d(x, \mathrm{id}_G) \leqslant Q$.

Write $\kappa(x,y) := \min(|\psi(xy^{-1})|, |\psi(yx^{-1})|)$. We will use the bound

(A.6) $|\psi(x) - \psi(y)| \ll Q^{O(1)}\kappa(x,y)(1 + \kappa(x,y) + |\psi(y)|)^{O(1)}.$

To prove this when $\kappa(x,y) = |\psi(xy^{-1})|$, we proceed much as in the proof of (A.11): set $x = zy$ and use Lemma A.3 to expand $\psi(x) - \psi(y) = \psi(zy) - \psi(y)$ as a polynomial in the coordinates of $v = \psi(y)$ and $w = \psi(z)$ which vanishes when $w = 0$. When $\kappa(x,y) = |\psi(yx^{-1})|$ we proceed similarly, setting $x = yz^{-1}$.

From (A.6), we see in particular that if $|\psi(y)| \leqslant 1$ and $\kappa(x,y) \leqslant 1$, then

$$|\psi(x)| \leqslant |\psi(y)| + CQ^C\kappa(x,y)$$

for some constant $C \geqslant 1$. Iterating this we see that if $x_0, \ldots, x_n$ are elements of G with $x_0 = \mathrm{id}_G$ and $\kappa(x_0, x_1) + \cdots + \kappa(x_{n-1}, x_n) \leqslant C^{-1}Q^{-C}$, then

$$|\psi(x_n)| \leqslant CQ^C(\kappa(x_0, x_1) + \cdots + \kappa(x_{n-1}, x_n)).$$

Inspecting the definition of d, we conclude that

(A.7) $|\psi(x)| \ll Q^{O(1)}d(x, \mathrm{id}_G)$ whenever $d(x, \mathrm{id}_G) \leqslant C^{-1}Q^{-C}$.

By right-invariance and symmetry of d, we can amplify this to

(A.8) $|\kappa(x,y)| \ll Q^{O(1)}d(x,y)$ whenever $d(x,y) \leqslant C^{-1}Q^{-C}$.

The estimate (A.7) is almost what we need, except that the bound on $d(x, \mathrm{id}_G)$ is too strict. To relax it, we argue as follows. To obtain (A.5), it

suffices to show that

$$|\psi(x_n)| \ll Q^{O(1)}(\kappa(x_0, x_1) + \cdots + \kappa(x_{n-1}, x_n))$$

whenever $x_0, \ldots, x_n \in G$ with $x_0 = \mathrm{id}_G$ and $\kappa(x_0, x_1) + \cdots + \kappa(x_{n-1}, x_n) \leqslant 2Q$ (say).

Using a greedy algorithm, split the path $(x_0, \ldots, x_n)$ into $O(Q^{O(1)})$ paths $(x_i, \ldots, x_j)$ with $\kappa(x_i, x_{i+1}) + \cdots + \kappa(x_{j-1}, x_j) \leqslant C^{-1}Q^{-C}$, plus $O(Q^{O(1)})$ singleton paths (x_i, x_{i+1}) with $C^{-1}Q^{-C} \leqslant \kappa(x_i, x_{i+1}) \leqslant 2Q$. Applying (A.8), we thus see that there exists a path $(y_0, \ldots, y_r)$ with $r = O(Q^{O(1)})$, $y_0 = \mathrm{id}_G$, and $y_r = x_n$, such that $\kappa(y_i, y_{i-1}) \ll Q^{O(1)}$ for all $1 \leqslant i \leqslant r$. In particular (using Lemma A.3) if we write $g_i := y_i y_{i-1}^{-1}$ for $1 \leqslant i \leqslant r$, then we see that $|\psi(g_i)| \ll Q^{O(1)}$. On the other hand, we have the telescoping product

$$x_n = g_r \ldots g_1.$$

Now if $g_1, \ldots, g_r \in G$ are any elements with $|\psi(g_i)| \leqslant t$ for all i, then

$$|\psi(g_1 \ldots g_r)| \ll (1+t)^{O(1)} r^{O(1)}.$$

This may be seen by applying Lemma A.3 repeatedly to expand the product out completely in coordinates. That the first coordinate is polynomially controlled is obvious, and it then follows that the second is also, and so on inductively. Applying this in the present situation gives $|\psi(x_n)| \ll Q^{O(1)}$, and similar arguments for each i give that in fact $|\psi(x_i)| \ll Q^{O(1)}$ uniformly for $0 \leqslant i \leqslant n$. Applying (A.6), we have

$$|\psi(x_i)| \leqslant |\psi(x_{i-1})| + O(Q^{O(1)} \kappa(x_{i-1}, x_i))$$

and (A.5) follows.

We have just established the special case $\mathcal{X} = \mathcal{X}'$, $y = \mathrm{id}_G$ of (A.4). We now deal with the case where $\mathcal{X} = \mathcal{X}'$ but y is arbitrary. Suppose then that $d(x, \mathrm{id}_G), d(y, \mathrm{id}_G) \leqslant Q$. Applying (A.5) we see that $|\psi(x)|, |\psi(y)| \ll Q^{O(1)}$. By Lemma A.3 we therefore have $|\psi(xy^{-1})| \ll Q^{O(1)}$, and hence by (A.3) it follows that $d(xy^{-1}, \mathrm{id}_G) \ll Q^{O(1)}$. Applying (A.5) once more, we see that

$$|\psi(xy^{-1})| \ll Q^{O(1)} d(xy^{-1}, \mathrm{id}_G),$$

which, since d is right-invariant, implies that

(A.9) $$|\psi(xy^{-1})| \ll Q^{O(1)} d(x, y).$$

The claimed result now follows immediately using (A.6).

