

RECURRENCE AND NON-UNIFORMITY OF BRACKET POLYNOMIALS

MATTHEW C. H. TOINTON

ABSTRACT. In his celebrated proof of Szemerédi’s theorem that a set of integers of positive density contains arbitrarily long arithmetic progressions, W. T. Gowers introduced a certain sequence of norms $\|\cdot\|_{U^2[N]} \leq \|\cdot\|_{U^3[N]} \leq \dots$ on the space of complex-valued functions on the set $[N]$. An important question regarding these norms concerns for which functions they are ‘large’ in a certain sense.

This question has been answered fairly completely by B. Green, T. Tao and T. Ziegler in terms of certain algebraic functions called *nilsequences*. In this work we show that more explicit functions called *bracket polynomials* have ‘large’ Gowers norm. Specifically, for a fairly large class of bracket polynomials, called *constant-free* bracket polynomials, we show that if ϕ is a bracket polynomial of degree $k - 1$ on $[N]$ then the function $f : n \mapsto e(\phi(n))$ has Gowers $U^k[N]$ -norm uniformly bounded away from zero.

We establish this result by first reducing it to a certain recurrence property of sets of constant-free bracket polynomials. Specifically, we show that if $\theta_1, \dots, \theta_r$ are constant-free bracket polynomials then their values, modulo 1, are all close to zero on at least some constant proportion of the points $1, \dots, N$.

The proof of this statement relies on two deep results from the literature. The first is work of V. Bergelson and A. Leibman showing that an arbitrary bracket polynomial can be expressed in terms of a so-called *polynomial sequence* on a nilmanifold. The second is a theorem of B. Green and T. Tao describing the quantitative distribution properties of such polynomial sequences.

In the special cases of the bracket polynomials $\phi_{k-1}(n) = \alpha_{k-1}n\{\alpha_{k-2}n\{\dots\{\alpha_1n\}\dots\}$, with $k \leq 5$, we give elementary alternative proofs of the fact that $\|\phi_{k-1}\|_{U^k[N]}$ is ‘large’, without reference to nilmanifolds. Here we write $\{x\}$ for the fractional part of x , chosen to lie in $(-1/2, 1/2]$.

1. INTRODUCTION

A remarkable theorem of E. Szemerédi [13] states that, for $\sigma > 0$, $k \in \mathbb{N}$ and $N \gg_{k,\sigma} 1$, every subset of $[N]$ of cardinality at least σN contains a k -term arithmetic progression. The first good bounds in this theorem were obtained in the celebrated proof of Szemerédi’s theorem by W. T. Gowers [3].

A key observation in Gowers’s work was that arithmetic progressions in a finite abelian group G can be detected using certain norms $\|\cdot\|_{U^2(G)} \leq \|\cdot\|_{U^3(G)} \leq \dots$ on the space of complex-valued functions on G . In general, the $U^k(G)$ -norm is helpful in detecting arithmetic progressions of length $k + 1$ in the group G , and this has led these norms to become one of the major tools in additive combinatorics. Gowers called them *uniformity norms*; they are now often called *Gowers uniformity norms*, or simply *Gowers norms*.

Date: July 21, 2014.

2010 Mathematics Subject Classification. 11B30, 37A17.

Key words and phrases. Gowers uniformity norm, bracket polynomial, nilmanifold.

Appeared in OJAC 2014. See analytic-combinatorics.org.

This article is licensed under a Creative Commons Attribution 4.0 International license.

Given a finite abelian group G , the Gowers norms $\|\cdot\|_{U^k(G)}$ are defined as follows. First, define the *multiplicative derivative* of a function $f : G \rightarrow \mathbb{C}$ by

$$\Delta_h^* f(x) := f(x+h) \overline{f(x)},$$

and abbreviate

$$\Delta_{h_1, \dots, h_k}^* f(x) := \Delta_{h_1}^* \dots \Delta_{h_k}^* f(x).$$

Then for each integer $k \geq 2$ define the *Gowers $U^k(G)$ -norm* by

$$\|f\|_{U^k(G)} := (\mathbb{E}_{x, h_1, \dots, h_k \in G} \Delta_{h_1, \dots, h_k}^* f(x))^{1/2^k}.$$

It can be shown that $\|\cdot\|_{U^k(G)}$ is indeed a norm, but we will not need this fact and so we omit its proof.

Szemerédi's theorem, of course, concerns arithmetic progressions in \mathbb{Z} or, more precisely, in $[N] := \{1, \dots, N\}$, neither of which is a finite group. However, it is also possible to define the Gowers norm of a function $f : [N] \rightarrow \mathbb{C}$, and this can then be applied in finding arithmetic progressions inside $[N]$. The following definition is reproduced from [10, §1].

Definition 1.1 (Gowers $U^k[N]$ -norm). *Given an integer $N > 0$ fix some other integer $\tilde{N} \geq 2^k N$. Define a function $\tilde{f} : \mathbb{Z}/\tilde{N}\mathbb{Z} \rightarrow \mathbb{C}$ by $\tilde{f}(x) = f(x)$ for $x \in [N]$ and $\tilde{f}(x) = 0$ otherwise. Then $\|f\|_{U^k[N]}$ is defined by*

$$(1) \quad \|f\|_{U^k[N]} := \|\tilde{f}\|_{U^k(\mathbb{Z}/\tilde{N}\mathbb{Z})} / \|1_{[N]}\|_{U^k(\mathbb{Z}/\tilde{N}\mathbb{Z})}.$$

Here, and throughout the present work, if X is a set then 1_X denotes the indicator function of X .

As is remarked in [10, §1], it is easy to see that the quantity (1) is independent of the choice of $\tilde{N} \geq 2^k N$, and so $\|\cdot\|_{U^k[N]}$ is well defined.

Denote by \mathcal{D} the unit disc $\{z \in \mathbb{C} : |z| \leq 1\}$. It turns out that when applying Gowers norms to finding arithmetic progressions in $[N]$ it is useful to have a classification of functions $f : [N] \rightarrow \mathcal{D}$ satisfying

$$(2) \quad \|f\|_{U^k[N]} \geq \delta.$$

A function satisfying (2) for a given δ is generally said to be *non-uniform*; a classification of such functions is the content of so-called *inverse conjectures* and *inverse theorems* for the Gowers norms.

It is easy to see that $\|f\|_{U^k[N]}$ is bounded above by 1 for every function $f : [N] \rightarrow \mathcal{D}$, and also that this bound is attained by the function $1_{[N]}$. In fact, there is a very natural broader class of functions attaining this upper bound, which we now describe. Given a function $\phi : [N] \rightarrow \mathbb{R}$ we denote the *discrete derivatives* of ϕ by

$$\Delta_h \phi(n) := \phi(n+h) - \phi(n),$$

and abbreviate

$$\Delta_{h_1, \dots, h_n} \phi := \Delta_{h_1} \dots \Delta_{h_n} \phi.$$

Adopting the standard convention that $e(x) := \exp(2\pi i x)$, we have

$$(3) \quad \Delta_h^* e(\phi(x)) = e(\Delta_h \phi(x)).$$

When ϕ is a polynomial of degree $k - 1$, this implies in particular that every term in the sum

$$(4) \quad \sum_{n, h_1, \dots, h_{k+1}} \Delta_{h_1, \dots, h_{k+1}}^* e(\phi(n))$$

is equal to 1, and so if $f : [N] \rightarrow \mathcal{D}$ is the function defined by setting $f(n) := e(\phi(n))$ then the Gowers norm $\|f\|_{U^k[N]}$ is equal to 1.

There are more exotic examples of functions $f : [N] \rightarrow \mathcal{D}$ satisfying (2). B. Green, T. Tao and T. Ziegler [10, Proposition 1.4] show that certain algebraic functions called *nilsequences*, which we define shortly, are non-uniform. Indeed, they demonstrate the stronger fact that any function that correlates with a nilsequence in a certain sense is non-uniform.

In order to define a nilsequence we must first recall that an *s-step nilmanifold* is the quotient G/Γ of an s -step nilpotent Lie group G by a discrete cocompact subgroup Γ . For example, if G is the *Heisenberg group*

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

and Γ is the discrete subgroup

$$\begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}$$

then it is straightforward to check that G/Γ has a fundamental domain in G defined by

$$(5) \quad \begin{pmatrix} 1 & (-1/2, 1/2] & (-1/2, 1/2] \\ 0 & 1 & (-1/2, 1/2] \\ 0 & 0 & 1 \end{pmatrix};$$

see, for example, [7, §1]. The subgroup Γ is therefore cocompact, and so G/Γ is a 2-step nilmanifold, called the *Heisenberg nilmanifold*.

Throughout this paper, when we write that G/Γ is a nilmanifold we assume that G is a connected, simply connected nilpotent Lie group.

A sequence y_n is said to be an *s-step nilsequence* if there exists an s -step nilmanifold G/Γ , elements $g, x \in G$ and a continuous function $F : G/\Gamma \rightarrow \mathbb{C}$ such that

$$y_n = F(g^n x \Gamma).$$

It turns out that this exhausts all the possibilities for non-uniform functions. Indeed, a remarkable inverse theorem for the Gowers norms, also due to Green, Tao and Ziegler [11], states, roughly, that if $f : [N] \rightarrow \mathcal{D}$ satisfies (2) then f correlates with a $(k - 1)$ -step nilsequence. We refer the reader to [11] for a precise formulation.

Thus we have a comprehensive, if not particularly explicit, classification of all the Gowers non-uniform functions $f : [N] \rightarrow \mathcal{D}$. It is noted in [11, §1], however, that more explicit formulations of the inverse conjectures for the Gowers norms are also possible. In the case of a cyclic group $\mathbb{Z}/N\mathbb{Z}$ of prime order, for example, [4, Theorem 10.9] gives a particularly concrete inverse theorem for the Gowers U^3 -norm.

In order to describe that result we require some notation. Here, and throughout this work, we denote by $\{x\}$ the fractional part of $x \in \mathbb{R}$, chosen to lie in $(-1/2, 1/2]$, and denote by $[x]$ the integer part $[x] := x - \{x\}$. Let N be a prime and let $k \geq 0$. Then [4, Theorem 10.9] says, roughly, that a function $f : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathcal{D}$ satisfies $\|f\|_{U^3(\mathbb{Z}/N\mathbb{Z})} \geq \delta$ if and only if there are

$\xi_1, \xi_2 \in \widehat{\mathbb{Z}/N\mathbb{Z}}$ and a real number α such that f correlates with the function $f' : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathcal{D}$ defined by

$$f'(x) := e(\alpha\{\xi_1 \cdot x\}\{\xi_2 \cdot x\}).$$

Again, we refer the reader to [4] for a precise statement.

The function $x \mapsto \alpha\{\xi_1 \cdot x\}\{\xi_2 \cdot x\}$ is an example of a so-called *bracket polynomial* on $\mathbb{Z}/N\mathbb{Z}$. It is also possible to define bracket polynomials on the set $[N]$, where N is now an arbitrary positive integer. Essentially these are functions like $\phi : n \mapsto 2n\{\sqrt{2}n^2\} + n^3$ that are constructed from genuine polynomials using the operations $+$, \cdot , $\{\cdot\}$; we give a precise definition in Section 2, and in particular clarify the notion of the *degree* of a bracket polynomial.

It turns out that bracket polynomials on $[N]$ arise quite naturally from sequences on nilmanifolds. To see this in the case of the bracket polynomial $\{\alpha n[\beta n]\}$, for example, let $g(n)$ be the sequence in the Heisenberg group given by

$$g(n) = \begin{pmatrix} 1 & -\alpha n & 0 \\ 0 & 1 & \beta n \\ 0 & 0 & 1 \end{pmatrix}.$$

It is straightforward to check that the image of $g(n)$ in the fundamental domain (5) is the element

$$\begin{pmatrix} 1 & \{-\alpha n\} & \{\alpha n[\beta n]\} \\ 0 & 1 & \{\beta n\} \\ 0 & 0 & 1 \end{pmatrix},$$

in which $\{\alpha n[\beta n]\}$ appears quite prominently as the upper-right entry. This in fact turns out to be a general phenomenon. Bergelson and Leibman [1] show that an arbitrary bracket polynomial can be expressed in terms of a nilmanifold in similar fashion. See Theorem 6.3 for more details.

Given the role played by bracket polynomials on $\mathbb{Z}/N\mathbb{Z}$ in the inverse theory for the $U^3(\mathbb{Z}/N\mathbb{Z})$ -norms, as well as the link between bracket polynomials and sequences on nilmanifolds due to Bergelson–Leibman and the link between sequences on nilmanifolds and Gowers norms due to Green–Tao–Ziegler, it is natural to ask what role bracket polynomials on $[N]$ play in the inverse theory for the $U^k[N]$ norms. The aim of this paper is to explore this role.

Our first, and principal, theorem states that a fairly large class of bracket polynomials ϕ give rise to non-uniform functions $e(\phi)$. This is the class of *constant-free* bracket polynomials. These are defined precisely in Section 2, but essentially a bracket polynomial is said to be constant free if it is constructed from genuine polynomials using the operations $+$, \cdot , $\{\cdot\}$, and each of the genuine polynomials used in this construction has zero constant term. Thus, for example, the bracket polynomial $\{\alpha n\{\beta n\}\} + \gamma n^2$ is constant free, but the bracket polynomial $\alpha n\{\beta n + \gamma\}$ is not.

We show, then, that if ϕ is a constant-free bracket polynomial of degree at most $k - 1$ then the quantity $\|e(\phi)\|_{U^k[N]}$ must be bounded away from zero. The bound obtained is uniform in N . It does depend on the bracket polynomial being considered, but only on its ‘shape’; we make this precise in Section 2 using the notion of a *bracket form*, but essentially this means, for example, that the bound obtained for the bracket polynomial $\{\alpha n\{\beta n\}\}$ is uniform across all choices of α and β .

Theorem 1.2 (Bracket polynomials are non-uniform; rough statement). *Let ϕ be a constant-free bracket polynomial of degree at most $k - 1$. Then there is some δ , depending only on the ‘shape’ of ϕ , such that*

$$\|e(\phi)\|_{U^k[N]} \geq \delta.$$

See Theorem 2.8 for a more precise statement. The ‘constant-free’ condition results from our use of Theorem 4.10; see also Remark 4.11.

At its simplest level, our proof of Theorem 1.2 rests on a preliminary result stating that bracket polynomials have a rather suggestive property called being *locally polynomial*.

Definition 1.3 (Locally polynomial). *Let $\phi : [N] \rightarrow \mathbb{R}$ be a function and let $B \subset [N]$. Then ϕ is said to be locally polynomial of degree $k - 1$ on B if whenever $n \in [N]$ and $h \in [-N, N]^k$ satisfy $n + \omega \cdot h \in B$ for all $\omega \in \{0, 1\}^k$ we have $\Delta_{h_1, \dots, h_k} \phi(n) = 0$.*

Indeed, we show in Section 3 that every bracket polynomial ϕ on $[N]$ is locally polynomial on some ‘large’ set $B_\phi \subset [N]$. The relevance of this to the study of Gowers norms lies in the identity (3). Just as this identity implied that the terms in the sum (4) were all equal to 1 when ϕ was a genuine polynomial, if ϕ is *locally polynomial* on a suitably large set then this suggests some bias towards 1 in the terms of the sum (4). This in turn suggests that $\|e(\phi)\|_{U^k[N]}$ should be bounded away from zero.

Unfortunately, it is not clear that one can proceed directly from the property of being locally polynomial to the property of having large Gowers norm, and so the results of Section 3 alone are not sufficient to prove Theorem 1.2. In Section 4, however, we show that a slightly stronger property, which we call being *strongly locally polynomial*, is sufficient to imply that a bracket polynomial is non-uniform.

It turns out that a certain recurrence property of bracket polynomials is sufficient to imply the property of being strongly locally polynomial, and hence to imply that a bracket polynomial is non-uniform. This is the principal motivation for our second theorem.

Theorem 1.4 (Recurrence of bracket polynomials; rough statement). *Let $\theta_1, \dots, \theta_r$ be constant-free bracket polynomials and let $\delta > 0$. Then there are some $\varepsilon > 0$ depending only on the ‘shapes’ of the θ_i , and $N_0 > 0$ depending on the ‘shapes’ of the θ_i and on δ , such that whenever $N \geq N_0$ the proportion of $n \in [N]$ for which $\{\theta_i(n)\} \in (-\delta, \delta)$ is at least ε .*

See Theorem 4.10 for a precise statement. We show in Section 4 that this is sufficient to imply Theorem 1.2; it is potentially also of interest in its own right.

Remark 1.5. *We show by example in Remark 4.11 that the constant-free condition is necessary in Theorem 1.4. Theorem 1.2, on the other hand, could conceivably remain true in the absence of that condition.*

In Section 6, we appeal to the work of Bergelson–Leibman showing that bracket polynomials can be expressed in terms of certain sequences on nilmanifolds, as well as to work of B. Green and T. Tao describing the distribution properties of such sequences, to establish Theorem 1.4, or rather the more precise Theorem 4.10.

