On combinatorial properties of nil–Bohr sets of integers and related problems

Jakub Konieczny
St Johns's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Abstract

Author's name: Jakub Konieczny.
(Under supervision of Ben Green.)

This thesis deals with five problems in additive combinatorics and ergodic theory. A brief introduction to this general area and a summary of included results is given in Chapter I.

In Chapter II, we consider sets of the form $\{n \in \mathbb{N}_0 \mid |p(n) \bmod 1| \leq \varepsilon(n)\}$, where p is a polynomial and $\varepsilon(n) \geq 0$. We obtain various conditions under which any sufficiently large integer can be represented as a sum of 2 or 3 elements of a given set of this form.

In Chapter III, we study the class of weakly mixing sets of integers, and prove that a certain class of polynomial equations can always be solved in such a set.

In Chapter IV, we show that any nil–Bohr set contains a certain type of additive pattern. Combined with earlier results of Host and Kra, this leads to a partial combinatorial characterisation of nil–Bohr sets.

In Chapter V, we study the combinatorial properties of generalised polynomials (expressions built from polynomials and the floor function). In contrast with results of Bergelson and Leibman, we show that if the set of integers where a given generalised polynomial takes a non-zero value has asymptotic density 0, then it does not contain any IP set. This leads to a partial characterisation of automatic sequences which are given by generalised polynomial formulas.

In Chapter VI, we estimate the Gowers norms of the Thue-Morse sequence and the Rudin-Shapiro sequence. This gives some of the simplest deterministic examples of sequences with small Gowers norms of all orders.

Acknowledgements

The author wishes to express his gratitude to the following people:

- to Ben Green for an endless supply of problems and methods for solving them;
- to Jakub Byszewski for the numerous conversations which were productive and enjoyable, and for the few which were both;
- to Tom Sanders, Roger Heath-Brown and David Conlon for comments on previous drafts of this thesis, especially Chapter II, and on writing in general;
- to Vitaly Bergelson for comments related to Chapters III and V, as well for his highly contagious vitality;
- to Bryna Kra for comments related to Chapters II, III, IV;
- to Alexander Fish for comments related to Chapter III;
- to Dominik Kwietniak for comments related to Chapter V and life in general;
- to Jean-Paul Allouche and Inger Håland-Knutson for comments related to Chapter V;
- to Tanja Eisner, Christian Mauduit, Clemens Müllner, and Aihua Fan for comments related to Chapter VI;
- to James Aaronson, Sean Eberhard, Sofia Lindqvist, Freddie Manners, Rudi Mrazović, Przemek Mazur and Aled Walker for many informal discussions;
- the organizers of the conference *New developments around* $\times 2 \times 3$ *conjecture and other classical problems in Ergodic Theory* in Cieplice, Poland in May 2016;
- to National Science Centre in Poland, Clarendon Fund and St John's College Kendrew Fund for providing generous funding;
- last but not least, to all my family and friends and especially to Magda Rusaczonek, without whom this thesis would never have come to be.

Contents

Notation

Below we list some of the notation used in this thesis, some of which is not entirely standard.

$\mathbb{N}$, $\mathbb{N}_0$: the sets of positive integers and non-negative integers, respectively;

$\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, ...: integers, rational numbers, real numbers, complex numbers, etc.;

$\mathcal{F}$, $\mathcal{F}_\emptyset$: the family of non-empty (resp. all) subsets of $\mathbb{N}$;

$\mathbb{T}$: the 1-dimensional torus $\mathbb{R}/\mathbb{Z}$;

$[N]$: the initial interval $\{0, 1, 2, \ldots, N-1\}$ of $\mathbb{N}_0$;

$[\![\Phi]\!]$: the Iverson bracket, equal 1 if Φ is a true sentence and 0 otherwise;

$\lfloor x \rfloor$, $\lceil x \rceil$, $\langle\!\langle x \rangle\!\rangle$: the best integer approximation of $x \in \mathbb{R}$ by integer from below, above and in absolute value $\lfloor x + 1/2 \rfloor$, respectively;

$\{x\}$, $\|x\|$: fractional part $x - \lfloor x \rfloor$ and absolute fractional part $|x - \langle\!\langle x \rangle\!\rangle|$ of $x \in \mathbb{R}$ respectively, also used for $x \in \mathbb{T}$;

1_X: the characteristic sequence $[\![x \in X]\!]$ of the set X;

$e(x)$: the function $e^{2\pi i x}$, $x \in \mathbb{R}$ or $x \in \mathbb{T}$;

$\nu_p(x)$: the p-adic valuation of $x \in \mathbb{Q}$, $\nu_p(p^t \frac{a}{b}) = t$ if $p \nmid a, b$;

$\deg p$, $\operatorname{lc} p$: the degree and the leading coefficient of a polynomial $p \in \mathbb{R}[x]$;

$A + B$: the sumset $\{a + b \mid a \in A,\ b \in B\}$ of two sets A and B;

kA: the k-fold sumset $A + A + \cdots + A$ (k times).

We use the symbol $\mathbb{E}$ borrowed from probability to denote averages:

$$\mathop{\mathbb{E}}_{x \in X} f(x) := \frac{1}{|X|} \sum_{x \in X} f(x).$$

We use standard asymptotic notation. We write $X \ll Y$ or $X = O(Y)$ if $|X| \leq cY$ for a universal constant $c > 0$. If c is allowed to depend on a parameter p, we denote it by a subscript: $X \ll_p Y$ or $X = O_p(Y)$. Conversely, we write $X \gg Y$ of $X = \Omega(Y)$ if $X \geq cY$. If $X \ll Y \ll X$, we write $X = \Theta(Y)$ or $X \sim Y$.

Similarly, for a variable v, we write $X = o_{v \to z}(Y)$ if $|X| \leq c(v)Y$ where $c(v) \to 0$ as $v \to z$. If v and z are clear from the context, we suppress the subscript $v \to z$. If c is allowed to depend on a parameter p, we denote this by subscript: $X = o_{p;v \to z}$ or $X = o_p(Y)$. Finally, we write $X = \omega(Y)$ if $X > c(v)Y$ where $c(v) \to \infty$ as $v \to z$.

We often write symbols $O(\cdot), \Omega(\cdot), \Theta(\cdot), o(\cdot), \omega(\cdot)$ to denote unspecified functions with the appropriate growth.

For a set $A \subset \mathbb{N}$, we define various notions of density. We list the ones which will be useful in this document:

$$d(A) = \lim_{n \to \infty} \frac{|A \cap [n]|}{n} \qquad \text{(natural/asymptotic density, if exists)},$$

$$\overline{d}(A) = \limsup_{n \to \infty} \frac{|A \cap [n]|}{n} \qquad \text{(upper natural density)},$$

$$d^*(A) = \limsup_{n \to \infty} \sup_m \frac{|A \cap [m, m+n)|}{n} \qquad \text{(upper Banach density)}.$$

Chapter I

Introduction

I.1 Background

I.1.1 Additive combinatorics

Unlike most other branches of mathematics, additive combinatorics escapes a clear definition. As aptly noted in [51], it is much easier to identify problems which lie within additive combinatorics than it is to give a satisfactory criterion for inclusion. Following this dictum, we shall not endeavour to give much background on additive combinatorics in general, but rather proceed to introduce the tools we use and problems we solve without further delay. Extensive introduction to this subject can be found in [104].

All results discussed in this thesis will have an additive combinatorial flavour, some to a greater degree than other. Most of the problems which we shall encounter concern the question of, given a set $A \subset \mathbb{N}$, deciding if A contains a solution to a specific Diophantine equation or an interesting combinatorial structure. Arguably, the earliest non-trivial result of this type is the celebrated Roth's theorem.

Theorem I.1.1 (Roth, [97]). *Let $A \subset \mathbb{N}$ be a set with positive upper Banach density $d^*(A) > 0$. Then, A contains infinitely many 3-term arithmetic progressions $n, n+m, n+2m$, $(m, n \in \mathbb{N})$.*

Countless generalisations and variations on the theme of this theorem exist, many of which we will encounter in the following chapters.

A distinguishing feature of additive combinatorics is in the tools it employs. While the problems we wish to tackle generally lie within number theory, combinatorics or possibly algebra, the tools are often very analytic, and originate from harmonic analysis or ergodic theory. Indeed, the classical proof of Roth's theorem proceeds by

a careful analysis of the Fourier transform of (the characteristic sequence of) the set under consideration.

In this thesis, we will be primarily interested in the connection with ergodic theory, and the relatively recent theory of higher order Fourier analysis. We proceed to introduce the necessary tools.

I.1.2 Basic definitions

Ergodic theory deals with the asymptotic behaviour of dynamical system, and more specifically with the relation between the "space averages" and "time averages". Our primary interest in ergodic theory stems from the fact that it has multiple applications to additive combinatorics. For the sake of completeness, we begin by introducing the relevant concepts, which can be found in any number of references such as [35], [96] or [62], [43].

A *measure preserving system* $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ consists of a compact metrizable space X, together with a probability measure μ on a Borel σ-algebra $\mathcal{B}$, and a $\mathcal{B}$-measurable transformation $T\colon X \to X$, such that $\mu(T^{-1}E) = \mu(E)$ for all $E \in \mathcal{B}$. The transformation T acts on functions on X by composition: $(Tf)(x) = f(Tx)$.