Finally we turn to the general case in which $\mathcal{X}$ and $\mathcal{X}'$ may be different. We start with the special case of (A.4) just proved, namely

(A.10) $$|\psi(x) - \psi(y)| \ll Q^{O(1)} d(x, y).$$

Applying (A.3) we obtain

$$d'(x, y) \ll Q^{O(1)} |\psi(x) - \psi(y)| \ll Q^{O(1)} d(x, y).$$

In particular, we have $d'(x, \mathrm{id}_G), d'(y, \mathrm{id}_G) \ll Q^{O(1)}$. A second application of (A.10), with $\mathcal{X}$ replaced by $\mathcal{X}'$, then gives

$$|\psi'(x) - \psi'(y)| \ll Q^{O(1)} d'(x, y) \ll Q^{O(1)} d(x, y).$$

This concludes the proof of Lemma A.4. $\qquad\square$

The metric d is right-invariant, that is to say $d(xg, yg) = d(x, y)$ for all $x, y, g \in G$. It is useful to have, in addition, the following approximate left-invariance property.

LEMMA A.5 (Approximate left-invariance of d). *Suppose that $Q \geqslant 2$ and that $\mathcal{X}$ is a Q-rational basis for $\mathfrak{g}$ satisfying the nesting condition (A.1). Suppose that $g, x, y \in G$ are elements with $|\psi(x)|, |\psi(y)|, |\psi(g)| \leqslant Q$. Then we have the bound*

$$d(gx, gy) \ll Q^{O(1)} d(x, y).$$

Proof. We start by observing that uniformly in $g, z \in G$ we have the bound

(A.11) $$|\psi(gzg^{-1})| \ll Q^{O(1)} (1 + |\psi(z)| + |\psi(g)|)^{O(1)} |\psi(z)|.$$

This follows by using Lemma A.3 to conclude that the components of $\psi(gzg^{-1})$ are polynomials of degree $O(1)$ with $Q^{O(1)}$-rational coefficients in the coordinates $v = \psi(g)$ and $w = \psi(z)$, and these polynomials all vanish when $w = 0$. Recall from Definition 2.2 that

(A.12) $$d(x, y) = \inf \Big\{ \sum_{i=0}^{n-1} \min(|\psi(x_{i-1} x_i^{-1})|, |\psi(x_i x_{i-1}^{-1})|) :$$
$$x_0, \ldots, x_n \in G; x_0 = x; x_n = y \Big\}.$$

We see, then, that the lemma will follow from (A.11) (taking $z = x_i x_{i-1}^{-1}$ or $x_{i-1} x_i^{-1}$) if we can show that the infimum may be taken over all those x_i, x_{i-1} that satisfy some bound $\min(|\psi(x_{i-1} x_i^{-1})|, |\psi(x_i x_{i-1}^{-1})|) \ll Q^{O(1)}$. But this follows from the inequality $d(x, y) \ll Q^{O(1)}$, which is an instant consequence of Lemma A.4. $\qquad\square$

We conclude this subsection by recording the following result.

LEMMA A.6 (Comparison lemma). *Suppose that $G' \subseteq G$ is a closed subgroup and that $\mathcal{X}, \mathcal{X}'$ are bases for $\mathfrak{g}, \mathfrak{g}'$ respectively which have the nesting property (A.1). Let $Q \geqslant 2$, and suppose that each X_i' is a Q-rational combination of the X_i. Then we have the bounds*

$$d'(x, y) \ll Q^{O(1)} d(x, y)$$

uniformly for all $x, y \in G'$ with $|\psi(x)|, |\psi(y)| \leqslant Q$ and

$$d(x, y) \ll Q^{O(1)} d'(x, y)$$

uniformly for all $x, y \in G'$ with $|\psi'(x)|, |\psi'(y)| \leqslant Q$.

Proof. We follow essentially the same argument used in the previous lemma. To prove the first bound, for example, replace (A.11) with the bound

$$|\psi'(z)| \ll Q^{O(1)}(1 + |\psi(z)|)^{O(1)}|\psi(z)|.$$

This follows immediately from Lemma A.2(ii), which guarantees that $\psi'(z)$ is a polynomial in the coordinates $\psi(z)$ that vanishes when $\psi(z) = 0$. □

Mal'cev bases. Suppose that G is a connected, simply-connected nilpotent Lie group with a filtration $G_\bullet$. Let us now introduce a discrete and cocompact subgroup Γ to the discussion. Throughout the paper we have assumed that G/Γ comes together with a special type of basis $\mathcal{X}$ called a *Mal'cev basis adapted to $G_\bullet$*, which is invoked whenever it is necessary to discuss the metric structure of G/Γ.