The appeal to the results of Bergelson–Leibman and Green–Tao in the proof of Theorem 1.2 renders the argument far from elementary. It is interesting to see for which bracket polynomials one can use elementary methods to establish Theorem 1.2. In Sections 7 and 8 we consider this problem in the model setting of the bracket polynomials $\phi_{k-1}(n)$ defined by $\alpha_{k-1} n \{ \alpha_{k-2} n \{ \dots \{ \alpha_1 n \} \dots \}$. Theorem 1.2 of course instantly tells us that

$$\|e(\phi_{k-1})\|_{U^k[N]} \gg_k 1;$$

In Sections 7 and 8 we arrive at this statement in the cases $k \leq 5$ by entirely elementary methods.

2. BRACKET POLYNOMIALS ON $[N]$

In this section we give formal definitions of some of the concepts we discussed in the introduction. In particular, we define bracket polynomials on $[N]$ precisely. The definitions of bracket polynomials are essentially already contained in the literature; see, for example, [1, 1.11-1.12]. Nonetheless, we give them in full detail here, in part so as to set notation, but also in order to introduce the related concept of a *bracket form*. The latter is necessary in order to make precise what we mean by ‘shape’ in Theorem 1.2.

Definition 2.1 (Bracket polynomials on $[N]$). *Bracket polynomials on $[N]$ are functions from $[N]$ to \mathbb{R} defined recursively as follows.*

- *A genuine polynomial ϕ of degree k is also a bracket polynomial of degree at most k .*
- *If $\phi : [N] \rightarrow \mathbb{R}$ is a bracket polynomial of degree at most k then the functions $-\phi : [N] \rightarrow \mathbb{R}$, defined by $(-\phi)(n) := -(\phi(n))$, and $\{\phi\} : [N] \rightarrow \mathbb{R}$, defined by $\{\phi\}(n) := \{\phi(n)\}$, are also bracket polynomials of degree at most k .*
- *If $\phi_1, \phi_2 : [N] \rightarrow \mathbb{R}$ are bracket polynomials of degree at most k_1, k_2 , respectively, then the function $\phi_1 \cdot \phi_2 : [N] \rightarrow \mathbb{R}$ defined by $\phi_1 \cdot \phi_2(n) := \phi_1(n)\phi_2(n)$ is a bracket polynomial of degree at most $k_1 + k_2$, and the function $\phi_1 + \phi_2 : [N] \rightarrow \mathbb{R}$ defined by $\phi_1 + \phi_2(n) := \phi_1(n) + \phi_2(n)$ is a bracket polynomial of degree at most $\max\{k_1, k_2\}$.*

If in this definition we restrict the genuine polynomials to those with zero constant term, the resulting functions are said to be constant-free bracket polynomials. Those bracket polynomials that do not use the $+$ operation are called elementary.

Remark 2.2. *It will almost always be the case that we will be interested only in the value of a bracket polynomial modulo 1, and so we might easily and naturally define bracket polynomials to be functions into \mathbb{R}/\mathbb{Z} . However, certain statements and proofs are slightly cleaner if we view them as functions into \mathbb{R} and then project to \mathbb{R}/\mathbb{Z} only when it comes to the final application.*

In order to make the statement of Theorem 1.2 precise, we need some way of defining what we mean by the ‘shape’ of a bracket polynomial. To that end, we first develop a definition that formalises this concept for genuine polynomials. The basic idea is to say that two polynomials ϕ_1 and ϕ_2 have the same ‘shape’ if there is some ‘polynomial’ in n with coefficients taken from the list of symbols $\alpha_1, \alpha_2, \dots$ such that both ϕ_1 and ϕ_2 can be obtained by replacing each symbol α_i by a real number. Thus, for example, the polynomials $3n^2$ and πn^2 would have the same ‘shape’ because they can each be realised by replacing the symbol α_1 in the ‘polynomial’ $\alpha_1 n^2$ by a real number. We shall call $\alpha_1 n^2$ a *polynomial form* and call $3n^2$ and πn^2 *realisations* of the polynomial form $\alpha_1 n^2$.

In fact, the definition of a polynomial form will need to be slightly more complicated than is suggested by the preceding paragraph. This is because in Appendix B it will be convenient for the set of polynomial forms to form a ring.

Definition 2.3 (Ring of polynomial forms). *Let $\alpha_1, \alpha_2, \dots$ be a countably infinite list of symbols; we shall call this list an alphabet. Define a monomial form in these symbols to be a string of the form $+\alpha_{i_1} \cdots \alpha_{i_t} n^k$ or $-\alpha_{i_1} \cdots \alpha_{i_t} n^k$, with $i_1 \leq \dots \leq i_t$ and $k \geq 0$ an integer.*

Define k to be the degree of such a monomial form. If $k \neq 0$ then we additionally say that $+\alpha_{i_1} \cdots \alpha_{i_t} n^k$ and $-\alpha_{i_1} \cdots \alpha_{i_t} n^k$ are constant-free monomial forms of degree k .

Now suppose that Φ_1, \dots, Φ_r is a finite list of monomial forms of degree at most k . Then the string $\Phi_1 + \dots + \Phi_r$ is said to be a polynomial form of degree at most k . If the Φ_i are all constant free then the string $\Phi_1 + \dots + \Phi_r$ is also said to be constant free.

We make the set of polynomial forms into a ring by defining multiplication on the set of monomial forms, and then extending it (uniquely) to the set of polynomial forms by requiring it to be distributive over addition. Specifically, we formally define multiplication on the symbols $+$ and $-$ by setting $(+\cdot+) = (-\cdot-) = +$ and $(+\cdot-) = (-\cdot+) = -$, and then if each of ϵ, ϵ' represents either $+$ or $-$ we define the product of the monomial forms $\epsilon \alpha_{i_1} \cdots \alpha_{i_t} n^k$ and $\epsilon' \alpha_{i'_1} \cdots \alpha_{i'_{t'}} n^{k'}$ to be

$$\epsilon \alpha_{i_1} \cdots \alpha_{i_t} n^k \cdot \epsilon' \alpha_{i'_1} \cdots \alpha_{i'_{t'}} n^{k'} = (\epsilon \cdot \epsilon') \alpha_{j_1} \cdots \alpha_{j_{t+t'}} n^{k+k'},$$

where the $\alpha_{j_1}, \dots, \alpha_{j_{t+t'}}$ are precisely the $\alpha_{i_1}, \dots, \alpha_{i_t}, \alpha_{i'_1}, \dots, \alpha_{i'_{t'}}$, only permuted so that $j_l \leq j_{l'}$ whenever $l < l'$.

From now on we drop the $+$ from the polynomial form $+\alpha_{i_1} \cdots \alpha_{i_t} n^k$ and write simply $\alpha_{i_1} \cdots \alpha_{i_t} n^k$.

Thus, for example, the strings $-\alpha_1 n$ and $-\alpha_2 n^3$ are polynomial forms, and their product is $\alpha_1 \alpha_2 n^4$.

Definition 2.4 (Realisation of a polynomial form). *Suppose that Φ is a polynomial form featuring the symbols $\alpha_1, \dots, \alpha_m$, and for each $i = 1, \dots, m$ let a_i be a real number. Then the polynomial obtained by replacing each instance of α_i in Φ with the real number a_i is said to be a realisation of the polynomial form Φ .*

Thus, for example, the string $\Phi = \alpha_1 n + \alpha_2 n^2 - \alpha_1 \alpha_2 n^3$ is a polynomial form of degree at most 3, and the polynomial $2n + 3n^2 - 6n^3$ is a realisation of Φ .

To convert this into a definition valid for bracket polynomials, let us make the following slightly more abstract version of Definition 2.1.

Definition 2.5 (Bracket expressions). *Let A be a set. Then the bracket expressions in the elements of A are certain strings of elements of A and the symbols $+, -, \cdot, \{, \}$, defined recursively as follows.*

- If $a \in A$ then the string a is a bracket expression in the elements of A .
- If b is a bracket expression in the elements of A then the strings $-b$ and $\{b\}$ are also a bracket expression in the elements of A .
- If b_1, b_2 are bracket expression in the elements of A then the strings $b_1 \cdot b_2$ and $b_1 + b_2$ are bracket expressions in the elements of A .

Remark 2.6. *Definition 2.5 is similar in spirit to that found in [1, §6.3].*

Definition 2.7 (Bracket forms). *Let A be the set of polynomial forms in the alphabet*

$$\alpha_1, \alpha_2, \dots$$

Then the bracket forms in the same alphabet are certain bracket expressions in the elements of A defined recursively as follows.

- A polynomial form of degree at most k is also a bracket form of degree at most k .
- If Φ is a bracket form of degree at most k then the expressions $-\Phi$ and $\{\Phi\}$ are also bracket forms of degree at most k .

- If Φ_1, Φ_2 are bracket forms of degree at most k_1, k_2 , respectively, then the expression $\Phi_1 \cdot \Phi_2$ is a bracket form of degree at most $k_1 + k_2$, and the expression $\Phi_1 + \Phi_2$ is a bracket polynomial of degree at most $\max\{k_1, k_2\}$.

If the polynomial forms appearing in this recursive construction of a bracket form are all constant free, then the resulting bracket form is also said to be constant free.

Now suppose that Φ is a bracket form featuring the symbols $\alpha_1, \dots, \alpha_m$, and for each $i = 1, \dots, m$ let a_i be a real number. Then the bracket polynomial obtained by replacing each instance of α_i in Φ with the real number a_i is said to be a realisation of the bracket form Φ .

Thus, for example, the string $\Phi = \{\alpha_1 n \{\alpha_2 n\}\}$ is a constant-free bracket form of degree at most 2, and the bracket polynomials $\{\sqrt{2}n \{\sqrt{3}n\}\}$ and $\{\pi n \{\frac{2}{7}n\}\}$ are realisations of Φ .

We are now in a position to state a more precise version of Theorem 1.2.

Theorem 2.8 (Bracket polynomials are non-uniform; precise statement). *Let Φ be a constant-free bracket form of degree at most $k - 1$. Then for every realisation ϕ of Φ we have*

$$\|e(\phi)\|_{U^k[N]} \gg_{\Phi} 1.$$

When considering a bracket polynomial such as $\phi : n \mapsto n\{\alpha n\{\beta n\}\{\gamma n\}\}$ it will be useful to have a way of referring to the simpler ‘bracketed’ components, in this case $\{\beta n\}$, $\{\gamma n\}$ and $\{\alpha n\{\beta n\}\{\gamma n\}\}$, from which ϕ is built up. We shall therefore call $\{\beta n\}$, $\{\gamma n\}$ and $\{\alpha n\{\beta n\}\{\gamma n\}\}$ the *bracket components* of ϕ . In general, we shall use the following definition, which is similar, though not identical, to that found in [1, §8].

Definition 2.9 (Bracket components of a bracket polynomial). *The set $C(\phi)$ of bracket components of a bracket polynomial ϕ will be a set of bracket polynomials defined recursively as follows.*

- If ϕ is a genuine polynomial then $C(\phi) = \emptyset$.
- If $\phi = -v$ for some bracket polynomial v then $C(\phi) = C(v)$.
- If $\phi = \{v\}$ for some bracket polynomial v then $C(\phi) = C(v) \cup \{v\}$.¹
- If $\phi = v_1 \cdot v_2$ or $v_1 + v_2$ for some bracket polynomials v_1, v_2 then $C(\phi) = C(v_1) \cup C(v_2)$.

The set $C(\Phi)$ of bracket components of a bracket form Φ will be a set of bracket forms defined analogously, as follows.

- If Φ is a polynomial form then $C(\Phi) = \emptyset$.
- If $\Phi = -\Theta$ for some bracket form Θ then $C(\Phi) = C(\Theta)$.
- If $\Phi = \{\Theta\}$ for some bracket form Θ then $C(\Phi) = C(\Theta) \cup \{\Theta\}$.
- If $\Phi = \Theta_1 \cdot \Theta_2$ or $\Theta_1 + \Theta_2$ for some bracket forms Θ_1, Θ_2 then $C(\Phi) = C(\Theta_1) \cup C(\Theta_2)$.

3. BRACKET POLYNOMIALS ARE LOCALLY POLYNOMIAL

We discussed in the introduction the relevance to the Gowers norms of the property of being locally polynomial. The aim of this section is to show that bracket polynomials have this property.

Proposition 3.1 (A bracket polynomial is locally polynomial on a set of positive density). *Let $\phi : [N] \mapsto \mathbb{R}$ be a bracket polynomial of degree at most k with $|C(\phi)| \leq c$. Then there exists a set $B \subset [N]$ of cardinality $\Omega_{k,c}(N)$ on which ϕ is locally polynomial of degree at most k .*

¹Note that $\{v\}$ here is the singleton containing v , not the fractional part of v .

The proof of this is straightforward, but will motivate much of what comes later. Before we embark on the main body of the proof, let us record the following trivial, but repeatedly useful, properties of fractional parts.

Lemma 3.2. *Let $x, y \in \mathbb{R}/\mathbb{Z}$ and let $\delta < 1/2$, and suppose that $\{x\}$ and $\{y\}$ both lie in a subinterval $I \subset (-1/2, 1/2]$ of width δ . Then*

- (i) $\{x\} - \{y\} = \{x - y\}$, and this quantity lies in the interval $(-\delta, \delta)$;
- (ii) if I is centred on 0 then we may additionally conclude that $\{x\} + \{y\} = \{x + y\}$ and that this quantity also lies in the interval $(-\delta, \delta)$.

The proof of Proposition 3.1 is essentially in two parts. In the first (more substantial) part of the argument we identify a collection of sets on which ϕ is locally polynomial.

Proposition 3.3. *Suppose that $\phi : [N] \rightarrow \mathbb{R}$ is a bracket polynomial of degree at most k . Then there exists a parameter $\delta \gg_k 1$ such that if for each $v \in C(\phi)$ we have an interval J_v of width at most δ inside $(-1/2, 1/2]$ then ϕ is locally polynomial of degree k on the set*

$$\{n \in [N] : \{v(n)\} \in J_v \text{ for all } v \in C(\phi)\}.$$

The second part of the proof of Proposition 3.1 is a simple pigeonholing argument, which we present as a lemma for ease of later reference, showing that at least one set of the form given by Proposition 3.3 must have cardinality $\Omega_{k,c}(N)$.

Lemma 3.4. *Let I be an interval in \mathbb{R} , let $A \subset \mathbb{N}$ be a finite set and let $g_1, \dots, g_l : A \rightarrow I$ be functions. Let $\delta_1, \dots, \delta_l < |I|$. Then there exist subintervals $J_1, \dots, J_l \subset I$ of widths $\delta_1, \dots, \delta_l$, respectively, such that*

$$|\{n \in A : g_i(n) \in J_i \text{ for } i = 1, \dots, l\}| \gg_l \frac{\delta_1 \cdots \delta_l |A|}{|I|^l}.$$

Proof. This is essentially contained in the first part of the proof of [14, Lemma 4.20]. For each $i = 1, \dots, l$ we divide I into $\lceil |I|/\delta_i \rceil$ subintervals: $\lfloor |I|/\delta_i \rfloor$ of length δ_i , and at most one, the remainder, of length less than δ_i .

Taking all products of these subintervals inside I^l , we divide I^l into $\prod_{i=1}^l \lceil |I|/\delta_i \rceil$ boxes of side lengths at most $\delta_1, \dots, \delta_l$. By the pigeonhole principle, one of these boxes must contain the images of at least

$$(6) \quad \frac{|A|}{\prod_{i=1}^l \lceil |I|/\delta_i \rceil}$$

elements of A under the map

$$(g_1, \dots, g_l) : A \rightarrow I^l.$$

Now (6) is certainly at least

$$\frac{\delta_1 \cdots \delta_l |A|}{2^l |I|^l},$$

and so the lemma is proved. \square

Proposition 3.3 follows more or less immediately from the following results, the proofs of which occupy the remainder of this section.