A morphism between m.p.s.'s $\mathcal{X} = (X, \mathcal{B}, \mu, T)$ and $\mathcal{Y} = (Y, \mathcal{C}, \nu, S)$ is a $(\mathcal{B}, \mathcal{C})$-measurable map $\pi\colon X \to Y$ such that $\pi \circ T = S \circ \pi$ and $\pi_*\mu = \nu$. In this context, $\mathcal{Y}$ is a *factor* and $\mathcal{X}$ is an *extension*. Any factor $\mathcal{Y}$ of $\mathcal{X}$ is uniquely determined, up to isomorphism, by the σ-algebra $\mathcal{B}' \subset \mathcal{B}$ generated by the sets $\pi^{-1}(F)$, $F \in \mathcal{C}$. In particular, we have the conditional expectation operation $\mathbb{E}(\cdot|\mathcal{Y})$, which we can view (with obvious identifications) as mapping $L^p(\mu)$ to a subspace of $L^p(\mu)$, $p \in [1, \infty]$.

A closely related concept is that of a *topological dynamical system*, $\mathcal{X} = (X, T)$ consisting of a compact topological space X and a continuous map $T\colon X \to X$. By the Krylov-Bogoliubov theorem, any dynamical system (X, T) on a metrizable space can be equipped with a Borel T-invariant probability measure μ, whence $(X, T, \mathcal{B}_X, \mu)$ is an m.p.s..

Arguably the most fundamental result in ergodic theory is the mean ergodic theorem, originally due to von Neumann.

Theorem I.1.2 (Mean ergodic theorem)**.** *Let $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ be an m.p.s., and let $\mathcal{I}$ denote the factor induced by the σ-algebra of T-invariant sets. Then, for any $f \in L^2(\mu)$ it holds that*

$$\mathop{\mathbb{E}}_{n<N} T^n f \to \mathbb{E}(f|\mathcal{I}) \text{ in } L^2(\mu) \text{ as } N \to \infty.$$

An ergodic theorem which is stronger, but less amenable to generalisations, is due to Birkhoff.

Theorem I.1.3 (Pointwise ergodic theorem)**.** *Let $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ be an m.p.s., and let $\mathcal{I}$ denote the factor induced by the σ-algebra of T-invariant sets. Then, for any $f \in L^1(\mu)$ it holds for μ-almost every point $x \in X$ that*

$$\mathop{\mathbb{E}}_{n<N} T^n f(x) \to \mathbb{E}(f|\mathcal{I})(x) \text{ as } N \to \infty.$$

an m.p.s. $\mathcal{X}$ is *ergodic* if for any $A, B \in \mathcal{B}$ we have that

$$\mathop{\mathbb{E}}_{n<N} \mu(A \cap T^{-n}B) \xrightarrow[N\to\infty]{} \mu(A)\mu(B),$$

and similarly $\mathcal{X}$ is *weakly mixing* if we have the stronger condition

$$\mathop{\mathbb{E}}_{n<N} \left|\mu(A \cap T^{-n}B) - \mu(A)\mu(B)\right| \xrightarrow[N\to\infty]{} 0.$$

Likewise, T is *totally ergodic* if T^n is ergodic for all $n \geq 1$. A point $x \in X$ is *generic* if for any $f \in \mathcal{C}(X)$ one has $\mathbb{E}_{n<N} T^n f(x) \to \int f d\mu$. It is a consequence of the ergodic theorem that μ-almost all points are generic.

In similar spirit, a t.d.s. $\mathcal{X} = (X, T)$ is transitive if there exists $x \in X$ whose orbit $\{T^n x \mid n \in \mathbb{Z}\}$ is dense in X. We say that $\mathcal{X}$ is minimal if for every point $x \in X$ the orbit $\{T^n x \mid n \in \mathbb{Z}\}$ is dense in X. (Equivalently, the only closed subsets $Y \subset X$ such that $T(Y) \subset Y$ are $Y = X$ and $Y = \emptyset$.) Accordingly, $\mathcal{X}$ is totally minimal if the system (X, T^n) is minimal for all $n \geq 1$. If $\mathcal{X}$ is minimal, $x \in X$, and $U \subset X$ is open, then the set $\{n \in \mathbb{Z} \mid T^n x \in U\}$ is syndetic, i.e. has bounded gaps (cf. [43, Thm. 1.15]). Finally, $\mathcal{X}$ is uniquely ergodic if it only has one T-invariant measure.

Theorem I.1.4 (Ergodic theorem for uniquely ergodic systems, [35], Theorem 4.10)**.** *Let (X, T) be a uniquely ergodic dynamical system with an invariant measure μ. Then, for any continuous function $f\colon X \to \mathbb{C}$, we have*

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^n x) \to \int_X f d\mu$$

as $N \to \infty$. Furthermore, the convergence is uniform in $x \in X$.

By a standard argument that bounds the characteristic function of a set from above and from below by continuous functions, this yields the following result, frequently used in the sequel.

Corollary I.1.5. *Let (X,T) be a uniquely ergodic dynamical system with an invariant measure μ. Then for any $x \in X$ and any $S \subset X$ with $\mu(\partial S) = 0$, the set $E = \{n \in \mathbb{N}_0 \mid T^n x \in S\}$ has Banach density $\mu(S)$.*

an m.p.s.'s $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ is *invertible* if T is invertible. Any m.p.s. $\mathcal{X}$ has a canonical invertible extension $\tilde{\mathcal{X}}$, which is often more convenient to work with. Passing to the invertible extension preserves many desirable properties of m.p.s.'s, such as ergodicity or weak mixing. Provided that T is continuous and surjective (as in often the case in practice), we may ensure that if $x \in \mathcal{X}$ is generic, and $\tilde{x}$ is a lift of x, then $\tilde{x}$ is generic as well.

I.1.3 Group actions

Even though we work almost exclusively work with m.p.s.'s $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ equipped with only one transformation T, a more general setup is possible. For a semigroup G, a measure preserving action of G on the probabilistic measure space $(X, \mathcal{B}, \mu)$ is a collection of μ-invariant maps T_g, $g \in G$, such that $T_{gh} = T_g \circ T_h$. Taking $G = \mathbb{N}$ and $T_n = T^n$ we recover the previous definition of m.p.s.; in the case when T is invertible we may take $G = \mathbb{Z}$. From now, we restrict to the case when G is a group.

In this more general context, averages over $n < N$ are replaced with averages along a Følner sequence. A sequence $(F_n)_{n=1}^{\infty}$ of finite subsets of a group G is Følner if for each $g \in G$,

$$\frac{|gF_n \triangle F_n|}{|F_n|} \xrightarrow[n\to\infty]{} 0,$$

where $A \triangle B$ denotes the symmetric difference $(A \setminus B) \cup (B \setminus A)$. For countable groups G, existence of a Følner sequence is equivalent to amenability. Many statements in ergodic theory classically proved for $\mathbb{Z}$-actions are equally valid for G-actions for an arbitrary (countable) amenable group G. For instance, the following follows by essentially the same argument as in the classical setup.

Theorem I.1.6 (Mean ergodic theorem for group actions)**.** *Let $(X, \mathcal{B}, \mu)$ be a probabilistic measure space equipped with a measure preserving action of a group G, let $(F_n)_{n=1}^{\infty}$ be a Følner sequence in G, and let $\mathcal{I}$ denote the factor induced by the σ-algebra of T-invariant sets. Then, for any $f \in L^2(\mu)$ it holds that*

$$\mathop{\mathbb{E}}_{g \in F_N} T_g f \to \mathbb{E}(f|\mathcal{I}) \text{ in } L^2(\mu) \text{ as } N \to \infty.$$

The obvious analogue for the pointwise ergodic theorem is known to be false in general, a similar theorem holds under additional restrictions on the Følner sequence, see [86].

I.1.4 Ergodic theory in additive combinatorics

While ergodic theory is a important branch of mathematics in its own right, it will interest us mostly as a powerful tool which can be applied to tackle various problems in additive combinatorics. Arguably the earliest and most influential such application is Furstenberg's proof of Szemerédi's theorem.

Theorem I.1.7 (Szemerédi)**.** *Let $A \subset \mathbb{N}$ be a set with positive upper Banach density. Then, for any $k \in \mathbb{N}$, there exist $n, m \in \mathbb{N}$ such that $n,\ n+m,\ n+2m, \dots n+(k-1)m \in A$.*

Phrased differently, the theorem asserts that any set of positive upper Banach density contains arithmetic progressions of arbitrary length. The first proof of this theorem, due to Szemeredi [102], is purely combinatorial and notoriously complicated.

A different argument due to Furstenberg [42], relies on ergodic theory, and derives Szemerédi's theorem as a simple consequence of the Multiple Recurrence Theorem, which we will state shortly.

The usefulness of ergodic theory in problems in additive combinatorics stems from the following principle.

Theorem I.1.8 (Furstenberg correspondence principle)**.** *Let $A \subset \mathbb{N}$ be a set with $d^*(A) > 0$. Then, there exists an m.p.s. $(X, \mathcal{B}, \mu, T)$ and a set $E \in \mathcal{B}$ with $\mu(E) = d^*(A)$, such that for any $r \in \mathbb{N}$ and any $m_i \in \mathbb{N}$, $1 \leq i \leq r$, we have*

$$d^*\left(\bigcap_{j=1}^{r}(A - m_j)\right) \geq \mu\left(\bigcap_{j=1}^{r} T^{-m_j}E\right). \tag{I.1}$$

In fact, we may choose X to be the closure of the orbit of $1_A \in \{0,1\}^{\mathbb{N}_0}$ under the shift operator $(Tx)(n) := x(n+1)$, and E to be the cylinder set $\{x \mid x(0) = 1\}$.