Let us recall from Section 2 the basic properties of these bases:

(i) For each $j = 0, \ldots, m - 1$, the subspace $\mathfrak{h}_j := \mathrm{Span}(X_{j+1}, \ldots, X_m)$ is a Lie algebra ideal in $\mathfrak{g}$, and hence $H_j := \exp \mathfrak{h}_j$ is a normal Lie subgroup of G.

(ii) For every i, $0 \leqslant i \leqslant s$, we have $G_i = H_{m-\dim(G_i)}$ (or equivalently, $\mathfrak{g}_i = \mathfrak{h}_{m-\dim(\mathfrak{g}_i)}$).

(iii) Each $g \in G$ can be written uniquely as $\exp(t_1 X_1) \ldots \exp(t_m X_m)$ for $t_1, \ldots, t_m \in \mathbb{R}$.

(iv) Γ consists precisely of those elements which, when written in the above form, all have $t_1, \ldots, t_m \in \mathbb{Z}$.

Mal'cev bases are not especially flexible in certain ways — for example it is not at all easy to take a Mal'cev basis on G/Γ and use it to construct one on $G^\square/\Gamma^\square$ as we had to do in the proof of Lemma 7.4. For additional flexibility it is convenient to introduce the notion of a *weak* basis for G/Γ. These are only ever used in the process of constructing actual Mal'cev bases with desirable properties.

Definition A.7 (Weak bases). Let $\mathcal{X} = \{X_1, \ldots, X_m\}$ be a basis for $\mathfrak{g}$. Let $Q \geqslant 2$ be a parameter. We say that $\mathcal{X}$ is a *Q-rational weak basis* for G/Γ if $\mathcal{X}$ is Q-rational (cf. Definition 2.4) and if we have $\frac{1}{q}\mathbb{Z}^m \subseteq \psi_{\exp}(\Gamma) \subseteq q\mathbb{Z}^m$ for some $q \leqslant Q$; that is to say the coordinates of $\log \Gamma$ relative to $\mathcal{X}$ are close to being integers.

Note carefully that $\log \Gamma$ is not necessarily a subgroup of $\mathfrak{g}$, as we saw in Section 5 in connection with the Heisenberg example.

We record some simple facts about weak bases.

LEMMA A.8 (Weak bases: simple facts). *Weak bases enjoy the following properties*:

(i) *Suppose that $\mathcal{X}$ is a Q-rational weak basis for G/Γ and that $\mathcal{X}' = \{X_1', \ldots, X_m'\}$ is another basis for $\mathfrak{g}$ with the property that each X_i' is a Q-rational combination of the X_i. Then $\mathcal{X}'$ is a $Q^{O(1)}$-rational weak basis for G/Γ.*

(ii) *Suppose that $\mathcal{X}$ is a Mal'cev basis adapted to some subgroup sequence $G_\bullet$; that is to say conditions* (i), (ii), (iii), *and* (iv) *from the start of the section are satisfied. Suppose that $\mathcal{X}$ is Q-rational. Then $\mathcal{X}$ is an $O(Q^{O(1)})$-rational weak basis for G/Γ.*

Proof. Part (i) is immediate. Part (ii) follows quickly from Lemma A.2.
□

The next proposition allows us to construct Mal'cev bases from weak bases. If $\mathcal{X}$ is a Mal'cev basis for G/Γ and if $G' \subseteq G$ is a subgroup, we say that G' is *Q-rational* if the Lie algebra $\mathfrak{g}'$ is generated by Q-rational combinations of the basis elements X_i.

PROPOSITION A.9 (Construction of Mal'cev bases). *Suppose that $\mathcal{X}$ is a Q-rational weak basis for G/Γ and that $G_\bullet$ is a filtration in which each subgroup G_i is Q-rational. Then there is a Mal'cev basis $\mathcal{X}' = \{X_1', \ldots, X_m'\}$ for G/Γ adapted to $G_\bullet$ in which each X_i' is a $Q^{O(1)}$-rational combination of the basis elements X_i. In particular, the Mal'cev basis $\mathcal{X}'$ is $Q^{O(1)}$-rational.*

Proof. Take a basis for $\mathfrak{g}_d$ consisting of Q-rational linear combinations of the X_i. By straightforward linear algebra this may be extended to a basis of $\mathfrak{g}_{d-1}$ consisting of $Q^{O(1)}$-rational combinations of the X_i. This in turn may be extended to a basis of $\mathfrak{g}_{d-2}$ and so on. In this fashion we obtain a basis $\mathcal{Y} = \{Y_1, \ldots, Y_m\}$ for $\mathfrak{g}$ as a vector space consisting of $Q^{O(1)}$-rational combinations of the X_i such that each $\mathfrak{g}_i$ equals $\mathrm{Span}(Y_{j+1}, \ldots, Y_m)$, where $j = m - m_i$. By Lemma A.8(i) we see that $\mathcal{Y}$ is a $Q^{O(1)}$-rational weak basis for G/Γ.

Since $[\mathfrak{g}, \mathfrak{g}_i] \subseteq \mathfrak{g}_{i+1}$ for all i, we see that the weak basis $\mathcal{Y}$ enjoys the nesting property; that is to say $[\mathfrak{g}, Y_j] \subseteq \mathrm{Span}(Y_{j+1}, \ldots, Y_m)$ for all j.