Lemma 3.5. *Let $B \subset \mathbb{N}$, let $v : \mathbb{N} \rightarrow \mathbb{R}$, and suppose that $\{v\}(B)$ is contained within an interval $J \subset (-1/2, 1/2]$ of width $\delta < 1/2$. Then $\Delta_h \{v\}(n) = \{\Delta_h v\}(n)$ for every $n \in B \cap (B - h)$. Furthermore, if v is locally polynomial of degree at most k on B and $\delta < 2^{-k}$ then $\{v\}$ is also locally polynomial of degree at most k on B .*

Lemma 3.6. *Let $B \subset \mathbb{N}$ and let $v_1, v_2 : \mathbb{N} \rightarrow \mathbb{R}$ be functions. Suppose that v_1, v_2 are locally polynomial of degree at most k_1, k_2 , respectively, on B . Then $v_1 + v_2$ is locally polynomial of degree at most $\max\{k_1, k_2\}$ on B and $v_1 \cdot v_2$ is locally polynomial of degree at most $k_1 + k_2$ on B .*

Proof of Lemma 3.5. By Lemma 3.2 (i) we have

$$\begin{aligned} \Delta_h\{v\}(n) &= \{v(n+h)\} - \{v(n)\} \\ &= \{v(n+h) - v(n)\} \\ &= \{\Delta_h v(n)\}, \end{aligned}$$

which was the first conclusion of the lemma. If $\delta < 2^{-k}$ then Lemma 3.2 (i) also implies that the image of $\{\Delta_h v\}$ is contained within the interval $(-2^{-k}, 2^{-k}]$, and by induction we may therefore assume that $\{\Delta_h v\}$ is locally polynomial of degree $k-1$ on $B \cap (B-h)$, and hence conclude that $\{v\}$ is locally polynomial of degree k on B . \square

Proof of Lemma 3.6. The assertion about $v_1 + v_2$ is immediate from the linearity of Δ_h . To prove the assertion about $v_1 \cdot v_2$, note that

$$\begin{aligned} \Delta_h(v_1 \cdot v_2)(n) &= v_1(n+h)v_2(n+h) - v_1(n)v_2(n) \\ &= (\Delta_h v_1(n) + v_1(n))(\Delta_h v_2(n) + v_2(n)) - v_1(n)v_2(n) \\ &= \Delta_h v_1(n)\Delta_h v_2(n) + \Delta_h v_1(n)v_2(n) + v_1(n)\Delta_h v_2(n), \end{aligned}$$

and so $\Delta_h(v_1 \cdot v_2) = \Delta_h v_1 \cdot \Delta_h v_2 + \Delta_h v_1 \cdot v_2 + v_1 \cdot \Delta_h v_2$. By induction each of these terms is locally polynomial of degree at most $k_1 + k_2 - 1$ on $B \cap (B-h)$, and so $\Delta_h(v_1 \cdot v_2)$ is locally polynomial of degree at most $k_1 + k_2 - 1$ on $B \cap (B-h)$ by the first part of the lemma. \square

4. STRONGLY LOCALLY POLYNOMIAL FUNCTIONS

Let $\phi : [N] \rightarrow \mathbb{R}$ be a function and define $f : [N] \rightarrow \mathbb{C}$ by $f(n) := e(\phi(n))$. In this section we develop a criterion for f to have a positive Gowers U^k -norm. Given a set $B \subset [N]$ of cardinality $\Omega(N)$ on which ϕ is locally polynomial of degree $k-1$ it is straightforward to check that $\|1_B \cdot f\|_{U^k[N]} \gg_k 1$. Indeed, given that

$$\begin{aligned} \|1_B \cdot f\|_{U^k[N]} &\gg_k \mathbb{E}_{n \in [N], h \in [-N, N]^k} (e(\Delta_{h_1, \dots, h_k} \phi(n)) \prod_{\omega \in \{0, 1\}^k} 1_B(n + \omega \cdot h)) \\ &= \mathbb{E}_{n \in [N], h \in [-N, N]^k} (\prod_{\omega \in \{0, 1\}^k} 1_B(n + \omega \cdot h)) \\ &= \mathbb{P}_{n \in [N], h \in [-N, N]^k} (n + \omega \cdot h \in B \text{ for all } \omega \in \{0, 1\}^k), \end{aligned}$$

it is an immediate consequence of the following lemma.

Lemma 4.1. *Suppose $B \subset [N]$ satisfies $|B| \geq \sigma N$ and for $j = 0, 1, 2, \dots$ write*

$$B_j := \{(n, h) \in [N] \times [-N, N]^j : n + \omega \cdot h \in B \text{ for all } \omega \in \{0, 1\}^j\}.$$

Then $|B_j| \gg_{j, \sigma} N^{j+1}$.

Proof. The claim is trivial for the case $j = 0$, and so we may proceed by induction, assuming that

$$(7) \quad |B_{j-1}| \gg_{j, \sigma} N^j.$$

For each $h \in [-N, N]^{j-1}$ write $C(h) := \{n \in [N] : n + \omega \cdot h \in B \text{ for all } \omega \in \{0, 1\}^{j-1}\}$ and set $r(h) := |C(h)|$; thus

$$(8) \quad |B_{j-1}| = \sum_{h \in [-N, N]^{j-1}} r(h).$$

Now note simply that $B_j = \{(n, (h_1, \dots, h_{j-1}, m - n)) : (h_1, \dots, h_{j-1}) \in [-N, N]^{j-1}; m, n \in C(h)\}$, and so $|B_j| = \sum_{h \in [-N, N]^{j-1}} r(h)^2$, which is at least $\Omega_j \left(\left(\sum_{h \in [-N, N]^{j-1}} r(h) \right)^2 / N^{j-1} \right)$ by Jensen's inequality. By (8) this is at least $\Omega_j (|B_{j-1}|^2 / N^{j-1})$, which in turn is at least $\Omega_{j,\sigma} (N^{j+1})$ by (7). \square

Our task, then, if we wish to prove that f is non-uniform, is to remove the 1_B from the expression $\|1_B \cdot f\|_{U^k[N]} \gg_k 1$. However, if we are to make use of the local behaviour of ϕ on B then the set B will have to play at least some role.

The key to reconciling this is the following.

Lemma 4.2. *Let $\phi : [N] \rightarrow \mathbb{R}$ be a function and define $f : [N] \rightarrow \mathbb{C}$ by $f(n) := e(\phi(n))$. Then for every set $B \subset [N]$ we have*

$$(9) \quad \|f\|_{U^k[N]} \gg_k \left| \mathbb{E}_{n \in [N], h \in [-N, N]^k} \left(e(\Delta_{h_1, \dots, h_k} \phi(n)) \prod_{\omega \in \{0, 1\}^k \setminus \{0\}} 1_B(n + \omega \cdot h) \right) \right|.$$

Proof. Given a finite abelian group G and functions $g_\omega : G \rightarrow \mathbb{C}$ indexed by $\omega \in \{0, 1\}^k$, the Gowers inner product $\langle g_\omega \rangle_{U^k(G)}$ is defined by

$$\langle g_\omega \rangle_{U^k(G)} := \mathbb{E}_{x \in G, h \in G^k} \prod_{\omega \in \{0, 1\}^k} \mathcal{C}^{|\omega|} g_\omega(x + \omega \cdot h).$$

Here we denote by \mathcal{C} the operation of complex conjugation, and by $|\omega|$ the number of entries of ω that are equal to 1. The Gowers–Cauchy–Schwarz inequality [14, (11.6)] states that

$$|\langle g_\omega \rangle_{U^k(G)}| \leq \prod_{\omega \in \{0, 1\}^k} \|g_\omega\|_{U^k(G)}.$$

Since $\|1_B \cdot f\|_{U^k[N]} \leq 1$, setting $g_0 = f$ and $g_\omega = 1_B \cdot f$ in this inequality, with $G = \mathbb{Z}/\tilde{N}\mathbb{Z}$ as in Definition 1.1, yields the desired result. \square

In order to conclude that f is non-uniform it will therefore be sufficient to find a set B for which we are able to place a lower bound on the right-hand side of (9). In this context it is natural to employ a slightly stronger notion than that of being locally polynomial.

Definition 4.3 (Strongly locally polynomial). *Let $\phi : [N] \rightarrow \mathbb{R}$ be a function and let $B \subset [N]$. Then ϕ is said to be strongly locally polynomial of degree $k - 1$ on B if whenever $n \in [N]$ and $h \in [-N, N]^k$ satisfy $n + \omega \cdot h \in B$ for all $\omega \in \{0, 1\}^k \setminus \{0\}$ we have $\Delta_{h_1, \dots, h_k} \phi(n) = 0$.*

Example 4.4. *Suppose that $\phi(n) = \alpha n \{\beta n\}$. Then ϕ is strongly locally quadratic on the set*

$$B := \{n \in [N] : \{\beta n\} \in (-1/16, 1/16)\}.$$

Indeed, suppose that $n \in [N]$ and $h \in [-N, N]^3$ satisfy $n + \omega \cdot h \in B$ for all $\omega \in \{0, 1\}^3 \setminus (0, 0, 0)$. Then Lemma 3.2 implies that $\{\beta n\} \in (-1/4, 1/4)$, and so for each $\omega \in \{0, 1\}^3$ (including $\omega = (0, 0, 0)$) the point $n + \omega \cdot h$ lies in

$$B' := \{n \in [N] : \{\beta n\} \in (-1/4, 1/4)\}.$$

Moreover, ϕ is locally quadratic on B' by Lemmas 3.5 and 3.6, and so $\Delta_{h_1, h_2, h_3} \phi(n) = 0$.

The reason for making this definition is the following result.

Proposition 4.5. *Let $\phi : [N] \rightarrow \mathbb{R}$ be a function and define $f : [N] \rightarrow \mathbb{C}$ by $f(n) := e(\phi(n))$. Suppose that ϕ is strongly locally polynomial of degree $k - 1$ on a set $B \subset [N]$ of cardinality σN . Then $\|f\|_{U^k[N]} \gg_{k,\sigma} 1$.*

Proof. This is a straightforward combination of Lemmas 4.1 and 4.2. \square

In light of Proposition 4.5, if we wish to prove that $\|f\|_{U^{k+1}[N]} \gg 1$ then it is sufficient to find a set $B \subset [N]$ satisfying $|B| \gg N$ on which ϕ is strongly locally polynomial of degree at most k . Unfortunately, we are not able just to take B to be an arbitrary set of the form given by Proposition 3.3, as the following example illustrates.

Example 4.6. *Suppose that $\phi(n) = \alpha \{\frac{1}{10}n\}$. Then Lemmas 3.5 and 3.6 imply that ϕ is locally linear on the set*

$$B = \{n \in [N] : \{\frac{1}{10}n\} \in (1/4, 1/2]\}.$$

However, if $n = 6$ and $h_1 = h_2 = -1$ then $n + h_1, n + h_2$ and $n + h_1 + h_2$ all lie in B , but $\Delta_{h_1, h_2} \phi(n) = -\alpha$. Hence ϕ is not strongly locally linear on B .

Nonetheless, provided the intervals J_ν appearing in the definition of a set of the form given by Proposition 3.3 are sufficiently small and are sufficiently far from the boundary of $(-1/2, 1/2]$, the problem exposed by Example 4.6 will not occur. Before we express this precisely, let us establish some notation. For functions $\nu_1, \dots, \nu_r : [N] \rightarrow \mathbb{R}$ and sets $S_1, \dots, S_r \subset (-1/2, 1/2]$ define the set

$$B_N^r(\nu_1, \dots, \nu_r; S_1, \dots, S_r) := \{n \in [N] : \{\nu_i(n)\} \in S_i \text{ for all } i\}.$$

We typically abuse notation slightly and write B_N instead of B_N^r . For $\varepsilon < 1/2$ we write I_ε for the interval $(-\varepsilon, \varepsilon)$.

Lemma 4.7. *Let $k \geq 1$ be an integer and set $c_k := 2^{-k}(2k + 1)^{-1}$. Suppose that $\phi : [N] \rightarrow \mathbb{R}$ is a bracket polynomial of degree at most k with bracket components ν_1, \dots, ν_m . Suppose further that*

$$(10) \quad \delta \leq c_k$$

and that

$$(11) \quad \varepsilon \geq k\delta,$$

and that J_1, \dots, J_m are intervals of width at most δ inside $I_{\frac{1}{2}-\varepsilon}$.

Then there is a set $B \subset [N]$ on which ϕ is locally polynomial of degree at most k and such that if $n \in [N]$, $h \in [-N, N]^{k+1}$ and $n + \omega \cdot h \in B_N(\nu_1, \dots, \nu_m; J_1, \dots, J_m)$ for every $\omega \in \{0, 1\}^{k+1} \setminus \{\mathbf{0}\}$ then $n + \omega \cdot h \in B$ for every $\omega \in \{0, 1\}^{k+1}$. In particular, ϕ is strongly locally polynomial of degree at most k on the set $B_N(\nu_1, \dots, \nu_m; J_1, \dots, J_m)$.

Before we prove Lemma 4.7, let us note that in combination with Proposition 4.5 it immediately implies the following result.

Proposition 4.8. *Let $k \geq 1$ be an integer and let c_k be as in Lemma 4.7. Suppose that $\phi : [N] \rightarrow \mathbb{R}$ is a bracket polynomial of degree at most k with bracket components ν_1, \dots, ν_m . Suppose further that $\delta \leq c_k$ and that $\varepsilon \geq k\delta$, and that J_1, \dots, J_m are intervals of width at most δ inside $I_{\frac{1}{2}-\varepsilon}$ such that*

$$|B_N(\nu_1, \dots, \nu_m; J_1, \dots, J_m)| \geq \sigma N.$$

Then, defining $f : [N] \rightarrow \mathbb{C}$ by $f(n) := e(\phi(n))$, we have $\|f\|_{U^{k+1}[N]} \gg_{k,\sigma} 1$.

Proof of Lemma 4.7. The lemma is trivial for genuine polynomials with $B = [N]$, and so we may assume that we are in one of three cases:

Case 1. $\phi = \{\theta\}$ with θ of degree at most k .

Case 2. $\phi = \theta_1 + \theta_2$ with θ_i of degree at most k_i and $k_1, k_2 \leq k$.

Case 3. $\phi = \theta_1 \cdot \theta_2$ with θ_i of degree at most k_i and $k_1 + k_2 \leq k$.

We proceed by induction on the number of operations required to construct ϕ . Therefore, in case 1 we may assume that there exists a set $B_0 \subset [N]$ on which θ is locally polynomial; in cases 2 and 3 we may assume that there exist a set $B_1 \subset [N]$ on which θ_1 is locally polynomial and a set $B_2 \subset [N]$ on which θ_2 is locally polynomial; and in each case we may assume that $n + \omega \cdot h \in B_i$ for every $\omega \in \{0, 1\}^{k+1}$ whenever $n + \omega \cdot h \in B_N(v_1, \dots, v_m; J_1, \dots, J_m)$ for every $\omega \neq \mathbf{0}$.

In case 1 we may assume that $\phi = v_m = \{\theta\}$. Suppose that $n + \omega \cdot h \in B_N(v_1, \dots, v_m; J_1, \dots, J_m)$ for every $\omega \neq \mathbf{0}$. Then $\Delta_h \theta(n) = 0$ by the inductive hypotheses, and so repeated application of Lemma 3.2 implies that $\{\theta(n)\} \in J'_m := J_m + (-k\delta, k\delta)$. The inequality (11) then implies that $J'_m \subset (-1/2, 1/2]$, whilst (10) and the definition of c_k imply that $|J'_m| \leq 2^{-k}$. Lemma 3.5 therefore implies that we may take $B = B_0 \cap B_N(\theta; J'_m)$.

In cases 2 and 3 we may simply take $B = B_1 \cap B_2$ by Lemma 3.6. \square

In the event that the bracket components v_i appearing in Proposition 4.8 are linear it is elementary to show that there exist intervals J_i satisfying the hypotheses of that proposition, with σ depending only on k and m . Indeed, one can even insist that the intervals J_i be centred at zero, as follows.

Lemma 4.9 (Sets of linear bracket polynomials are strongly recurrent). *Let $\alpha_1, \dots, \alpha_r \in \mathbb{R}$, let $\delta > 0$ and let $v_i(n) := \{\alpha_i n\}$. Then*

$$|B_N(v_1, \dots, v_r; I_\delta, \dots, I_\delta)| \gg_{r, \delta} N.$$

Proof. A pigeonholing argument similar to that used in [14, Lemma 4.20] gives points $\xi_1, \dots, \xi_r \in \mathbb{R}$ and a subset $A \subset [N]$ satisfying $|A| \gg_{r, \delta} N$ such that whenever $n \in A$ we have $\|\alpha_j n - \xi_j\|_{\mathbb{R}/\mathbb{Z}} < \delta/2$ for every $j = 1, \dots, r$.

Following [14, Lemma 4.20], note that if $\|\alpha_j n - \xi_j\|_{\mathbb{R}/\mathbb{Z}} < \delta/2$ and $\|\alpha_j n' - \xi_j\|_{\mathbb{R}/\mathbb{Z}} < \delta/2$ for every $j = 1, \dots, r$ then by the triangle inequality we have $\|\alpha_j(n - n')\|_{\mathbb{R}/\mathbb{Z}} < \delta$ for every $j = 1, \dots, r$. Therefore, writing m for the maximum element of A , the set $(m - A) \setminus \{0\}$ is contained in $B_N(v_1, \dots, v_r; I_\delta, \dots, I_\delta)$, and so

$$|B_N(v_1, \dots, v_r; I_\delta, \dots, I_\delta)| \geq |m - A| - 1 \gg_{r, \delta} N.$$

\square

Combined with Proposition 4.8, this immediately implies Theorem 2.8 in the case that every bracket component of Φ is linear. To conclude Theorem 2.8 in general, we require the following generalisation of Lemma 4.9.

Theorem 4.10 (Recurrence of bracket polynomials; precise statement). *Let $\Theta_1, \dots, \Theta_r$ be constant-free bracket forms and suppose that $\theta_1, \dots, \theta_r$ are realisations of $\Theta_1, \dots, \Theta_r$, respectively. Let $\delta > 0$. Then, provided N is sufficiently large in terms of $\Theta_1, \dots, \Theta_r$ and δ , we have*

$$|B_N(\theta_1, \dots, \theta_r; I_\delta, \dots, I_\delta)| \gg_{\Theta_1, \dots, \Theta_r, \delta} N.$$

Remark 4.11. *The restriction here to constant-free bracket forms is necessary. For example, in the case $r = 1$, if Θ_1 were the bracket form $\{1/2 + \alpha n\}$ then for $c \ll N^{-1}$ the realisation $\{1/2 + cn\}$ of Θ_1 would not satisfy the theorem.*

We prove Theorem 4.10 in Section 6.

Compared to Lemma 4.9, the proof of which was very straightforward, Theorem 4.10 appears to be rather deep, in that our proof makes use of two major results from the literature. The first is work of Bergelson and Leibman [1] that allows us to express a bracket polynomial in terms of a so-called *polynomial sequence* on a nilmanifold. The second is a difficult theorem of Green and Tao [7] describing the distribution of such polynomial sequences.

