Ergod. Th. & Dynam. Sys., (2024), **44**, 2621–2648

doi:10.1017/etds.2023.112

Bracket words along Hardy field sequences

JAKUB KONIECZNY †‡ and CLEMENS MÜLLNER §

† *Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, F-69622 Villeurbanne Cedex, France*

‡ *Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3QD, UK*

(e-mail: jakub.konieczny@gmail.com)

§ *Institut für Diskrete Mathematik und Geometrie, TU Wien, Wiedner Hauptstr. 8–10, 1040 Wien, Austria*

(e-mail: clemens.muellner@tuwien.ac.at)

(*Received* 5 *March* 2023 *and accepted in revised form* 22 *October* 2023)

Abstract. We study bracket words, which are a far-reaching generalization of Sturmian words, along Hardy field sequences, which are a far-reaching generalization of Piatetski-Shapiro sequences $\lfloor n^c \rfloor$. We show that sequences thus obtained are deterministic (that is, they have subexponential subword complexity) and satisfy Sarnak's conjecture.

Key words: generalized polynomial, Sturmian word, subword complexity, nilsequence, Möbius orthogonality

2020 Mathematics Subject Classification: 11B50, 11J54 (Primary); 11L20, 11N37, 37A17, 68R15 (Secondary)

Contents

1 Introduction 2621
2 Hardy fields 2624
3 Parametric generalized polynomials 2630
4 Nilmanifolds 2635
5 Möbius orthogonality 2639
Acknowledgements 2647
References 2647

1. *Introduction*

One of the key results in a recent paper [**DDM+22**] by Deshouillers *et al* states that for each $m \in \mathbb{N}$ and $c > 1$, the subword complexity of the sequence $(\lfloor n^c \rfloor \bmod m)_{n=0}^{\infty}$ grows at most polynomially, which in particular shows that this sequence is deterministic. The

philosophy behind this result is the following: if we take a regularly growing function $((\lfloor n^c \rfloor)_{n=0}^{\infty})$ and apply a very simple rule to it (taking the residue modulo m), then the resulting sequence is still quite simple (in this case it has polynomial subword complexity). In this paper we vastly generalize both main aspects of this result, that is, we replace $(\lfloor n^c \rfloor)_{n=0}^{\infty}$ with Hardy sequences and we replace taking the residue modulo m by applying a bracket word.

Sturmian words are among the simplest and most extensively studied classes of infinite words over a finite alphabet. One of their defining properties is extremely low subword complexity. Recall that the subword complexity of an infinite word $\mathbf{a} = (a(n))_{n=0}^{\infty}$ over a finite alphabet Σ is the function $p_{\mathbf{a}}$ which assigns to each integer N the number $p_{\mathbf{a}}(N)$ of words $w \in \Sigma^N$ which appear in $\mathbf{a}$. If there exists at least one value of N such that $p_{\mathbf{a}}(N) \leq N$ then a must be eventually periodic, in which case $p_{\mathbf{a}}$ is bounded. If $\mathbf{a}$ is a Sturmian word then $p_{\mathbf{a}}(N) = N + 1$ for all N, which in light of the remark above is the least subword complexity possible for a word that is not eventually periodic. In [AK23] Adamczewski and the first-named author studied a generalization of Sturmian words obtained by considering letter-to-letter codings of finite-valued generalized polynomials, which they dubbed *bracket words*. A *generalized polynomial* is an expression built from the usual polynomials using addition, multiplication and the integer part function. More precisely, generalized polynomials from $\mathbb{Z}$ to $\mathbb{R}$ are the smallest class of sequences that contain the usual polynomials, and such that if $g, h \colon \mathbb{Z} \to \mathbb{R}$ are generalized polynomials then so are $g + h$, $g \cdot h$ and $\lfloor g \rfloor$. A bracket word is an infinite word $\mathbf{a} = (a(n))_{n=0}^{\infty}$ over some finite alphabet Σ which takes the form $a(n) = \varphi(g(n))$ for some generalized polynomial g such that $g(\mathbb{Z})$ is finite and some map $\varphi \colon \mathbb{Z} \to \Sigma$. For instance, Sturmian words (up to letter-to-letter coding) take the form

$$a(n) = \lfloor \alpha(n+1) + \beta \rfloor - \lfloor \alpha n + \beta \rfloor$$

with $\alpha \in (0,1) \setminus \mathbb{Q}$ and $\beta \in (0,1)$ (possibly with the integer part $\lfloor \cdot \rfloor$ replaced by the ceiling $\lceil \cdot \rceil$), and hence are special cases of bracket words. One of the main results of [AK23] is a polynomial bound on subword complexity of bracket words: $p_{\mathbf{a}}(N) \ll N^C$ for a constant C (dependent on $\mathbf{a}$).

In [DDM+22], Deshouillers *et al* investigated synchronizing automatic sequences along Piatetski-Shapiro sequences $(\lfloor n^c \rfloor)_{n=0}^{\infty}$, where $c > 1$. A special case which plays a crucial role in the argument is when the synchronizing automatic sequence is periodic, in which case they obtained a polynomial bound on the subword complexity.

As a joint extension of the two lines of investigation discussed above, we investigate bracket words along Piatetski-Shapiro sequences. In fact, we can deal with a considerably larger class of Hardy field functions with polynomial growth, which in addition to n^c $(c > 1)$ include logarithmic-exponential expressions such as $\alpha n^c + \alpha' n^{c'}$ and $n^c \log^{c'} n$, as well as some more complicated expressions such as $\log(n!)$. Our first result is a bound on the subword complexity.

THEOREM A. *Let $\mathbf{a} = (a(n))_{n \in \mathbb{Z}}$ be a (two-sided) bracket word over the alphabet Σ and let $f \colon \mathbb{R}_+ \to \mathbb{R}$ be a Hardy field function with polynomial growth. Then the subword complexity of $(a(\lfloor f(n) \rfloor)_{n=0}^{\infty}$ is bounded by $\exp(O(N^{\delta}))$ for some $0 < \delta < 1$.*

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

The study of (special) automatic sequences along Piatetski-Shapiro sequences $\lfloor n^c \rfloor$ has a long history. We mention results by Mauduit and Rivat [**MR95**, **MR05**], Deshouillers, Drmota and Morgenbesser [**DDM12**], Spiegelhofer [**Spi15**, **Spi20**], and Spiegelhofer and the second-named author [**MS17**]. Interestingly, two very different situations can appear. On the one hand, the Thue–Morse sequence along Piatetski-Shapiro sequences (for $1 < c < 3/2$) is normal; in particular it has maximal subword complexity. On the other hand, synchronizing automatic sequences along Piatetski-Shapiro sequences are very far from normal; they have subexponential subword complexity. One natural generalization of automatic sequences are morphic sequences. These are letter-to-letter codings of fixed points of substitutions. A prominent morphic sequence is the Fibonacci word which is the fixed point of the substitution $0 \mapsto 01$, $1 \mapsto 0$. Moreover, this sequence is also a Sturmian word and many interesting morphic sequences are also Sturmian words (see, for example, [**KMPS18**]). Thus, we obtain as a very special case (one of) the first results for morphic sequences along Piatetski-Shapiro sequences.

It follows from Theorem A that the sequence $(a(\lfloor f(n) \rfloor))_{n=0}^{\infty}$ is deterministic, meaning that it has subexponential subword complexity. A conjecture of Sarnak [**Sar11**] asserts that each deterministic sequence should be orthogonal to the Möbius function, given by

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n \text{ is the product of } k \text{ distinct primes;} \\ 0 & \text{if } n \text{ is divisible by a square.} \end{cases}$$

This conjecture is wide open in general. However, it has been resolved in a number of special cases [**Bou13**, **BSZ13**, **DDM15**, **DK15**, **eAKL16**, **eALdlR14**, **FKPLM16**, **GT12a**, **Gre12**, **KPL15**, **LS15**, **MR10**, **MR15**, **Mül17**, **Pec18**, **Vee16**]; see also the recent survey articles [**DLMR**, **FKPL18**]. Of particular importance to the current paper is Möbius orthogonality for nilsequences [**GT12a**], which was recently strengthened to short intervals [**MSTT22**]. As we discuss later in this paper, this is closely connected to bracket words thanks to the work of Bergelson and Leibman [**BL07**]. Our second result is the Möbius orthogonality for bracket words along Hardy field functions.

THEOREM B. *Let* $\mathbf{a} = (a(n))_{n\in\mathbb{Z}}$ *be a (two-sided)* $\mathbb{R}$*-valued bracket word and let* $f : \mathbb{R}_+ \to \mathbb{R}$ *be a Hardy field function with polynomial growth. Then*

$$\frac{1}{N}\sum_{n=1}^{N} \mu(n)a(n) \to 0 \quad \text{as } N \to \infty. \tag{1}$$

Remark 1.1. We point out that using similar techniques, it is possible to obtain a slightly stronger result. Firstly, instead of the bracket word, we could work with a bounded generalized polynomial; in fact, each bounded generalized polynomial can be approximated in the supremum norm by finite-valued ones, which allows for a straightforward reduction. Secondly, since all of the key ingredients in the proof of Theorem B are quantitative, one can obtain an explicit rate of convergence to 0 in (1). We leave the details to the interested reader.

Theorem B is closely related to Möbius orthogonality for nilsequences, that is, sequences that can be obtained by evaluating a continuous function along an orbit of a

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

point in a nilsystem. The connection between generalized polynomials and nilsequences was established by Bergelson and Leibman [**BL07**], who showed that bounded generalized polynomials can be represented by evaluating a piecewise polynomial function along an orbit in a nilsystem (see Theorem 4.2 for details).

The fact that nilsequences are orthogonal to the Möbius function was established by Green and Tao [**GT12a**] as a part of their programme of understanding additive patterns in the primes. In fact, [**GT12a**] already contains an outline of the proof of Möbius orthogonality for bounded generalized polynomials, although some technical details are left out.

In order to obtain a result for a bracket word along a Hardy field function, we split the range of summation into intervals where the Hardy field function under consideration can be efficiently approximated by polynomials. We are then left with the task of establishing cancellation in each of these intervals. A key ingredient is Möbius orthogonality for nilsequences in short intervals, Theorem 5.3, recently established in [**MSTT22**]. The main technical difficulty of our argument lies in extending Theorem 5.3 to piecewise constant (and hence necessarily not continuous) functions with semialgebraic pieces, which we accomplish in §5.2.

1.1. *Plan of the paper.* In §2 we recall some basic definitions and results about Hardy fields. Moreover, we study Taylor polynomials of functions from a Hardy field generalizing the corresponding part in [**DDM$^+$22**]. This allows us to locally replace functions from a Hardy field with polynomials. Thus, we need to be able to work with polynomials with varying coefficients. To do so, we study in §3 parametric generalized polynomials, building on and refining results obtained in [**AK23**]. These tools allow us to prove Theorem A. In §4 we present some basics on nilmanifolds and discuss the connection to generalized polynomials. Then in §5 we recall a result on Möbius orthogonality for nilsequences in short intervals. This is the final result that we need to prove Theorem B. One naturally arising difficulty is to translate the result on Möbius orthogonality for smooth functions to piecewise polynomial functions.

1.2. *Notation.* We use $\mathbb{N} = \{1, 2, \dots\}$ to denote the set of positive integers and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For $N \in \mathbb{N}$, we let $[N] = \{0, 1, \dots, N-1\}$. For a non-empty finite set X and a map $f\colon X \to \mathbb{R}$, we use the symbol $\mathbb{E}$ borrowed from probability theory to denote the average $\mathbb{E}_{x\in X} f(x) = (1/|X|) \sum_{x\in X} f(x)$.

2. *Hardy fields*

In this section we discuss functions from a Hardy field which have polynomial growth. In particular, we study how the Taylor polynomial of f can be used to describe $\lfloor f(n) \rfloor$. Therefore, we first gather some basic results on Hardy fields. Then we discus the uniform distribution of polynomials modulo $\mathbb{Z}$. Finally, we study properties of Taylor polynomials and prove the main theorem of this section, namely Theorem 2.11.

2.1. *Preliminaries.* We start by gathering the basic facts and results on Hardy fields. For further discussion we refer, for example, to [**Bos94**, **Fra09**].

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

Let $\mathcal{B}$ be the collection of equivalence classes of real-valued functions defined on some half-line (c, ∞), where we identify two functions if they agree eventually. (The equivalence classes just defined are often called *germs of functions*. We choose instead to refer to elements of $\mathcal{B}$ as functions, with the understanding that all the operations defined and statements made for elements of $\mathcal{B}$ are considered only for sufficiently large values of $t \in \mathbb{R}$.) A *Hardy field* H is a subfield of the ring $(\mathcal{B}, +, \cdot)$ that is closed under differentiation, meaning that H is a subring of $\mathcal{B}$ such that for each $0 \neq f \in H$, the inverse $1/f$ exists and belongs to H, f is differentiable and $f' \in H$. We let $\mathcal{H}$ denote the union of all Hardy fields. If $f \in \mathcal{H}$ is defined on $[0, \infty)$ (one can always choose such a representative of f) we call the sequence $(f(n))_{n=0}^{\infty}$ a *Hardy sequence*.

We note that choosing different representatives of the same germ of a function f changes the number of subwords of length N of $a(\lfloor f(n) \rfloor)$ by at most an additive constant. As a consequence, the asymptotic behaviour of the subword complexity of $a(\lfloor f(n) \rfloor)$ depends only on the germ of f.