We now convert this basis $\mathcal{Y}$ into the desired Mal'cev basis by choosing $X_m' = c_m Y_m, \ldots, X_1' = c_1 Y_1$ in turn so that
(A.13)
$$\mathrm{Span}(Y_{i+1}, \ldots, Y_m) \cap \Gamma = \{\exp(n_{i+1}X_{i+1}') \ldots \exp(n_m X_m') : n_{i+1}, \ldots, n_m \in \mathbb{Z}\}$$
for $i = m - 1, \ldots, 0$. Such a basis $\mathcal{X}'$ has all of the properties (i), (ii), (iii), and (iv) required to qualify as a Mal'cev basis. Suppose this is done for $i = j$. Since $\mathcal{Y}$ is a $Q^{O(1)}$-rational weak basis for G/Γ, we see that
$$\left(\mathrm{Span}(Y_j, \ldots, Y_m) \cap \Gamma \right) / \mathrm{Span}(Y_{j+1}, \ldots, Y_m)$$
is generated by $\overline{\exp(c_j Y_j)}$ for some $c_j \in \mathbb{Q}$ with heights bounded by $Q^{O(1)}$. Taking $X_j' := c_j Y_j$, we see that (A.13) holds for $i = j - 1$ too.   □

For applications (for example in the proof of Lemma 7.4) it is convenient to have the following variant of the above proposition.

PROPOSITION A.10 (Mal'cev bases of subnilmanifolds). *Suppose that* $\mathcal{X} = \{X_1, \ldots, X_m\}$ *is a Q-rational Mal'cev basis for G/Γ adapted to a filtration $G_\bullet$. Suppose that $G' \subseteq G$ is a Q-rational subgroup of G and, furthermore, that $G'_\bullet$ is a filtration on G' in which each of the groups G'_i is Q-rational(with respect to the basis $\mathcal{X}$). Write $\Gamma' := \Gamma \cap G'$. Then G'/Γ' has a Mal'cev basis $\mathcal{X}' = \{X'_1, \ldots, X'_{m'}\}$ adapted to $G'_\bullet$ in which each X'_i is a $Q^{O(1)}$-rational combination of the X_i.*

Proof. One simply observes that by linear algebra there is a basis $\mathcal{Y} = \{Y_1, \ldots, Y_{m'}\}$ for $\mathfrak{g}'$ together with an extension $\tilde{\mathcal{Y}} = \{Y_1, \ldots, Y_m\}$ to a basis for $\mathfrak{g}$ such that each of the Y_i is a $Q^{O(1)}$-rational combination of the X_i. By Lemma A.8, $\tilde{\mathcal{Y}}$ is a weak basis for G/Γ, and therefore $\mathcal{Y}$ is a weak basis for G'/Γ'. The result now follows from Proposition A.9 applied to this weak basis. $\square$

Rationality. We now record some simple results about rational points in nilmanifolds G/Γ. Recall Definition 1.11: $g \in G$ is *rational* if $g^r \in \Gamma$ for some integer $r > 0$. Recall also the quantitative version of this, Definition 1.17: $g \in G$ is *Q-rational* if $g^r \in \Gamma$ for some integer r, $0 < r \leqslant Q$.

LEMMA A.11 (Properties of rational points). *Suppose that $\mathcal{X}$ is a Q-rational Mal'cev basis adapted to some subgroup sequence $G_\bullet$, where $Q \geqslant 2$.*
   (i) *If $\gamma \in G$, then γ is rational if and only if $\psi(\gamma) \in \mathbb{Q}^m$.*
   (ii) *The set of rational points in G is a group.*
   (iii) *If $\gamma \in G$ is Q-rational, then $\psi(\gamma) \in \frac{1}{Q'}\mathbb{Z}^m$ for some Q', $1 \leqslant Q' \ll Q^{O(1)}$, which does not depend on g.*
   (iv) *If $\gamma \in G$ is such that $\psi(\gamma) \in \frac{1}{Q}\mathbb{Z}^m$, then γ is $O(Q^{O(1)})$-rational.*
   (v) *If γ, γ' are Q-rational, then $\gamma\gamma'$ and γ^{-1} are $O(Q^{O(1)})$-rational.*

Proof. If γ is rational, then by definition there exists $r \geqslant 1$ such that $\gamma^r \in \Gamma$, and thus $\psi(\gamma^n) \in \mathbb{Z}^m$ whenever n is a multiple of r. Now from Lemma 6.7 we know that the coordinates $\psi(g^n)$ are all polynomials of degree $O(1)$; these vanish at zero and take integer values at multiples of r. By the Lagrange interpolation formula we conclude that all the coefficients of these polynomials are rational and so, in particular, we have $\psi(\gamma) \in \mathbb{Q}^m$.

Suppose conversely that $\psi(\gamma) \in \mathbb{Q}^m$. Then by Lemma A.3 we see that each of $\psi(\gamma^2)$, $\psi(\gamma^3), \ldots$ also lies in $\mathbb{Q}^m$. By another application of Lemma 6.7 and the Lagrange interpolation formula, we conclude that each coordinate of $\psi(\gamma^n)$ is a polynomial with rational coefficients that vanishes at zero. In particular it is easy to see that by choosing $r \in \mathbb{N}$ suitably we may ensure that $\psi(\gamma^r) \in \mathbb{Z}^m$, which of course implies that $\gamma^r \in \Gamma$.