Of course, it may well be that there is an elementary proof of Theorem 4.10, or at least of some variant of it that is still strong enough to imply Theorem 2.8. In Sections 7 and 8 we give elementary arguments establishing weak versions of Theorem 4.10 that are sufficient to prove Theorem 2.8 in certain simple cases.

5. BASES AND COORDINATES ON NILMANIFOLDS

Our aim now is to prove Theorem 4.10. As we remarked at the end of the last section, the proof makes use of results of Bergelson and Leibman [1] that allow us to express a bracket polynomial in terms of a so-called *polynomial sequence* on a nilmanifold, and results of Green and Tao [7] that describe the behaviour of such a sequence.

Even just to state these results requires a fair amount of background and notation concerning nilmanifolds, which we introduce in this section. This allows us to state the results of Bergelson–Leibman and Green–Tao in the next section, where we also prove Theorem 4.10.

At this point let us recall our convention, which applies throughout this paper, that when we write that G/Γ is a nilmanifold we assume that G is a *connected* and *simply connected* nilpotent Lie group. This is consistent with a standing assumption in [7], for example.

We start this section by introducing Mal’cev bases and coordinates on nilmanifolds. These are standard concepts in the study of nilpotent Lie groups and nilmanifolds, and are well documented in the literature; the reader may consult [2], for example, for more detailed background.

Definition 5.1 (Mal’cev basis of a nilpotent Lie algebra [2, §1.1.13]). *Let \mathfrak{g} be an m -dimensional nilpotent Lie algebra, and let $\mathcal{X} = \{X_1, \dots, X_m\}$ be a basis for \mathfrak{g} over \mathbb{R} . Then \mathcal{X} is said to be a Mal’cev basis for \mathfrak{g} if for each $j = 0, \dots, m$ the subspace \mathfrak{h}_j spanned by the vectors X_{j+1}, \dots, X_m is a Lie algebra ideal in \mathfrak{g} .*

In the event that \mathfrak{g} is the Lie algebra of a connected, simply connected nilpotent Lie group G , we sometimes say that \mathcal{X} is a Mal’cev basis for G .

Remark 5.2. *It follows from [2, Theorem 1.1.13] that every nilpotent Lie algebra admits a Mal’cev basis. We will not need this general fact in this paper, however, since we deal only with explicit bases that can easily be verified to be Mal’cev bases.*

We can use a Mal’cev basis for a connected, simply connected nilpotent Lie group G to place a coordinate system on G , using the following result.

Proposition 5.3. *Let G be a connected, simply connected, m -dimensional nilpotent Lie group with Lie algebra \mathfrak{g} . Then for every $g \in G$ there is a unique m -tuple $(t_1, \dots, t_m) \in \mathbb{R}^m$ such that*

$$(12) \quad g = \exp(t_1 X_1) \cdots \exp(t_m X_m).$$

Proof. It follows from [2, Theorem 1.2.1 (a)] and [2, Proposition 1.2.7 (c)] that every element in $\exp \mathfrak{g}$ can be expressed uniquely in the form (12), and from [2, Theorem 1.2.1 (a)] that $\exp \mathfrak{g} = G$. \square

This allows us to make the following definition.

Definition 5.4 (Mal'cev coordinates). *Let G be a connected, simply connected, m -dimensional nilpotent Lie group with Lie algebra \mathfrak{g} , and let $g \in G$. Then we call the t_i appearing in the expression (12) the Mal'cev coordinates of g . We define the Mal'cev coordinate map $\psi = \psi_{\mathcal{X}} : G \rightarrow \mathbb{R}^m$ by*

$$\psi(g) := (t_1, \dots, t_m).$$

Definition 5.5 (Mal'cev basis for a nilmanifold). *Let G/Γ be an m -dimensional nilmanifold. Then a Mal'cev basis \mathcal{X} for G is said to be compatible with Γ if Γ consists precisely of those elements whose Mal'cev coordinates are all integers. We also indicate this by saying simply that \mathcal{X} is a Mal'cev basis for G/Γ .*

In the event that \mathcal{X} is a Mal'cev basis for G/Γ , for each $g \in G$ there is a unique $z \in \Gamma$ for which the coordinates of gz all lie in $(-1/2, 1/2]$; see, for example, [7, Lemma A.14]. In this case, we call the coordinates of gz the nilmanifold coordinates of g , and define the nilmanifold coordinate map $\chi = \chi_{\mathcal{X}} : G \rightarrow (-1/2, 1/2]^m$ by

$$\chi(g) := \psi(gz).$$

These definitions are somewhat technical, so at this point the reader may find it instructive to consider the following example.

Example 5.6. *Let G/Γ be the Heisenberg nilmanifold. Let*

$$X_1 = \log \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad X_2 = \log \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad X_3 = \log \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and let $\mathcal{X} = \{X_1, X_2, X_3\}$. It is straightforward to check that

$$\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} = \exp(yX_1) \exp(xX_2) \exp(zX_3),$$

and so \mathcal{X} is a Mal'cev basis for G/Γ and its Mal'cev coordinate map $\psi_{\mathcal{X}}$ satisfies

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

On the other hand, if we change the order of X_1, X_2, X_3 , setting

$$Y_1 = X_2 \quad Y_2 = X_1 \quad Y_3 = X_3$$

and setting $\mathcal{Y} = \{Y_1, Y_2, Y_3\}$, then we have

$$\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} = \exp(xY_1) \exp(yY_2) \exp((z - xy)Y_3).$$

Thus \mathcal{Y} is also a Mal'cev basis for G/Γ , but the Mal'cev coordinate map $\psi_{\mathcal{Y}}$ satisfies

$$\psi_{\mathcal{Y}} \left(\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} \right) = (x, y, z - xy).$$

In particular, note that changing the order of the basis elements does not simply change the order of the Mal'cev coordinates.

For the purposes of this paper we will need to consider slightly more specific Mal'cev bases than those we have defined so far.

Definition 5.7 (Filtration of a nilpotent group). *Let G be a nilpotent group. A filtration G_\bullet of G is a sequence of closed connected subgroups*

$$G = G_0 = G_1 \supset G_2 \supset \cdots \supset G_d \supset G_{d+1} = \{1\}$$

with the property that $[G_i, G_j] \subset G_{i+j}$ for all integers $i, j \geq 0$. We define the degree of G_\bullet to be the minimal integer d such that $G_{d+1} = \{1\}$.

For example, the lower central series is a filtration with degree equal to the nilpotency class of the group.

Definition 5.8 (Mal'cev basis adapted to a filtration [7, Definition 2.1]). *Let G/Γ be an m -dimensional nilmanifold and let G_\bullet be a filtration for G . A Mal'cev basis $\mathcal{X} = \{X_1, \dots, X_m\}$ for G/Γ is said to be adapted to G_\bullet if for each $i = 1, \dots, d$ we have $G_i = \exp \mathfrak{h}_{m-\dim G_i}$. Here $\mathfrak{h}_j = \text{Span}\{X_{j+1}, \dots, X_m\}$ as in Definition 5.1.*

Remark 5.9. *According to a result of Mal'cev [12], every nilmanifold admits a Mal'cev basis adapted to the lower central series. It is easy to check that the Mal'cev bases for the Heisenberg group that we considered in Example 5.6 are both Mal'cev bases adapted to the lower central series.*

We close this section by introducing some higher-step variants of the Heisenberg nilmanifold, and some Mal'cev bases for them. We denote by T_p the group of $(p+1) \times (p+1)$ real upper-triangular matrices with every diagonal element equal to 1; thus, for example, T_2 is the Heisenberg group. Define Z_p to be the subgroup of T_p consisting of those matrices having only integer entries. The quotient T_p/Z_p is then a p -step nilmanifold. We define a Mal'cev basis for T_p/Z_p as follows.

Definition 5.10 (Standard basis for an upper-triangular nilmanifold). *Let $p \in \mathbb{N}$, and let U be the set of all elements of T_p that have one non-diagonal entry equal to 1, and every other non-diagonal entry equal to zero. Now for $i = 1, \dots, p$ define U_i to be the set of elements of U in which the unique non-diagonal non-zero entry is at a distance i from the main diagonal; more precisely, if the unique non-diagonal non-zero entry of $A \in U$ is the (j, k) entry then A belongs to U_{k-j} . Note, therefore, that U is the union of the U_i , and that for each $l = 1, \dots, p+1$ and each $i = 1, \dots, p$ the set U_i contains at most one element with a non-diagonal non-zero entry in row l .*

Then we define the standard basis \mathcal{X}_p for T_p/Z_p to consist of the elements of U , ordered such that if $i < j$ then every element of U_i appears before every element of U_j , and such that if $j < k$ then an element of U_i whose non-diagonal non-zero entry lies in row j appears before any element of U_i whose non-diagonal non-zero entry lies in row k .

More generally, let $r \in \mathbb{N}$, and for each $i = 1, \dots, p$ and each $j = 1, \dots, r$ define the subset $U_{i,j}$ of the direct product T_p^r to be the set

$$\{(A_1, \dots, A_r) \in T_p^r : A_j \in U_i; A_k = 1 \text{ for all } k \neq j\}.$$

Then we define the standard basis $\mathcal{X}_{p,r}$ for T_p^r/Z_p^r to consist of those elements belonging to the union of the $U_{i,j}$, ordered such that if $i < i'$ and j, j' are arbitrary then every element of

$U_{i,j}$ appears before every element of $U_{i',j'}$; such that if $j < j'$ and i is arbitrary then every element of $U_{i,j}$ appears before every element of $U_{i,j'}$; and such that if $A, B \in U_i$ and A appears before B in the basis \mathcal{X}_p then the element $(1, \dots, 1, A, 1, \dots, 1)$ of $U_{i,j}$ appears before the element $(1, \dots, 1, B, 1, \dots, 1)$ of $U_{i,j}$ in $\mathcal{X}_{p,r}$.

Thus, for example, the standard basis for T_2^2 consists of the elements

$$\begin{aligned} & \left(\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right), \\ & \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right), \\ & \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right), \\ & \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \right), \\ & \left(\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right), \\ & \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right), \end{aligned}$$

in that order.

Remark 5.11. *It is straightforward to check that the standard basis for T_p/Z_p is indeed a Mal'cev basis, adapted to the lower central series [1, §5].*

The definition of a Mal'cev basis \mathcal{X} of a nilpotent Lie algebra \mathfrak{g} requires that the vector subspaces \mathfrak{h}_j are Lie algebra ideals, which is to say that

$$[\mathfrak{g}, \mathfrak{h}_j] \subset \mathfrak{h}_j \quad (j = 0, \dots, m-1).$$

If \mathcal{X} is a Mal'cev basis for a nilmanifold adapted to a filtration, however, then it obeys the stronger property that

$$(13) \quad [\mathfrak{g}, \mathfrak{h}_j] \subset \mathfrak{h}_{j+1} \quad (j = 0, \dots, m-1);$$

here, we adopt the convention that $\mathfrak{h}_m = \{0\}$. In [7], (13) is called the *nesting property*. The nesting property turns out to be an important technical condition for various results from [7, Appendix A] that we use repeatedly in this paper. However, when we apply these results it is not always the case that we are applying them to a Mal'cev basis adapted to a filtration. It is therefore useful to introduce the following definition.

Definition 5.12 (Nested Mal'cev basis). *Let \mathfrak{g} be an m -dimensional nilpotent Lie algebra. A Mal'cev basis for \mathfrak{g} that satisfies (13) is called a nested Mal'cev basis.*

We close this section by defining what it means for a basis for \mathfrak{g} to be *rational* in a quantitative sense, which is an essential concept for understanding the results of Green and Tao that we present in the next section.

Definition 5.13 (Quantitative rationality). *The height of a rational number x is defined to be $\max\{|a|, |b|\}$ if $x = a/b$ in reduced form. If $Y, X_1, \dots, X_r \in \mathbb{R}^m$ then we say that Y is a Q -rational combination of the X_i if there are rationals q_i of height at most Q such that $Y = q_1 X_1 + \dots + q_r X_r$.*

Definition 5.14 (Rationality of a basis). *Let \mathfrak{g} be a nilpotent Lie algebra with basis $\mathcal{X} = \{X_1, \dots, X_m\}$. We say that \mathcal{X} is Q -rational if the structure constants c_{ijk} appearing in the relations*

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

are all rational of height at most Q .

6. POLYNOMIAL SEQUENCES ON NILMANIFOLDS

In this section we introduce results of Bergelson and Leibman [1] and Green and Tao [7] that allow us to prove Theorem 4.10. We start by describing the work of Bergelson and Leibman.

Recall from the introduction that the bracket polynomial $\{\alpha n[\beta n]\}$ arises naturally from the sequence

$$(14) \quad g(n) = \begin{pmatrix} 1 & -\alpha n & 0 \\ 0 & 1 & \beta n \\ 0 & 0 & 1 \end{pmatrix}$$

in the Heisenberg nilmanifold. Remarkably, Bergelson and Leibman show that *every* bracket polynomial arises in a similar way.

Definition 6.1 (Polynomial mappings and polynomial forms). *A map $\rho : \mathbb{Z} \rightarrow T_p^r$ is said to be a polynomial mapping of degree at most k if there are polynomials $\phi_{i,j,l}$ of degree at most k such that for each $n \in \mathbb{Z}$ the (i, j) -entry of the matrix $\rho_l(n)$ is given by $\phi_{i,j,l}(n)$. If the $\phi_{i,j,l}$ all have zero constant term then ρ is said to be a constant-free polynomial mapping.*

Now let $\Phi_{i,j,l}$ be polynomial forms of degree at most k in the sense of Definition 2.3. Then the r -tuple $P = (P_1, \dots, P_r)$ of $(p+1) \times (p+1)$ matrices whose diagonal entries are all 1, and such that every above-diagonal (i, j) -entry of P_l is equal to the polynomial form $\Phi_{i,j,l}$, is said to be a polynomial form of degree at most k on T_p^r . If every $\Phi_{i,j,l}$ is constant free then P is also said to be constant free.

Finally, for each i, j, l let $\phi_{i,j,l}$ be a realisation of $\Phi_{i,j,l}$. Let $\rho : \mathbb{Z} \rightarrow T_p^r$ be the polynomial mapping defined by setting the (i, j) -entry of $\rho_l(n)$ equal to $\phi_{i,j,l}(n)$. Then ρ is said to be a realisation of the polynomial form P .

Remark 6.2. *If the α and β appearing in (14) are taken to be elements of some alphabet (as opposed to real numbers) then the matrix $g(n)$ can be viewed as a polynomial form of degree at most 1 on T_2 . It follows from [1, §5.8] that for every p, r the nilmanifold coordinates $\chi(g)$ of an element $g \in T_p^r$ with respect to the standard basis are given by bracket expressions in the entries of the matrices appearing in g . Thus, in particular, if P is a polynomial form on T_p^r then each coordinate $\chi(P)_i$ naturally defines a bracket form.*

Theorem 6.3 (Bergelson–Leibman [1]). *Let $\Theta_1, \dots, \Theta_r$ be constant-free bracket forms. Then there exist $p \geq 1$, a constant-free polynomial form P on T_p^r , and a nested Mal'cev basis $\mathcal{Y} = \{Y_1, \dots, Y_m\}$ for T_p^r/Z_p^r such that each element Y_i of \mathcal{Y} is equal to either an element X_j of the standard basis \mathcal{X} or its inverse $-X_j$, and such that for every $i = 1, \dots, r$ we have either $\{\Theta_i\} = \chi_{\mathcal{Y}}(P)_{m-r+i}$ or $\{-\Theta_i\} = \chi_{\mathcal{Y}}(P)_{m-r+i}$.*

Theorem 6.3 is not stated exactly in this way in Bergelson and Leibman's paper, but it can be read out of the work contained therein. In particular, Bergelson and Leibman express concrete bracket polynomials in terms of concrete polynomial mappings, whereas

Theorem 6.3 expresses bracket forms in terms of polynomial forms. The reason for this modification is to make it clear that all implied constants appearing in our subsequent work are uniform across all realisations of a given bracket form. In Appendix B we offer a brief discussion of how to obtain Theorem 6.3 from the original work of Bergelson and Leibman.

It turns out to be useful to note, as we do in Lemma 6.6 below, that polynomial mappings into nilmanifolds are examples of slightly more specific objects called *polynomial sequences* on nilmanifolds. We define these now.

Definition 6.4 (Polynomial sequence in a nilpotent group). *Let G be a nilpotent group with a filtration G_\bullet and let $g : \mathbb{Z} \rightarrow G$ be a sequence. For $h \in \mathbb{Z}$ define $\partial_h g(n) := g(n+h)g(n)^{-1}$. Then g is said to be a polynomial sequence with respect to the filtration G_\bullet if $\partial_{h_1} \dots \partial_{h_i} g$ takes values in G_i for all $i \in \mathbb{N}$ and $h_1, \dots, h_i \in \mathbb{Z}$.*

Example 6.5. *In the Heisenberg group T_2 , the sequence*

$$\begin{pmatrix} 1 & \alpha_1 n & \beta n^2 \\ 0 & 1 & \alpha_2 n \\ 0 & 0 & 1 \end{pmatrix}$$

is a polynomial sequence with respect to the lower central series. In the group T_3 the sequence

$$\begin{pmatrix} 1 & \alpha_1 n & \beta_1 n^2 & \gamma n^3 \\ 0 & 1 & \alpha_2 n & \beta_2 n^2 \\ 0 & 0 & 1 & \alpha_3 n \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

is a polynomial sequence with respect to the lower central series. More generally, if $g(n)$ is a sequence inside the group T_p defined by a matrix, each of whose entries is a polynomial in n of degree at most its distance from the main diagonal, then g is a polynomial sequence with respect to the lower central series. We leave it to the reader to verify this fact.