A *logarithmic-exponential function* is any real-valued function on a half-line (c, ∞) that can be constructed from the identity map $t \mapsto t$ using basic arithmetic operations $+, -, \times, \div$, the logarithmic and the exponential functions, and real constants. For example, $t^2 + 5t$, $t^{\sqrt{2}+\sqrt{3}}$, $e^{(\log t)^2}$ and $e^{\sqrt{\log t}}/\sqrt{t^2+1}$ are all logarithmic-exponential functions. Every logarithmic-exponential functions belongs to $\mathcal{H}$, and so do some other classical functions such as Γ, ζ and $t \mapsto \sin(1/t)$.

For real-valued functions f and g on (c, ∞) such that $g(t)$ is non-zero for sufficiently large t, we write $f(t) \prec g(t)$ if $\lim_{t\to\infty} f(t)/g(t) = 0$, $f(t) \sim g(t)$ if $\lim_{t\to\infty} f(t)/g(t)$ is a non-zero real number, and $f(t) \ll g(t)$ if there exists $C > 0$ such that $|f(t)| \leq C|g(t)|$ for all large t. For completeness, we let $0 \sim 0$ and $0 \ll 0$.

We state the following well-known facts as lemmas.

LEMMA 2.1. *Let $f \in \mathcal{H}$ be a function that is not eventually zero. Then f is eventually strictly positive or negative. If f is not eventually constant, then f is eventually strictly monotone.*

Proof. Since f is not eventually 0, there exists the inverse function $1/f$; in particular, $f(t) \neq 0$ for t large enough. Now, the first part follows from continuity of f. The second part follows directly from the first part by considering f'. □

LEMMA 2.2. *Let H be a Hardy field and let $f, g \in H$. Then one of the following holds: $f \prec g$, $f \sim g$ or $f \succ g$.*

Proof. If g is eventually zero, the situation is trivial, so assume that this is not the case. Since f/g is eventually monotone, the limit $\lim_{t\to\infty} |f(t)|/|g(t)| \in \mathbb{R} \cup \{\infty\}$ exists. If the limit is infinite then $f \succ g$. If the limit is zero then $f \prec g$. If the limit is finite and non-zero then $f \sim g$. □

Definition 2.3. We say that f has *polynomial growth* if there exists $n \in \mathbb{N}$ such that $f(t) \prec t^n$.

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

We will make use of the following estimates for the derivatives of functions with polynomial growth.

LEMMA 2.4. [**Fra09**, Lemma 2.1] *Let $f \in \mathcal{H}$ be a function with polynomial growth. Then at least one of the following statements holds.*

(i) $f(t) \prec t^{-n}$ *for all* $n \in \mathbb{N}$.
(ii) $f(t) \to c \neq 0$ *as* $t \to \infty$ *for some constant* c.
(iii) $f(t)/(t(\log t)^2) \prec f'(t) \ll f(t)/t$.

LEMMA 2.5. *Let $f \in \mathcal{H}$ be a function such that $f(t) \prec t^{-n}$ for all $n \in \mathbb{N}$. Then also $f^{(\ell)}(n) \prec t^{-n}$ for all $\ell, n \in \mathbb{N}$.*

Proof. Reasoning inductively, it is enough to consider the case where $\ell = 1$. Suppose, for the sake of contradiction, that $|f'(t)| \gg t^{-n}$ for some $n \in \mathbb{N}$. Since $f(t) \to 0$ as $t \to \infty$ and since f is eventually monotone, for sufficiently large t we have

$$|f(t)| = \int_t^\infty |f'(s)|\, ds \gg \int_t^\infty s^{-n}\, ds \gg t^{-n+1},$$

contradicting the assumption on f. □

LEMMA 2.6. *Let $f \in \mathcal{H}$ and assume that $f(t) \ll t^k$ for some $k \in \mathbb{Z}$. Then $f^{(\ell)}(t) \ll t^{k-\ell}$ for each $\ell \in \mathbb{N}$.*

Proof. Reasoning inductively, it is enough to consider the case where $\ell = 1$. We consider the three possibilities in Lemma 2.4. If $f(t) \prec t^{-n}$ for all $n \in \mathbb{N}$ then the claim is trivially true by Lemma 2.5. If $f'(t) \ll f(t)/t$ then $f'(t) \ll t^{k-1}$, as required. Finally, suppose that $f(t) \to c \neq 0$ as $n \to \infty$. Clearly, in this case $k \geq 0$. We may decompose $f(t) = \overline{f}(t) + c$, where $\overline{f}(t) = f(t) - c$ and $\overline{f}(t) \prec 1$. Repeating the reasoning with $\overline{f}$ in place of f, we conclude that $f'(t) = \overline{f}'(t) \ll t^{-1} \ll t^{k-1}$. □

Remark 2.7. For each $f \in \mathcal{H}$ and each logarithmic-exponential function g, there exists a Hardy field H such that $f, g \in H$ (see, for example, [**Bos94**]). Hence, it follows from Lemma 2.2 that for each $f \in \mathcal{H}$ there exists $k_0(f) \in \mathbb{Z} \cup \{-\infty, +\infty\}$ such that, for $k \in \mathbb{Z}$ we have: $f(t) \prec t^k$ if $k > k_0(f)$, $f(t) \succ t^k$ if $k < k_0(f)$, and, if $k_0(f)$ is finite, $f(t) \ll t^{k_0(f)}$. Lemma 2.6 implies that $k_0(f^{(\ell)}) \leq k_0(f) - \ell$ (with the convention that $\pm\infty - \ell = \pm\infty$).

2.2. *Uniform distribution of polynomials.* In this subsection we recall a result about the uniform distribution of polynomials modulo $\mathbb{Z}$ which we need for the next subsection about Taylor polynomials. It is well known that a polynomial distributes uniformly modulo $\mathbb{Z}$ if and only if at least one (non-constant) coefficient is irrational. The following proposition is a quantitative version of this statement.

First we need to specify the way we quantify how uniformly distributed a sequence $a(n)$ mod $\mathbb{Z}$ is. Let $(x_1, \ldots, x_N)$ be a finite sequence of real numbers. Its *discrepancy* is defined by

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

$$D_N(x_1, \dots, x_N) = \sup_{0\le\alpha\le\beta\le 1} \left| \frac{\#\{n \le N : \alpha \le \{x_n\} < \beta\}}{N} - (\beta - \alpha) \right|. \tag{2}$$

Thus, we have the necessary prerequisites to state the following proposition.

PROPOSITION 2.8. [DDM+22, Proposition 5.2] *Suppose that $g : \mathbb{Z} \to \mathbb{R}$ is a polynomial of degree d, which we write as*

$$g(n) = \beta_0 + n\beta_1 + \cdots + n^d\beta_d.$$

Furthermore, let $\delta \in (0, 1/2)$. Then either the discrepancy of $(g(n) \bmod \mathbb{Z})_{n\in[N]}$ is smaller than δ, or else there is an integer $1 \le \ell \ll \delta^{-O_d(1)}$ such that

$$\max_{1\le j\le d} N^j \|\ell\beta_j\| \ll \delta^{-O_d(1)}.$$

This proposition is a direct consequence of [GT12b, Proposition 4.3], who attribute this result to Weyl.

2.3. *Taylor expansions.* For any germ $f \in \mathcal{H}$ we consider a representative that is defined on $[1, \infty)$ and also call it f. Then, for any $x \in (1, \infty)$ and $\ell \in \mathbb{N}_0$, we can consider the length-ℓ Taylor expansion of f at the point x,

$$f(x + y) = P_{x,\ell}(y) + R_{x,\ell}(y), \tag{3}$$

$$P_{x,\ell}(y) := f(x) + yf'(x) + \cdots + \frac{y^{\ell-1}}{(\ell-1)!} f^{(\ell-1)}(x), \tag{4}$$

$$R_{x,\ell}(y) := \frac{y^\ell}{\ell!} f^{(\ell)}(x + \xi_\ell(N, h)) \quad \text{where } \xi_\ell(x, y) \in [0, y]. \tag{5}$$

PROPOSITION 2.9. *Let $k \in \mathbb{Z}$, $\ell \in \mathbb{N}_0$, and let $f \in \mathcal{H}$ be a function with $f(t) \ll t^k$. Then the error term $R_{x,\ell}(y)$ in the Taylor expansion (3)–(5) satisfies*

$$R_{x,\ell}(y) \ll y^\ell x^{k-\ell}$$

uniformly for all $x \ge 1$ and $0 \le y \le x$, where the implied constant only depends on f and ℓ.

Proof. Combining (5) and Lemma 2.6. we have

$$y^{-\ell} R_{x,\ell}(y) \ll \sup_{\xi\in[0,y]} f^{(\ell)}(x + \xi) \ll \sup_{\xi\in[0,y]} (x + \xi)^{k-\ell} = \begin{cases} x^{k-\ell} & \text{if } k < \ell, \\ (x + y)^{k-\ell} & \text{if } k \ge \ell. \end{cases}$$

Assuming that $x \ge y$, the two estimates are equivalent. □

LEMMA 2.10. *Let $k \in \mathbb{N}$ and let f be a k times continuously differentiable function defined on an open interval $I \subseteq \mathbb{R}$. Suppose that $f^{(k)}(t)$ has constant sign on I. Then f changes monotonicity on I at most $k - 1$ times.*

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

Proof. If $f^{(k)}(t)$ is constant zero for all $t \in I$, then f is a polynomial of degree at most $k-1$ and the statement is trivially true. Thus, we assume without loss of generality that $f^{(k)}(t) > 0$ for all $t \in I$. Let us assume for the sake of contradiction that f changes monotonicity at least k times. Thus, f' has at least k zeros in I. It follows from the mean value theorem that f'' has at least $k-1$ zeros in I. Inductively applying this reasoning shows that $f^{(k)}$ has at least one zero in I, giving the desired contradiction. □

THEOREM 2.11. *Let $k, \ell \in \mathbb{N}$ be integers with $k < \ell$, and let $f \in \mathcal{H}$ be a function satisfying $f(t) \ll t^k$, and let $P_{N,\ell}$ and $R_{N,\ell}$ be given by (3)–(5). Then there exists some $0 < \eta < 1$ (only depending on ℓ) such that for any $H \in \mathbb{N}$, the formula*

$$e_N(h) := \lfloor f(N+h) \rfloor - \lfloor P_{N,\ell}(h) \rfloor, \quad 0 \le h < H, \tag{6}$$

defines at most $\exp(O(H^\eta))$ different functions $e_N : [H] \to \mathbb{Z}$ for $N \in \mathbb{N}$. Moreover, for each N, at least one of the following statements holds.

(i) *N is small: $N = O(H^{(\ell+\eta)/(\ell-k)})$.*
(ii) *e_N is sparse: there are at most $O(H^\eta)$ values of $h \in [H]$ such that $e_N(h) \neq 0$.*
(iii) *e_N is structured: there exists a partition of $[H]$ into $O(H^\eta)$ arithmetic progressions with step $O(H^\eta)$ on which e_N is constant.*

(In the theorem above, the constants implicit in the $O(\cdot)$ notation are allowed to depend on k, ℓ and f.)

Proof. We define $\varepsilon = H^{-\eta_0}$ for some $\eta_0 > 0$ which only depends on ℓ and will be specified later. Let $N \in \mathbb{N}$. Recall that by Proposition 2.9, we have

$$|R_{N,\ell}(h)| \le \varepsilon \quad \text{for all } 0 \le h < H \tag{7}$$

unless $N \ll \varepsilon^{-1/(\ell-k)} H^{\ell/(\ell-k)} = H^{(\ell+\eta_0)/(\ell-k)}$. Thus, the values of N such that (7) is false contribute only $O(H^{O(1)})$ different sequences e_N, and we may freely assume that N is large enough that (7) holds. In this case we have $e_N : [H] \to \{-1, 0, 1\}$. Additionally, by Lemma 2.1 we may also assume that $f^{(\ell)}(x) \neq 0$ for all $x \ge N$. As a consequence of (7), for each $0 \le h < H$, if

$$\varepsilon < \{P_{N,\ell}(h)\} < 1 - \varepsilon \tag{8}$$

then $\lfloor f(N+h) \rfloor = \lfloor P_{N,\ell}(h) \rfloor$ and hence $e_N(h) = 0$.

Let $\alpha_0, \dots, \alpha_{\ell-1}$ denote the coefficients of $P_{N,\ell}$:

$$P_{N,\ell}(h) = \alpha_0 + \alpha_1 h + \cdots + \alpha_{\ell-1} h^{\ell-1}.$$

By Proposition 2.8, we distinguish two cases.

(I) $(P_{N,\ell}(h))_{h \in [H]}$ has discrepancy at most ε.
(II) There exists $1 \le q \ll \varepsilon^{-O(1)}$ such that $\max_{0 \le j < \ell} H^j \|q\alpha_j\| \ll \varepsilon^{-O(1)}$.

In the first case, it follows that the number of $h \in [H]$ such that (8) does not hold is at most $3\varepsilon H$. Thus, e_N is sparse, that is, it has at most $3\varepsilon H \ll H^{1-\eta_0}$ non-zero entries.

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

It remains to estimate the number of the sequences e_N of this type. Using a standard estimate $\binom{n}{k} \le n^k/k! < (en)^k/k^k$, we find

$$\begin{aligned}\log\left(\sum_{0\le j\le 3\varepsilon H}\binom{H}{j}2^j\right) &\ll \log(3\varepsilon H) + \log\binom{H}{3\varepsilon H} + 3\varepsilon H\\ &\ll \log(3H^{1-\eta_0}) + 3\varepsilon H\log(e3H^{1-\eta_0}) + 3H^{1-\eta_0}\\ &\ll_{\eta_0} H^{1-\eta_0/2}.\end{aligned}$$

Thus the number of distinct sequences e_N is bounded by $\exp(O(H^{1-\eta_0/2}))$, which gives the desired result as long as $1-\eta_0/2 \le \eta$.