Part (ii) follows immediately from (i) and Lemma A.3.

Claims (iii)–(v) follow by repeating the above arguments, but keeping track of the heights of all the rational numbers involved; the key point is that the group operations, as well as Lagrange interpolation, are all polynomial in nature and so all heights will be $O(Q^{O(1)})$. We omit the routine details.   □

Let us now recall the notion of a rational sequence, also given in Definition 1.11. A sequence $\gamma : \mathbb{Z} \to G$ is rational if $\gamma(n)\Gamma$ is rational for all n, and it is Q-rational if $\gamma(n)\Gamma$ is rational for all n. The next lemma records some useful properties of rational *polynomial* sequences.

LEMMA A.12 (Properties of rational polynomial sequences). *Suppose that* $\gamma : \mathbb{Z} \to G$ *is a polynomial sequence of degree d.*

(i) *Suppose that γ is rational. Then $\gamma(n)\Gamma$ is periodic.*

(ii) *Suppose that there is a Q-rational Mal'cev basis $\mathcal{X}$ for G/Γ and that γ is Q-rational. Then $\gamma(n)\Gamma$ is periodic with period $\ll Q^{O(1)}$.*

Proof. (i). Let $\mathcal{X}$ be any Mal'cev basis for G/Γ. By Lemma 6.7 the coordinates $\psi(\gamma(n))$ are all polynomials of degree $O(1)$, and by the previous lemma and the Lagrange interpolation formula they all have rational coefficients. Clearing denominators, we thus find some q such that $\psi(\gamma(n)) \in \frac{1}{q}\mathbb{Z}^m$ for all integers n. By Lemma A.3 we see that there is some $q' \in \mathbb{N}$ such that, for any $r \in \mathbb{Z}$, we have $\psi(\gamma(n+r)\gamma(n)^{-1}) \in \frac{r}{qq'}\mathbb{Z}^m$. Thus $\gamma(n)\Gamma$ is indeed periodic, with period qq'.

Part (ii) is proved in exactly the same way, once again taking care to keep track of the heights of all rationals involved.   □

We leave the formulation and proof of the multidimensional version of this lemma (that is, concerning maps $\gamma : \mathbb{Z}^t \to G$) to the reader; only trivial modifications are required.

The next result, stating that conjugates of rational subgroups by rational elements are rational, is not needed in the present paper. It is required in the companion paper [12].

LEMMA A.13 (Rational conjugates). *Suppose that $\mathcal{X} = \{X_1, \ldots, X_m\}$ is a Q-rational Mal'cev basis for G/Γ adapted to some filtration. Suppose that $\gamma \in G$ is Q-rational and additionally that the coordinates $\psi(\gamma)$ are all bounded in magnitude by Q. Suppose that $G' \subseteq G$ is a Q-rational subgroup. Then the conjugate $\gamma G' \gamma^{-1}$ is $Q^{O(1)}$-rational.*

Proof. Set $H := \gamma G' \gamma^{-1}$, and let $\mathfrak{h}$ be the corresponding Lie algebra. Recall from basic Lie theory the identity

$$\log(\gamma \exp(X)\gamma^{-1}) = \mathrm{Ad}(\gamma)X,$$

where $\mathrm{Ad}(\gamma) : \mathfrak{g} \to \mathfrak{g}$ is the adjoint automorphism of $\mathfrak{g}$ associated to the element $\gamma \in G$. For the purposes of this argument, all we need is the following immediate consequence of this identity: if $X'_1, \ldots, X'_{m'}$ is a basis for the Lie algebra $\mathfrak{g}'$, then the elements

$$\tilde{X}_i := \log(\gamma \exp(X'_i)\gamma^{-1})$$

are a basis for $\mathfrak{h}$. By assumption we may choose the X'_i to be Q-rational combinations of the X_i. It then follows from Lemmas A.2 and A.3 that each $\tilde{X}_i$ is a $Q^{O(1)}$-rational combination of the X_i.

Fundamental domain and reduction. The next lemma provides a description of G/Γ in terms of coordinates relative to any Mal'cev basis $\mathcal{X}$.

LEMMA A.14 (Reducing to the fundamental domain). *Let $\mathcal{X}$ be a Mal'cev basis adapted to some subgroup sequence $G_\bullet$. Suppose that $g \in G$. Then we may write $g = \{g\}[g]$ in a unique way, where $\psi(\{g\}) \in [0,1)^m$ and $[g] \in \Gamma$.*

Proof. Recall Lemma A.3, which describes the multiplication on G in coordinates relative to $\mathcal{X}$. Using this we may iteratively construct $\gamma_m, \gamma_{m-1}, \ldots, \gamma_1 \in \Gamma$ in such a way that coordinates $i+1, \ldots, m$ of $\psi(g\gamma_m \ldots \gamma_i)$ all lie in the interval $[0,1)$.