In particular, if we are able to show that the measure on the intersection appearing on left-hand side of (I.1) is positive, then we may conclude that A contains the pattern $\{n + m_1, \dots, n + m_r\}$. Hence, Szemerédi's theorem is a consequence of the following result.

Theorem I.1.9 (Multiple Recurrence)**.** *Let $(X, \mathcal{B}, \mu, T)$ be an m.p.s.. Then, for any $E \subset X$ with $\mu(E) > 0$, and for any $k \in \mathbb{N}$,*

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mu\left(\bigcap_{j=0}^{k-1} T^{-jn}E\right) > 0, \tag{I.2}$$

and hence in particular the measure of the intersection above is positive for infinitely many $n \in \mathbb{N}$.

Many generalisations of the Szemerédi's Theorem exist. A theorem of Sárközy [99] (see also [42], [43]) asserts that in sets of positive Banach density one can find patterns such as $n, n+m^2$. In approximately the same time, but different direction, a result of Furstenberg and Katznelson [44] pertains to configurations in higher dimensions, showing that a set $A \subset \mathbb{N}^r$ of positive Banach density contains the configuration $n + mF$, where $F \subset \mathbb{Z}^r$ is any finite set.

Returning to the polynomial in a single dimension, Bergelson and Leibman [14] were able to improve Sárközy's theorem to several polynomials vanishing at 0. This result was ultimately improved by these authors and Lesigne [18] to deal with intersective families of polynomials. A sequence $p_i(x) \in \mathbb{Z}[x]$, $1 \leq i \leq r$ is *intersective* if for any integer k there exists $n_k \in \mathbb{N}$ such that $k \mid p_i(n_k)$ for all $1 \leq i \leq r$.

Theorem I.1.10 (Bergelson-Leibman-Lesigne)**.** *Let $A \subset \mathbb{N}$ be a set with positive Banach density, and let $p_i \in \mathbb{Z}[x]$ be an intersective family of polynomials with $p_i(n) \to \infty$ as $n \to \infty$. Then, there exists $n, m \in \mathbb{N}$ such that $m, m+p_1(n), \ldots, m+p_r(n) \in A$.*

In light of the Furstenberg's Correspondence Principle, it is natural to view a set of integers $A \subset \mathbb{N}$ as a set of recurrence for dynamical system, i.e. a set of the form $A = \{n \in \mathbb{N} \mid T^n x \in E\}$ where $(X, \mathcal{B}, \mu, T)$ is an m.p.s., $E \subset X$, $\mu(E) > 0$ and $x \in E$. One hopes that nice properties of the underlying m.p.s. will translate into nice combinatorial properties of the set A. This idea will feature prominently in Chapters II, III and IV.

I.1.5 IP sets

The class of IP sets plays an important role in additive combinatorics and ergodic theory. Before elaborating more on why it is so, we introduce some terminology.

Throughout this section, let $\mathcal{F}$ denote the set of all finite nonempty subsets $\alpha \subset \mathbb{N}$ (which we will somewhat perversely denote by lowercase Greek letters). For sets $\alpha, \beta \in \mathcal{F}$, we write $\alpha < \beta$ if $a < b$ for all $a \in \alpha, b \in \beta$. For a sequence $(n_i)_{i=1}^{\infty}$, we define the corresponding set of partial sums

$$\mathrm{FS}(n_i) = \left\{ \sum_{i \in \alpha} n_i \;\middle|\; \alpha \in \mathcal{F} \right\}. \tag{I.3}$$

For brevity of notation, it is convenient to define in this context $n_\alpha := \sum_{i \in \alpha} n_i$ for $\alpha \in \mathcal{F}$, so that $\mathrm{FS}(n_i) = \{n_\alpha \mid \alpha \in \mathcal{F}\}$; this is consistent with the natural inclusion $\mathbb{N} \ni i \mapsto \{i\} \in \mathcal{F}$.

Definition I.1.11 (IP set)**.** A set $E \subset \mathbb{N}$ is called an IP *set* if it contains a set of the form $\mathrm{FS}(n_i)$ for some sequence of natural numbers $(n_i)_{i\in\mathbb{N}}$. Similarly, $E \subset \mathbb{N}$ is an IP_+ set if it contains $E_0 + a$ for some $a \in \mathbb{N}_0$ and some IP set E_0.

We remark that since we view IP as a notion of largeness, we only require E to contain a set of finite sums (as opposed to being equal to such a set). This is consistent with usage e.g. in [13], but different from the original definition in [46].

The class of IP sets is partition regular, i.e. whenever an IP set E is written as a union of finitely many subsets $E = E_1 \cup E_2 \cup \ldots \cup E_r$, at least one of the sets E_i is an IP set. This is the statement of Hindman's Theorem (see [63], or [7]).

For a class $\mathcal{C}$ of subsets of $\mathbb{N}$, we define the *dual* class $\mathcal{C}^*$ by declaring that $B \in \mathcal{C}^*$ if and only if for every $A \in \mathcal{C}$ the sets A and B intersect non-trivially: $A \cap B \neq \emptyset$ (see e.g. [43, Section 9.1]). In particular, a set $E \subset \mathbb{N}$ is called an IP^* *set* if it non-trivially intersects any IP set.

It is clear by definition that for any class $\mathcal{C}$, the dual class $\mathcal{C}^*$ is closed under taking supersets. The operation of taking the dual reverses the inclusion: if $\mathcal{C} \subset \mathcal{D}$ then $\mathcal{C}^* \supset \mathcal{D}^*$. If $\mathcal{C}$ is partition regular, then $\mathcal{C}^*$ is easily seen to be closed under finite intersections, but $\mathcal{C}^*$ will not generally be partition regular [43, Lemma 9.5]. If additionally $\emptyset \notin \mathcal{C}$ then $\mathcal{C} \subset \mathcal{C}^*$ [43, Lemma 9.4].

It follows that an IP^* set is an IP set, an intersection of two IP^* sets is an IP^* set, and an intersection of an IP^* set and an IP set is an IP set. (For the first and the third statement, let A be an IP^* set and let B be an IP set. Apply Hindman's theorem to the partitions $\mathbb{N} = A \cup (\mathbb{N} \setminus A)$ and $B = (B \cap A) \cup (B \setminus A)$. The second statement then easily follows.)

One reason for the interest in IP sets in the context of dynamics is that there exists a closely related notion of convergence, which may play a similar role to Cesaro averages in classical ergodic theory.

To introduce the aforementioned notion of convergence, it is more convenient to define for IP rings, which are the analogues of IP sets in $\mathcal{F}$. More precisely, a family $\mathcal{G} \subset \mathcal{F}$ of subsets of $\mathbb{N}$ is called an IP *ring* if there exists a sequence $\beta = (\beta_i)_{i\in\mathbb{N}} \subset \mathcal{F}$ with β_i pairwise disjoint[1] and such that

$$\mathcal{G} = \mathrm{FU}(\beta_i) = \left\{ \bigcup_{i\in\alpha} \beta_i \;\middle|\; \alpha \in \mathcal{F} \right\}.$$

[1]Note that we only require β_i to be pairwise disjoint. A similar definition is sometimes made with a stronger condition $\max \beta_i < \min \beta_{i+1}$. We do not follow this approach here.

Note that an IP ring, equipped with the operation of disjoint union, is a partial semigroup isomorphic to $\mathcal{F}$. An equivalent version of Hindman's theorem says that whenever an IP ring $\mathcal{G}$ is written as a union of finitely many subfamilies $\mathcal{G} = \mathcal{G}_1 \cup \mathcal{G}_2 \cup \ldots \cup \mathcal{G}_r$, at least one of the subfamilies $\mathcal{G}_i$ contains an IP ring (cf. [11], [7].)

There is a natural notion of convergence for sequences indexed by $\mathcal{F}$, or more generally by an IP ring $\mathcal{G}$. Let X be a topological space and let $(x_\alpha)_{\alpha\in\mathcal{F}}$ be a sequence of points of X. We say that the sequence $(x_\alpha)_{\alpha\in\mathcal{F}}$ converges to a limit $x \in X$ along an IP ring $\mathcal{G}$ and we write

$$\text{IP}-\lim_{\alpha\in\mathcal{G}} x_\alpha = x$$

if for any neighbourhood U of x there exists $\alpha_0 \in \mathcal{F}$ such that for all $\alpha \in \mathcal{G}$ with $\alpha > \alpha_0$ we have $x_\alpha \in U$. Limits along IP rings are closely related to limits along idempotent ultrafilters in $\mathbb{N}$; see Appendix B for more details.

Let X be a compact topological space and let $(x_\alpha)_{\alpha\in\mathcal{F}}$ be a sequence of points of X. It is a corollary of Hindman's theorem and a diagonal argument ([43, Theorem 8.14] or [45, Theorem 1.3]) that there exists an IP ring $\mathcal{G}$ and a point $x \in X$ such that $\text{IP}-\lim_{\alpha\in\mathcal{G}} x_\alpha = x$.

The analogue of the mean ergodic theorem now takes the following form.