In the second case we split $[H]$ into arithmetic progressions with common difference $q \ll \varepsilon^{-O_\ell(1)}$. This allows us to write (for $0 \le m < q$)

$$\begin{aligned}P_{N,\ell}(qh+m) &= \alpha_0 + (qh+m)\alpha_1 + \cdots + (qh+m)^{\ell-1}\alpha_{\ell-1}\\ &= \beta_0 + h\beta_1 + \cdots + h^{\ell-1}\beta_{\ell-1}.\end{aligned}$$

The defining property of q implies that

$$\max_{1\le j<\ell} H^j\|\beta_j\| \ll \varepsilon^{-O_\ell(1)}.$$

In particular, we can write

$$\beta_j = z_j + s_j,$$

where $z_j \in \mathbb{Z}$ and $|s_j| \ll H^{-j}\cdot\varepsilon^{-O_\ell(1)}$ for $0 \le j < \ell$. Putting everything together, we find

$$f(N+qh+m) = Q(h) + r(h) + R_{N,\ell}(qh+m),$$

where

$$\begin{aligned}Q(h) &= z_0 + hz_1 + \cdots + h^{\ell-1}z_{\ell-1},\\ r(h) &= s_0 + hs_1 + \cdots + h^{\ell-1}s_{\ell-1}.\end{aligned}$$

In particular, Q is a polynomial of degree at most $\ell-1$ with integer coefficients and $P_{N,\ell}(qh+m) = Q(h) + r(h)$. Moreover, $|r(h)| \ll \varepsilon^{-O_\ell(1)}$ for all $h \in [0, H/q]$. Since $|R_{N,\ell}(h)| \le \varepsilon$, we see that

$$\lfloor f(N+qh+m)\rfloor \ne \lfloor P_{N,\ell}(qh+m)\rfloor$$

holds exactly if either

$$\begin{aligned}\{r(h)\} \le \varepsilon \quad &\text{and} \quad \{r(h)+R_{N,\ell}(qh+m)\} \ge 1-\varepsilon, \quad \text{or}\\ \{r(h)\} \ge 1-\varepsilon \quad &\text{and} \quad \{r(h)+R_{N,\ell}(qh+m)\} \le \varepsilon.\end{aligned} \tag{9}$$

In the first case $e_N(qh+m) = 1$ and in the second case $e_N(qh+m) = -1$. Since $r(h)$ is a polynomial of degree at most $\ell-1$, it changes monotonicity at most $\ell-2$ times. Since the

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

ℓth derivative of $r(h) + R_{N,\ell}(qh+m) = f(N+qh+m) - P_{N,\ell}(qh+m) + r(h)$ has constant sign, by Lemma 2.10 it changes monotonicity at most $\ell - 1$ times on the interval $[0, H/q]$. Hence, we can decompose $[0, H/q]$ into at most $2\ell - 2$ intervals $I_1, \dots, I_p$ on which $r(h)$ and $r(h) + R_{N,\ell}(qh+m)$ are both monotone. As $|r(h)| \ll \varepsilon^{-O_\ell(1)}$, we can further subdivide each of the intervals I_j into $O(\varepsilon^{-O_\ell(1)})$ subintervals such that for each subinterval, each of the inequalities is either true on the entire subinterval or false on the entire subinterval. As a consequence, e_N is structured, that is, e_N is constant on each subinterval. Thus, we have found a decomposition of $[H]$ into $O(\varepsilon^{-O_\ell(1)})$ arithmetic progressions on which e_N is constant. We can write $O(\varepsilon^{-O_\ell(1)}) = O(H^{C\eta_0})$ for some $C = C(\ell) > 0$. Using the rough estimate H^3 for the number of arithmetic sequences contained in $[H]$, we can bound the number of sequences e_N which arise this way by

$$(H^3)^{O(H^{C\eta_0})} = \exp(O(H^{C\eta_0} \log H)) = \exp(O_{\eta_0}(H^{(C+1)\eta_0}).$$

It remains to choose $\eta_0 = (C+2)^{-1}$ and $\eta = 1 - (2(C+2))^{-1}$ to finish the proof. □

3. *Parametric generalized polynomials*

In this section we discuss parametric generalized polynomials, building on and refining results obtained in [**AK23**]. In particular, we show that for any parametrized general polynomial that takes values in $[M]$, we can assume that the parameters belong to $[0, 1)^J$ for some finite set J (Proposition 3.5). This allows us to show a polynomial bound on the number of subwords of bracket words along polynomials of a fixed degree (Corollary 3.7). At the end of the section we give the proof of Theorem A.

Let $d \in \mathbb{N}$. Generalized polynomial (GP) maps from $\mathbb{R}^d$ to $\mathbb{R}$ are the smallest family $\mathcal{G}$ such that (1) all polynomial maps belong to $\mathcal{G}$; (2) if $g, h \in \mathcal{G}$ then also $g + h, g \cdot h \in \mathcal{G}$ (with operations defined pointwise); (3) if $g \in \mathcal{G}$ then also $\lfloor g \rfloor \in \mathcal{G}$, where $\lfloor g \rfloor$ is defined pointwise by $\lfloor g \rfloor(x) = \lfloor g(x) \rfloor$. We note that GP maps are also closed under the operation of taking the fractional part, given by $\{g\} = g - \lfloor g \rfloor$. For sets $\Omega \subseteq \mathbb{R}^d$ and $\Sigma \subseteq \mathbb{R}$ (for example, $\Omega = \mathbb{Z}^d$, $\Sigma = \mathbb{Z}$), by a GP map $g\colon \Omega \to \Sigma$ we mean the restriction $\widetilde{g}|_\Omega$ to Ω of a GP map $\widetilde{g}\colon \mathbb{R}^d \to \mathbb{R}$ such that $\widetilde{g}(\Omega) \subseteq \Sigma$. We point out that, unlike in the case of polynomials, the lift $\widetilde{g}$ is not uniquely determined by g, unless $\Omega = \mathbb{R}^d$.

In [**AK23**], we introduced the notion of a *parametric GP map* $\mathbb{Z} \to \mathbb{R}$ with a finite index set I, which (modulo some notational conventions) is essentially the same as a GP map $\mathbb{R}^I \times \mathbb{Z} \to \mathbb{R}$. For instance, the formula

$$g_{\alpha,\beta}(n) = \lfloor \alpha n \lfloor \beta n \rfloor + \sqrt{2} n^2 \rfloor \quad (\alpha, \beta \in \mathbb{R})$$

defines a GP map $\mathbb{Z} \to \mathbb{R}$ (or, strictly speaking, a family of GP maps) parametrized by $\mathbb{R}^2$. Formally, a *parametric GP map with index set I* or a *GP map parametrized by $\mathbb{R}^I$* is a map $\mathbb{R}^I \to \mathbb{R}^{\mathbb{Z}}$, $\alpha \mapsto g_\alpha$, such that the combined map $\mathbb{R}^I \times \mathbb{Z} \to \mathbb{R}$, $(\alpha, n) \mapsto g_\alpha(n)$, is a GP map.

Here we will need a marginally more precise notion, where the set of parameters takes the form $\mathbb{R}^{I_{\text{real}}} \times \mathbb{Z}^{I_{\text{int}}} \times [0,1)^{I_{\text{frac}}}$ rather than $\mathbb{R}^I$. Let $I_{\text{real}}, I_{\text{int}}, I_{\text{frac}}$ be pairwise disjoint

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

finite sets and put $I = I_{\text{real}} \cup I_{\text{int}} \cup I_{\text{frac}}$. Then a *GP map parametrized by* $\mathbb{R}^{I_{\text{real}}} \times \mathbb{Z}^{I_{\text{int}}} \times [0,1)^{I_{\text{frac}}}$ is the restriction of a GP map parametrized by $\mathbb{R}^{I_{\text{real}}} \times \mathbb{R}^{I_{\text{int}}} \times \mathbb{R}^{I_{\text{frac}}}$ (as defined above) to $\mathbb{R}^{I_{\text{real}}} \times \mathbb{Z}^{I_{\text{int}}} \times [0,1)^{I_{\text{frac}}}$. We note that in the case where $I_{\text{int}} = I_{\text{frac}} = \emptyset$, the new definition is consistent with the previous one.

In [AK23] we defined the operations of addition, multiplication and the integer part for parametric GP maps, not necessarily indexed by the same set. Roughly speaking, if $I \subseteq J$ are finite sets then we can always think of a GP map parametrized by $\mathbb{R}^I$ as a GP map parametrized by $\mathbb{R}^J$, with trivial dependence on the parameters in $\mathbb{R}^{J \setminus I}$. Thus, if $g_\bullet$ and $h_\bullet$ are GP maps parametrized by $\mathbb{R}^I$ and $\mathbb{R}^J$ respectively, then we can think of both $g_\bullet$ and $h_\bullet$ as GP maps parametrized by $\mathbb{R}^{I \cup J}$, which gives us a natural way to define the (pointwise) sum and product $g_\bullet + h_\bullet$ and $g_\bullet \cdot h_\bullet$. We refer to [AK23] for a formal definition. This construction directly extends to GP maps parametrized by $\mathbb{R}^{I_{\text{real}}} \times \mathbb{Z}^{I_{\text{int}}} \times [0,1)^{I_{\text{frac}}}$.

Definition 3.1. Let $g_\bullet$ and $h_\bullet$ be two GP maps parametrized by $\mathbb{R}^{I_{\text{real}}} \times \mathbb{Z}^{I_{\text{int}}} \times [0,1)^{I_{\text{frac}}}$ and $\mathbb{R}^{J_{\text{real}}} \times \mathbb{Z}^{J_{\text{int}}} \times [0,1)^{J_{\text{frac}}}$, respectively. Then we say that $h_\bullet$ *extends* $g_\bullet$, denoted $h_\bullet \succeq g_\bullet$, if there exists a GP map $\varphi\colon \mathbb{R}^{I_{\text{real}}} \times \mathbb{R}^{I_{\text{int}}} \times \mathbb{R}^{I_{\text{frac}}} \to \mathbb{R}^{J_{\text{real}}} \times \mathbb{R}^{J_{\text{int}}} \times \mathbb{R}^{J_{\text{frac}}}$ such that

- $\varphi(\mathbb{R}^{I_{\text{real}}} \times \mathbb{Z}^{I_{\text{int}}} \times [0,1)^{I_{\text{frac}}}) \supset \mathbb{R}^{I_{\text{real}}} \times \mathbb{Z}^{I_{\text{int}}} \times [0,1)^{I_{\text{frac}}}$, and
- $g_\alpha = h_{\varphi(\alpha)}$ for all $\alpha \in \mathbb{R}^{I_{\text{real}}} \times \mathbb{Z}^{I_{\text{int}}} \times [0,1)^{I_{\text{frac}}}$.

We use different notation $h_\bullet \succeq g_\bullet$ than in [AK23] in order to avoid confusion with the symbol $\succ$ extensively used in §2. In [AK23] we obtained a polynomial bound on the number of possible prefixes of a given GP map parametrized by $[0,1)^I$.

THEOREM 3.2. [AK23, Theorem 15.3] *Let $g_\bullet : \mathbb{Z} \to \mathbb{Z}$ be a GP map parametrized by $[0,1)^I$ for some finite set I. Then there exists a constant C such that, as $N \to \infty$, we have*

$$|\{g_\alpha|_{[N]} \mid \alpha \in [0,1)^I\}| = O(N^C). \tag{10}$$

Above, the implicit constant depends only on $g_\bullet$.

Our next goal is to obtain a similar bound for the number of prefixes of a bounded GP map parametrized by $\mathbb{R}^I$. Even though we are ultimately interested in bounded GP maps, Proposition 3.4 concerning unbounded GP maps is more amenable to proof by structural induction. We will use the following induction scheme.

PROPOSITION 3.3. [AK23, Proposition 13.9] *Let $\mathcal{G}$ be a family of parametric GP maps from $\mathbb{Z}$ to $\mathbb{Z}$ with index sets contained in $\mathbb{N}$. Suppose that $\mathcal{G}$ has the following closure properties.*

(i) *All GP maps $\mathbb{Z} \to \mathbb{Z}$ belong to $\mathcal{G}$.*
(ii) *For every $g_\bullet$ and $h_\bullet \in \mathcal{G}$, we have that $g_\bullet + h_\bullet \in \mathcal{G}$ and $g_\bullet \cdot h_\bullet \in \mathcal{G}$.*
(iii) *For every $g_\bullet \in \mathcal{G}$, $\mathcal{G}$ contains all the parametric GP maps $g'_\bullet\colon \mathbb{Z} \to \mathbb{Z}$ satisfying $g_\bullet \succeq g'_\bullet$.*

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

(iv) *For every pair of disjoint finite sets $I \subseteq \mathbb{N}$, $J \subseteq \mathbb{N}$, and every sequence of parametric GP maps $h_\bullet^{(i)} \in \mathcal{G}$, $i \in I$, with index set J, $\mathcal{G}$ contains the parametric GP map $g_\bullet$ defined by*

$$g_{\alpha,\beta}(n) = \left\lfloor \sum_{i\in I} \alpha_i h_\beta^{(i)}(n) \right\rfloor, \quad n \in \mathbb{Z},\ \alpha \in \mathbb{R}^I,\ \beta \in \mathbb{R}^J.$$

Then $\mathcal{G}$ contains all parametric GP maps $\mathbb{Z} \to \mathbb{Z}$ with index sets contained in $\mathbb{N}$.