The uniqueness also follows easily from Lemma A.3: if $\psi(x\gamma), \psi(x) \in [0,1)^m$, then we may equate coefficients of $\psi(\gamma)$ starting at the right to deduce that $\gamma = \mathrm{id}_G$. $\qquad\square$

Metrics on nilmanifolds. Let $\mathcal{X}$ be a Mal'cev basis for some nilmanifold G/Γ. Recall from Definition 2.2 the manner in which we used the metric $d = d_{\mathcal{X}}$ on G to define a "metric" on G/Γ via

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

We can now prove that d really is a metric on G/Γ (and thus the inverted commas above can be dispensed with).

LEMMA A.15 (Nondegeneracy of metric). *Suppose that $\mathcal{X}$ is a rational Mal'cev basis for a nilmanifold G/Γ, adapted to some filtration. Suppose that $d(x\Gamma, y\Gamma) = 0$. Then $x \equiv y(\mathrm{mod}\,\Gamma)$.*

Proof. Since the metric d on G is right-invariant, we have

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

It suffices to show that the inf here is a actually a minimum, to which end we need only show that for any M, there are just finitely many $\gamma \in \Gamma$ with $d(x, y\gamma) \leqslant M$. By Lemma (A.5) this assumption implies that $d(y^{-1}x, \gamma) \leqslant M'$, for some M' depending on M, the rationality of the Mal'cev basis $\mathcal{X}$, and the size of the coordinates of x and y. This in turn implies that $d(\mathrm{id}_G, \gamma) \leqslant M''$

which, in view of Lemma A.4, implies that $|\psi(\gamma)| \leqslant M'''$. But if $\gamma \in \Gamma$, then the coordinates $\psi(\gamma)$ are all integers, so the result follows. $\square$

LEMMA A.16 (Nilmanifolds are bounded). *Let $Q \geqslant 2$, and suppose that $\mathcal{X}$ is a Q-rational Mal'cev basis for a nilmanifold G/Γ (with respect to some filtration). Then $d(x\Gamma, y\Gamma) \ll Q^{O(1)}$ uniformly in $x, y \in G$.*

Proof. By Lemma A.14 we may choose γ and γ' so that $|\psi(x\gamma)|, |\psi(y\gamma')| \leqslant 1$. The claim now follows immediately from Lemma A.4. $\square$

The final result of this appendix is not used in this paper but is required in Section 2 of the companion paper [12].

LEMMA A.17 (Comparison of metrics on nilmanifolds). *Let $Q \geqslant 2$. Suppose that $G' \subseteq G$ is a closed subgroup and that $\mathcal{X}, \mathcal{X}'$ are Q-rational Mal'cev bases for G/Γ and G'/Γ' respectively such that each X_i' is a Q-rational combination of the X_i. Let d and d' be the metrics induced on G/Γ and G'/Γ' respectively. Then for any $x, y \in G'$, we have*

$$d'(x\Gamma', y\Gamma') \ll Q^{O(1)} d(x\Gamma, y\Gamma)$$

and

$$d(x\Gamma, y\Gamma) \ll Q^{O(1)} d'(x\Gamma', y\Gamma').$$

Proof. We prove the second inequality first. By the proof of Lemma A.15 there is some $\gamma' \in \Gamma'$ such that $d'(x\Gamma', y\Gamma') = d'(x, y\gamma')$. Here we may assume, using Lemma A.14, that $|\psi'(x)|, |\psi'(y)| \leqslant 1$. By Lemma A.16 we have $d'(x, y\gamma') \leqslant Q^{O(1)}$, and therefore by Lemma A.4 and the triangle inequality we have $d'(\mathrm{id}_{G'}, y\gamma') \ll Q^{O(1)}$. By a second application of Lemma A.4 it follows that $|\psi'(y\gamma')| \ll Q^{O(1)}$. By Lemma A.6 we therefore have $d(x, y\gamma') \ll Q^{O(1)} d'(x, y\gamma')$. Since $\Gamma' \subseteq \Gamma$, this implies that

$$d(x\Gamma, y\Gamma) \leqslant d(x, y\gamma') \ll Q^{O(1)} d'(x, y\gamma') = Q^{O(1)} d'(x\Gamma', y\Gamma'),$$

which is the second inequality claimed.

To prove the first inequality we make the same initial manoeuvres. That is, we may assume that $|\psi(x)|, |\psi(y)| \leqslant 1$ and that there is some $\gamma \in \Gamma$ such that $d(x\Gamma, y\Gamma) = d(x, y\gamma)$. Let C be a constant to be specified later. If $d(x, y\gamma) \geqslant Q^{-C}$ then, by Lemma A.16, the bound is trivial. Suppose, then, that $d(x, y\gamma) < Q^{-C}$. This is an assertion to the effect that γ lies "near" G'. We will use the rationality properties of the coordinates of Γ to conclude from this that γ must actually lie in G'.