The polynomial sequences given in Example 6.5 are of course also polynomial mappings into T_p . However, not every polynomial mapping is a polynomial sequence with respect to the lower central series, as can be seen by considering, for example, the mapping

$$\begin{pmatrix} 1 & \alpha n^2 \\ 0 & 1 \end{pmatrix}$$

into T_1 . It turns out, however, that every polynomial mapping into T_p^r is a polynomial sequence with respect to *some* filtration.

Lemma 6.6. *Let $k, p, r \in \mathbb{N}$. Then there is a filtration G_\bullet of T_p^r of degree at most $O_{k,p}(1)$ with respect to which every polynomial mapping $\rho : \mathbb{Z} \rightarrow T_p^r$ of degree at most k is a polynomial sequence, and such that the standard basis \mathcal{X} for T_p^r / Z_p^r is a Mal'cev basis adapted to G_\bullet .*

We prove Lemma 6.6 shortly, but first we note the following statement, which is a key ingredient of Lemma 6.6.

Lemma 6.7. *Let $k, p, r \in \mathbb{N}$. Then there is a some $d \in \mathbb{N}$ depending only on k and p such that if ρ is an arbitrary polynomial mapping of degree at most k into T_p^r then the derivatives $\partial_{h_{d+1}} \dots \partial_{h_1} \rho$ are all trivial.*

The proof of Lemma 6.7 is a straightforward exercise, but given its importance to this paper we present it in full in Appendix C.

Proof of Lemma 6.6. Let G'_\bullet be the lower central series of T_p^r , which is a filtration of degree p , and let d be the natural number given by Lemma 6.7. Following the procedure outlined in the paragraphs following [7, Corollary 6.8], define a finer filtration G_\bullet of degree pd by setting $G_i = G'_{\lfloor i/d \rfloor}$. Then ρ is a polynomial sequence for the filtration G_\bullet , and so the first conclusion of the lemma is proved.

The fact that \mathcal{X} is a Mal'cev basis adapted to G_\bullet follows straightforwardly from the fact (noted in Remark 5.11) that it is a Mal'cev basis adapted to G'_\bullet , and from the fact that each G_i is equal to some G'_j . \square

Our objective in this section is to prove Theorem 4.10, which is a recurrence result for bracket polynomials, modulo 1. Moreover, in light of Theorem 6.3 and Lemma 6.6 the study of bracket polynomials reduces, in a sense, to the study of polynomial sequences on nilmanifolds. This suggests that it would be useful to understand the distribution of such polynomial sequences. We now describe deep work of Green and Tao investigating precisely this.

We start by defining a metric on a nilmanifold.

Definition 6.8 (Metrics on nilmanifolds). *Given a nilmanifold G/Γ with a rational nested Mal'cev basis \mathcal{X} , we define a metric $d = d_{G,\mathcal{X}}$ on G by taking the largest metric such that $d(x, y) \leq |\psi(xy^{-1})|$ for all $x, y \in G$; here, and throughout this paper, $|\cdot|$ denotes the ℓ^∞ -norm on $\mathbb{R}^{\dim G}$. We also define a metric $d = d_{G/\Gamma,\mathcal{X}}$ on G/Γ by*

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

Remark 6.9. *It is shown in [7, Lemma A.15] that this is a metric on G/Γ . Note that, although the hypotheses of that lemma include the assumption that \mathcal{X} is a Mal'cev basis adapted to some filtration, all that is used is that the elements of Γ have integer coordinates, and that \mathcal{X} is rational and nested (the assumption that \mathcal{X} is nested being necessary in order to apply [7, Lemmas A.4 and A.5]).*

This definition is rather abstract. However, we never need to calculate it explicitly, and the only properties we require are detailed in [7, Appendix A]. The interested reader may find a more explicit formulation of this metric in [7, Definition 2.2].

One immediate property of the metric d is that it is *right invariant*, in the sense that

$$(15) \quad d(x, y) = d(xg, yg)$$

for every $x, y, g \in G$. Another is that the metric d is symmetric at the identity, in the sense that

$$(16) \quad d(x, 1) = d(x^{-1}, 1)$$

for every $x \in G$.

Once we have a metric on G/Γ we are able to make the following definitions.

Definition 6.10 (Lipschitz norm). *Define the Lipschitz norm $\|\cdot\|_{\text{Lip}}$ on the space of Lipschitz functions $f : G/\Gamma \rightarrow \mathbb{C}$ by*

$$\|f\|_{\text{Lip}} := \|f\|_\infty + \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)},$$

Definition 6.11 (Equidistribution). *Let G/Γ be a nilmanifold, and write μ for the unique normalised Haar measure on G/Γ . A sequence $(g(n)\Gamma)_{n \in \mathbb{Z}}$ is said to be equidistributed in G/Γ if for every continuous function $f : G/\Gamma \rightarrow \mathbb{C}$ we have*

$$\mathbb{E}_{n \in \mathbb{Z}} f(g(n)\Gamma) = \int_{G/\Gamma} f d\mu.$$

Now let $\delta > 0$ be a parameter and let $Q \subset \mathbb{Z}$ be an arithmetic progression of length N . A sequence $(g(n)\Gamma)_{n \in Q}$ is said to be δ -equidistributed in G/Γ if for every Lipschitz function $f : G/\Gamma \rightarrow \mathbb{C}$ we have

$$\left| \mathbb{E}_{n \in Q} f(g(n)\Gamma) - \int_{G/\Gamma} f d\mu \right| \leq \delta \|f\|_{\text{Lip}}.$$

We say that $(g(n)\Gamma)_{n \in Q}$ is totally δ -equidistributed if $(g(n)\Gamma)_{n \in Q'}$ is δ -equidistributed for every subprogression $Q' \subset Q$ of length at least δN .

The key result of Green and Tao shows that *every* polynomial sequence g on a nilmanifold G/Γ has an ‘equidistributed’ component, in a certain precise sense. More specifically, their result allows us to factor an arbitrary polynomial sequence g on a nilmanifold G/Γ as a product $\varepsilon g' \gamma$, in which g' is δ -equidistributed on some subnilmanifold G'/Γ' of G , in which ε is ‘almost constant’ in a certain sense, and in which γ is periodic with fairly short period. Thus, ignoring for the moment the effects of the ‘almost constant’ sequence ε , we see that g is roughly equidistributed on the union of a small number of translates of the subnilmanifold G'/Γ' .

For this to make sense, we must first define what we mean by a ‘subnilmanifold’.

Definition 6.12 (Rational subgroups and subnilmanifolds). *Let G/Γ be a nilmanifold with Mal'cev basis $\mathcal{X} = \{X_1, \dots, X_m\}$. Suppose that G' is a closed connected subgroup of G . We say that G' is Q -rational relative to \mathcal{X} if the Lie algebra \mathfrak{g}' has a basis consisting of Q -rational combinations of the X_i . In this case, the subgroup Γ' , defined to be $G' \cap \Gamma$, is a discrete cocompact subgroup of G' , and so G'/Γ' is a nilmanifold. We call G'/Γ' a subnilmanifold of G/Γ .*

We must also define a way in which a polynomial sequence can be ‘almost constant’.

Definition 6.13 (Smooth sequences). *Let $M, N \geq 1$. Let G/Γ be a nilmanifold with Mal'cev basis \mathcal{X} , and let d be the metric on G associated to \mathcal{X} . Let $(\varepsilon(n))_{n \in \mathbb{Z}}$ be a sequence in G . Then we say that $(\varepsilon(n))_{n \in \mathbb{Z}}$ is (M, N) -smooth if*

$$d(\varepsilon(n), 1) \leq M \quad \text{and} \quad d(\varepsilon(n), \varepsilon(n-1)) \leq M/N$$

for all $n \in [N]$.

We can now finally state precisely the factorisation theorem of Green and Tao.

Theorem 6.14 (Green–Tao [7, Theorem 1.19]). *Let $M_0, N > 0$ and let $A > 0$. Let G/Γ be a nilmanifold with a filtration G_\bullet of degree d and let \mathcal{X} be an M_0 -rational Mal'cev basis for G/Γ adapted to G_\bullet . Let $g : \mathbb{Z} \rightarrow G$ be a polynomial sequence such that $g(0) = 1$. Then there exist an integer M with $M_0 \leq M \leq M_0^{O_{A,G,G_\bullet}(1)}$; a rational subgroup $G' \subset G$; a Mal'cev basis \mathcal{X}' for G'/Γ' that is adapted to some filtration of G' and 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 $\varepsilon, g', \gamma : \mathbb{Z} \rightarrow G$ satisfying the following conditions:*

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

- (ii) g' takes values in G' and the finite sequence $(g'(n)\Gamma')_{n \in [N]}$ is totally $1/M^A$ -equidistributed in G'/Γ' with respect to \mathcal{X}' ;
- (iii) $(\gamma(n)\Gamma')_{n \in \mathbb{Z}}$ is periodic with period at most M .
- (iv) $\varepsilon(0) = g'(0) = \gamma(0) = 1$

Remarks on the proof. The statement of [7, Theorem 1.19] is not quite the same as the statement of Theorem 6.14, in that there is no assumption that $g(0) = 1$ and, correspondingly, there is no conclusion that $\varepsilon(0) = g'(0) = \gamma(0) = 1$. One sees that the former implies the latter on inspection of the proof of [7, Theorem 1.19]. In fact, [7, Theorem 1.19] is an instance of [7, Theorem 10.2]. This in turn is obtained by repeated application of [7, Theorem 9.2], which establishes [7, Theorem 10.2] in a certain special case.

The first part of the proof of [7, Theorem 9.2] reduces to the case in which the polynomial sequence under consideration takes value 1 at 0. Once in that case, it is straightforward to verify that the sequences ε, g', γ arising from the proof also take the value 1 at 0. Indeed, the sequences ε and γ satisfy

$$\psi(\gamma(n)) = \sum_{j>0} v_j \binom{n}{j} \quad \text{and} \quad \psi(\varepsilon(n)) = \sum_{j>0} v'_j \binom{n}{j}.$$

Here, the v_j, v'_j are certain real numbers, the values of which are superfluous for the purposes of this discussion since when $n = 0$ the binomial coefficients appearing in the sums all take the value zero, and so $\gamma(0)$ and $\varepsilon(0)$ must both equal the identity. The condition $g = \varepsilon g' \gamma$ then implies that $g'(0)$ is also the identity, as claimed.

One could therefore simply require by definition that a polynomial sequence takes value 1 at 0, without affecting the truth of [7, Theorem 9.2]. The deduction of [7, Theorem 10.2] would proceed in exactly the same way as in [7], but with the additional conclusion that all polynomial sequences arising as a result would take the value 1 at 0.

The reader may also note that [7, Theorem 1.19] does not say explicitly that the Mal'cev basis \mathcal{X}' for G'/Γ' is adapted to a filtration of G' . However, this apparent omission is simply because the nomenclature of that paper is not quite the same as in this paper, in that in [7] Mal'cev bases are, by definition, always adapted to some filtration. In fact, one sees from the proof of [7, Theorem 10.2] that the filtration of G' to which \mathcal{X}' is adapted is given by $G_\bullet \cap G'$. \square

Remark 6.15. *A slightly more careful inspection of the proof of [7, Theorem 10.2] reveals that, even in the absence of any assumption on $g(0)$, one can conclude that $\varepsilon(0)$ lies in the fundamental domain of G/Γ and that $\gamma(0) \in \Gamma$, with $\varepsilon(0)\gamma(0) = g(0)$. Thus, in particular, if $g(0)$ is the identity then so too are $\varepsilon(0)$, $g'(0)$ and $\gamma(0)$.*

The following lemma, the proof of which we defer until Appendix A, gives an idea of how we will use the factorisation theorem of Green and Tao to deduce recurrence results for polynomial sequences.

Lemma 6.16. *Let $M \geq 2$. Let G/Γ be an m -dimensional nilmanifold with an M -rational nested Mal'cev basis \mathcal{X} , and let d be the metric associated to \mathcal{X} . Let $\rho \leq 1$ and $x \in G$, and suppose that $g : [N] \rightarrow G$ is η -equidistributed in G/Γ . Then a proportion of at least*

$$\frac{\rho^m}{M^{O(m)}} - \frac{3\eta}{\rho}$$

of the points $(g(n)\Gamma)_{n \in [N]}$ lie in the ball $\{y\Gamma : d(y\Gamma, x\Gamma) \leq \rho\}$.

Lemma 6.16, of course, shows that the component g' of the polynomial sequence g given by Theorem 6.14 is recurrent in a certain sense. In the context of proving Theorem 4.10, however, it is the sequence g itself that we will need to be recurrent. The following lemma allows us to obtain recurrence of g from recurrence of g' . Again, we defer the proof until Appendix A.

Lemma 6.17. *Let $\rho \leq 1$ and σ be parameters. Let G/Γ be a nilmanifold with an M -rational nested Mal'cev basis \mathcal{X} , suppose that G' is a rational subgroup of G , and suppose that \mathcal{X}' is a nested Mal'cev basis for G'/Γ' in which each element is an M -rational combination of the elements of \mathcal{X} . Suppose that $\varepsilon \in G$ satisfies $d(\varepsilon, 1) \leq \sigma$, and that $\gamma \in \Gamma$. Finally, suppose that g is an element of G' such that $d'(g\Gamma', \Gamma') \leq \rho$. Then $d(\varepsilon g\gamma\Gamma, \Gamma) \leq M^{O(1)}\rho + \sigma$.*

An immediate issue with combining Theorems 6.3 and 6.14 is that the Mal'cev basis \mathcal{Y} given by Theorem 6.3 is not necessarily adapted to the filtration G_\bullet given by Lemma 6.6. However, the following result shows that the coordinate system associated to \mathcal{Y} is at least comparable to the metric associated to the standard basis, which is a Mal'cev basis adapted to G_\bullet .

Lemma 6.18. *Let $\mathcal{Y} = \{Y_1, \dots, Y_m\}$ be a nested Mal'cev basis for T_p^r/Z_p^r in which each element Y_i is equal to either an element X_j of the standard basis \mathcal{X} or its inverse $-X_j$. Then the nilmanifold coordinate system $\chi_{\mathcal{Y}}$ associated to \mathcal{Y} , and the metric d associated to the standard basis $\mathcal{X} = \{X_1, \dots, X_m\}$, satisfy*

$$|\chi_{\mathcal{Y}}(x)| \ll_{p,r} d(x\Gamma, \Gamma)$$

for every $x \in T_p^r$.

The proofs of Lemmas 6.16, 6.17 and 6.18 all essentially proceed by piecing together various results from [7, Appendix A]. We present the details in Appendix A. Modulo these proofs, it is now a fairly straightforward matter to combine the results of this section to prove Theorem 4.10, as follows. Recall that Proposition 4.8 and Theorem 4.10 combine to give Theorem 2.8.

Proof of Theorem 4.10. Apply Theorem 6.3 and Lemma 6.18 to obtain a constant-free polynomial form P on a nilmanifold $G/\Gamma := T_p^r/Z_p^r$ with a nested Mal'cev basis

$$\mathcal{Y} = \{Y_1, \dots, Y_m\}$$

such that

$$(17) \quad \{\pm\Theta_i(n)\} = \chi_{\mathcal{Y}}(P(n))_{m-r+i},$$

and such that the nilmanifold coordinate map $\chi_{\mathcal{Y}}$ associated to \mathcal{Y} and the metric d associated to the standard basis $\mathcal{X} = \{X_1, \dots, X_m\}$ satisfy $|\chi_{\mathcal{Y}}(x)| \ll_{p,r} d(x\Gamma, \Gamma)$ for every $x \in G$. In fact, since p and r depend only on $\Theta_1, \dots, \Theta_r$ we have

$$(18) \quad |\chi_{\mathcal{Y}}(x)| \ll_{\Theta_1, \dots, \Theta_r} d(x\Gamma, \Gamma).$$

Applying Lemma 6.6, let G_\bullet be a filtration of G to which the standard basis is adapted, and with respect to which every realisation of P is a polynomial sequence.