PROPOSITION 3.4. *Let $g_\bullet\colon \mathbb{Z} \to \mathbb{Z}$ be a GP map parametrized by $\mathbb{R}^I$ for a finite set I. Then there exist finite sets J, K and a GP map $\widetilde{g}_\bullet\colon \mathbb{Z} \to \mathbb{Z}$ parametrized by $\mathbb{Z}^J \times [0,1)^K$ such that $\widetilde{g}_\bullet \succeq g_\bullet$ and $\widetilde{g}_\bullet$ takes the form*

$$\widetilde{g}_{a,\beta} = \sum_{j\in J} a_j h_\beta^{(j)}, \quad a \in \mathbb{Z}^J,\ \beta \in [0,1)^K,$$

where for each $j \in J$, $h_\bullet^{(j)}\colon \mathbb{Z} \to \mathbb{Z}$ is a GP map parametrized by $[0,1)^K$.

Proof. (i) If $g\colon \mathbb{Z} \to \mathbb{Z}$ is a fixed GP map (that is, if $I = \emptyset$) then we can simply take $\widetilde{g} = g$.

(ii) Suppose that the conclusion holds for $g_\bullet, h_\bullet\colon \mathbb{Z} \to \mathbb{Z}$, and let the corresponding extensions $\widetilde{g}_\bullet$ and $\widetilde{h}_\bullet$ be given by

$$\widetilde{g}_{a,\beta} = \sum_{j\in J} a_j h_\beta^{(j)}, \quad a \in \mathbb{Z}^J,\ \beta \in [0,1)^K,$$
$$\widetilde{h}_{c,\delta} = \sum_{l\in L} c_l h_\delta^{(l)}, \quad c \in \mathbb{Z}^L,\ \delta \in [0,1)^M.$$

We may freely assume that the index sets J, K, L, M are pairwise disjoint. We will show that the conclusion also holds for $g_\bullet + h_\bullet$ and $g_\bullet \cdot h_\bullet$. In the case of $g_\bullet + h_\bullet$ it is enough to combine the sums representing $\widetilde{g}_{a,\beta}$ and $\widetilde{h}_{c,\delta}$ into a single sum. In the case of $g_\bullet \cdot h_\bullet$, we take

$$\widetilde{f}_{e,(\beta,\delta)} = \sum_{j\in J,\ l\in L} e_{j,l}(h_\beta^{(j)} \cdot h_\delta^{(l)}), \quad e \in \mathbb{Z}^{J\times L},\ (\beta,\delta) \in [0,1)^{K\times M}.$$

Then $\widetilde{f}$ has the required form and (taking $e_{j,l} = a_j c_l$) we see that $\widetilde{f}_\bullet \succeq \widetilde{g}_\bullet \cdot \widetilde{h}_\bullet \succeq g_\bullet \cdot h_\bullet$.

(iii) Suppose that the conclusion holds for $g_\bullet$ and that $g_\bullet \succeq g'_\bullet$. Then the conclusion also holds for $g'_\bullet$ because the relation of being an extension is transitive.

(iv) Suppose that $I \subseteq \mathbb{N}$, $J \subseteq \mathbb{N}$ are disjoint finite sets, $h_\bullet^{(i)}$ are GP maps parametrized by $\mathbb{R}^J$ which satisfy the conclusion for each for $i \in I$, and $g_\bullet$ is the parametric GP map defined by

$$g_{\alpha,\beta}(n) := \left\lfloor \sum_{i\in I} \alpha_i h_\beta^{(i)}(n) \right\rfloor, \quad n \in \mathbb{Z},\ \alpha \in \mathbb{R}^I,\ \beta \in \mathbb{R}^J.$$

Let the extensions of $h^{(i)}$ be given by

$$\widetilde{h}_{c,\delta}^{(i)} = \sum_{l\in L} c_l f_\delta^{(i,l)}, \quad c \in \mathbb{Z}^L,\ \delta \in [0,1)^M.$$

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

(Note that without loss of generality we may use the same index sets L and M for each $i \in I$.) We will show that the conclusion is satisfied for $g_\bullet$. We observe that we have the equality

$$\left\lfloor \sum_{i\in I} \alpha_i \widetilde{h}^{(i)}_{c,\delta} \right\rfloor = \left\lfloor \sum_{i\in I,\, l\in L} \alpha_i c_l f^{(i,l)}_\delta \right\rfloor = \sum_{i\in I,\, l\in L} \lfloor \alpha_i c_l \rfloor f^{(i,l)}_\delta + \left\lfloor \sum_{i\in I,\, l\in L} \{\alpha_i c_l\} f^{(i,l)}_\delta \right\rfloor.$$

This motivates us to define

$$\widetilde{g}_{e,\delta,\phi} := \sum_{i\in I,\, l\in L} e_{i,l} f^{(i,l)}_\delta + e_\diamond \left\lfloor \sum_{i\in I,\, l\in L} \phi_{i,l} f^{(i,l)}_\delta \right\rfloor,$$
$$e \in \mathbb{Z}^{I\times L\cup\{\diamond\}},\ \phi \in [0,1)^{I\times L},\ \delta \in [0,1)^M,$$

where $\diamond$ is some index that does not belong to $I \times J$. Letting also

$$f^{(\diamond)}_{\delta,\phi} := \left\lfloor \sum_{i\in I,\, l\in L} \phi_{i,l} f^{(i,l)}_\delta \right\rfloor, \quad \phi \in [0,1)^{I\times L},\ \delta \in [0,1)^M,$$

we see that $\widetilde{g}_\bullet$ takes the required form and (setting $\phi_{i,l} = \{\alpha_i c_l\}$ and $e_\diamond = 1$) we have $\widetilde{g}_\bullet \succeq g_\bullet$.

Combining the closure properties proved above, we infer from Proposition 3.3 that the conclusion holds for all parametric GP maps. □

PROPOSITION 3.5. *Let $M \in \mathbb{N}$ and let $g_\bullet\colon \mathbb{Z} \to [M]$ be a GP map parametrized by $\mathbb{R}^I$ for a finite set I. Then there exists a GP map $\widetilde{g}_\bullet\colon \mathbb{Z} \to [M]$ parametrized by $[0,1)^J$ for a finite set J such that $\widetilde{g}_\bullet \succeq g_\bullet$.*

Proof. Let $\widetilde{g}^{(0)}_\bullet \succeq g_\bullet$ be the parametric GP from Proposition 3.4, and let

$$\widetilde{g}^{(0)}_{a,\beta} = \sum_{j\in J} a_j h^{(j)}_\beta, \quad a \in \mathbb{Z}^J,\ \beta \in [0,1)^K.$$

Since the value of $g_{\alpha,\beta}(n)$ is completely determined by its residue modulo M, we expect that it is enough to consider the values of a with $a \in [M]^J$. This motivates us to put

$$\widetilde{g}_{\alpha,\beta} = \sum_{j\in J} \lfloor M\alpha_j \rfloor h^{(j)}_\beta, \quad \alpha \in [0,1)^J,\ \beta \in [0,1)^K.$$

Let $\phi\colon \mathbb{Z}^I \to \mathbb{Z}^J$ and $\psi\colon \mathbb{Z}^I \to \mathbb{R}^K$ be GP maps such that $g_\alpha = \widetilde{g}^{(0)}_{\phi(\alpha),\psi(\alpha)}$. Let $\theta\colon \mathbb{Z}^I \to [0,1)^J$ be given by $\theta(\alpha) := \{\phi(\alpha)/M\}$ (with fractional part taken coordinatewise). Then

$$\widetilde{g}^{(0)}_{\phi(\alpha),\beta}(n) \equiv \widetilde{g}_{\theta(\alpha),\beta}(n) \bmod M \quad \text{for all } n \in \mathbb{Z},\ \alpha \in \mathbb{R}^I,\ \beta \in [0,1)^K.$$

Since $g_\bullet$ takes values in $[M]$, it follows that

$$g_\alpha(n) = \widetilde{g}^{(0)}_{\phi(\alpha),\psi(\alpha)}(n) \equiv \widetilde{g}_{\theta(\alpha),\psi(\alpha)}(n) \bmod M \quad \text{for all } n \in \mathbb{Z},\ \alpha \in \mathbb{R}^I.$$

Replacing $\widetilde{g}_\bullet$ with $M \cdot \{\widetilde{g}_\bullet/M\}$ if necessary, we may further ensure that $\widetilde{g}_\bullet$ takes values in $[M]$. As a consequence, $\widetilde{g}_\bullet \succeq g_\bullet$, as needed. □

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

PROPOSITION 3.6. *Let $\mathbf{a} = (a(n))_{n\in\mathbb{Z}}$ be a (two-sided) bracket word over a finite alphabet Σ, and let $g_\bullet\colon \mathbb{Z} \to \mathbb{Z}$ be a GP map parametrized by $\mathbb{R}^I$ for some finite set I. Then there exists a constant $C > 0$ such that, as $N \to \infty$, we have*

$$|\{(a(g_\alpha(n)))_{n=0}^{N-1} \mid \alpha \in \mathbb{R}^I\}| = O(N^C).$$

Above, the implicit constant depends on $\mathbf{a}$ *and* $g_\bullet$.

Proof. Let $M := |\Sigma|$. We may freely assume that $\Sigma = [M]$, in which case a is a GP map by [**AK23**, Lemma 5.2]. Thus, $a \circ g_\bullet$ is a GP map parametrized by $\mathbb{R}^I$ and taking values in $[M]$. By Proposition 3.5, there exists a GP map $\widetilde{g}_\bullet$ parametrized by $[0, 1)^J$ for a finite set J such that $\widetilde{g}_\bullet \succeq a \circ g_\bullet$. Thus, it suffices to show that, for a certain $C > 0$, the number of words $(\widetilde{g}_\alpha(n))_{n=0}^{N-1}$ for $\alpha \in [0, 1)^J$ is $O(N^C)$ as $N \to \infty$. This is precisely Theorem 3.2. □

As a special case, we obtain a bound on the number of subsequences of bracket words along polynomials of a given degree.

COROLLARY 3.7. *Let $\mathbf{a} = (a(n))_{n\in\mathbb{Z}}$ be a (two-sided) bracket word over a finite alphabet Σ and let $d \in \mathbb{N}$. Then there exists a constant $C > 0$ such that, as $N \to \infty$, we have*

$$|\{(a(\lfloor p(n)\rfloor))_{n=0}^{N-1} \mid p \in \mathbb{R}_{\le d}[x]\}| = O(N^C),$$

where the implied constant depends only on $\mathbf{a}$ *and* d.

Thus we are now in a position to prove Theorem A.

Proof of Theorem A. We aim to estimate the number of subwords of length H of $(a(\lfloor f(n)\rfloor))_{n=0}^{\infty}$, that is, we count words of the form

$$(a(\lfloor f(N)\rfloor), \dots, a(\lfloor f(N + H - 1)\rfloor)) = (a(\lfloor f(N + h)\rfloor))_{h=0}^{H-1}$$

for $N \in \mathbb{N}$. Since f has polynomial growth, there exists $k \in \mathbb{N}$ such that $f(t) \ll t^k$. We choose $\ell \ge k + 1$ and apply Theorem 2.11 to find some $0 < \eta < 1$ such that for any $H \in \mathbb{N}$ at least one of statements (i)–(iii) in Theorem 2.11 holds, where

$$e_N(h) := \lfloor f(N + h)\rfloor - \lfloor P_{N,\ell}(h)\rfloor, \quad 0 \le h < H,$$

and $P_{N,\ell}$ is the Taylor polynomial of f (see (4)). We distinguish the three possible cases. Obviously (i) contributes at most $O(H^{\ell+1})$ different words. For (ii) we first consider $a(\lfloor P_{N,\ell}(h)\rfloor)_{h=0}^{H-1}$. By Corollary 3.7 this word is contained in a set of size $O(H^C)$. By assumption $a(\lfloor f(N + h)\rfloor) \ne a(\lfloor P_{N,\ell}(h)\rfloor)$ for at most $O(H^\eta)$ values of $h \in [H]$, which can be chosen in $\binom{H}{O(H^\eta)}$ ways For each position h with $a(\lfloor f(N + h)\rfloor) \ne a(\lfloor P_{N,\ell}(h)\rfloor)$ we have at most $|\Sigma|$ possibilities for the value of $a(\lfloor f(N + h)\rfloor)$. In total, we can estimate the number of subwords of length H in this case (up to a constant) by

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

$$H^C \cdot \binom{H}{O(H^\eta)} \cdot |\Sigma|^{O(H^\eta)} \le H^C \cdot H^{O(H^\eta)} \cdot |\Sigma|^{O(H^\eta)}$$
$$= \exp(C \log H + O((\log H) \cdot H^\eta) + O((\log |\Sigma|) \cdot H^\eta))$$
$$= \exp(O_{C,\eta}(H^{(1+\eta)/2})).$$

In the last case (iii) we decompose $[H]$ into $O(H^\eta)$ arithmetic progressions on which e_N is constant. We let these arithmetic progressions be denoted by $P_1, \dots, P_s$. As there are at most H^3 arithmetic progressions contained in $[H]$ we can bound the number of possible different decompositions by $(H^3)^{O(H^\eta)}$. On every such progression there exists a polynomial q (which is either $P_{N,\ell}$, $P_{N,\ell}+1$ or $P_{N,\ell}-1$) such that $a(\lfloor f(N+h)\rfloor) = a(\lfloor q(h)\rfloor)$. As a polynomial along an arithmetic progression is again a polynomial, by Corollary 3.7 we can bound the number of subwords appearing along some P_j by H^C. In total, we can estimate the number of subwords of length H in this case by

$$(H^3)^{O(H^\eta)} \cdot (H^C)^{O(H^\eta)} = \exp((C+3)\log(H) \cdot O(H^\eta))$$
$$= \exp(O_{C,\eta}(H^{(1+\eta)/2})).$$

This finishes the proof for $\delta = (1+\eta)/2 < 1$. □

4. *Nilmanifolds*

In this section we recall some basic definitions and results on nilmanifolds and discuss the connection to generalized polynomials which goes back to the work of Bergelson and Leibman [**BL07**].