By Lemmas A.5 and A.3 we obtain $d(z, \gamma) \ll Q^{O(1)-C}$, where $z := y^{-1}x$. Since $d(z, \mathrm{id}_G) \ll Q^{O(1)}$, we have $d(\gamma, \mathrm{id}_G) \ll Q^{O(1)}$, and so by Lemma A.4 it follows that $|\psi(z) - \psi(\gamma)| \ll Q^{O(1)-C}$. It follows from this and Lemma A.2 that

$$(A.14) \qquad\qquad |\psi_{\exp}(z) - \psi_{\exp}(\gamma)| \ll Q^{O(1)-C}.$$

Now G' is defined, in exponential or type I coordinates, as the intersection of the kernels of $O(1)$ linear forms with rational coefficients of height $O(Q^{O(1)})$. The coordinates $\psi(\gamma)$ are integers and so the type I coordinates $\psi_{\exp}(\gamma)$ are, by Lemma A.2, rationals of height $O(Q^{O(1)})$. The element z, of course, lies in G'. If C is chosen sufficiently large, it follows from these observations and (A.14) that indeed γ lies in G' and hence in Γ'.

We now have that $d(x, y\gamma') \ll Q^{O(1)}$, where $\gamma' = \gamma$ lies in Γ'. One final application of Lemma A.6 implies that $d'(x, y\gamma') \ll Q^{O(1)}d(x, y\gamma')$, from which it of course follows that

$$d'(x\Gamma', y\Gamma') \leqslant d'(x, y\gamma') \ll Q^{O(1)}d(x, y\gamma') = Q^{O(1)}d(x\Gamma, y\Gamma).$$

This concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

References

[1] L. AUSLANDER, L. GREEN, and F. HAHN, *Flows on Homogeneous Spaces*, *Ann. Math. Stud.* **53**, Princeton Univ. Press, Princeton, N.J., 1963. MR 0167569. Zbl 0106.36802.

[2] V. BERGELSON, B. HOST, and B. KRA, Multiple recurrence and nilsequences, *Invent. Math.* **160** (2005), 261–303. MR 2138068. Zbl 1087.28007. http://dx.doi.org/10.1007/s00222-004-0428-6.

[3] J. BOURGAIN, On triples in arithmetic progression, *Geom. Funct. Anal.* **9** (1999), 968–984. MR 1726234. Zbl 0959.11004. http://dx.doi.org/10.1007/s000390050105.

[4] L. J. CORWIN and F. P. GREENLEAF, *Representations of Nilpotent Lie Groups and their Applications. Part I. Basic Theory and Examples*, *Cambridge Stud. Adv. Math.* **18**, Cambridge Univ. Press, Cambridge, 1990. MR 1070979. Zbl 0704.22007.

[5] M. EINSIEDLER, G. MARGULIS, and A. VENKATESH, Effective equidistribution for closed orbits of semisimple groups on homogeneous spaces, *Invent. Math.* **177** (2009), 137–212. MR 2507639. Zbl 1176.37003. http://dx.doi.org/10.1007/s00222-009-0177-7.

[6] H. FURSTENBERG, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton Univ. Press, Princeton, N.J., 1981. MR 0603625. Zbl 0459.28023.

[7] ———, Nonconventional ergodic averages, in *The Legacy of John von Neumann* (Hempstead, NY, 1988), *Proc. Sympos. Pure Math.* **50**, Amer. Math. Soc., Providence, RI, 1990, pp. 43–56. MR 1067751. Zbl 0711.28006.

[8] B. GREEN, Generalising the Hardy-Littlewood method for primes, in *International Congress of Mathematicians. Vol. II* (Madrid, 2006), Eur. Math. Soc., Zürich, 2007, pp. 373–399. MR 2275602. Zbl 1157.11007.

[9] B. GREEN and T. TAO, An inverse theorem for the Gowers $U^3(G)$ norm, *Proc. Edinb. Math. Soc.* **51** (2008), 73–153. MR 2391635. Zbl 1202.11013. http://dx.doi.org/10.1017/S0013091505000325.

[10] B. GREEN and T. TAO, Quadratic uniformity of the Möbius function, *Ann. Inst. Fourier* (*Grenoble*) **58** (2008), 1863–1935. MR 2473624. Zbl 1160.11017. http://dx.doi.org/10.5802/aif.2401.

[11] _____, Linear equations in primes, *Ann. of Math.* **171** (2010), 1753–1850. MR 2680398. Zbl 05712763. http://dx.doi.org/10.4007/annals.2010.171.1753.

[12] _____, The Möbius function is strongly orthogonal to nilsequences, *Ann. of Math.* **175** (2012), 541–566. http://dx.doi.org/10.4007.annals.2012.175.2.3.

[13] L. W. GREEN, Spectra of nilflows, *Bull. Amer. Math. Soc.* **67** (1961), 414–415. MR 0126504. Zbl 0099.39104. http://dx.doi.org/10.1090/S0002-9904-1961-10650-2.

[14] P. HALL, A contribution to the theory of groups of prime power order, *Proc. London Math. Soc.* **36** (1933), 29–95. Zbl 0007.29102.

[15] B. HOST and B. KRA, Nonconventional ergodic averages and nilmanifolds, *Ann. of Math.* **161** (2005), 397–488. MR 2150389. Zbl 1077.37002. http://dx.doi.org/10.4007/annals.2005.161.397.

[16] _____, Uniformity seminorms on ℓ^∞ and applications, *J. Anal. Math.* **108** (2009), 219–276. MR 2544760. Zbl 1183.37011. http://dx.doi.org/10.1007/s11854-009-0024-1.