Theorem I.1.12 (Mean IP ergodic theorem)**.** *Let $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ be an m.p.s., and let $\mathcal{I}$ denote the factor induced by the σ-algebra of T-invariant sets. Then, for any sequence $(n_i)_{i=1}^\infty$ of integers and any $f \in L^2(\mu)$, there exists an* IP *ring $\mathcal{G}$ such that*

$$\text{IP}-\lim_{\alpha\in\mathcal{G}} T^{n_\alpha} f = \mathbb{E}(f|\mathcal{I}),$$

in the weak topology on $L^2(\mu)$.

Likewise, IP sets occur naturally in topological dynamics. The following is sometimes known as the IP Recurrence Theorem. Recall that a topological dynamical system (X, T) is distal if for $x, y \in X$, $x \neq y$, $\inf_{n\in\mathbb{N}} \text{dist}(T^n x, T^n y)$ is strictly positive, and nilrotations are distal.

Theorem I.1.13 ([43], Lemma 9.10)**.** *Let (X, T) be a minimal distal topological dynamical system. Then for every $x \in X$ and every neighbourhood U of x, the set of return times $\{n \in \mathbb{N} \mid T^n(x) \in U\}$ is an* IP* *set.*

I.1.6 Nilsystems

In recent years, it has become clear that among various measure preserving systems, the class of *nilrotations* plays a particularly significant role in ergodic theory, and its connections with additive combinatorics.

Let G be a (s-step) nilpotent Lie group, and let $\Gamma < G$ a discrete subgroup, which is cocompact in the sense that G/Γ is compact. Then, the quotient space $X = G/\Gamma$, which carries the structure of a smooth manifold, is a (s-step) *nilmanifold*. The group G has a natural action on the quotient X, given by $T_g(h\Gamma) = gh\Gamma$; in analogy with rotations on the tori, any of the maps T_g is called a *nilrotation*. There exists a unique Borel measure denoted by μ_X, namely the Haar measure, which is invariant under the action of G. The quotient space $X = G/\Gamma$ together with a nilrotation T_g, and equipped with the Borel σ-algebra $\mathcal{B}_X$ and the Haar measure μ_X is a (s-step) *nilrotation*. Note that to specify a nilsystem it is enough to specify G/Γ and T_g, so we often remove other data from the notation. If (X, T) is a (s-step) nilsystem and $f\colon X \to \mathbb{C}$ is a continuous map, then the sequence $n \mapsto f(T^n x_0)$ is a (s-step) nilsequence.

By a result of Parry [95], the properties of minimality, unique ergodicity, and ergodicity with respect to the measure μ_X are equivalent. (In fact, all this conditions can be verified on the maximal torus $G/[G, G]\Gamma$.) Furthermore, the action of G on X is distal, i.e. for $x, y \in X$, $x \neq y$, $\inf_{g\in G} \operatorname{dist}(gx, gy)$ is strictly positive.

The importance of nilsystem in modern ergodic theory stems largely from the fact that *characteristic factors* for convergence of averages such as (I.2) turn out to be inverse limits of nilsystems. A closely related fact in additive combinatorics is that a set is Gowers uniform precisely when it does not correlate with a nilsequence of bounded complexity (see Section I.1.11 for details).

More precisely, a factor $\mathcal{Z}$ of an m.p.s.'s $\mathcal{X} = (X, \mathcal{B}, \mu, T)$ is *characteristic* for the convergence of the averages

$$\mathop{\mathbb{E}}_{n<N} f_0 \cdot T^n f_1 \cdot \cdots \cdot T^{kn} f_k \tag{I.4}$$

if for any $f_0, f_1, \ldots, f_k \in L^\infty(X)$, the asymptotic behaviour of the averages (I.4) does not change if we replace f_i with $\mathbb{E}(f_i|\mathcal{Z})$, i.e. it holds that

$$\left| \mathop{\mathbb{E}}_{n<N} \prod_{i=0}^{k} T^{in} f_i - \mathop{\mathbb{E}}_{n<N} \prod_{i=0}^{k} \mathbb{E}(T^{in} f_i|\mathcal{Z}) \right| \to 0 \text{ as } N \to \infty.$$

The Host-Kra-Ziegler factors we first introduced in order to study the convergence of multiple ergodic averages; see [65] and [109]. They are most easily described in

terms of the Host-Kra-Gowers norms $\|\cdot\|_k$. Define inductively $\|f\|_0 = \int f d\mu$ for $f \in L^\infty(\mu)$, and for $k \geq 0$

$$\|f\|_{k+1}^{2^{k+1}} = \lim_{N\to\infty} \mathop{\mathbb{E}}_{n<N} \|f \cdot T^n f\|_k^{2^k},$$

where $\|f\|_{k+1} \geq 0$. It can be checked that this definition is well posed, and indeed defines a norm for $k \geq 1$. The factor $\mathcal{Z}_k$ is then characterised by the property that $\mathbb{E}(f|\mathcal{Z}_k) = 0$ if and only if $\|f\|_{k+1} = 0$, where $f \in L^\infty(\mu)$.

Theorem I.1.14. *The factor $\mathcal{Z}_k$ is characteristic for the convergence of the averages* (I.4)*, and is the inverse limit of $(k-1)$-step nilsystems.*

I.1.7 Nil-Bohr sets

The notion of a Bohr (or Bohr_0) set is classical and well-studied. A set $A \subset \mathbb{N}$ is said to be a Bohr set if it contains the preimage of an open, non-empty set U through the natural embedding of $\mathbb{N}$ in the Bohr compactification of $\mathbb{Z}$, usually denoted $b\mathbb{Z}$. Accordingly, A is a Bohr_0 set if additionally $0 \in U$.

While very satisfying from the categorical point of view, the above definition gives limited idea of what a Bohr set looks like. A more concrete description is possible. Namely, a set is Bohr if it contains a non-empty set of the form $\{n \in \mathbb{N} \mid n\alpha \in U\}$ where $\alpha \in \mathbb{T}^m = \mathbb{R}^m/\mathbb{Z}^m$ and $U \subset \mathbb{T}^m$ is open; A is Bohr_0 if additionally $0 \in U$. Hence, Bohr_0 sets can be viewed dynamically as a return-times sets for the point $0 \in \mathbb{T}^m$, under the rotation $T\colon x \mapsto x + \alpha$.

(Note that we construe being a Bohr set as a notion of largeness, hence above we only insist on containment, rather than equality. In context when precise structure is important, different definitions are used, see e.g. [104, Section 4.4].)

With the advent of higher-order Fourier analysis, a natural analogue of the class of Bohr sets has come into view. The role of the circle rotations in classical Fourier analysis is now played by *nilrotations*, which we discuss in Introduction I.1.6.

We now define a set $A \subset \mathbb{N}$ to be a Nil_d–Bohr set, in analogy to the abelian case, if it contains a non-empty set of the form

$$\{n \in \mathbb{N} \mid g^n\Gamma \in U\}$$

where G/Γ is a d-step nilmanifold, and $U \subset G/\Gamma$ is open. If additionally $e\Gamma \in U$ then A is a Nil_d–Bohr_0 set.

A useful example to keep in mind are sets of the form:

$$\{n \in \mathbb{N} \mid \|p(n)\| \leq \varepsilon\},$$

where $p \in \mathbb{R}[x]$ is a polynomial with at least one irrational non-constant coefficient. In general, such sets are Nil_d–Bohr. If additionally $p(0) = 0$ then they are Nil_d–Bohr_0. (This can be seen by a classical construction, which is discussed for instance in [16, Section 0.16]; a similar construction is used in Chapter V.)

I.1.8 Filtered groups and polynomial sequences

When studying dynamics on a nilmanifold $X = G/\Gamma$, we usually deal with "linear" maps of the form $n \mapsto g^n\Gamma$ from $\mathbb{N}_0$ to a nilmanifold G/Γ. It turns out that it natural to work with the more general class of *polynomial* maps, which we will presently define. The reason to assume this more general approach is that the "linear" sequences lack various closure properties, such as closure under products and pointwise limits. For an accessible introduction, we refer to [50] or [103].

Systematic study of polynomial sequences $\mathbb{Z} \to G$ was initiated by Leibman [80] (the results generalise easily to sequences $H \to G$ for abelian H), although some early results were obtained by Lazard [79] and others. Polynomial sequences $H \to G$ for H nilpotent are studied in [81]. The notion of a polynomial sequence whose domain is a partial semigroup, such as the partial semigroup of finite subsets of $\mathbb{N}$ equipped with disjoint union $(\mathcal{F}, \cup)$, does not explicitly appear until later, but can easily be gleaned from [80]. From a slightly different perspective, polynomial sequences $\mathcal{F} \to G$ appear in [15].

Finally, the idea of measuring the "degree" of a polynomial sequence by means of a filtration appears in [55] in context of sequences $\mathbb{Z} \to G$ (or $\mathbb{Z} \to G/\Gamma$), although most results straightforwardly generalise to different domains such as $\mathcal{F}$. To the best of our knowledge, the precise definition of a polynomial sequence $\mathcal{F} \to G$ we shall use first appears in print in the work of Zorin-Kranich [111, 110].

A Lie *prefiltration* on a Lie group G is a descending sequence $G_\bullet$ of Lie subgroups of G such that $G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots$ and for any i, j we have the nesting condition $[G_i, G_j] \subseteq G_{i+j}$. We always assume that the prefiltration terminates at some point, in the sense that for some d we have $G_{d+1} = \{e_G\}$. The least such d is the *length* of the filtration. A Lie *filtration* is a prefiltration $G_\bullet$ such that additionally $G_0 = G_1 = G$. Since we have no need to consider filtrations which are not Lie, we will usually omit this adjective. We may also keep the group G implicit, and assume that $G = G_0$.