4.1. *Basic definitions.* In this section we very briefly introduce definitions and basic facts related to nilmanifolds and nilpotent dynamics. Throughout this section, we let G denote an s-step nilpotent Lie group of some dimension D. We assume that G is connected and simply connected. We also let $\Gamma < G$ denote a subgroup that is discrete and cocompact, meaning that the quotient space G/Γ is compact. The space $X = G/\Gamma$ is called an *s-step nilmanifold*. A *degree- d filtration* on G is a sequence $G_\bullet$ of subgroups

$$G = G_0 = G_1 \ge G_2 \ge G_3 \ge \cdots$$

such that $G_{d+1} = \{e_G\}$ (and hence $G_i = \{e_G\}$ for all $i > d$) and for each i, j we have $[G_i, G_j] \subseteq G_{i+j}$, where $[G_i, G_j]$ is the group generated by the commutators $[g,h] = ghg^{-1}h^{-1}$ with $g \in G_i$, $h \in G_j$. A standard example of a filtration is the *lower central series* given by $G_{(0)} = G_{(1)} = G$ and $G_{(i+1)} = [G, G_{(i)}]$ for $i \ge 1$.

A *Mal'cev basis* compatible with Γ and $G_\bullet$ is a basis $\mathcal{X} = (X_1, X_2, \dots, X_D)$ of the Lie algebra $\mathfrak{g}$ of G such that:

(i) for each $0 \le j \le D$, the subspace $\mathfrak{h}_j := \operatorname{span}(X_{j+1}, X_{j+2}, \dots, X_D)$ is a Lie algebra ideal in $\mathfrak{g}$;

(ii) for each $0 \le i \le d$, each $g \in G_i$ has a unique representation as $g = \exp(t_{D(i)+1} X_{t_{D(i)+1}}) \cdots \exp(t_{D-1} X_{D-1}) \exp(t_D X_D)$, where $D(i) := \operatorname{codim} G_i$ and $t_j \in \mathbb{R}$ for $D(i) < j \le D$;

(iii) Γ is the set of all products $\exp(t_1 X_1) \exp(t_2 X_2) \cdots \exp(t_D X_D)$ with $t_j \in \mathbb{Z}$ for $1 \le j \le D$.

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

If the Lie bracket is given in coordinates by

$$[X_i, X_j] = \sum_{k=1}^{D} c_{i,j}^{(k)} X_k,$$

where all of the constants $c_{i,j}^{(k)}$ are rationals with height at most M, then we will say that the complexity of $(G, \Gamma, G_\bullet)$ is at most M. We recall that the height of a rational number a/b is $\max(|a|, |b|)$ ($a \in \mathbb{Z}$, $b \in \mathbb{N}$, $\gcd(a, b) = 1$).

We will usually keep the choice of the Mal'cev basis implicit, and assume that each filtered nilmanifold under consideration comes equipped with a fixed choice of Mal'cev basis. The Mal'cev basis $\mathcal{X}$ induces bijective coordinate maps $\tau\colon X \to [0, 1)^D$ and $\widetilde{\tau}\colon G \to \mathbb{R}^D$, such that

$$x = \exp(\tau_1(x)X_1)\exp(\tau_2(x)X_2)\cdots\exp(\tau_D(x)X_D)\Gamma, \quad x \in X,$$
$$g = \exp(\widetilde{\tau}_1(g)X_1)\exp(\widetilde{\tau}_2(g)X_2)\cdots\exp(\widetilde{\tau}_D(g)X_D), \quad g \in G.$$

The Mal'cev basis also induces a natural choice of a right-invariant metric on G and a metric on X. We refer to [**GT12b**, Definition 2.2] for a precise definition. Keeping the dependence on $\mathcal{X}$ implicit, we will use the symbol d to denote either of those metrics.

The space X comes equipped with the Haar measure μ_X, which is the unique Borel probability measure on X invariant under the action of G: $\mu_X(gE) = \mu_X(E)$ for all measurable $E \subseteq X$ and $g \in G$. When there is no risk of confusion, we write dx as a shorthand for $d\mu_X(x)$.

A map $g\colon \mathbb{Z} \to G$ is polynomial with respect to the filtration $G_\bullet$, denoted $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$, if it takes the form

$$g(n) = g_0 g_1^{n} \cdots g_d^{\binom{n}{d}},$$

where $g_i \in G_i$ for all $0 \le i \le d$ (cf. [**GT12b**, Lemma 6.7]; see also [**GT12b**, Definition 1.8] for an alternative definition). Although it is not immediately apparent from the definition above, polynomial sequences with respect to a given filtration form a group and are preserved under passing to an arithmetic progression (that is, if $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ and $g'(n) := g(An + B)$ for some $A, B \in \mathbb{Z}$ then $g' \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$).

4.2. *Semialgebraic geometry.* A basic semialgebraic set $S \subseteq \mathbb{R}^D$ is a set given by a finite number of polynomial equalities and inequalities:

$$S = \{x \in \mathbb{R}^d \mid P_1(x) > 0, \ldots, P_n(x) > 0, Q_1(x) = 0, \ldots, Q_m(x) = 0\}. \tag{11}$$

A semialgebraic set is a finite union of basic semialgebraic sets. In a somewhat ad hoc manner, we define the complexity of the basic semialgebraic set S given by (11) to be the sum $\sum_{i=1}^{n} \deg P_i + \sum_{j=1}^{m} \deg Q_j$ of degrees of polynomials appearing in its definition. (Strictly speaking, we take the infimum over all representations of S in the form (11).) We also define the complexity of a semialgebraic set

$$S = S_1 \cup S_2 \cup \cdots \cup S_r, \tag{12}$$

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

represented to be the finite union of basic semialgebraic sets S_i as the sum of complexities of S_i. (Again, we take the infimum over all representations (12).)

Using the Mal'cev coordinates to identify the nilmanifold X with $[0, 1)^D$, we extend the notion of a semialgebraic set to subsets of X. A map $F\colon X \to \mathbb{R}$ is piecewise polynomial if there exists a partition $X = \bigcup_{i=1}^{r} S_i$ into semialgebraic pieces and polynomial maps $\Phi_i\colon \mathbb{R}^D \to \mathbb{R}$ such that $F(x) = \Phi_i(\tau(x))$ for each $1 \le i \le r$ and $x \in S_i$. One can check that these notions are independent of the choice of basis, although strictly speaking we will not need this fact.

4.3. *Quantitative equidistribution.* The Lipschitz norm of a function $F\colon X \to \mathbb{R}$ is defined as

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

A sequence $(x_n)_{n=0}^{N-1}$ in X is *δ-equidistributed* if for each Lipschitz function $F\colon X \to \mathbb{R}$ we have

$$\left| \mathop{\mathbb{E}}_{n<N} F(x_n) - \int_X F(x)\, dx \right| \le \delta \|F\|_{\mathrm{Lip}}.$$

In the case where $X = [0, 1]$ this notion is highly connected to the discrepancy of a sequence (see (2)). In fact, for $\delta > 0$ small enough we have that $(x_n)_{n=0}^{N-1}$ has discrepancy δ if and only if it is $\delta^{O(1)}$ distributed. One direction follows immediately from the Koksma–Hlawka inequality and the other direction can be found for example in the proof of [**DDM+22**, Proposition 5.2].

More restrictively, $(x_n)_{n=0}^{N-1}$ is *totally δ-equidistributed* if for each arithmetic progression $P \subseteq [N]$ of length at least δN we have

$$\left| \mathop{\mathbb{E}}_{n\in P} F(x_n) - \int_X F(x)\, dx \right| \le \delta \|F\|_{\mathrm{Lip}}.$$

A sequence $(\varepsilon_n)_{n=0}^{N-1}$ in G is *(M, N)-smooth* if $d(\varepsilon_n, e_G) \le M$ and $d(\varepsilon_n, \varepsilon_{n+1}) \le M/N$ for all $n \in [N-1]$. A group element $\gamma \in G$ is *Q-rational* if $\gamma^r \in \Gamma$ for some positive integer $r \le Q$. A point $x \in G/\Gamma$ is *Q-rational* if it takes the form $x = \gamma\Gamma$ for some Q-rational $\gamma \in G$. A sequence $(x_n)_{n=0}^{N-1}$ in X is Q-rational if each point x_n is Q-rational.

THEOREM 4.1. [**GT12b**, Theorem 1.19] *Let $C > 0$ be a constant. Let G be a connected, simply connected nilpotent Lie group of dimension D, let $\Gamma < G$ be a lattice, let $G_\bullet$ be a nilpotent filtration on G of length d, and assume that the complexity of $(G, \Gamma, G_\bullet)$ is at most M_0. Then for each $N \in \mathbb{N}$ and each polynomial sequence $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ there exists an integer M with $M_0 \le M \ll M_0^{O_{C,d,D}(1)}$ and a decomposition $g(n) = \varepsilon(n)g'(n)\gamma(n)$ $(n \in \mathbb{Z})$, where $\varepsilon, g', \gamma \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ and*

(i) *the sequence $(\varepsilon(n))_{n=0}^{N-1}$ is (M, N)-smooth;*

(ii) *the sequence $(\gamma(n)\Gamma)_{n=0}^{N-1}$ is M-rational and periodic with period less than or equal to M;*

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

(iii) *there is a group $G' < G$ with Mal'cev basis $\mathcal{X}'$ in which each element is an M-rational combination of elements of $\mathcal{X}$ such that $g'(n) \in G'$ for all $n \in \mathbb{Z}$, and the sequence $(g'(n)\Gamma')_{n=0}^{N-1}$ is totally $1/M^C$-equidistributed in G'/Γ', where $\Gamma' = \Gamma \cap G'$.*

4.4. *Generalized polynomials.* The connection between nilmanifolds and generalized polynomials was first elucidated by Bergelson and Leibman [**BL07**].

THEOREM 4.2. [**BL07**] *Let $f : \mathbb{Z} \to [0, 1)$ be a sequence. Then the following conditions are equivalent.*

(i) *f is a GP map.*

(ii) *There exist a connected, simply connected nilpotent Lie group G, a lattice $\Gamma < G$, $g \in G$ and a piecewise polynomial map $F : G/\Gamma \to [0, 1)$ such that $f(n) = F(g^n\Gamma)$ for all $n \in \mathbb{Z}$.*

(iii) *There exist a connected, simply connected nilpotent Lie group G of some dimension D, a lattice $\Gamma < G$, a compatible filtration $G_\bullet$, a polynomial sequence $g \in \mathrm{poly}(\mathbb{Z}, G_\bullet)$ and an index $1 \le j \le D$ such that $f(n) = \tau_j(g(n)\Gamma)$ for all $n \in \mathbb{Z}$.*

Remark 4.3. Strictly speaking, [**BL07**] does not include the assumption that G should be connected and simply connected. However, this requirement can be ensured by replacing G with a larger group. (cf. the 'lifting argument' in [**Fra09**, pp. 368] and also [**BL07**, Theorem A*]). The cost of this operation is that in (ii) one may not assume that the action of g on G/Γ is minimal, but we do not need this assumption.

In our applications, we will need to simultaneously represent maps of the form $f(\lfloor p(n) \rfloor)$ where f is a fixed GP map and p is a polynomial which is allowed to vary. Such a representation is readily obtained from Theorem 4.2.

THEOREM 4.4. *Let $f : \mathbb{Z} \to \mathbb{R}$ be a bounded GP map and let $d \in \mathbb{N}$. Then there exist a connected, simply connected nilpotent Lie group G, a lattice $\Gamma < G$, a filtration $G_\bullet$, and a piecewise polynomial map $F : G/\Gamma \to \mathbb{Z}$ such that for each polynomial $p(x) \in \mathbb{R}[x]$ with $\deg p \le d$ there exists $g_p \in \mathrm{poly}(G_\bullet)$ such that for all $n \in \mathbb{Z}$ we have $f(\lfloor p(n) \rfloor) = F(g_p(n)\Gamma)$.*

Proof. By Theorem 4.2, there exist a nilmanifold $G^{(0)}/\Gamma^{(0)}$ together with a piecewise polynomial map $F^{(0)} : G^{(0)}/\Gamma^{(0)} \to \mathbb{R}$, and a group element $g_0 \in G^{(0)}$ such that $f(n) = F^{(0)}(g_0^n\Gamma)$ for all $n \in \mathbb{Z}$. Following the strategy in [**Fra09**, Lemma 4.1], let $G := G^{(0)} \times \mathbb{R}$ and $\Gamma := \Gamma^{(0)} \times \mathbb{Z}$, and let $F : G/\Gamma \to \mathbb{R}$ be given by $F(t + \mathbb{Z}, h\Gamma^{(0)}) := F^{(0)}(g_0^{-\{t\}}h\Gamma^{(0)})$ for $t \in \mathbb{R}$ and $h \in G^{(0)}$. This construction guarantees that F is piecewise polynomial and for all $t \in \mathbb{R}$ we have

$$F(t + \mathbb{Z}, g_0^t\Gamma) = F^{(0)}(g_0^{\lfloor t \rfloor}\Gamma) = f(\lfloor t \rfloor).$$

For $p \in \mathbb{R}[x]$ and $n \in \mathbb{Z}$ let $g_p(n) := (p(n), g_0^{p(n)})$. Then g_α is polynomial with respect to the filtration $G_\bullet$ given by $G_i = G_{(\lfloor i/d \rfloor)}$, where $(G_{(j)})_j$ denotes the lower central series, and we have $f(\lfloor p(n) \rfloor) = F(g_p(n)\Gamma)$ for all $n \in \mathbb{Z}$. □

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

5. *Möbius orthogonality*

5.1. *Main result.* In this section, we discuss Möbius orthogonality of bracket words along Hardy field sequences. Our main result is Theorem B, which we restate below.