[17] B. KRA, From combinatorics to ergodic theory and back again, in *International Congress of Mathematicians. Vol. III*, Eur. Math. Soc., Zürich, 2006, pp. 57–76. MR 2275670. Zbl 1118.37010.

[18] M. LAZARD, Sur les groupes nilpotents et les anneaux de Lie, *Ann. Sci. Ecole Norm. Sup.* **71** (1954), 101–190. MR 0088496. Zbl 0055.25103.

[19] A. LEIBMAN, Polynomial sequences in groups, *J. Algebra* **201** (1998), 189–206. MR 1608723. Zbl 0908.20029. http://dx.doi.org/10.1006/jabr.1997.7269.

[20] _____, Polynomial mappings of groups, *Israel J. Math.* **129** (2002), 29–60. MR 1910931. Zbl 1007.20035. http://dx.doi.org/10.1007/BF02773152.

[21] _____, Pointwise convergence of ergodic averages for polynomial actions of $\mathbb{Z}^d$ by translations on a nilmanifold, *Ergodic Theory Dynam. Systems* **25** (2005), 215–225. MR 2122920. Zbl 1080.37004. http://dx.doi.org/10.1017/S0143385704000227.

[22] _____, Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold, *Ergodic Theory Dynam. Systems* **25** (2005), 201–213. MR 2122919. Zbl 1080.37003. http://dx.doi.org/10.1017/S0143385704000215.

[23] _____, Orbit of the diagonal in the power of a nilmanifold, *Trans. Amer. Math. Soc.* **362** (2010), 1619–1658. MR 2563743. Zbl 1191.37016. http://dx.doi.org/10.1090/S0002-9947-09-04961-7.

[24] A. I. MAL'CEV, On a class of homogeneous spaces, *Izvestiya Akad. Nauk. SSSR. Ser. Mat.* **13** (1949), 9–32. MR 0028842. Zbl 0034.01701.

[25] D. W. MORRIS, *Ratner's Theorems on Unipotent Flows*, *Chicago Lectures in Math.*, University of Chicago Press, Chicago, IL, 2005. MR 2158954. Zbl 1069.22003.

[26] W. PARRY, Dynamical systems on nilmanifolds, *Bull. London Math. Soc.* **2** (1970), 37–40. MR 0267558. Zbl 0194.05601. http://dx.doi.org/10.1112/blms/2.1.37.

[27] J. PETRESCO, Sur les commutateurs, *Math. Z.* **61** (1954), 348–356. MR 0066380. Zbl 0056.25505. http://dx.doi.org/10.1007/BF01181351.

[28] M. RATNER, Raghunathan's topological conjecture and distributions of unipotent flows, *Duke Math. J.* **63** (1991), 235–280. MR 1106945. Zbl 0733.22007. http://dx.doi.org/10.1215/S0012-7094-91-06311-8.

[29] N. A. SHAH, Invariant measures and orbit closures on homogeneous spaces for actions of subgroups generated by unipotent elements, in *Lie Groups and Ergodic Theory* (Mumbai, 1996), *Tata Inst. Fund. Res. Stud. Math.* **14**, Tata Inst. Fund. Res., Bombay, 1998, pp. 229–271. MR 1699367. Zbl 0951.22006.

[30] E. SZEMERÉDI, Regular partitions of graphs, in *Problèmes Combinatoires et Théorie des Graphes* (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), *Colloq. Internat. CNRS* **260**, CNRS, Paris, 1978, pp. 399–401. MR 0540024. Zbl 0413.05055.

[31] T. TAO, Obstructions to uniformity and arithmetic patterns in the primes, *Pure Appl. Math. Q.* **2** (2006), 395–433. MR 2251475. Zbl 1105.11032.

[32] T. TAO and V. VU, *Additive Combinatorics*, *Cambridge Stud. Adv. Math.*, Cambridge Univ. Press, Cambridge, 2010. MR 2289012. Zbl 1127.11002. http://dx.doi.org/10.1017/CBO9780511755149.

[33] R. C. VAUGHAN, *The Hardy-Littlewood Method*, second ed., *Cambridge Tracts in Math.* **125**, Cambridge Univ. Press, Cambridge, 1997. MR 1435742. Zbl 0868.11046.

[34] A. VENKATESH, Spectral theory of automorphic forms, a very brief introduction, in *Equidistribution in Number Theory, an Introduction* (A. GRANVILLE and Z. RUDNICK, eds.), *NATO Sci. Ser. II. Math. Phys. Chem.* **237**, Kluwer Acad. Publ., Dordrecht, 2007, Proceedings of the NATO Advanced Study Institute on Equidistribution in Number Theory, Montreal, Canada, 11–22 July 2005, pp. 245–260. MR 2290502. Zbl 1181.11040.

[35] T. ZIEGLER, Universal characteristic factors and Furstenberg averages, *J. Amer. Math. Soc.* **20** (2007), 53–97. MR 2257397. Zbl 1198.37014. http://dx.doi.org/10.1090/S0894-0347-06-00532-7.

CENTRE FOR MATHEMATICAL SCIENCES, CAMBRIDGE, ENGLAND
E-mail: b.j.green@dpmms.cam.ac.uk

UNIVERSITY OF CALIFORNIA, LOS ANGELES, LOS ANGELES, CA 90095-1596
E-mail: tao@math.ucla.edu