The most important example of a filtration to bear in mind is the lower central series given by $G_0 = G_1 = G$ and $G_{i+1} := [G, G_i]$ (where $[G, H]$ denotes the group generated by elements of the form $[g, h]$ with $g \in G,\ h \in H$). Each of these groups G_i

is normal. If G is d-step nilpotent, then the length of this filtration is d. Conversely, if G has a filtration of length d, then G is at most d-step nilpotent. (Of course, in general the length of a filtration may be greater than the nilpotency class of the group.) If $G_\bullet$ is a prefiltration, we may construct a prefiltration $G'_i := G_{i+1}$ $(i \geq 0)$ on G_1. (Note that the nesting property is clear.) We denote this new prefiltration by $G_{\bullet+1}$. Images and preimages of (pre-)filtrations (by morphisms of Lie groups) are again (pre-)filtrations of (at most) the same length.

The polynomial maps taking values in a nilpotent group G classically have $\mathbb{Z}$ (or $\mathbb{Z}^m$) as their domain. However, a more general definition is possible, where the domain is an arbitrary partial semigroup. This definition is applicable to the semigroup of finite subsets of $\mathbb{N}$ (with the operation of disjoint union), which we exploit in Chapter IV (and to a lesser extent in Chapter V).

Recall that a partial semigroup $(\mathcal{S}, *)$ is a set $\mathcal{S}$ equipped with a binary operation $*$ defined of a subset of $\mathcal{S} \times \mathcal{S}$, such that $\alpha * (\beta * \gamma) = (\alpha * \beta) * \gamma$, whenever both sides are defined. By the usual abuse of notation, we usually refer to the set $\mathcal{S}$ alone as a semigroup, keeping the operation $*$ implicit.

Definition I.1.15 (Polynomial sequence; Def. 1.21 in [110])**.** Let $\mathcal{S}$ be an partial semigroup, and let $G_\bullet$ be a prefiltration on a nilpotent Lie group G. Then a map $g\colon \mathcal{S} \to G$ is declared to be polynomial (with respect to $G_\bullet$) if either $G = \{e_G\}$ is trivial and $g(\alpha) = e_G$ is the constant sequence, or if G is non-trivial and for every $\beta \in \mathcal{S}$ there exists a $G_{\bullet+1}$ polynomial $D_\beta g$ such that

$$D_\beta g(\alpha) = g(\alpha)^{-1} g(\alpha * \beta) \tag{I.5}$$

whenever $\alpha * \beta$ is defined. The set of all such polynomial maps is denoted by $\mathrm{poly}(\mathcal{S} \to G_\bullet)$. The length of the prefiltration my occasionally be referred to as the degree of the polynomial.

When $\mathcal{S} = \mathcal{F}$ or $\mathcal{S} = \mathbb{N}$, and $g \in \mathrm{poly}(\mathcal{S} \to G_\bullet)$ is such that $g(\emptyset) = e_G$, then the symmetric derivative

$$\Delta_\beta g(\alpha) = D_\beta f(\alpha) g(\beta)^{-1} = g(\alpha)^{-1} g(\alpha \cup \beta) g(\beta)^{-1} \qquad (\text{for } \beta \cap \alpha = \emptyset) \tag{I.6}$$

is a polynomial with respect to $G_{\bullet+1}$, which has the added advantage that $\Delta_\beta g(\emptyset) = e_G$. This property can also be used for as definition of polynomials, when we restrict to maps with $g(\emptyset) = e_G$. (See [110, Sec. 1.3] for details.)

Example I.1.16. If G is d-step nilpotent, then the sequence $\mathbb{N} \to G$ given by $g(n) = a^n$, $(a \in G)$, is polynomial with respect to the lower central series: indeed, $D_m g(n) = a^m$ and $D_{m_2} D_{m_1} g(n) = e$. We will shortly see that the sequence $g'(n) = a^n b^n$ $(a, b \in G)$ is also polynomial with respect to the same filtration.

Example I.1.17. Let $G = \mathbb{R}$, equipped with length d filtration $G_0 = G_1 = \cdots = G_d = \mathbb{R}$, $G_{d+1} = \cdots = \{0\}$. Then polynomial sequences $\mathbb{N} \to \mathbb{R}$ are precisely the polynomials in the conventional sense, that is sequences of the form $p(n) = \sum_{i=0}^{d} n^i a_i$ where $a_i \in \mathbb{R}$.

By the same token, for any sequence of integers (n_i) and polynomial $p\colon \mathbb{N} \to \mathbb{R}$, the map $\mathcal{F} \ni \alpha \mapsto p(n_\alpha) \in \mathbb{R}$ is polynomial with respect to the aforementioned filtration.

Example I.1.18. Let $G_\bullet$ be a prefiltration on a nilpotent Lie group G, and let $g \in \mathrm{poly}(\mathbb{N} \to G_\bullet)$. Then, for any sequence $(n_i)_i$, the map $\mathcal{F} \ni \alpha \mapsto g(n_\alpha) \in G$ is polynomial. This is a special instance of a general fact that composition of a polynomial with a morphism of partial semigroups is again a polynomial.

A principal advantage of working with polynomial maps is that they are closed under products. In a slightly more specific context, this is the main result of [80], in this form it appears in [110, Thm. 1.23] (see also Section IV.2.4).

Theorem I.1.19 (Lazard–Leibman)**.** *Let $G_\bullet$ be a prefiltration and $\mathcal{S}$ a partial semigroup. Then* $\mathrm{poly}(\mathcal{S} \to G_\bullet)$ *is a group under pointwise multiplication.*

Historically, the original definition polynomial maps stated that for any polynomial $q \in \mathbb{Q}[x]$ with $q(\mathbb{Z}) \subset \mathbb{Z}$ and any $g \in G$, the map $n \mapsto g^{q(n)}$ should be polynomial. Restricting to polynomial sequences on $\mathbb{Z}$, we have the following characterisation of polynomials, relating the two approaches. Yet another point of view on polynomial maps using Host-Kra cube groups will be exploited in Chapter IV.

Proposition I.1.20. *Let G be a nilpotent Lie group with a filtration $G_\bullet$ of length d. Then,* $\mathrm{poly}(\mathbb{Z} \to G_\bullet)$ *consists precisely of the maps $g(n)$ of the form*

$$g(n) = g_0^{\binom{n}{0}} g_1^{\binom{n}{1}} g_2^{\binom{n}{2}} \dots g_d^{\binom{n}{d}}$$

for some $g_i \in G_i$, $0 \le i \le d$.

We are fundamentally interested not in maps $\mathrm{poly}(\mathcal{S} \to G_\bullet)$, but rather in their projections onto G/Γ, where Γ is a discrete cocompact subgroup. In order for such maps to be well behaved, we need to assume that $G_\bullet$ is compatible with Γ. We will say that $G_\bullet$ is Γ-*rational* if for any i the discrete subgroup $\Gamma_i := G_i \cap \Gamma$ is cocompact in G_i. It was shown by Mal'cev[88] that the lower central series is Γ-rational with respect to any choice of Γ, under the additional assumption that G is simply connected (which we take to mean in particular connected). For the sake of brevity, if $G_\bullet$ is a Lie filtration of length d and Γ is a discrete cocompact subgroup of $G = G_0$ such that $G_\bullet$ is Γ-rational, we will say that $G_\bullet/\Gamma$ is a d-step nilmanifold.

Definition I.1.21. Let $\mathcal{S}$ be a partial semigroup and let $G_\bullet/\Gamma$ be a nilmanifold. Then a map $\bar{g}\colon \mathcal{S} \to G/\Gamma$ is declared to be polynomial (with respect to $G_\bullet$) if and only if $\bar{g}$ takes the form $\bar{g} = g \circ \pi$ where $g \in \mathrm{poly}(\mathcal{S} \to G_\bullet)$ and $\pi\colon G \to G/\Gamma$ is the standard projection $h \mapsto h\Gamma$. We denote the set of all such polynomial maps by $\mathrm{poly}(\mathcal{S} \to G_\bullet/\Gamma)$.

A crucial feature of polynomial sequences $\mathrm{poly}(\mathbb{Z} \to G_\bullet/\Gamma)$ is they constitute a compact set (in the topology of pointwise convergence), in contrast to linear sequences. A similar fact for $\mathrm{poly}(\mathcal{F} \to G_\bullet/\Gamma)$, where $\mathcal{F}$ denotes the semigroup of finite subsets of $\mathbb{N}$, will play an essential role in Chapter IV.

I.1.9 Mal'cev coordinates and generalised polynomials

When working with a nilmanifold G/Γ, it is often convenient to introduce coordinates on G and G/Γ which are compatible with the nilpotent structure. As will become clear shortly, polynomial sequences in G/Γ in these coordinates are intimately connected with generalised polynomials.

Throughout this section, we will restrict our attention to Lie groups G such that the connected component of identity G° is *simply connected*. For a more detailed introduction and proofs of various claims, we refer to [16] and [88].

Let G be a connected simply connected nilpotent Lie group. For each $g \in G$, there is a unique one-parameter subgroup $\{g^t\}_{t\in\mathbb{R}}$ of G, i.e. a continuous homomorphism $\mathbb{R} \to G$, $t \mapsto g^t$ with $g^1 = g$.