THEOREM 5.1. *Let* $\mathbf{a} = (a(n))_{n\in\mathbb{Z}}$ *be a (two-sided)* $\mathbb{R}$*-valued bracket word and let* $f : \mathbb{R}_+ \to \mathbb{R}$ *be a Hardy field function with polynomial growth. Then*

$$\frac{1}{N}\sum_{n=1}^{N} \mu(n)a(\lfloor f(n)\rfloor) \to 0 \quad \text{as } N \to \infty. \tag{13}$$

As usual, we will use Taylor expansion to approximate the restriction of $f(n)$ to an interval with a polynomial sequence, and then use Theorem 2.11 to control the error term involved in computing $\lfloor f(n)\rfloor$. The sequence $a(\lfloor f(n)\rfloor)$ can then be represented on a nilmanifold by Bergelson–Leibman machinery. As the next step, we require a suitable result on Möbius orthogonality in short intervals. In §5.2, we will prove the following theorem, which is closely related to [**MSTT22**, Theorem 1.1(i)]. Below, we let $\mathcal{AP}$ denote the set of all arithmetic progressions in $\mathbb{Z}$.

THEOREM 5.2. *Let* G *be a connected, simply connected nilpotent Lie group, let* $\Gamma < G$ *be a lattice, let* $G_\bullet$ *be a filtration on* G*, assume that* $G_\bullet$ *and* Γ *are compatible, and let* $F : G/\Gamma \to \mathbb{R}$ *be a finite-valued piecewise polynomial map. Let* N, H *be integers with* $N^{0.626} \le H \le N$*. Then*

$$\sup_{g\in\mathrm{poly}(G_\bullet)} \sup_{P\in\mathcal{AP}} \left| \mathop{\mathbb{E}}_{h<H} 1_P(h)\mu(N+h)F(g(h)\Gamma)\right| = o_{N\to\infty}(1), \tag{14}$$

where the rate of convergence may depend on $G, \Gamma, G_\bullet$ *and* F*.*

Proof of Theorem 5.1 assuming Theorem 5.2. Applying a dyadic decomposition, it will suffice to show that

$$\mathop{\mathbb{E}}_{N\le n<2N} \mu(n)a(\lfloor f(n)\rfloor) \to 0 \quad \text{as } N \to \infty. \tag{15}$$

Fix a small $\varepsilon > 0$. We will show that for all sufficiently large N we have

$$\left| \mathop{\mathbb{E}}_{N\le n<2N} \mu(n)a(\lfloor f(n)\rfloor)\right| \ll \varepsilon. \tag{16}$$

Splitting the average in (16) into intervals of length $\lceil (2N)^{0.7}\rceil$, we see that (16) will follow once we show that for sufficiently large N and for H satisfying $N^{0.7} \le H < N$ we have

$$\left| \mathop{\mathbb{E}}_{h<H} \mu(N+h)a(\lfloor f(N+h)\rfloor)\right| \ll \varepsilon. \tag{17}$$

Pick an integer $k \in \mathbb{N}$ such that $f(t) \ll t^k$, and let $\ell = 10k$. By Theorem 2.11, we have

$$\lfloor f(N+h)\rfloor = \lfloor P_N(h)\rfloor + e_N(h), \tag{18}$$

where P_N is a polynomial of degree (at most) ℓ and one of the conditions (i)–(iii) in Theorem 2.11 holds. In the case (i) we have $N \ll_\varepsilon H^{10/9} \le N^{7/9}$, which implies that $N = O_\varepsilon(1)$. Assuming that N is sufficiently large, we may disregard this case.

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

In case (ii) we have $\mathbb{E}_{h<H}|e_N(h)| < \varepsilon$, and as a consequence

$$\mathop{\mathbb{E}}_{h<H} \mu(N+h)a(\lfloor f(N+h)\rfloor) = \mathop{\mathbb{E}}_{h<H} \mu(N+h)a(\lfloor P_N(h)\rfloor) + O(\varepsilon). \tag{19}$$

By Theorem 4.4, there exist a connected and simply connected nilpotent Lie group G, a lattice $\Gamma < G$, a filtration $G_\bullet$ and a finite-valued piecewise polynomial map $F\colon G/\Gamma \to \mathbb{Z}$ such that for each polynomial P of degree at most ℓ there exists $g \in \mathrm{poly}(G_\bullet)$ such that $a(\lfloor P(h)\rfloor) = F(g(h)\Gamma)$; it is crucial that we have the same $(G, \Gamma, G_\bullet)$ for all polynomials P. In particular,

$$\Big|\mathop{\mathbb{E}}_{h<H} \mu(N+h)a(\lfloor P_N(h)\rfloor)\Big| \le \sup_{g\in\mathrm{poly}(G_\bullet)} \Big|\mathop{\mathbb{E}}_{h<H} \mu(N+h)a(F(g(h)\Gamma)\Big|. \tag{20}$$

By Theorem 5.2 (which is uniform in g), for sufficiently large N the expression in (20) is bounded by ε. Inserting this bound into (19) yields (17).

In case (iii), passing to an arithmetic progression, we may replace e_N with a constant sequence:

$$\Big|\mathop{\mathbb{E}}_{h<H} \mu(N+h)a(\lfloor f(N+h)\rfloor)\Big| \tag{21}$$

$$\ll_\varepsilon \max_{P\in\mathcal{AP}} \max_{e\in\{-1,0,1\}} \Big|\mathop{\mathbb{E}}_{h<H} \mu(N+h)1_P(h)a(\lfloor P_N(h)\rfloor + e)\Big|. \tag{22}$$

To finish the argument, it suffices to apply Theorem 5.2 similarly to the previous case. □

5.2. *Short intervals.* The remainder of this section is devoted to proving Theorem 5.2. We will derive it from closely related estimates for correlations of the Möbius function with nilsequences in short intervals. Recall that we let $\mathcal{AP}$ denote the set of all arithmetic progressions in $\mathbb{Z}$.

THEOREM 5.3. (Corollary of [**MSTT22**, Theorem 1.1(i)]) *Let N, H be integers with $N^{0.626} \le H \le N$ and let $\delta \in (0, 1/2)$. Let G be a connected, simply connected nilpotent Lie group of dimension D, let $\Gamma < G$ be a lattice, let $G_\bullet$ be a nilpotent filtration on G of length d, and assume that the complexity of $(G, \Gamma, G_\bullet)$ is at most $1/\delta$. Let $F\colon G/\Gamma \to \mathbb{C}$ be a function with Lipschitz norm at most $1/\delta$. Then for each $A > 0$ we have the bound*

$$\sup_{g\in\mathrm{poly}(G_\bullet)} \sup_{P\in\mathcal{AP}} \Big|\mathop{\mathbb{E}}_{h<H} \mu(N+h)1_P(h)F(g(h)\Gamma)\Big| \ll_A \frac{(1/\delta)^{O_{d,D}(1)}}{\log^A N}. \tag{23}$$

This theorem is almost the ingredient that we need, except that in our application the function F is not necessarily continuous (much less Lipschitz). Instead, F is a finite-valued piecewise polynomial function, meaning that there exists a partition $G/\Gamma = \bigcup_{i=1}^r S_i$ into semialgebraic pieces and constants $c_i \in \mathbb{R}$ such that for each $x \in X$ and $1 \le i \le r$, $F(x) = c_i$ if and only if $x \in S_i$. In this case, it is enough to consider each of the level sets separately. It is clear that Theorem 5.2 will follow from the following more precise result.

THEOREM 5.4. *Let N, H be integers with $N^{0.626} \le H \le N$ and let $\delta \in (0, 1/2)$. Let G be a connected, simply connected nilpotent Lie group of dimension D, let $\Gamma < G$ be a lattice, let $G_\bullet$ be a nilpotent filtration on G of length d, and assume that the complexity of*

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

$(G, \Gamma, G_\bullet)$ is at most $1/\delta$. Let $S \subseteq G/\Gamma$ be a semialgebraic set with complexity at most E. Then for each $A \geq 1$ we have the bound

$$\sup_{g \in \mathrm{poly}(G_\bullet)} \sup_{P \in \mathcal{AP}} \left| \mathop{\mathbb{E}}_{h<H} \mu(N+h) 1_P(h) 1_S(g(h)\Gamma) \right| \ll_A \frac{(1/\delta)^{O_{d,D,E}(1)}}{\log^A N}. \tag{24}$$

In the case where $(g(n)\Gamma)_n$ is highly equidistributed in G/Γ, we will derive Theorem 5.4 directly from Theorem 5.3. In fact, we will obtain a slightly stronger version, given in Proposition 5.6 below. Then we will deduce the general case of Theorem 5.4 using the factorization theorem from [**GT12b**]. In order to avoid unnecessarily obfuscating the notation, from this point onwards we will allow all implicit constants to depend on the parameters d, D and E; thus, for instance, the term on the right-hand side of (24) will be more succinctly written as $(1/\delta)^{O(1)}/\log^A N$.

5.3. *Equidistributed case.* Before we proceed, we will need the following technical lemma.

LEMMA 5.5. *Let $d, D \in \mathbb{N}$, and let $\mathcal{V}$ denote the vector space of all polynomial maps $P\colon [0,1)^D \to \mathbb{R}$ of degree at most d.*

(i) *There is a constant $C > 1$ (dependent on d, D) such that for $P \in \mathcal{V}$ given by*

$$P(x) = \sum_{\alpha \in \mathbb{N}_0^D} a_\alpha \prod_{i=1}^{D} x_i^{\alpha_i},$$

we have the inequalities $C^{-1}\|P\|_\infty \leq \max_\alpha |a_\alpha| \leq C\|P\|_\infty$.

(ii) *For each $P \in \mathcal{V}$ and for each $\delta \in (0,1)$ we have*

$$\lambda(\{x \in [0,1)^D \mid |P(x)| < \delta^d \|P\|_\infty\}) \ll_{d,D} \delta. \tag{25}$$

Proof. Item (i) follows from the fact that each pair of norms on the finite-dimensional vector space $\mathcal{V}$ are equivalent. For item (ii) we proceed by induction with respect to D. Multiplying P by a scalar, we may assume that $\|P\|_\infty = 1$.

Suppose first that $D = 1$. We proceed by induction on d. If $d = 1$ then P is an affine function $P(x) = ax + b$, and the claim follows easily. Assume that $d \geq 2$ and that the claim has been proved for $d-1$. By item (i), at least one of the coefficients of P has absolute value $\gg_{d,D} 1$. In fact, we may assume that this coefficient is not the constant term, since otherwise for all $x \in [0,1)$ we would have $P(x) \in (\frac{99}{100}P(0), \frac{101}{100}P(0))$ and hence the set in (25) would be empty for sufficiently small δ. Thus, $\|P'\|_\infty \gg_{d,D} 1$. By the inductive assumption,

$$\lambda(\{x \in [0,1) \mid |P'(x)| < \delta^{d-1}\}) \ll_d \delta. \tag{26}$$

Thus, it will suffice to show that

$$\lambda(\{x \in [0,1) \mid |P(x)| < \delta^d,\ |P'(x)| > \delta^{d-1}\}) \ll_d \delta. \tag{27}$$

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

For each interval $I \subseteq [0, 1)$ such that $P'(x)$ has constant sign for $x \in I$, we have

$$\lambda(\{x \in I \mid |P(x)| < \delta^d,\ |P'(x)| > \delta^{d-1}\}) \ll \delta. \tag{28}$$

Since $[0, 1)$ can be divided into $O(d)$ intervals where P is monotone, (27) follows.

Suppose now that $D \geq 2$ and the claim has been proved for all $D' < D$. Reasoning as above, we infer from item (i) that P has a coefficient with absolute value $\gg_{d,D} 1$ other than the constant. We may expand P as

$$P(y, t) = \sum_{i=0}^{d} t^i Q_i(y), \quad y \in [0, 1)^{D-1},\ t \in [0, 1),$$

where the Q_i are polynomials in $D - 1$ variables of degree $d - i$. Changing the order of variables if necessary, we may assume that there exists j with $1 \leq j \leq d$ such that Q_j has a coefficient $\gg_{d,D} 1$, and hence $\|Q_j\|_\infty \gg_{d,D} 1$. For $k \in \mathbb{N}$, let us consider the set

$$E_k := \{(y, t) \in [0, 1)^D \mid |P(y, t)| < \delta^d,\ 2^{-k} \leq |Q_j(y)| < 2^{-k+1}\}.$$

The set in (28) is the disjoint union $\bigcup_{k=1}^{\infty} E_i$, so our goal is to show that

$$\sum_{k=1}^{\infty} \lambda(E_k) \ll_{d,D} \delta. \tag{29}$$

Fix a value of k. By the inductive assumption, as long as $j \neq d$, we have

$$\lambda(\{y \in [0, 1)^{D-1} \mid |Q_j(y)| < 2^{-k+1}\}) \ll_{d,D} 2^{-k/(d-j)}. \tag{30}$$

(If $j = d$, the set in (30) is empty for all sufficiently large k, and the reasoning simplifies.) For each $y \in [0, 1)^{D-1}$ such that $2^{-k} \leq |Q_j(y)| < 2^{-k+1}$, by the inductive assumption (for $D = 1$) we have

$$\lambda(\{t \in [0, 1) \mid |P(y, t)| < \delta^d\}) \ll_{d,D} 2^{k/d}\delta. \tag{31}$$

Combining (30) and (31) yields

$$\lambda(E_k) \ll_{d,D} 2^{-kj/d(d-j)}\delta \leq 2^{-k/d^2}\delta. \tag{32}$$

Summing (32) gives (29) and finishes the argument. □

We are now ready to prove a variant of Theorem 5.2 concerning highly equidistributed polynomial sequences on nilmanifolds. For technical reasons which will become clear in the next section, it will be convenient to consider more general type of averages, where instead of a factor of the form $1_S(g(h)\Gamma)$ with semialgebraic $S \subseteq G/\Gamma$ we have a factor of the form $1_S(h/H, g(h)\Gamma)$ with semialgebraic $S \subseteq (\mathbb{R}/\mathbb{Z}) \times (G/\Gamma)$; thus, in addition to the highly equidistributed sequence $g(h)\Gamma$, we keep track of how large h is compared to H.