Let $A > 0$ be a constant to be chosen later but depending only on $\Theta_1, \dots, \Theta_r$ and δ . Fix arbitrary realisations $\theta_1, \dots, \theta_r$ of $\Theta_1, \dots, \Theta_r$, respectively. Let g be a realisation of P such that

$$(19) \quad \{\pm\theta_i(n)\} = \chi_{\mathcal{Y}}(g(n))_{m-r+i}$$

for $i = 1, \dots, r$; such a realisation exists by (17). Note in particular that the fact that P is constant free implies that $g(0) = 1$. Applying Theorem 6.14 with $M_0 = 2$ therefore gives an integer M with $2 \leq M \ll_{A,G,G_\bullet} 1$; a rational subgroup $G' \subset G$; a Mal'cev basis \mathcal{X}' for G'/Γ' adapted to some filtration of G' and 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 $\varepsilon, g', \gamma : \mathbb{Z} \rightarrow G$ that satisfy conditions (i), (ii), (iii) and (iv) of Theorem 6.14. In fact, since A depends only on $\Theta_1, \dots, \Theta_r$ and δ , and since G and G_\bullet depend only on $\Theta_1, \dots, \Theta_r$, we may assume that

$$(20) \quad 2 \leq M \ll_{\Theta_1, \dots, \Theta_r, \delta} 1.$$

Condition (iii) states that $(\gamma(n)\Gamma)_{n \in \mathbb{Z}}$ is periodic with period q , say, with $q \leq M$. This last inequality, combined with the upper bound (20) on M , implies that in order to prove the proposition it is sufficient to show that a fraction $\Omega_{\Theta_1, \dots, \Theta_r, \delta}(1)$ of the points in $q\mathbb{N} \cap [N]$ belong to the set $B_N(\theta_1, \dots, \theta_r; I_\delta, \dots, I_\delta)$, and so we may restrict attention to $q\mathbb{N} \cap [N]$ if we wish. Let us do so, replacing each θ_i by the function $\hat{\theta}_i$ defined by $\hat{\theta}_i(n) = \theta_i(qn)$; replacing g by the sequence \hat{g} defined by $\hat{g}(n) = g(qn)$; and replacing N by $\lfloor N/q \rfloor$. Once these replacements are made, conditions (i), (ii), (iii) and (iv) of Theorem 6.14 become the following conditions:

- (i) $\varepsilon : \mathbb{Z} \rightarrow G$ is (M, N) -smooth;
- (ii) g' takes values in G' , and for any progression $Q \subset [N]$ of length at least N/M^{A-1} the sequence $(g'(n)\Gamma')_{n \in Q}$ is $1/M^A$ -equidistributed in G'/Γ' with respect to \mathcal{X}' ;
- (iii) γ takes values in Γ ;
- (iv) $\varepsilon(0) = g'(0) = \gamma(0) = 1$.

Since A depends only on $\Theta_1, \dots, \Theta_r$ and δ , the inequality (20) implies that we may restrict attention to the subsequence $[N/M^{A-2}]$. We may therefore replace N by $\lfloor N/M^{A-2} \rfloor$, and hence replace conditions (i) and (ii) by the following:

- (i) $d(\varepsilon(n), 1) \leq 1/M^{A-3}$ for every $n \in [N]$;
- (ii) g' takes values in G' , and the finite sequence $(g'(n)\Gamma')_{n \in [N]}$ is $1/M^A$ -equidistributed in G'/Γ' with respect to \mathcal{X}' .

Let $\rho \leq 1$ be a constant to be determined later. Lemma 6.16 and condition (ii) together imply that

$$(21) \quad \mathbb{P}_{n \in [N]}(d'(g'(n)\Gamma', \Gamma') \leq \rho) \geq \frac{\rho^m}{M^{O(m)}} - \frac{3}{\rho M^A}.$$

Condition (i), on the other hand, combines with Lemma 6.17 to imply that whenever $n \in [N]$ and $d'(g'(n)\Gamma', \Gamma') \leq \rho$ we have $d(\varepsilon(n)g'(n)\gamma(n)\Gamma, \Gamma) \leq M^{O(1)}\rho + 1/M^{A-3}$. It therefore follows from (21) that

$$\mathbb{P}_{n \in [N]}(d(g(n)\Gamma, \Gamma) \leq M^{O(1)}\rho + 1/M^{A-3}) \geq \frac{\rho^m}{M^{O(m)}} - \frac{3}{\rho M^A}.$$

In light of (18) and the lower bound (20) on M , this in turn implies that there exists a constant $C \geq 1$ depending only on $\Theta_1, \dots, \Theta_r$ such that

$$\mathbb{P}_{n \in [N]} \left(|\chi_{\mathcal{Y}}(g(n))| \leq M^C \left(\rho + \frac{1}{M^A} \right) \right) \geq \frac{\rho^m}{M^{Cm}} - \frac{3}{\rho M^A}.$$

Setting $\rho = \delta/2M^C$ therefore implies that

$$\mathbb{P}_{n \in [N]} \left(|\chi_{\mathcal{Y}}(g(n))| \leq \frac{\delta}{2} + M^{C-A} \right) \geq \frac{\delta^m}{2^m M^{2Cm}} - \frac{6M^{C-A}}{\delta}.$$

Thanks to the lower bound (20) on M , by setting A sufficiently large in terms of m and C , which depend only on $\Theta_1, \dots, \Theta_r$, and δ , we may therefore conclude that

$$\mathbb{P}_{n \in [N]} (|\chi_{\mathcal{Y}}(g(n))| \leq \delta) \geq \frac{\delta^m}{M^{O_{\Theta_1, \dots, \Theta_r}(m)}}.$$

It follows from (19), the upper bound (20) on M and the fact that m depends only on $\Theta_1, \dots, \Theta_r$ that

$$\mathbb{P}_{n \in [N]} (\|\theta_i(n)\|_{\mathbb{R}/\mathbb{Z}} \leq \delta \text{ for each } i = 1, \dots, r) \gg_{\Theta_1, \dots, \Theta_r, \delta} 1,$$

and so the proposition is proved. \square

7. WEAK RECURRENCE OF BRACKET POLYNOMIALS

If one includes the work of Bergelson–Leibman and Green–Tao that we used to prove Theorem 4.10, the proof of Theorem 2.8 is extremely long and difficult. In this section and Section 8 we investigate the extent to which we can prove similar results using only elementary methods.

We concentrate our attention on an explicit model setting. Specifically, we consider the bracket polynomials ϕ_{k-1} defined by $\phi_{k-1}(n) = \alpha_{k-1}n\{\alpha_{k-2}n\{\dots\{\alpha_1n\}\dots\}$. Theorem 2.8 instantly tells us that

$$(22) \quad \|e(\phi_{k-1})\|_{U^k[N]} \gg_k 1.$$

Our elementary methods will allow us to prove (22) directly in the cases $k \leq 5$.

For $k \leq 3$ there is already nothing more to do. Indeed, the case $k = 2$ is trivial, whilst the case $k = 3$ follows from Proposition 4.8 and Lemma 4.9. When $k \geq 4$, however, ϕ_{k-1} has the non-linear bracket component $\{\alpha_2n\{\alpha_1n\}\}$, and so Lemma 4.9 does not apply to the set of bracket components of ϕ_{k-1} . The cases $k = 4, 5$ therefore require some more work. We treat the case $k = 4$ in this section, and then prove the case $k = 5$ in Section 8.

Theorem 4.10 is a very general result, but it is also somewhat stronger than is strictly necessary to prove Theorem 2.8. For that purpose it would in fact be sufficient to establish recurrence of bracket polynomials in a weaker sense.

Definition 7.1 (Weak recurrence (modulo 1)). *A set $\{v_1, \dots, v_m\}$ of bracket polynomials on $[N]$ will be said to be (ε, λ) -weakly recurrent (modulo 1) if*

$$|B_N(v_1, \dots, v_m; I_{\frac{1}{2}-\varepsilon}, \dots, I_{\frac{1}{2}-\varepsilon})| \geq \lambda N.$$

Whilst the first notion of recurrence that we discussed required a positive fraction of the values of $\{v_i(n)\}$ to be very close to zero, weak recurrence requires only that a positive fraction of the values of $\{v_i(n)\}$ are not too close to $1/2$.

An elementary weak-recurrence result for arbitrary finite sets of bracket polynomials would give an elementary proof that bracket polynomials were non-uniform, thanks to the following result.

Proposition 7.2. *Let $\phi : [N] \rightarrow \mathbb{R}$ be a bracket polynomial of degree at most k . Suppose that $C(\phi) = \{v_1, \dots, v_m\}$ is (ε, λ) -weakly recurrent (modulo 1). Then there exist intervals*

$J_1, \dots, J_m \subset (-1/2, 1/2]$ such that $|B_N(\nu_1, \dots, \nu_m; J_1, \dots, J_m)| \gg_{k,m,\varepsilon,\lambda} N$, and such that ϕ is strongly locally polynomial of degree at most k on $B_N(\nu_1, \dots, \nu_m; J_1, \dots, J_m)$.

Proof. By weak recurrence we have

$$|B_N(\nu_1, \dots, \nu_m; I_{\frac{1}{2}-\varepsilon}, \dots, I_{\frac{1}{2}-\varepsilon})| \geq \lambda N.$$

Let c_k be the constant given by Lemma 4.7, and set $\delta = \min\{c_k, \varepsilon k^{-1}\}$. Applying Lemma 3.4 to ν_1, \dots, ν_m with $A = B_N(\nu_1, \dots, \nu_m; I_{\frac{1}{2}-\varepsilon}, \dots, I_{\frac{1}{2}-\varepsilon})$, with $\delta_1 = \dots = \delta_m = \delta$ and $I = I_{\frac{1}{2}-\varepsilon}$ gives intervals J_1, \dots, J_m of width δ inside $I_{\frac{1}{2}-\varepsilon}$ such that

$$|B_N(\nu_1, \dots, \nu_m; J_1, \dots, J_m)| \gg_{k,m,\varepsilon,\lambda} N,$$

Lemma 4.7 then implies the desired result. \square

Remark 7.3. Combining Proposition 7.2 with Proposition 4.5 shows that if $f(n) := e(\phi(n))$ for some ϕ of degree at most k with (ε, λ) -weakly recurrent $C(\phi) = \{\nu_1, \dots, \nu_m\}$ then

$$\|f\|_{U^{k+1}_{[N]}} \gg_{k,m,\varepsilon,\lambda} 1.$$

Proving weak recurrence (modulo 1) for an arbitrary bracket polynomial without appealing to the work of Bergelson–Leibman and Green–Tao appears to be somewhat difficult. However, building on Lemma 4.9, we are at least able to make some progress in the case that the set of bracket polynomials under consideration has at most one non-linear member.

Proposition 7.4 (Bracket linears and their product are weakly recurrent). *Let $k, m, r \in \mathbb{Z}$ with $0 \leq m \leq k, r$ and let $\alpha_0, \dots, \alpha_r \in \mathbb{R}$. For $i = 1, \dots, r$ let $\nu_i : n \mapsto \{\alpha_i n\}$ be a linear bracket polynomial. Let $\phi(n) := \alpha_0 n^{k-m} \prod_{i=1}^m \nu_i(n)$. Then for $\varepsilon > 0$ and $\varepsilon' \ll_{k,r} 1$ we have*

$$|B_N(\phi, \nu_1, \dots, \nu_r; I_{\frac{1}{2}-\varepsilon'}, I_\varepsilon, \dots, I_\varepsilon)| \gg_{k,r,\varepsilon} N.$$

We in fact prove Proposition 7.4 in the following form, which easily implies Proposition 7.4 when combined with Lemma 4.9.

Proposition 7.5. *Let $k, m, r \in \mathbb{Z}$ with $0 \leq m \leq k, r$ and let $\alpha_0, \dots, \alpha_r \in \mathbb{R}$. For $i = 1, \dots, r$ let $\nu_i : n \mapsto \{\alpha_i n\}$ be a linear bracket polynomial. Let $\phi(n) := \alpha_0 n^{k-m} \prod_{i=1}^m \nu_i(n)$. Let $\varepsilon \ll_k 1$ and δ be parameters. Then there exists a real number $\eta \in (0, 1)$, depending only on k and ε , such that if there is some interval $J \subset (-1/2, 1/2]$ of width at most δ for which*

$$|B_N(\phi, \nu_1, \dots, \nu_r; J, I_\varepsilon, \dots, I_\varepsilon)| \geq (1 - \eta) |B_N(\nu_1, \dots, \nu_r; I_\varepsilon, \dots, I_\varepsilon)|$$

then

$$|B_N(\phi, \nu_1, \dots, \nu_r; I_{O_k(\delta)}, I_{O_k(\varepsilon)}, \dots, I_{O_k(\varepsilon)})| \gg_{k,r,\varepsilon} N.$$

We make use of two lemmas. We will say that a bracket polynomial is of degree *exactly* k if it is of degree at most k but not of degree at most $k - 1$.

Lemma 7.6. *Suppose ϕ is an elementary bracket polynomial of degree exactly k with bracket components ν_1, \dots, ν_m . Then there exists \hat{c}_k such that if $\varepsilon \leq \hat{c}_k$ and $n, n + h, \dots, n + kh$ all lie in $B_N(\nu_1, \dots, \nu_m; I_\varepsilon, \dots, I_\varepsilon)$ we have*

$$(\Delta_h)^k \phi(n) = k! \phi(h).$$

The proof is a simple induction and left as an exercise to the reader.

Lemma 7.7. *Let $k, m, r \in \mathbb{Z}$ with $0 \leq m \leq k, r$ and let $\alpha_0, \dots, \alpha_r \in \mathbb{R}$. For $i = 1, \dots, r$ let $v_i : n \mapsto \{\alpha_i n\}$ be a linear bracket polynomial. Let $\phi(n) := \alpha_0 n^{k-m} \prod_{i=1}^m v_i(n)$. Let \hat{c}_k be as in Lemma 7.6, and let $\varepsilon \leq \hat{c}_k$ and $\delta > 0$ be parameters. Suppose that $n \in [N]$ and $h > 0$ and that there exists an interval $J \subset (-1/2, 1/2]$ of width at most δ such that $n, n+h, \dots, n+kh$ all lie in $B_{\lfloor N/k! \rfloor}(\phi, v_1, \dots, v_r; J, I_\varepsilon, \dots, I_\varepsilon)$. Then*

$$k!h \in B_N(\phi, v_1, \dots, v_r; I_{O_k(\delta)}, I_{O_k(\varepsilon)}, \dots, I_{O_k(\varepsilon)}).$$

Proof. The fact that $n, n+h, \dots, n+kh \in B_{\lfloor N/k! \rfloor}(\phi, v_1, \dots, v_r; J, I_\varepsilon, \dots, I_\varepsilon)$ implies in particular that $n, n+h, \dots, n+kh \in B_N(v_1, \dots, v_m; I_\varepsilon, \dots, I_\varepsilon)$, and so Lemma 7.6 implies that

$$(23) \quad (\Delta_h)^k \phi(n) = k! \phi(h).$$

However, the fact that $\phi(n+jh) \in J$ for all j implies, by Lemma 3.2 (i) and induction, that

$$(\Delta_h)^k \phi(n) \in I_{O_k(\delta)},$$

which combined with (23) of course implies that

$$(24) \quad k! \phi(h) \in I_{O_k(\delta)}.$$

Now the fact that $v_i(n)$ and $v_i(n+h)$ both lie in I_ε for all i implies, by Lemma 3.2 (i), that

$$(25) \quad v_i(h) \in I_{2\varepsilon} \text{ for all } i,$$

which in turn implies, by Lemma 3.2 (ii), that $v_i(k!h) = k!v_i(h)$ for all i . This implies that $\phi(k!h) = (k!)^k \phi(h)$, which combined with (24) and Lemma 3.2 (ii) gives

$$(26) \quad \phi(k!h) \in I_{O_k(\delta)}.$$

Furthermore, (25) and Lemma 3.2 (ii) imply that

$$(27) \quad v_i(k!h) \in I_{O_k(\varepsilon)} \text{ for all } i.$$

Combining (26) and (27) yields the desired result. \square

Proof of Proposition 7.5. By Lemma 7.7 it suffices to find $\Omega_{k,r,\varepsilon}(N)$ values of h for each of which there exists at least one progression $n, n+h, \dots, n+kh$ of common difference h contained within $B_{\lfloor N/k! \rfloor}(\phi, v_1, \dots, v_r; J, I_\varepsilon, \dots, I_\varepsilon)$. We can certainly find many values of h for which there exist such progressions contained within $B_{\lfloor N/k! \rfloor}(v_1, \dots, v_r; I_\varepsilon, \dots, I_\varepsilon)$; indeed, by Lemma 4.9 we have

$$(28) \quad |B_{\lfloor N/(k+1)! \rfloor}(v_1, \dots, v_r; I_{\varepsilon/(k+1)}, \dots, I_{\varepsilon/(k+1)})| \gg_{k,r,\varepsilon} N,$$

and if $h \in B_{\lfloor N/(k+1)! \rfloor}(v_1, \dots, v_r; I_{\varepsilon/(k+1)}, \dots, I_{\varepsilon/(k+1)})$ then for every $n \in B_{\lfloor N/(k+1)! \rfloor}(v_1, \dots, v_r; I_{\varepsilon/(k+1)}, \dots, I_{\varepsilon/(k+1)})$ we have $n, n+h, \dots, n+kh \in B_{\lfloor N/k! \rfloor}(v_1, \dots, v_r; I_\varepsilon, \dots, I_\varepsilon)$. Writing $a = a_{k,r,\varepsilon}$ for the constant implicit in (28), so that

$$|B_{\lfloor N/(k+1)! \rfloor}(v_1, \dots, v_r; I_{\varepsilon/(k+1)}, \dots, I_{\varepsilon/(k+1)})| > aN,$$

we may conclude that for at least $\Omega_{k,r,\varepsilon}(N)$ values of h we have at least aN values of n for which $n, n+h, \dots, n+kh \in B_{\lfloor N/k! \rfloor}(v_1, \dots, v_r; I_\varepsilon, \dots, I_\varepsilon)$.