Let $G_\bullet$ be the lower central series; similar constructions are possible for arbitrary filtration, but we fix one choice for the sake of concreteness. The subgroups G_i are closed and G_i/G_{i+1} are finite dimensional $\mathbb{R}$-vector spaces with the action of $\mathbb{R}$ given by one-parameter subgroups.

Let $l_i = \dim_{\mathbb{R}} G_i/G_{i+1}$, $k_i = \sum_{j=1}^{i} l_j$, $0 \le i \le d$. Then G has a *Mal'cev basis*, i.e. elements $e_1, \dots, e_k \in G$, $k = k_d$, such that $e_{k_{i-1}+1}, \dots, e_{k_i} \in G_i$ and their images in G_i/G_{i+1} constitute a basis of G_i/G_{i+1} as a $\mathbb{R}$-vector space.

It follows easily that any element $g \in G$ can be written uniquely in the form

$$g = e_1^{t_1} \cdots e_k^{t_k}, \quad t_i \in \mathbb{R}.$$

Furthermore, it is possible to choose a Mal'cev basis to be compatible with Γ, i.e. so that $g = e_1^{t_1} \cdots e_k^{t_k}$ is in Γ if and only if $t_1, \dots, t_k \in \mathbb{Z}$. We will always assume that our Mal'cev bases are compatible with Γ. A choice of a Mal'cev basis determines a diffeomorphism

$$\tilde{\tau} \colon G \to \mathbb{R}^k, \quad e_1^{t_1} \cdots e_k^{t_k} \mapsto (t_1, \dots, t_k).$$

Under this identification, multiplication in G is given by polynomial formulae in variables t_i with rational coefficients. These polynomials take integer values if all the variables are integers.

Let $[0,1)^k = \{(t_1, \dots, t_k) \in \mathbb{R}^k \mid 0 \le t_i < 1, 1 \le i \le k\}$ be the unit cube. The preimage $D = \tilde{\tau}^{-1}([0,1)^k) \subset G$ is the *fundamental domain*. For any $g \in G$, there is a unique choice of elements $\{g\} \in D, [g] \in \Gamma$ (called the *fractional part* and the *integral part* of g) such that $g = \{g\}\,[g]$. Note that the elements $\{g\}$ and $[g]$ depend on the choice of the Mal'cev basis. Since the Mal'cev basis is compatible with Γ, the map $g \mapsto \tilde{\tau}(\{g\})$ factors to a bijection

$$\tau \colon G/\Gamma \to [0,1)^k,$$

which will play a crucial rôle. While τ is not continuous, its inverse τ^{-1} is.

The choice of the Mal'cev basis also induces a natural choice of a metrics on G and G/Γ, (cf. [54, Definition 2.2]). The metric d_G on G is characterised as the largest metric with $d_G(g,h) \le \|\tilde{\tau}(gh^{-1})\|_\infty$, and the metric $d_{G/\Gamma}$ on G/Γ is given by $d_{G/\Gamma}(g\Gamma, h\Gamma) = \inf_{\gamma \in \Gamma} d_G(g, h\Gamma)$.

A *horizontal character* on $X = G/\Gamma$ is a non-trivial morphism of groups $\eta \colon G \to \mathbb{R}/\mathbb{Z}$ with $\Gamma \subset \eta^{-1}(0)$, and is necessarily of the form $\eta(g) = \sum_{i=1}^{l_1} a_i \tau_i(g\Gamma)$, where $\tau = (\tau_1, \dots, \tau_k)$ and $a_i \in \mathbb{Z}$ are not all 0. The *norm* of η, denoted $\|\eta\|$, is by definition the Euclidean norm $\|(a_i)\|_2$.

A set $A \subset \mathbb{R}^k$ is called semialgebraic if it is a finite union of subsets of $\mathbb{R}^k$ given by a system of finitely many polynomial equalities and inequalities; in particular, A is Borel and either A has nonempty interior or it is of Lebesgue measure zero. A map $p \colon A \to \mathbb{R}^k$ defined on a semialgebraic set $A \subset \mathbb{R}^k$ is called *piecewise polynomial* if

there is a decomposition $A = A_1 \cup \ldots \cup A_s$ of A into a finite union of semialgebraic sets such that $p|_{A_i} \colon A_i \to \mathbb{R}^l$ is a polynomial mapping restricted to A_i. We call a map $f \colon G \to \mathbb{R}^l$ a *polynomial map* if the composed map $f \circ \tilde{\tau}^{-1} \colon \mathbb{R}^l \to \mathbb{R}^k$ is a polynomial map. While the map $\tilde{\tau} \colon G \to \mathbb{R}^k$ depends on the choice of a Mal'cev basis, the concept of a polynomial map on G does not.

A subset $A \subset X = G/\Gamma$ is called *semialgebraic* if its image $\tau(A) \subset [0,1)^k$ is semialgebraic. We call a map $p \colon X \to \mathbb{R}^l$ *piecewise polynomial* if it takes the form $p = q \circ \tau$, where $q \colon [0,1)^k \to \mathbb{R}^l$ is a piecewise polynomial map. As before, the concept of a piecewise polynomial map on X does not depend on the choice of a Mal'cev basis. Note that if $p \colon X \to \mathbb{R}^l$ is a piecewise polynomial map and $g \in G$, then so is $q \colon X \to \mathbb{R}^l$ given by $q(h\Gamma) = p(gh\Gamma)$.

We now extend the above definitions to not necessarily connected simply connected Lie groups G. A Mal'cev basis of G is simply a Mal'cev basis of its connected component G°. Similarly, we define piecewise polynomial maps $p \colon X \to \mathbb{R}^l$ in terms of their restrictions to the connected components of X by demanding that for each $g \in G$ the map $G^\circ/(\Gamma \cap G^\circ) \to \mathbb{R}^l$, $h(\Gamma \cap G^\circ) \mapsto p(gh\Gamma)$, is piecewise polynomial. (It is enough to verify this condition for a single g in each class $gG^\circ \in G/G^\circ$.) We extend the notion of the fractional and the integral part to the case of possibly disconnected G, but we define and use it only if the nilmanifold G/Γ is nevertheless connected. In this case the map $G^\circ \to G/\Gamma$ is surjective and induces a diffeomorphism $G^\circ/\Gamma' \to G/\Gamma$, where $\Gamma' = \Gamma \cap G^\circ$. Thus, we may define the fractional part with respect to G°/Γ', and define the integral part of $g \in G$ so that $g = \{g\}\,[g]$. (We can do this since these notions do not involve dynamics, and depend only on the nilmanifold; in dynamical considerations, we are still interested in the action of the (disconnected) group G on the nilmanifold $X = G/\Gamma \simeq G^\circ/\Gamma'$.)

As mentioned before, there is a close relation between generalised polynomials an coordinates of orbits in nilsystems.

Generalised polynomials are sequences which can be obtained from standard polynomials by allowing the use of the floor function. More precisely, the generalised polynomials from $\mathbb{Z}$ to $\mathbb{R}$ are defined to be the smallest class of functions $\mathbb{Z} \to \mathbb{R}$ containing the usual polynomial sequences $\mathbb{R}[x]$ and closed under addition, multiplication, and the operation of taking the integer part. Hence, generalised polynomials include sequences such as $f(x) = \left\lfloor \sqrt{3} \left\lfloor \sqrt{2}n^2 + \pi \right\rfloor + \sqrt{5}n^2 \right\rfloor$.

The connection between generalised polynomials and nilsystems is made precise by the main result of [16, Theorem A].

Theorem I.1.22 (Bergelson-Leibman)**.**

(i) *If $X = G/\Gamma$ is a nilmanifold, $g \in G$ acts on X by left translations, $p\colon X \to \mathbb{R}$ is a piecewise polynomial map, and $x \in X$, then $u\colon \mathbb{Z} \to \mathbb{R}$ given by $u(n) = p(g^n x), n \in \mathbb{Z}$, is a bounded generalised polynomial.*

(ii) *If $u\colon \mathbb{Z} \to \mathbb{R}$ is a bounded generalised polynomial, then there exists a nilmanifold $X = G/\Gamma$, $g \in G$ whose action on X by left translations is ergodic (equivalently: minimal), a piecewise polynomial map $p\colon X \to \mathbb{R}$, and $x \in X$ such that $u(n) = p(g^n x)$, $n \in \mathbb{Z}$.*

While the above theorem concerns "linear" orbits, we note that since in Mal'cev coordinates multiplication is given by polynomial formulas, an apparently stronger statement is true. Namely, if $G_\bullet/\Gamma$ is a nilmanifold and $g \in \mathrm{poly}(\mathbb{Z} \to G/\Gamma)$, and $p\colon G/\Gamma \to \mathbb{R}$ is a piecewise polynomial map, then $p \circ g$ is a generalised polynomial.

I.1.10 Equidistribution

A sequence of points $(x_n)_{n=1}^\infty$ in a compact topological space X is *equidistributed* with respect to a measure μ if for test function any $f \in \mathcal{C}(X)$ we have

$$\mathop{\mathbb{E}}_{n<N} f(x_n) \to \int_X f d\mu \text{ as } N \to \infty.$$

For instance, an easy consequence of the pointwise ergodic theorem is that if $\mathcal{X}$ is an m.p.s., then for μ-a.e. $x \in X$, the sequence $(T^n x)_{n=1}^\infty$ is equidistributed.