PROPOSITION 5.6. *Let N, H be integers with $N^{0.626} \leq H \leq N$ and let $\delta \in (0, 1/2)$. Let G be a connected, simply connected nilpotent Lie group of dimension D, let $\Gamma < G$ be a lattice, let $G_\bullet$ be a nilpotent filtration on G of length d, and assume that the complexity of $(G, \Gamma, G_\bullet)$ is at most $1/\delta$. Let $S \subseteq (\mathbb{R}/\mathbb{Z}) \times (G/\Gamma)$ be a semialgebraic set with complexity at most E. Then for each $A \geq 1$, there exists $B = O(A)$ such that*

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

$$\sup_{\substack{g\in\mathrm{poly}(G_\bullet)\\ \widetilde{\delta}\text{-t.e.d.}}}\ \sup_{P\in\mathcal{AP}}\left|\mathop{\mathbb{E}}_{h<H}\mu(N+h)1_P(h)1_S\left(\frac{h}{H},g(h)\Gamma\right)\right|\ll_A\frac{(1/\delta)^{O(1)}}{\log^A N},\tag{33}$$

where $\widetilde{\delta}:=1/\log^B N$ *and the supremum is taken over all polynomial sequences* g *such that* $(g(h)\Gamma)_{h=0}^H$ *is totally* $\widetilde{\delta}$*-equidistributed (t.e.d.).*

Proof. We may freely assume that $\delta\geq 1/\log^A N$, since otherwise there is nothing to prove. In particular, $\delta=\log^{O(A)} N$ and $1/\delta=O(\log^A N)$. Decomposing S into a bounded number of pieces, we may assume that S is a basic semialgebraic set. We will assume that $\operatorname{int} S\neq\emptyset$; the case where $\operatorname{int} S=\emptyset$ can be handled using similar methods and is somewhat simpler. Thus, S takes the form

$$S=\{(t,x)\in(\mathbb{R}/\mathbb{Z})\times(G/\Gamma)\mid P_1(t,x)>0,\ P_2(t,x)>0,\ \dots,\ P_r(t,x)>0\},\tag{34}$$

where $r=O(1)$ and P_i are polynomial maps (under identification of $(\mathbb{R}/\mathbb{Z})\times(G/\Gamma)$ with $[0,1)^{1+D}$) with $\deg P_i=O(1)$ for $1\leq i\leq r$. Scaling, we may assume that $\|P_i\|_\infty=1$ for all $1\leq i\leq r$. Let τ_1 denote Mal'cev coordinates on $(\mathbb{R}/\mathbb{Z})\times(G/\Gamma)$, given by $\tau_1(t,x)=(t,\tau(x))$, where we identify $[0,1)$ with $\mathbb{R}/\mathbb{Z}$ in the standard way. Furthermore, splitting S further and applying a translation if necessary, we may assume that $\tau_1(S)\subseteq(\frac{1}{10},\frac{9}{10})^{1+D}$, implying in particular that τ_1 is continuous in a neighbourhood of S.

Let $\eta\in(0,\delta)$ be a small positive quantity, to be specified in the course of the argument, and let $\Psi,\Psi'\colon\mathbb{R}\to[0,1]$ be given by

$$\Psi(t)=\begin{cases}0 & \text{if } t<0,\\ t/\eta & \text{if } t\in[0,\eta],\\ 1 & \text{if } t>\eta,\end{cases}\qquad \Psi'(t)=\begin{cases}0 & \text{if } |t|>2\eta,\\ 2-|t|/\eta & \text{if } |t|\in[\eta,2\eta],\\ 1 & \text{if } |t|<\eta.\end{cases}$$

It is clear that $\|\Psi\|_{\mathrm{Lip}}=\|\Psi'\|_{\mathrm{Lip}}=1/\eta$. Let $\Psi_\square\colon[0,1)^{1+D}\to[0,1]$ be an $O(1)$-Lipschitz function with $\Psi_\square(t,u)=1$ if $(t,u)\in(\frac{1}{10},\frac{9}{10})^{1+D}$ and $\Psi_\square(t,u)=0$ if $(t,u)\notin(\frac{1}{20},\frac{19}{20})^{1+D}$. For $1\leq i\leq r$, put

$$F_i(t,x)=\Psi(P_i(t,x)),\quad F_i'(t,x)=\Psi_\square(\tau_1(t,x))\Psi'(P_i(t,x)),$$
$$F(t,x)=\prod_{i=1}^r F_i(t,x),\quad F'(t,x)=\min\left(\sum_{i=1}^r F_i'(t,x),1\right).$$

It is routine (although tedious) to verify that F and F' are $1/\eta^{O(1)}$-Lipschitz (cf. [**GT12b**, Lemma A.4]); this follows from the aforementioned bounds on the Lipschitz norms of Ψ and Ψ' and the fact that the derivatives of the polynomials P_i are bounded by $O(1)$ on $[0,1)^{D+1}$, which follows, for example, from Lemma 5.5. Directly from the definitions, we see that for each $t\in\mathbb{R}/\mathbb{Z}$ and $x\in G/\Gamma$ we have $F(t,x)=1_S(t,x)$ or $F'(t,x)=1$. It follows that

$$\left|\mathop{\mathbb{E}}_{h<H}\mu(N+h)1_P(h)1_S\left(\frac{h}{H},g(h)\Gamma\right)\right|\leq\left|\mathop{\mathbb{E}}_{h<H}\mu(N+h)1_P(h)F\left(\frac{h}{H},g(h)\Gamma\right)\right|+\mathop{\mathbb{E}}_{h<H}F'\left(\frac{h}{H},g(h)\Gamma\right).\tag{35}$$

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

In order to estimate either of the summands in (35), we begin by dividing the interval $[H]$ into $O(1/\alpha)$ subintervals with lengths between αH and $2\alpha H$, where

$$\alpha := (\log^A N \max(\|F\|_{\mathrm{Lip}}, \|F'\|_{\mathrm{Lip}}, 1))^{-1} = \eta^{O(1)}/\log^A N \tag{36}$$

To estimate the first summand, we note that for each such subinterval $[k, k+H')$ (where $\alpha H \le H' < 2\alpha H < H$), for each $h \in [k, k+H')$, we have

$$\begin{aligned} F\left(\frac{h}{H}, g(h)\Gamma\right) &= F\left(\frac{k}{H}, g(h)\Gamma\right) + O\left(\frac{H'}{H}\|F\|_{\mathrm{Lip}}\right) \\ &= F\left(\frac{k}{H}, g(h)\Gamma\right) + O\left(\frac{1}{\log^A N}\right). \end{aligned} \tag{37}$$

Applying Theorem 5.3 to each subinterval, for each constant $C \ge 1$ we obtain

$$\left| \mathop{\mathbb{E}}_{h<H} \mu(N+h) 1_P(h) F(g(h)\Gamma) \right| \ll_C \frac{1}{\log^A N} + \frac{1/\eta^{O(1)}}{\log^{C-A} N}. \tag{38}$$

Let us now consider the second summand. We have, similarly to (37),

$$F'\left(\frac{h}{H}, g(h)\Gamma\right) = F'\left(\frac{k}{H}, g(h)\Gamma\right) + O\left(\frac{1}{\log^A N}\right).$$

For now, let us assume that $\alpha > \widetilde{\delta}$, which we will verify at the end of the argument. We conclude from the fact that $(g(h)\Gamma)_{h=0}^{H-1}$ is totally $\widetilde{\delta}$-equidistributed that

$$\mathop{\mathbb{E}}_{h\in[k,k+H')} F'\left(\frac{h}{H}, g(h)\Gamma\right) = \int_{G/\Gamma} F'\left(\frac{k}{H}, x\right) dx + \frac{\widetilde{\delta}}{\eta^{O(1)}} + O\left(\frac{1}{\log^A N}\right), \tag{39}$$

where we use dx as a shorthand for $d\mu_{G/\Gamma}(x)$. Taking the weighted average of (39) over all subintervals, we conclude that

$$\mathop{\mathbb{E}}_{h<H} F'\left(\frac{h}{H}, g(h)\Gamma\right) = \int_{[0,1)} \int_{G/\Gamma} F'(t,x)\, dxdt + \frac{\widetilde{\delta}}{\eta^{O(1)}} + O\left(\frac{1}{\log^A N}\right). \tag{40}$$

Applying Lemma 5.5(ii) to estimate the measure of the support of F_i' for each $1 \le i \le r$, we conclude that

$$\int_{[0,1)} \int_{G/\Gamma} F'(t,x)\, dxdt \ll \eta^{1/O(1)}. \tag{41}$$

Thus, we may choose $\eta = 1/\log^{O(A)} N$ such that

$$\int_{[0,1)} \int_{G/\Gamma} F'(t,x)\, dxdt \le \frac{1}{\log^A N}, \tag{42}$$

which allows us to simplify (40) to

$$\mathop{\mathbb{E}}_{h<H} F'\left(\frac{h}{H}, g(h)\Gamma\right) = O\left(\frac{1}{\log^{B-O(A)} N}\right) + O\left(\frac{1}{\log^A N}\right). \tag{43}$$

Combining (38) and (43) with (35), we conclude that

$$\left| \mathop{\mathbb{E}}_{h<H} \mu(N+h) 1_P(h) 1_S(g(h)\Gamma) \right| \ll_C \frac{1}{\log^{C-O(A)} N} + \frac{1}{\log^{B-O(A)} N} + \frac{1}{\log^A N}. \tag{44}$$

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

Letting C and B be sufficiently large multiples of A, we conclude that

$$\Big| \mathop{\mathbb{E}}_{h<H} \mu(N+h) 1_P(h) 1_S(g(h)\Gamma) \Big| \ll_A \frac{1}{\log^A N}, \tag{45}$$

as needed. Note that choosing B as a large multiple of A also guarantees that $\alpha = 1/\log^{O(A)} N > \widetilde{\delta} = 1/\log^B N$. □

5.4. *General case.* We now have all the ingredients necessary to complete the proof.

Proof of Theorem 5.4. The argument is very similar to the proof of Theorem 1.1, assuming [**GT12a**, Proposition 2.1]. As the first step, we apply the factorization theorem [**GT12b**, Theorem 1.19], Theorem 4.1 above, with $M_0 = \log N$ and parameter C to be determined in the course of the argument. We conclude that there exists an integer M with $\log N \le M \ll \log^{O_C(1)} N$ such that g admits a factorization of the form

$$g(h) = \varepsilon(h) g'(h) \gamma(h), \tag{46}$$

where ε is (M, N)-smooth, γ is M-rational, and g' takes values in a rational subgroup $G' < G$ which admits a Mal'cev basis $\mathcal{X}'$ where each element is an M-rational combination of elements of $\mathcal{X}$, and $(g'(h)\Gamma)_{h=0}^{H-1}$ is totally $1/M^C$-equidistributed in $G'/(\Gamma \cap G')$ (with respect to the metric induced by $\mathcal{X}'$).

With the same reasoning as in [**GT12a**], we conclude that $(\gamma(h)\Gamma)_h$ is a periodic sequence with some period $q \le M$, and for each $0 \le j < q$ and $h \equiv j \bmod q$ we have $\gamma(h)\Gamma = \gamma_j \Gamma$ for some $\gamma_j \in G$ with coordinates $\tau(\gamma_j)$ that are rationals with height much less than $M^{O(1)}$. Splitting the average in (24) into subprogressions, it will suffice to show that for each residue $0 \le j < q$ modulo q, and for each arithmetic progression $Q \subseteq q\mathbb{Z} + j$ with diameter at most N/M, we have

$$\Big| \mathop{\mathbb{E}}_{h<H} \mu(N+h) 1_Q(h) 1_S(\varepsilon(h) g'(h) \gamma_j \Gamma) \Big| \ll_A \frac{(1/\delta)^{O(1)}}{M^2 \log^A N}. \tag{47}$$

The key difference between our current work and the corresponding argument in [**GT12a**] is that 1_S is not continuous and hence in (47) we cannot replace $\varepsilon(h)$ with a constant and hope that the value of the average will remain approximately unchanged. Instead, we will use an argument of a more algebraic type. We note that, as a consequence of invariance of the metric on G under multiplication on the right, for each $h, h' \in Q$ we have

$$d(\varepsilon(h) g'(h) \gamma_j, \varepsilon(h') g'(h) \gamma_j) = d(\varepsilon(h), \varepsilon(h')) = O(1).$$

Let us fix $k \in Q$ and put $\varepsilon'(h) = \varepsilon(h)\varepsilon(k)^{-1}$. Then $d(\varepsilon'(h), e_G) = O(1)$ and $g(h)\Gamma = \varepsilon(h) g'(h) \gamma_j \Gamma = \varepsilon'(h)\varepsilon(k) g'(h) \gamma_j \Gamma$.