Fix such an h . Now $B_{\lfloor N/k! \rfloor}(\phi, v_1, \dots, v_r; J, I_\varepsilon, \dots, I_\varepsilon)$ contains all but ηN of the points in $B_{\lfloor N/k! \rfloor}(v_1, \dots, v_r; I_\varepsilon, \dots, I_\varepsilon)$, and each element of $B_{\lfloor N/k! \rfloor}(v_1, \dots, v_r; I_\varepsilon, \dots, I_\varepsilon)$ can belong to at most $k+1$ progressions $n, n+h, \dots, n+kh$, and so $B_{\lfloor N/k! \rfloor}(\phi, v_1, \dots, v_r; J, I_\varepsilon, \dots, I_\varepsilon)$ contains at least all but $(k+1)\eta N$ of the progressions $n, n+h, \dots, n+kh$ in $B_{\lfloor N/k! \rfloor}(v_1, \dots, v_r; I_\varepsilon, \dots, I_\varepsilon)$.

In particular, if we fix $\eta < a/(k+1)$ then the set $B_{\lfloor N/k \rfloor}(\phi, v_1, \dots, v_r; J, I_\varepsilon, \dots, I_\varepsilon)$ contains at least one such progression. The value of h was chosen arbitrarily from a set of cardinality $\Omega_{k,r,\varepsilon}(N)$, and so the proposition is proved. \square

Proof of (22) in the case $k = 4$. The function ϕ_2 is of the form required in order to apply Proposition 7.4, and so the bracket components of the function ϕ_3 are weakly recurrent (modulo 1). The case $k = 4$ then follows from Remark 7.3. \square

Remark 7.8. *More generally, and by an identical proof, the function*

$$f : n \mapsto e \left(\gamma n^t \left\{ \beta n^r \prod_{i=1}^m \{ \alpha_i n \} \right\}^s \right)$$

satisfies $\|f\|_{U^{s(r+m)+t+1}[N]} \gg_{m,r,s,t} 1$.

8. APPROXIMATELY LOCALLY POLYNOMIAL FUNCTIONS

We can push slightly further than Section 7 and prove the $k = 5$ case of (22) by relaxing the definition of being locally polynomial of degree $k - 1$. The most obvious modification is to require the k th derivatives to vanish only modulo 1, since it is only their value modulo 1 that will affect the quantity $e(\Delta_{h_1, \dots, h_k} \phi(n))$. Indeed, as was remarked in the introduction, we have been considering bracket polynomials as functions into \mathbb{R} , rather than into \mathbb{R}/\mathbb{Z} , only because it made some of the proofs cleaner in earlier sections.

Another natural way in which it is possible to weaken the definition is not even to require the derivatives to vanish (modulo 1), but instead to require that $\|\Delta_{h_1, \dots, h_k} \phi(n)\|_{\mathbb{R}/\mathbb{Z}} < \delta$ for some $\delta \ll 1$, as this would still be sufficient to introduce some bias into the sum $\mathbb{E}_{n \in [N], h \in [-N, N]^k} e(\Delta_{h_1, \dots, h_k} \phi(n))$.

Definition 8.1 (Approximately locally polynomial (modulo 1)). *Let $\phi : [N] \rightarrow \mathbb{R}$ be a function and let $B \subset [N]$. Then ϕ is said to be δ -approximately locally polynomial of degree $k - 1$ (modulo 1) on B if whenever $n \in [N]$ and $h \in [-N, N]^k$ satisfy $n + \omega \cdot h \in B$ for all $\omega \in \{0, 1\}^k$ we have*

$$(29) \quad \|\Delta_{h_1, \dots, h_k} \phi(n)\|_{\mathbb{R}/\mathbb{Z}} \leq \delta.$$

If (29) holds whenever $n + \omega \cdot h \in B$ for all $\omega \in \{0, 1\}^k \setminus \{\mathbf{0}\}$ then ϕ is said to be strongly δ -approximately locally polynomial of degree at most $k - 1$ (modulo 1) on B .

The utility of making this definition lies in the following result.

Proposition 8.2. *Let $\phi : [N] \rightarrow \mathbb{R}$ be a function and define $f : [N] \rightarrow \mathbb{C}$ by $f(n) := e(\phi(n))$. Let $\delta \in [0, 1/4)$ be a parameter and suppose that ϕ is strongly δ -approximately locally polynomial of degree at most k (modulo 1) on some set $B \subset [N]$ with $|B| \gg N$. Then $\|f\|_{U^{k+1}[N]} \gg_{k,\delta} 1$.*

Proof. The definition of being strongly δ -approximately locally polynomial of degree k (modulo 1) on B implies that $\text{Re}(e(\Delta_{h_1, \dots, h_{k+1}} \phi(n))) \gg_\delta 1$ whenever $n + \omega \cdot h \in B$ for each $\omega \in \{0, 1\}^{k+1} \setminus \{\mathbf{0}\}$, and so

$$\begin{aligned} & \left| \mathbb{E}_{n \in [N], h \in [-N, N]^{k+1}} (e(\Delta_{h_1, \dots, h_{k+1}} \phi(n)) \prod_{\omega \in \{0, 1\}^{k+1} \setminus \{\mathbf{0}\}} 1_B(n + \omega \cdot h)) \right| \\ & \gg_\delta \mathbb{P}_{n \in [N], h \in [-N, N]^{k+1}} (n + \omega \cdot h \in B \text{ for all } \omega \in \{0, 1\}^{k+1} \setminus \{\mathbf{0}\}). \end{aligned}$$

However, this last quantity is trivially at least $\mathbb{P}_{n \in [N], h \in [-N, N]^{k+1}}(n + \omega \cdot h \in B \text{ for all } \omega \in \{0, 1\}^{k+1})$, which by Lemma 4.1 is at least $\Omega_k(1)$. An application of Lemma 4.2 therefore completes the proof. \square

Lemma 8.3. *Suppose that $v : [N] \rightarrow \mathbb{R}$ is a bracket polynomial that is strongly locally polynomial of degree $k - 1$ on $A \subset [N]$, and define $\phi(n) := \lambda n \{v(n)\}$. Let $J \subset (-1/2, 1/2]$ be an interval with $|J| \leq 2^{-k}$. Then whenever $n + \omega \cdot h \in B_N(v; J) \cap A$ for all $\omega \in \{0, 1\}^{k+1} \setminus \{\mathbf{0}\}$ we have $\Delta_{h_1, \dots, h_{k+1}} \phi(n) \in \{q \lambda n : q \in \mathbb{Z}, |q| \ll_k 1\}$.*

Corollary 8.4. *Suppose that $v : [N] \rightarrow \mathbb{R}$ is a bracket polynomial that is strongly locally polynomial of degree $k - 1$ on $A \subset [N]$. Let $J \subset (-1/2, 1/2]$ be an interval with $|J| \leq 2^{-k}$, and let $\delta > 0$ be a parameter. Then the bracket polynomial $\phi : [N] \rightarrow \mathbb{R}$ defined by $\phi(n) := \lambda n \{v(n)\}$ is strongly $O_k(\delta)$ -approximately locally polynomial of degree k on $B_N(\lambda n, v; I_\delta, J) \cap A$.*

Proof of Lemma 8.3. Assume that

$$(30) \quad n + \omega \cdot h \in B_N(v; J) \cap A \text{ for all } \omega \in \{0, 1\}^{k+1} \setminus \{\mathbf{0}\}.$$

We have

$$(31) \quad \Delta_{h_1, \dots, h_{k+1}} \phi(n) = \sum_{\omega \in \{0, 1\}^{k+1}} (-1)^{k+1-|\omega|} \lambda(n + \omega \cdot h) \{v(n + \omega \cdot h)\}.$$

Splitting the right-hand side of (31), we see that $\Delta_{h_1, \dots, h_{k+1}} \phi(n)$ is equal to

$$(32) \quad \begin{aligned} & \lambda n \sum_{\omega \in \{0, 1\}^{k+1}} (-1)^{k+1-|\omega|} \{v(n + \omega \cdot h)\} \\ & + \sum_{i=1}^{k+1} \lambda h_i \sum_{\omega: \omega_i=1} (-1)^{k+1-|\omega|} \{v(n + \omega \cdot h)\}. \end{aligned}$$

However, the final sum of (32) is equal to $\Delta_{h_1, \dots, h_{i-1}, h_{i+1}, \dots, h_{k+1}} \{v\}(n + h_i)$, which vanishes because $n + \omega \cdot h \in B_N(v; J) \cap A$ for every ω with $\omega_i = 1$ and because $\{v\}$ is locally polynomial of degree $k - 1$ on $B_N(v; J) \cap A$ by Lemma 3.5 and the hypothesis that $|J| \leq 2^{-k}$. We therefore have

$$(33) \quad \Delta_{h_1, \dots, h_{k+1}} \phi(n) = \lambda n \sum_{\omega \in \{0, 1\}^{k+1}} (-1)^{k+1-|\omega|} \{v(n + \omega \cdot h)\} = \lambda n \Delta_{h_1, \dots, h_{k+1}} \{v\}(n).$$

Now n may not belong to $B_N(v; J) \cap A$, and so we cannot similarly conclude that

$$\Delta_{h_1, \dots, h_{k+1}} \{v\}(n) = 0.$$

However, by (30) and the assumption that v is strongly locally polynomial on A we can conclude that $\Delta_{h_1, \dots, h_{k+1}} v(n) = 0$, and it is clear that $\Delta_{h_1, \dots, h_{k+1}} \{v\}(n)$ and $\Delta_{h_1, \dots, h_{k+1}} v(n)$ differ by an integer and that $|\Delta_{h_1, \dots, h_{k+1}} \{v\}(n)| \ll_k 1$. Hence $\Delta_{h_1, \dots, h_r} \{v\}(n) \in \{q \in \mathbb{Z} : |q| \ll_k 1\}$, which combined with (33) yields the desired result. \square

Proof of (22) in the case $k = 5$. Recall that

$$\phi_{k-1}(n) = \alpha_{k-1} n \{ \alpha_{k-2} n \{ \dots \{ \alpha_1 n \} \dots \} \}.$$

By Proposition 7.4 there exist $\varepsilon \ll 1$ and $\varepsilon' \gg 1$ such that

$$|B_N(\phi_2, \phi_1, \alpha_4 n; I_{1/2-\varepsilon'}, I_\varepsilon, I_\varepsilon)| \gg N,$$

and so a similar argument to Proposition 7.2 implies that there is some interval J inside $(-1/2, 1/2]$ such that

$$|B_N(\phi_2, \phi_1, \alpha_4 n; J, I_\varepsilon, I_\varepsilon)| \gg N$$

and such that ϕ_3 is strongly locally polynomial of degree 3 on $B_N(\phi_2, \phi_1, \alpha_4 n; J, I_\varepsilon, I_\varepsilon)$.

By Corollary 8.4 there exists $\delta \gg 1$ such that if J' is an interval in $(-1/2, 1/2]$ of width δ then ϕ_4 is, say, $1/10$ -approximately strongly locally polynomial of degree 4 (modulo 1) on $B_N(\phi_3, \phi_2, \phi_1, \alpha_4 n; J', J, I_\varepsilon, I_\varepsilon)$. Applying the pigeonhole principle to the elements of $B_N(\phi_2, \phi_1, \alpha_4 n; J, I_\varepsilon, I_\varepsilon)$ we can obtain such an interval whilst ensuring that

$$|B_N(\phi_3, \phi_2, \phi_1, \alpha_4 n; J', J, I_\varepsilon, I_\varepsilon)| \gg N.$$

Proposition 8.2 then completes the proof of the theorem. \square

Remark 8.5. *An identical proof shows, more generally, that the function*

$$f : n \mapsto e \left(\lambda n \left\{ \gamma n^t \left\{ \beta n^r \prod_{i=1}^m \{\alpha_i n\} \right\}^s \right\} \right)$$

satisfies $\|f\|_{U^{s(r+m)+t+2}[N]} \gg_{m,r,s,t} 1$.

APPENDIX A. COORDINATES, METRICS AND EQUIDISTRIBUTION IN NILMANIFOLDS

The aim of this appendix is to prove Lemmas 6.16, 6.17 and 6.18. Throughout, where \mathcal{X} and \mathcal{X}' are Mal'cev bases for a nilmanifold G/Γ we write d for the metrics on G and G/Γ , and ψ for the coordinates, associated to \mathcal{X} ; we write d' and ψ' , respectively, for the metrics and coordinates associated to \mathcal{X}' .

As we remarked in Section 6, the lemmas we are about to prove essentially follow by combining various results from [7, Appendix A]. The notation of that work is identical to ours, and so the results we cite can be read directly from [7, Appendix A] without difficulty. We therefore refer to these results by number only, without restating them here.

We repeatedly use the observation, made in the proof of [7, Lemma A.15], that if G/Γ is a nilmanifold then for every $x, y \in G$ there is some $z \in \Gamma$ such that $d(x\Gamma, y\Gamma) = d(x, yz)$.

We begin by recalling and proving Lemma 6.16.

Lemma 6.16. *Let $M \geq 2$. Let G/Γ be an m -dimensional nilmanifold with an M -rational nested Mal'cev basis \mathcal{X} , and let d be the metric associated to \mathcal{X} . Let $\rho \leq 1$ and $x \in G$, and suppose that $g : [N] \rightarrow G$ is η -equidistributed in G/Γ . Then a proportion of at least*

$$\frac{\rho^m}{M^{O(m)}} - \frac{3\eta}{\rho}$$

of the points $(g(n)\Gamma)_{n \in [N]}$ lie in the ball $\{y\Gamma : d(y\Gamma, x\Gamma) \leq \rho\}$.

We start by bounding from below the measure of a metric ball in G/Γ . Here and throughout this appendix we write $B_\rho(x)$ for the ball $\{y\Gamma : d(y\Gamma, x\Gamma) \leq \rho\}$.

Lemma A.1. *Suppose $x \in G$ and let $\rho \in (0, 1)$ be a parameter. Then*

$$\mu(B_\rho(x)) \geq \frac{\rho^m}{M^{O(m)}}.$$

Proof. In this proof we appeal [7, Lemma A.14]. The reader may note that the hypothesis of that lemma includes the assumption that \mathcal{X} is adapted to some filtration of G . However, the only place this is used is in invoking [7, Lemma A.3], which assumes only the weaker property of being nested. We are therefore free to apply [7, Lemma A.14] in the context of Lemma 6.16.

Let $\delta \in (0, 1)$ be a parameter to be determined later. Set $B' = \{y \in G : |\psi(y) - \psi(x)| \leq \delta\}$. By [7, Lemma A.14] we may assume that $|\psi(x)| \leq 1$, and so [7, Lemma A.4] implies that there is an absolute constant C such that for every $y \in B'$ we have

$$d(y, x) \leq M^C |\psi(y) - \psi(x)| \leq M^C \delta.$$

Setting $\delta = \rho/M^C$ therefore implies that $B'\Gamma \subset B_\rho(x)$, and in particular that $\mu(B) \geq \mu(B'\Gamma)$. It is a straightforward exercise to verify that for sufficiently small δ we have

$$\mu(B'\Gamma) = \mu(B') = (2\delta)^m \geq \frac{\rho^m}{M^{Cm}},$$

and so the lemma is proved. \square

Proof of Lemma 6.16. Define a non-negative function $f : G/\Gamma \rightarrow \mathbb{R}$ by

$$f(y\Gamma) = \max\left\{0, 1 - \left(\frac{2}{\rho}\right) d(y\Gamma, B_{\rho/2}(x))\right\}.$$

Since f takes the value 1 on $B_{\rho/2}(x)$ we have

$$\int_{G/\Gamma} f \geq \mu(B_{\rho/2}(x)) \geq \frac{\rho^m}{M^{O(m)}}$$

by Lemma A.1. Observe also that f is Lipschitz with Lipschitz norm $1 + 2/\rho$, and so the η -equidistribution of g therefore implies that

$$\mathbb{E}_{n \in [N]} f(g(n)\Gamma) \geq \frac{\rho^m}{M^{O(m)}} - \eta \left(1 + \frac{2}{\rho}\right) \geq \frac{\rho^m}{M^{O(m)}} - \frac{3\eta}{\rho}.$$

The fact that f is bounded by 1 and supported on $B_\rho(x)$ therefore yields the desired result. \square

We now recall and prove Lemma 6.17.

Lemma 6.17. *Let $\rho \leq 1$ and σ be parameters. Let G/Γ be a nilmanifold with an M -rational nested Mal'cev basis \mathcal{X} , suppose that G' is a rational subgroup of G , and suppose that \mathcal{X}' is a nested Mal'cev basis for G'/Γ' in which each element is an M -rational combination of the elements of \mathcal{X} . Suppose that $\varepsilon \in G$ satisfies $d(\varepsilon, 1) \leq \sigma$, and that $\gamma \in \Gamma$. Finally, suppose that g is an element of G' such that $d'(g\Gamma', \Gamma') \leq \rho$. Then $d(\varepsilon g\gamma\Gamma, \Gamma) \leq M^{O(1)}\rho + \sigma$.*

Proof. The fact that $d'(g\Gamma', \Gamma') \leq \rho$ implies that there exists $z \in \Gamma'$ such that

$$(34) \quad d'(gz, 1) \leq \rho.$$

An application of [7, Lemma A.4] therefore implies that $|\psi'(gz)| \leq M^{O(1)}\rho$, and so (34) and [7, Lemma A.6] combine to give

$$(35) \quad d(gz, 1) \leq M^{O(1)}\rho.$$

The right-invariance of d (15) implies that $d(\varepsilon gz, 1) = d(\varepsilon, (gz)^{-1})$, and so the symmetry of d about the identity (16) and the triangle inequality imply that

$$(36) \quad d(\varepsilon gz, 1) \leq d(\varepsilon, 1) + d((gz)^{-1}, 1) = d(\varepsilon, 1) + d(gz, 1).$$

The left-hand side of (36) is equal to $d(\varepsilon g\gamma\Gamma, \Gamma)$, since $\gamma, z \in \Gamma$, whilst the right-hand side is at most $M^{O(1)}\rho + \sigma$ by (35) and the assumption on ε , and so the lemma is proved. \square

Finally, let us recall and prove Lemma 6.18.