In applications, it is often useful to have quantitative control over the quality of equidistribution. For a sequence $(x_n)_{n=1}^\infty \in X$ of points in a compact metric space endowed with a probability measure μ, we shall say, following terminology e.g. in [54], that (x_n) is (δ, N)-*equidistributed* if and only if for each $f \in \mathrm{Lip}(X; \mathbb{R})$ we have

$$\left| \mathop{\mathbb{E}}_{n<N} f(x_n) - \int_X f d\mu \right| \leq \delta \left\| f \right\|_{\mathrm{Lip}}.$$

Here, $\left\| f \right\|_{\mathrm{Lip}}$ denotes the Lipschitz norm $\left\| f \right\|_{\mathrm{Lip}} = \left\| f \right\|_\infty + \sup_{x,y \in X} \frac{|f(x)-f(y)|}{d_X(x,y)}$, and $\mathrm{Lip}(X; \mathbb{R}) \subset C(X; \mathbb{R})$ denotes the space of those f with $\left\| f \right\|_{\mathrm{Lip}} < \infty$.

Although in many applications we are ultimately interested in density, equidistribution often turns out to be easier to work with. Of course, not every dense sequence is equidistributed. However, equidistribution implies density, if we allow for a slight change in the parameters. The following observation is elementary.

Observation. Suppose that X is a d-dimensional compact smooth manifold equipped with a Riemannian metric and with a measure μ arising from a volume form. Then there exists a constant $c > 0$, such that the following is true. Let $\delta > 0$, and suppose that a sequence x_n is $(c\delta^d, N)$-equidistributed. Then for any $x \in X$, there exists n such that $d_X(x, x_n) < \delta$.

Proof. Let x_n be (δ, N)-equidistributed, and suppose that x_n is not $(\tilde{\delta}, N)$-dense. Thus, we can choose y such that $d(x_n, y) \geq \tilde{\delta}$ for all $n < N$. We may consider the function f given by $f(x) := \max(0, \tilde{\delta} - d(x, y))$. It is easy to verify that $\int_X f d\mu \geq c\tilde{\delta}^d$, where c depends only on X. We also have $\|f\|_{\text{Lip}} \leq 1$ and $f(x_n) = 0$ for any $n < N$. Equidistribution now implies:

$$c\tilde{\delta}^d \leq \left| \int_X f d\mu - \mathop{\mathbb{E}}_{n<N} f(x_n) \right| \leq \delta.$$

Hence $\tilde{\delta} \ll \delta^{1/d}$, and consequently the sequence x_n is $(O(\delta^{1/d}), N)$-dense. □

It is a classical result of Weyl that lack of equidistribution of a polynomial orbit on the torus can always be explained by a rational obstruction. We have the following classical theorem, which we state in a slightly unusual for to make the relation to the quantitative form more visible. (see [35, Thm. 1.4]).

Theorem I.1.23 (Weyl equidistribution)**.** *For any d there exist a family of constants $N_0(p, \delta)$ such that the following is true.*

Let $p(n) = (p_i(n))_{i=1}^d$ be a polynomial sequence in $\mathbb{T}^d$. Suppose that $p(n)$ is not (δ, N)-equidistributed. Then either $N < N_0(p, \delta)$, or there exists $k \in \mathbb{Z}^d \setminus \{0\}$ such that if $\sum_i k_i p_i = \sum_j \alpha_j n^j$, then $\alpha_j \in \mathbb{Z}$ for all j.

We will need a quantitative version of the above theorem. The following result is a special case of the quantitative equidistribution theorem for nilmanifolds of Green and Tao, [54, Theorem 1.16], which will suffice for all of our applications.

Theorem I.1.24 (Quantitaive Leibman Theorem)**.** *Let $G_\bullet/\Gamma$ be a nilmanifold. Then, there exists a constant $C > 0$ such that the following is true.*

Let $g \in \text{poly}(\mathbb{Z} \to G_\bullet)$, and suppose that $g(n)\Gamma$ is not (δ, N)-equidistributed. Then, there exists a horizontal character $\eta \colon G \to \mathbb{T}$ such that $|\eta| \ll \delta^{-C}$ and

$$\|\eta(g(n)) - \eta(g(n+1))\| \ll \delta^{-C}/N,$$

for all $n < N$.

In particular, we have the following equidistribution result for polynomials, also mentioned in [54].

Theorem I.1.25 (Quantitative Weyl Theorem)**.** *For any d, r there exists a constant C such that the following is true.*

Let $p(n) = (p_i(n))_{i=1}^d$ be a polynomial sequence in $\mathbb{T}^d$ with $\deg p = r$. Suppose that $p(n)$ is not (δ, N)-equidistributed. Then there exists $k \in \mathbb{Z}^d \setminus \{0\}$ such that $k_i \ll \frac{1}{\delta^C}$ and if we write $\sum_i k_i p_i = \sum_j \alpha_j n^j$ then $\|\alpha_j\| \ll \frac{1}{N^j \delta^C}$.

I.1.11 Higher order Fourier analysis

The uniformity norms, known also as the Gowers' norms, were first introduced by Gowers in his work on an new proof of Szemeredi's theorem [48]. These norms now play a crucial role in additive combinatorics. For exposition of the relevant theory, we refer to [50] and [103].

Definition I.1.26. Fix $s \in \mathbb{N}$. For $N \in \mathbb{N}$ and $f\colon [N] \to \mathbb{C}$, the s-th Gowers uniformity norm of f is defined by

$$\|f\|_{U^s[N]}^{2^s} := \mathop{\mathbb{E}}_{n,\mathbf{h}} \prod_{\omega \in \{0,1\}^s} \mathrm{C}^{|\omega|} f(n + \omega \cdot \mathbf{h}) \tag{I.7}$$

where $\omega \cdot \mathbf{h} = \sum_{i=1}^s \omega_i h_i$, C denotes complex conjugation, and the expectation is taken over all $n \in \mathbb{Z}$, $\mathbf{h} \in \mathbb{Z}^s$ for which the cube $\{n + \omega \cdot \mathbf{h} \mid \omega \in \{0,1\}^s\}$ is contained in $[N]$.

A sequence is (informally) said to be Gowers uniform of order s if its $U^s[N]$-norm is small. It is not immediately obvious that the definition is well posed. However, we have the following basic facts.

Proposition I.1.27. *Let $s \in \mathbb{N}$ and $N \in \mathbb{N}$ and $f, g\colon [N] \to \mathbb{C}$. Then the following hold.*

(i) The expression on the right hand side of (I.7) *is ≥ 0, and hence $\|f\|_{U^s[N]}$ is well defined.*

(ii) We have the triangle inequality $\|f + g\|_{U^s[N]} \leq \|f\|_{U^s[N]} + \|g\|_{U^s[N]}$.

(iii) We have the nesting of consecutive norms: $\|f\|_{U^s[N]} \ll \|f\|_{U^{s+1}[N]}$, where the implicit constant depends only on s.

The reason why Gowers norms are so useful in additive combinatorics is that they control the count of various linear patterns. Which of the norms is needed to control a particular pattern depends on its complexity, which we currently define following [53]. (For the origin of the name, see [49]; for the origin of the Generalised von Neumann Theorem see [52].)

Definition I.1.28 (Cauchy-Schwarz complexity)**.** The Cauchy-Schwarz complexity of a system of affine maps $\varphi_1, \dots, \varphi_t \colon \mathbb{Z}^d \to \mathbb{Z}$ is the least integer s such that for each $1 \leq i \leq t$ there exists a partition of $\{\varphi_j \mid j \neq i\}$ into $s+1$ cells $C_0, \dots, C_s$, such that φ_i does not lie in the affine-linear span of C_r for any $0 \leq r \leq s$. If no such s exists, then the Cauchy-Schwarz complexity of $\varphi_1, \dots, \varphi_t$ is ∞.

Theorem I.1.29 (Generalised von Neumann)**.** *Let $\varphi_1, \dots, \varphi_t \colon \mathbb{Z}^d \to \mathbb{Z}$ be a system of affine maps with Cauchy-Schwarz complexity s, and let $f_1, \dots, f_t \colon [N] \to \mathbb{C}$ be functions $|f_i| \leq 1$ for $1 \leq i \leq t$. Then,*

$$\left| \mathop{\mathbb{E}}_{\mathbf{n}} \prod_{i=1}^{t} f_i(\varphi_i(\mathbf{n})) \right| \ll \min_{1 \leq i \leq t} \|f_i\|_{U^{s+1}[N]},$$

where the average is taken over all $\mathbf{n} \in \mathbb{Z}^d$ with $\varphi_i(\mathbf{n}) \in [N]$ for $1 \leq i \leq t$.

The Inverse Theorem for Gower norms is a landmark result, connecting Gowers norms and nilsequences. Informally speaking, it asserts that for a bounded sequence $f \colon [N] \to \mathbb{C}$ the following conditions are equivalent:

(i) the Gowers norm $\|f\|_{U^s[N]}$ is bounded away from 0;

(ii) there exists a $(s-1)$-step nilsequence φ of bounded complexity such that the correlation $\mathbb{E}_{n \in [N]} f(n)\varphi(n)$ is bounded away from 0.

Above, all bounds are allowed to depend on s, but are uniform in N. We refrain from properly stating the Inverse Theorem, since that would necessitate defining complexity of a nilsequence, which we prefer to avoid.