Let $\Omega \subseteq G$ be a bounded semialgebraic set such that $\varepsilon'(h) \in \Omega$ for all $h \in Q$. For instance, we may take Ω to be the preimage of a certain ball with radius $1/\delta^{O(1)}$ under $\widetilde{\tau}$. Let also $\Pi := \widetilde{\tau}^{-1}([0,1)^D)$ denote the standard fundamental domain for G/Γ. Consider the set

$$R = \{(g_1, g_2) \in \Omega \times \Pi \mid g_1 g_2 \Gamma \in S\}.$$

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

We may decompose R as

$$R = \bigcup_{\gamma \in \Gamma} R_\gamma \quad \text{where } R_\gamma = \{(g_1, g_2) \in \Omega \times \Pi \mid g_1 g_2 \Gamma \in S,\ g_1 g_2 \gamma \in \Pi\}. \tag{48}$$

Using the quantitative bounds in [GT12b, Lemmas A.2 and A.3], we see that for each $\gamma \in \Gamma$ such that $R_\gamma \neq \emptyset$ we have $|\widetilde{\tau}(\gamma)| = O(1/\delta^{O(1)})$. Hence, the union in (48) involves $O(1/\delta^{O(1)})$ non-empty terms, and in particular is finite. Each of the sets R_γ is semialgebraic with complexity $O(1)$. Moreover, since ε' is a polynomial map of bounded degree, for each $\gamma \in \Gamma$ the set

$$T_\gamma = \{(t, x) \in [0, 1) \times \Pi \mid (\varepsilon'(tH), x) \in R_\gamma\}$$

is also semialgebraic with complexity $O(1)$. Hence, (47) will follow once we show that for each semialgebraic set $T \subseteq [0, 1) \times G/\Gamma$ with bounded complexity we have

$$\left| \mathop{\mathbb{E}}_{h<H} \mu(N + h) 1_Q(h) 1_T\left(\frac{h}{H}, \varepsilon(k) g'(h) \gamma_j \Gamma\right) \right| \ll_A \frac{(1/\delta)^{O(1)}}{M^2 \log^A N}. \tag{49}$$

Following [GT12a], we put $\widetilde{G}' := \gamma_j^{-1} G' \gamma_j$, $\Lambda := \Gamma \cap \widetilde{G}'$ and $\widetilde{g}'(n) := \gamma_j^{-1} g'(n) \gamma_j$. Let also $D' = \dim G'$, let σ and $\widetilde{\sigma}$ denote the coordinate maps on $\widetilde{G}'/\Lambda$ and $\widetilde{G}'$ respectively, and let $\Delta = \widetilde{\sigma}^{-1}([0, 1)^{D'})$ denote the fundamental domain. Then $\widetilde{g}'$ is a polynomial sequence with respect to the filtration $\widetilde{G}'_\bullet$ given by $\widetilde{G}'_i = \gamma_j^{-1} G'_i \gamma_j$. We have a well-defined map $\iota\colon \widetilde{G}'/\Lambda \to G/\Gamma$ given by

$$\iota(x\Lambda) = \varepsilon(k) \gamma_j x \Gamma.$$

Thus, for all $h \in [H]$ we have

$$\varepsilon(k) g'(h) \gamma_j \Gamma = \iota(\widetilde{g}'(h)\Lambda).$$

As discussed in [GT12b], the Lipschitz norm of the map ι is $O(M^{O(1)})$ and the sequence $(\widetilde{G}'(h)\Lambda)_{h=0}^{H-1}$ is $1/M^{\lambda C + O(1)}$-equidistributed, where $\lambda > 0$ is a constant dependent only on d and D.

For each $\gamma \in \Gamma$, the map ι is a polynomial on the semialgebraic set $\Delta \cap \iota^{-1}(\Pi\gamma)$. The estimate on the Lipschitz norm of ι implies that Δ can be partitioned into $M^{O(1)}$ semialgebraic sets with complexity $O(1)$ such that on each of the pieces ι is a polynomial of degree $O(1)$ (using the coordinates $\widetilde{\tau}$ and $\widetilde{\sigma}$). Applying the corresponding partition in (49), we see that it will suffice to show that for each semialgebraic set $T \subseteq (\mathbb{R}/\mathbb{Z}) \times (\widetilde{G}'/\Lambda)$ with bounded complexity and for each constant $A' > 0$ we have

$$\left| \mathop{\mathbb{E}}_{h<H} \mu(N + h) 1_Q(h) 1_T\left(\frac{h}{H}, g_j(h)\Lambda\right) \right| \ll_{A,A'} \frac{(1/\delta)^{O(1)}}{M^{A'} \log^A N}. \tag{50}$$

Bearing in mind that $M \geq \log N$, it will suffice to show that

$$\left| \mathop{\mathbb{E}}_{h<H} \mu(N + h) 1_Q(h) 1_T\left(\frac{h}{H}, g_j(h)\Lambda\right) \right| \ll_A \frac{(1/\delta)^{O(1)}}{M^A}. \tag{51}$$

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

We are now in position to apply Proposition 5.6 on $\widetilde{G}'/\Lambda$. The complexity of $(\widetilde{G}', \Lambda, \widetilde{G}'_\bullet)$ is $1/\delta'$, where $\delta' = 1/M^{O(1)}$. The largest exponent A' with which Proposition 5.6 is applicable to $(\widetilde{g}'(h))_{h=0}^{H-1}$ satisfies $\log^{A'} N \gg M^{\mu C}$ for a constant $\mu \gg 1$, leading to

$$\left| \mathop{\mathbb{E}}_{h<H} \mu(N+h) 1_Q(h) 1_T\left(\frac{h}{H}, g_j(h)\Lambda\right)\right| \ll_C \frac{1}{M^{\mu C - O(1)}}. \tag{52}$$

In order to derive (51) it is enough to let C be a sufficiently large multiple of A. □

Acknowledgements. The authors wish to thank Michael Drmota for many insightful discussions, for suggesting this problem, and also for inviting the first-named author to Vienna for a visit during which this project started; and Fernando Xuancheng Shao for helpful comments on Möbius orthogonality of nilsequences. The authors are also grateful to the anonymous referee for careful reading of the paper and for thoughtful corrections.

During the initial work on this paper, the first-named author worked within the framework of the LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon, within the 'Investissements d'Avenir' programme (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). Currently, he works at the University of Oxford and is supported by UKRI Fellowship EP/X033813/1. The second-named author is supported by the Austrian-French project 'Arithmetic Randomness' between FWF and ANR (grant numbers I4945-N and ANR-20-CE91-0006).

References

[AK23] B. Adamczewski and J. Konieczny. Bracket words: a generalisation of Sturmian words arising from generalised polynomials. *Trans. Amer. Math. Soc*. **376**(7) (2023), 4979–5044

[BL07] V. Bergelson and A. Leibman. Distribution of values of bounded generalized polynomials. *Acta Math.* **198**(2) (2007), 155–230.

[Bos94] M. D. Boshernitzan. Uniform distribution and Hardy fields. *J. Anal. Math.* **62** (1994), 225–240.

[Bou13] J. Bourgain. On the correlation of the Moebius function with rank-one systems. *J. Anal. Math.* **120** (2013), 105–130.

[BSZ13] J. Bourgain, P. Sarnak and T. Ziegler. Disjointness of Moebius from horocycle flows. *From Fourier Analysis and Number Theory to Radon Transforms and Geometry (Developments in Mathematics, 28)*. Eds. H. M. Farkas, R. C. Gunning, M. I. Knopp and B. A. Taylor. Springer, New York, 2013, pp. 67–83.

[DDM12] J.-M. Deshouillers, M. Drmota and J. F. Morgenbesser. Subsequences of automatic sequences indexed by $\lfloor n^c \rfloor$ and correlations. *J. Number Theory* **132**(9) (2012), 1837–1866.

[DDM15] J.-M. Deshouillers, M. Drmota and C. Müllner. Automatic sequences generated by synchronizing automata fulfill the Sarnak conjecture. *Studia Math.* **231** (2015), 83–95.

[DDM+22] J.-M. Deshouillers, M. Drmota, C. Müllner, A. Shubin and L. Spiegelhofer. Synchronizing automatic sequences along Piatetski-Shapiro sequences. *Preprint*, 2022, arXiv:2211.01422, *Israel J. Math.* accepted.

[DK15] T. Downarowicz and S. Kasjan. Odometers and Toeplitz systems revisited in the context of Sarnak's conjecture. *Studia Math.* **229**(1) (2015), 45–72.

[DLMR] M. Drmota, M. Lemanczyk, C. Müllner and J. Rivat. Some recent developments on the Sarnak Conjecture. *Panoramas et Syntheses*. Société Mathématique de France, accepted.

[eALdlR14] E. H. el Abdalaoui, M. Lemańczyk and T. de la Rue. On spectral disjointness of powers for rank-one transformations and Möbius orthogonality. *J. Funct. Anal.* **266**(1) (2014), 284–317.

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press

[eAKL16] E. H. el Abdalaoui, S. Kasjan and M. Lemańczyk. 0–1 sequences of the Thue–Morse type and Sarnak's conjecture. *Proc. Amer. Math. Soc.* **144**(1) (2016), 161–176.

[FKPL18] S. Ferenczi, J. Kułaga-Przymus and M. Lemańczyk. Sarnak's conjecture: what's new. *Ergodic Theory and Dynamical Systems in their Interactions with Arithmetics and Combinatorics*. Eds. S. Ferenczi, J. Kułaga-Przymus and M. Lemańczyk. Springer, Cham, 2018.

[FKPLM16] S. Ferenczi, J. Kułaga-Przymus, M. Lemanczyk and C. Mauduit. Substitutions and Möbius disjointness. *Ergodic Theory, Dynamical Systems, and the Continuing Influence of John C. Oxtoby (Contemporary Mathematics, 678)*. Eds. J. Auslander, A. Johnson and C. E. Silva. American Mathematical Society, Providence, RI, 2016.

[Fra09] N. Frantzikinakis. Equidistribution of sparse sequences on nilmanifolds. *J. Anal. Math.* **109** (2009), 353–395.

[Gre12] B. Green. On (not) computing the Möbius function using bounded depth circuits. *Combin. Probab. Comput.* **21**(6) (2012), 942–951.

[GT12a] B. Green and T. Tao. The Möbius function is strongly orthogonal to nilsequences. *Ann. of Math. (2)* **175**(2) (2012), 541–566.

[GT12b] B. Green and T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)* **175**(2) (2012), 465–540.

[KMPS18] K. Klouda, K. Medková, E. Pelantová and Š. Starosta. Fixed points of Sturmian morphisms and their derivated words. *Theoret. Comput. Sci.* **743** (2018), 23–37.

[KPL15] J. Kułaga-Przymus and M. Lemańczyk. The Möbius function and continuous extensions of rotations. *Monatsh. Math.* **178**(4) (2015), 553–582.

[LS15] J. Liu and P. Sarnak. The Möbius function and distal flows. *Duke Math. J.* **164**(7) (2015), 1353–1399.

[MR95] C. Mauduit and J. Rivat. Répartition des fonctions q-multiplicatives dans la suite $\left(\left[n^c\right]\right)_{n\in\mathbb{N}}$, $c > 1$. *Acta Arith.* **71**(2) (1995), 171–179.

[MR05] C. Mauduit and J. Rivat. Propriétés q-multiplicatives de la suite $\lfloor n^c \rfloor$, $c > 1$. *Acta Arith.* **118**(2) (2005), 187–203.

[MR10] C. Mauduit and J. Rivat. Sur un problème de Gelfond: la somme des chiffres des nombres premiers. *Ann. of Math. (2)* **171**(3) (2010), 1591–1646.

[MR15] C. Mauduit and J. Rivat. Prime numbers along Rudin–Shapiro sequences. *J. Eur. Math. Soc. (JEMS)* **17**(10) (2015), 2595–2642.

[MS17] C. Müllner and L. Spiegelhofer. Normality of the Thue–Morse sequence along Piatetski-Shapiro sequences. II. *Israel J. Math.* **220**(2) (2017), 691–738.

[MSTT22] K. Matomäki, X. Shao, T. Tao and J. Teräväinen. Higher uniformity of arithmetic functions in short intervals I. All intervals. *Forum Math. Pi* **11** (2023), E29.

[Mül17] C. Müllner. Automatic sequences fulfill the Sarnak conjecture. *Duke Math. J.* **166**(17) (2017), 3219–3290.

[Pec18] R. Peckner. Möbius disjointness for homogeneous dynamics. *Duke Math. J.* **167**(14) (2018), 2745–2792.

[Sar11] P. Sarnak. Three lectures on the Mobius function randomness and dynamics, 2011. https://www.math.ias.edu/files/wam/2011/PSMobius.pdf.

[Spi15] L. Spiegelhofer. Normality of the Thue–Morse sequence along Piatetski-Shapiro sequences. *Q. J. Math.* **66**(4) (2015), 1127–1138.

[Spi20] L. Spiegelhofer. The level of distribution of the Thue–Morse sequence. *Compos. Math.* **156**(12) (2020), 2560–2587.

[Vee16] W. A. Veech. Möbius orthogonality for generalized Morse–Kakutani flows. *Amer. J. Math.* **139** (2017), 1157–1203.

https://doi.org/10.1017/etds.2023.112 Published online by Cambridge University Press