Lemma 6.18. *Let $\mathcal{Y} = \{Y_1, \dots, Y_m\}$ be a nested Mal'cev basis for T_p^r/Z_p^r in which each element Y_i is equal to either an element X_j of the standard basis \mathcal{X} or its inverse $-X_j$. Then the nilmanifold coordinate map $\chi_{\mathcal{Y}}$ associated to \mathcal{Y} , and the metric d associated to the standard basis $\mathcal{X} = \{X_1, \dots, X_m\}$, satisfy*

$$|\chi_{\mathcal{Y}}(x)| \ll_{p,r} d(x\Gamma, \Gamma)$$

for every $x \in T_p^r$.

Proof. Let $c < 1$ be an absolute constant to be determined later. Since $|\chi_{\mathcal{Y}}(x)| \ll_{p,r} 1$ for every $x \in G$, it is sufficient to prove the lemma under the additional assumption that

$$(37) \quad d(x\Gamma, \Gamma) < c < 1.$$

Let z be an element of Γ satisfying

$$(38) \quad d(xz, 1) = d(x\Gamma, \Gamma).$$

The assumptions on \mathcal{Y} imply in particular that each element of \mathcal{Y} is a 1-rational combination of elements of the standard basis, and *vice versa*, and so [7, Lemma A.4] combines with (37) and (38) to imply that there is an absolute constant C such that

$$(39) \quad |\psi_{\mathcal{Y}}(xz)| \leq Cd(x\Gamma, \Gamma).$$

Setting $c = 1/2C$, condition (37) therefore implies that $|\psi_{\mathcal{Y}}(xz)| < 1/2$, which in particular implies that

$$\chi_{\mathcal{Y}}(x) = \psi_{\mathcal{Y}}(xz),$$

and so the lemma follows from (39). \square

APPENDIX B. BERGELSON AND LEIBMAN'S CHARACTERISATION OF BRACKET POLYNOMIALS

The purpose of this appendix is to sketch how Theorem 6.3 can be read out of the work of Bergelson and Leibman [1]. Let us begin, then, by recalling the statement of Theorem 6.3.

Theorem 6.3 (Bergelson–Leibman [1]). *Let $\Theta_1, \dots, \Theta_r$ be constant-free bracket forms. Then there exist $p \geq 1$, a constant-free polynomial form P on T_p^r , and a nested Mal'cev basis $\mathcal{Y} = \{Y_1, \dots, Y_m\}$ for T_p^r/Z_p^r such that each element Y_i of \mathcal{Y} is equal to either an element X_j of the standard basis \mathcal{X} or its inverse $-X_j$, and such that for every $i = 1, \dots, r$ we have $\{\pm\Theta_i\} = \chi_{\mathcal{Y}}(P)_{m-r+i}$.*

This essentially follows from [1, Proposition 6.9]. Indeed, it is shown in [1, §6.8] how, given a Mal'cev basis \mathcal{Y} of T_p , the nilmanifold coordinates $\chi_{\mathcal{Y}}(g)_i$ of a matrix $g \in T_p$ can be defined equivalently as formal bracket expressions in the entries of g ; [1, Proposition 6.9] then states that if \mathcal{A} is a commutative ring, and b is an arbitrary bracket expression in the elements of \mathcal{A} , then there is some T_p/Z_p with Mal'cev basis $\mathcal{Y} = \{Y_1, \dots, Y_m\}$, and some upper-triangular matrix with elements of \mathcal{A} as entries, such that $\{\pm b\} = \chi(A)_m$. To prove Theorem 6.3, therefore, we essentially just apply this result with \mathcal{A} as the ring of constant-free polynomial forms.

There are, however, some issues with this deduction.

- (1) In [1] fractional parts are taken to lie in $[0, 1)$, whereas in the present work they lie in $(-1/2, 1/2]$.

- (2) Whilst it is explicit in [1, Proposition 6.9] each element Y_i of the Mal'cev basis \mathcal{Y} is equal to either an element X_j of the standard basis \mathcal{X} or its inverse $-X_j$, it is not stated explicitly that the Y_i are ordered in such a way that \mathcal{Y} is nested.
- (3) Applying [1, Proposition 6.9] gives only a single bracket polynomial in terms of a polynomial mapping into T_p/Z_p , rather than an r -tuple of bracket polynomials in terms of a polynomial mapping into T_p^r/Z_p^r .

It is straightforward to check that the change in the range of the fractional part operation does not affect the truth of Theorem 6.3; in particular, the calculations in [1, §5.9] proceed in exactly the same way. Point (1) is therefore of no concern.

Point (2) is also of no concern, since the Mal'cev basis defined implicitly in [1, Proposition 6.9] is, in fact, nested. This is a consequence of the fact that the basis elements are taken in a *legal order* in the sense of [1, §5.7].²

Point (3) is straightforward to overcome. So far, for each $i = 1, \dots, r$ we have a nilmanifold T_{p_i}/Z_{p_i} of dimension m_i , say; a nested Mal'cev basis \mathcal{Y}_i for T_{p_i}/Z_{p_i} consisting of elements of the standard basis and their inverses; and a polynomial form P_i on T_{p_i} such that $\{\pm\Theta_i\} = \chi_{\mathcal{Y}_i}(P_i)_{m_i}$. We can define a nested Mal'cev basis \mathcal{Y}' for the direct product $T_{p_1} \otimes \dots \otimes T_{p_r}$ by simply taking the elements of \mathcal{Y}_1 in order, followed by the elements of \mathcal{Y}_2 in order, and so on up until we finally take the elements of \mathcal{Y}_r in order. Note that $\chi_{\mathcal{Y}'} = (\chi_{\mathcal{Y}_1}, \dots, \chi_{\mathcal{Y}_r})$, and so in particular we have $\{-\Theta_i\} = \chi_{\mathcal{Y}'}(P_i)_{m_1+\dots+m_i}$.

This leaves two further issues.

- (4) Theorem 6.3 requires the p_i all to be equal.
- (5) Theorem 6.3 requires that the Θ_i are expressed in terms of the *last* r coordinates of some polynomial form.

We resolve point (4) really only for convenience in the main body of the paper. The proof of Theorem 2.8 would proceed almost identically in the event that $T_{p_1} \otimes \dots \otimes T_{p_r}$ appeared in place of T_p^r in the conclusion of Theorem 6.3, but having T_p^r makes some of our notation slightly cleaner. In fact, if we were concerned with optimising the implied constant in the conclusion of Theorem 2.8 then it would be preferable to allow $T_{p_1} \otimes \dots \otimes T_{p_r}$ in place of T_p^r . However, we are not concerned with the exact bounds in Theorem 2.8, and so we prove Theorem 6.3 as stated.

In any case, it is not difficult to obtain T_p^r in place of $T_{p_1} \otimes \dots \otimes T_{p_r}$. The key observation is that if q is a multiple of p then there is an obvious embedding $\iota_{p,q} : T_p \hookrightarrow T_q$ such that the image $\iota_{p,q}(Z_p)$ is a subset of Z_q . For example, the Heisenberg group T_2 embeds into T_4 via the map $\iota_{2,4} : T_2 \hookrightarrow T_4$ defined by

$$\iota_{2,4} \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & x & 0 & z \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & y \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

²Mal'cev bases in [1] are, by definition, adapted to the lower central series [1, §1.2]. However, taking the basis elements in a legal order in the sense of [1, §5.7] does not guarantee that the resulting basis is adapted to the lower central series, as can be seen by considering the order defined in [1, §5.5] in the case $d = 4$. Being in a legal order does, however, guarantee that the basis is a nested Mal'cev basis in the sense we have defined in this paper.

and the subgroup $\iota_{2,4}(Z_2)$ is equal to

$$\begin{pmatrix} 1 & 0 & \mathbb{Z} & 0 & \mathbb{Z} \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \mathbb{Z} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Moreover, and crucially, if $\mathcal{W} = \{W_1, \dots, W_k\}$ is a nested Mal'cev basis for T_p/Z_p consisting entirely of elements of the standard basis and their inverses, then it is possible to choose a nested Mal'cev basis $\mathcal{W}^{(q)}$ for T_q/Z_q consisting entirely of elements of the standard basis and their inverses, and that includes the elements $W_i^{(q)} := \log \iota_{p,q}(\exp W_i)$ in the same order that they appear in \mathcal{W} .³ The upshot of this is that the non-zero coordinates of an element $\iota_{p,q}(g)$ with respect to $\mathcal{W}^{(q)}$ in T_q will be the same as the coordinates of g with respect to \mathcal{W} in T_p . This implies that if $z \in Z_p$ and gz belongs to the fundamental domain of T_p/Z_p then $\iota_{p,q}(gz)$ belongs to the fundamental domain of T_q/Z_q , and so the non-zero entries of $\chi_{\mathcal{W}^{(q)}}(\iota_{p,q}(g))$ are equal to the non-zero entries of $\chi_{\mathcal{W}}(g)$.

Set q as the lowest common multiple of the p_i , and write m for the dimension of the group T_q . Set $P'_i = \iota_{p_i,q}(P_i)$ and $P = (P'_1, \dots, P'_r)$. Define a basis \mathcal{Y} for T_q^r by taking the bases $\mathcal{Y}_i^{(q)}$ in order, but with the elements $Y_m^{(q)}, \dots, Y_r^{(q)}$ moved to the right so that they are now the last r elements of the basis. Note that these basis elements are central, and so this last operation affects neither the property of being a nested Mal'cev basis nor the corresponding coordinates, and resolves point (5) above. We then have

$$\{\pm\Theta_i\} = (\chi_{\mathcal{Y}}(P)_{rm-r+i}),$$

as required by Theorem 6.3.

APPENDIX C. BASIC PROPERTIES OF POLYNOMIAL MAPPINGS

The main purpose of this appendix is to prove Lemma 6.7, which we now recall.

Lemma 6.7. *Let $k, p, r \in \mathbb{N}$. Then there is a some $d \in \mathbb{N}$ depending only on k and p such that if ρ is an arbitrary polynomial mapping of degree at most k into T_p^r then the derivatives $\partial_{h_{d+1}} \dots \partial_{h_1} \rho$ are all trivial.*

The proof of Lemma 6.7 rests on the following basic properties of polynomial mappings into T_p^r .

Lemma C.1. *Let ρ, σ be polynomial mappings into T_p of degree at most k, k' , respectively. Then*

- (i) *the product mapping $\rho\sigma$ taking n to $\rho(n)\sigma(n)$ is a polynomial mapping of degree at most $k + k'$;*
- (ii) *the inverse mapping ρ^{-1} taking n to $\rho(n)^{-1}$ is a polynomial mapping of degree $O_{k,p}(1)$.*

³The basis $\mathcal{W}^{(q)}$ is not uniquely defined in this way. We simply choose a basis arbitrarily from all those nested Mal'cev bases consisting of elements of the standard basis and their inverses in which the elements $\log \iota_{p,q}(\exp W_i)$ appear in the desired order.

Proof. The first assertion is trivial. The second is also straightforward; we present the details for completeness.

We claim that each entry $(\rho^{-1}(n))_{ij}$ with $i < j \leq p + 1$ is a polynomial of degree at most $O_{k,i}(1)$, which is clearly sufficient to prove the second assertion. We prove this claim by induction on $p - i$; thus for any fixed i we may assume that the claim holds for all values of j , for all greater values of i .

By definition of ρ^{-1} , for $i < j \leq p + 1$ we have

$$\sum_{t=1}^p \rho(n)_{it} \rho^{-1}(n)_{tj} = 0,$$

but since the diagonal entries of each matrix $\rho(n)$ and $\rho(n)^{-1}$ are 1, and the below-diagonal entries are 0, this reduces to

$$\rho(n)_{ij} + \rho^{-1}(n)_{ij} + \sum_{t=i+1}^{j-1} \rho(n)_{it} \rho^{-1}(n)_{tj} = 0.$$

This implies that

$$\rho^{-1}(n)_{ij} = -\rho(n)_{ij} - \sum_{t=i+1}^{j-1} \rho(n)_{it} \rho^{-1}(n)_{tj},$$

which is, by induction, a polynomial of degree at most $O_{k,i}(1)$, as claimed. \square

Proof of Lemma 6.7. It clearly suffices to prove the lemma in the case $r = 1$.

Denote by $T_p(l)$ the subgroup of T_p consisting of those matrices whose non-diagonal entries at a distance at most l from the main diagonal are zero. Thus, for example, $T_p(0) = T_p$ and $T_p(p) = \{\text{Id}\}$. We claim that there is some $d \in \mathbb{N}$ depending only on k, l and p such that if ρ is an arbitrary polynomial mapping of degree at most k into T_p whose image lies in $T_p(l)$ then the derivatives $\partial_{h_{d+1}} \dots \partial_{h_1} \rho$ are all trivial. This is clearly sufficient to prove the lemma.

We prove this claim by induction on $p - l$; thus, for any fixed l , we may assume that the claim holds for all greater values of l .

The group operation of $T_p(l)$ restricted to the entries at a distance exactly $l + 1$ from the main diagonal is simply addition in each entry. Therefore, if ρ is a polynomial mapping of degree at most k into T_p whose image lies in $T_p(l)$, then every derivative $\partial_{h_{k+1}} \dots \partial_{h_1} \rho$ lies in $T_p(l + 1)$. Moreover, by Lemma C.1 its other entries are all polynomials of degree at most $O_{k,p}(1)$, and so $\partial_{h_{k+1}} \dots \partial_{h_1} \rho$ is a polynomial mapping of degree at most $O_{k,p}(1)$ whose image lies in $T_p(l + 1)$. The claim, and hence the lemma, therefore follows by induction. \square

ACKNOWLEDGEMENTS

For some of the time during which this work was carried out the author was supported by an EPSRC doctoral training grant, awarded by the Department of Pure Mathematics and Mathematical Statistics in Cambridge, and a Bye-Fellowship from Magdalene College, Cambridge.

It is a pleasure to thank Tim Gowers and Ben Green for helpful and stimulating conversations, and Emmanuel Breuillard, Tom Sanders and an anonymous referee for careful readings of and detailed comments on earlier versions of this paper.

REFERENCES

- [1] V. Bergelson. and A. Leibman *Distribution of values of bounded generalized polynomials*, Acta Math. 198(2) (2007), pp. 155–230.
- [2] L. J. Corwin and F. P. Greenleaf, *Representations of nilpotent Lie groups and their applications. Part 1: Basic theory and examples*. Cambridge Stud. Adv. Math., 18 (1990).
- [3] W. T. Gowers, *A new proof of Szemerédi’s theorem*, Geom. Funct. Anal., 11 (2011), pp. 465–588.
- [4] B. J. Green and T. C. Tao, *An inverse theorem for the Gowers U^3 -norm, with applications*, Proc. Edinburgh Math. Soc., 51(1) (2008) pp. 73–153.
- [5] B. J. Green and T. C. Tao, *Quadratic uniformity of the Möbius function*, Annales de l’Institut Fourier, 58(6) (2008), pp. 1863–1935.
- [6] B. J. Green and T. C. Tao, *Linear equations in primes*, Annals of Math., 171(3) (2010), pp. 1753–1850.
- [7] B. J. Green and T. C. Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Annals of Math., 175(2) (2012), pp. 465–540.
- [8] B. J. Green and T. C. Tao, *The Möbius function is strongly orthogonal to nilsequences*, Annals of Math., 175(2) (2012), pp. 541–566.
- [9] B. J. Green and T. C. Tao, *Yet another proof of Szemerédi’s theorem*, in *An irregular mind*, Bolyai Soc. Math. Stud., 21 (2010), pp. 335–342.
- [10] B. J. Green, T. C. Tao and T. Ziegler, *An inverse theorem for the Gowers $U^4[N]$ -norm*, Glasgow Math. J., 53 (2011), pp. 1–50.
- [11] B. J. Green, T. C. Tao and T. Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$ -norm*, to appear in Annals of Math.. arXiv:1009.3998.
- [12] A. Mal’cev, *On a class of homogeneous spaces*, Izvestiya Akad. Nauk SSSR, Ser Mat., 13 (1949), pp. 9–32.
- [13] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith., 27 (1975), pp. 299–345.
- [14] T. C. Tao and V. H. Vu, *Additive combinatorics*. Cambridge Stud. Adv. Math., 105 (2006).

CENTRE FOR MATHEMATICAL SCIENCES, UNIVERSITY OF CAMBRIDGE, WILBERFORCE ROAD, CAMBRIDGE CB3 0WB, UNITED KINGDOM

E-mail address: M.Tointon@dpmms.cam.ac.uk