I.2 Overview

Each chapter of this thesis is fairly self-contained, and corresponds more or less precisely to one of the papers by the author.

Chapter II

Chapter II, based on [73] deals with additive properties of sets of the form

$$A = \left\{ n \in \mathbb{N}_0 \;\middle|\; |p(n) \bmod 1| \leq \varepsilon(n) \right\},$$

where $p \in \mathbb{R}[x]$ and $\varepsilon \colon \mathbb{N}_0 \to \mathbb{R}_{>0}$. Specifically, we are asking if these sets are bases of order 2 or, barring that, if the sumset $A+A$ has density 1. There are various regimes where these questions can be asked, but arguably the most interesting case is when $p(n) = \alpha n^2$ and $\varepsilon(n) = \varepsilon$ is constant, to which we will for now restrict.

We study the close connection which exists between largeness of the sumset $A+A$, equidistribution of the sequence $(p(n) \bmod 1)$, and good rational approximations of α (or — of the coefficients of $p(n)$ in higher degrees). In fact, we are able to show that there is an almost one-to-one correspondence between elements in the complement of $A + A$ and good rational approximations to α of a certain specific form. Establishing this connection involves combining some elementary number theory with known equidistribution results for polynomial orbits on tori, specifically the quantitative equidistribution theorem of Green and Tao.

Once the importance of the rational approximations of α becomes clear, it is natural to apply the theory of continued fractions. We are able to use it to show that, in a variety of situations, the sets A are *not* bases of order 2, which initially came as a surprise. However, as a result of the fact that good rational approximations tend to occur rather sparsely, we also show that $A+A$ has full density, provided that α is irrational. Finally, by carefully rigging the expansion of α, we are able to show that A can be bases of order 2 for arbitrarily small ε.

There are many other results obtained in this section. In the case when α is a quadratic surd, we are able to obtain a much more detailed picture of the situation, ultimately owning to the extremely simple continued fraction expansions, and the resulting simple structure of the good rational approximations. These results have a more algorithmic flavour, and exploit some elementary algebraic number theory.

We also study the question of A being bases of order 3, which is interesting when we allow $\varepsilon(n)$ to tend to 0 slowly with $n \to \infty$. Because we are now dealing with equations in three variables, the circle method becomes applicable, and we are able to allow $\varepsilon(n) \sim n^{-\delta}$ for sufficiently small $\delta \leq \delta_0(\alpha)$.

Moving to the case when the polynomial $p(n)$ appearing in the definition of A has a larger degree, the situation becomes significantly simpler. This is because lack of equidistribution of a higher-degree polynomial imposes more stringent conditions on

the coefficients. Consequently, the generic behaviour of A changes: we now show that the sets A usually are bases of order 2. It is still possible for the A's to not be bases of order 2, for instance when $p(n)$ is too well approximable by rational polynomials.

Chapter III

Chapter III is based on [76]. It deals with weakly mixing sets of integers (defined therein), and in particular shows that certain classes of polynomial equations admit solutions inside a weakly mixing set. This direction is largely inspired by earlier work in Chapter II. We show, among other things, that if A is a weakly mixing set of integer, and $p(n)$ is a polynomial of odd degree, then any sufficiently large integer N has a representation $N = n_1 + n_2$ where $p(n_1) + m,\ p(n_2) + m,\ m \in A$ for some m. A similar result also holds for $p(n)$ of even degree, although the claim may fail for N belonging to a set of zero density.

This chapter extensively uses methods of ergodic theory. In fact, our first step is to reduce the issue of solvability of systems of equations to proving a certain uniform convergence theorem for polynomial averages in weakly mixing dynamical systems. For instance, in order to prove the aforementioned result that any large N has a representation $N = n_1 + n_2$ where $p(n_1) + m,\ p(n_2) + m,\ m \in A$, we are led to study the averages such as

$$\mathop{\mathbb{E}}_{n<N} f_0 \cdot T^{p(n)} f_1 \cdot T^{p(N-n)} f_2,$$

where $(X, T, \mathcal{B}, \mu)$ is a weakly mixing measure preserving system and $f_0, f_1, f_2 \in L^2(\mu)$. More generally, we study averages of the form

$$\mathop{\mathbb{E}}_{n<N} f_0 \cdot T^{p_1^{(N)}(n)} f_1 \cdot T^{p_2^{(N)}(n)} f_2, \qquad (*)$$

where $p_i^{(N)}(n)$ are polynomials which depend on N.

Convergence of averages such as $(*)$, but with the polynomials $p_1(n), p_2(n)$ independent of N, is a classical result of Bergelson, established by means of PET induction. By a careful adaptation of Bergelson's methods, we are able to prove the uniform version of this theorem, under some natural constraints on the polynomials $p_i^{(N)}$. This is sufficient establish convergence of $(*)$, and hence to deal with the case of odd degree polynomials in the original motivation.

In order to deal with polynomials of even degrees, we work with a more constrained class of polynomial sequences, whose coefficients are again given by polynomial formulas: $p_1^{(N)}(n) = p(n),\ p_2^{(N)}(n) = p(N-n)$ is one such example. We show that then

a uniform convergence theorem holds after removing a zero-density set of N's, which translates into convergence in $(*)$.

Chapter IV

Chapter IV is based on [74]. It provides a partial combinatorial characterisation of nil–Bohr sets. Specifically, it is proved that d-step nil–Bohr sets are closely related to a purely combinatorial class of SG_k^* sets (which are a variant of IP sets) for an appropriate choice of k.

The connection between nil–Bohr sets and SG_k^* was first discovered by Host and Kra who proved that any SG_d^* contains large pieces of a d-step nil–Bohr set. We prove the reverse implication: a d-step nil–Bohr set is necessarily SG_k^* for some k, which can be effectively bounded in terms of d.

We are naturally led to study the class of polynomial maps from the semigroup of the finite subsets of $\mathbb{N}$ into a nilmanifold. Our main result is a recurrence statement for such polynomials, from which the characterisation mentioned above follows easily. The proof relies on a good understanding of polynomial maps to the torus, and on an inductive argument exploiting the structure of a nilmanifold as a tower of extensions.

Chapter V

Chapter V is based on [24] and [23], which is joint work with Jakub Byszewski. It stemmed from the seemingly easy question: *Are there any non-trivial examples of automatic sequences given by generalised polynomials?*. Here, a generalised polynomial is an expression built from standard polynomials and the floor function, using standard arithmetic operations, and an automatic sequence is one produced by a finite computing device. Both classes of sequences include all sequences which are periodic starting from some point, and the question pertains to existence of other examples.

With relative ease, we obtain a density-1 result, stating that if a sequence is both automatic and generalised polynomial, then it is periodic away from some exceptional set whose asymptotic density is equal to 0. This is proved using the basic properties of nilrotations, and the close connection between generalised polynomials and nilsequences.

When it comes to the original question, we obtain a conditional result: For each k either there are no non-trivial base k automatic sequences given by generalised polynomials, or the characteristic sequence of the powers of k is given by a generalised

polynomial. Unconditionally, we obtain a rather rigid structural description of the exceptional set mentioned above.

This is achieved by relating automatic sequences and generalised polynomials to the class of IPS sets we introduce in analogy with IP sets. On one hand, we show that if a set is given by a generalised polynomial and contains an IPS set, then it needs to have positive density, which may be seen as complementing results of Bergelson and Leibman asserting combinatorial richness of generalised polynomial sets. On the other hand, we are able to give a complete description of all $\{0,1\}$-valued automatic sequences whose set of 1's does not contain an IPS set. The main theorem then follows by combining these two statements and the density-1 result.

We also give several examples of sequences which are somewhat unexpectedly given by generalised polynomials. Specifically, we show that the characteristic sequence of the set of Tribonacci numbers ($T_0 = 0,\ T_1 = T_2 = 1,\ T_{i+3} = T_{i+2} + T_{i+1} + T_i$) is generalised polynomial (the same result for Fibonacci numbers is much simpler). Moreover, the same conclusion holds for any set $\{n_i \mid i \geq 0\} \subset \mathbb{N}$ whose elements satisfy $n_{i+1} \geq n_i^{1+c}$ for a constant $c > 0$.

Chapter VI

Chapter V is based on [75]. In this chapter, we estimate the Gowers norms of two classical automatic sequences: the Thue-Morse and the Rudin-Shapiro sequence.

The Thue-Morse $t(n)$ sequence takes the value $+1$ or -1, depending on whether the number of 1's in the binary representation of n is even or odd. The Rudin-Shapiro sequence $r(n)$ is defined similarly, except that instead of 1's we count pairs of consecutive 1's. Various pseudorandomness properties of these sequences have been extensively studied, and we continue this line of inquiry by proving that both of these sequences are Gowers uniform. More precisely, for each s we have the estimates $\|t\|_{U^s[N]} = O(N^{-c})$ and $\|r\|_{U^s[N]} = O(N^{-c})$ for some $c > 0$.

As a consequence, we obtain asymptotics for the count of various linear patterns in these sequences, with a power saving in the error term. We point out that these are possibly the simplest sequences whose all Gowers norms are known to be small.

Similar results can be obtained for sequences counting different patterns by analogous arguments. We also conjecture that the same conclusion should hold for much more general classes of automatic sequences.

Chapter II

Nil–Bohr type sets as bases for the positive integers

II.1 Introduction

In this chapter we investigate additive properties of sets of the form

$$A = \{n \in \mathbb{N}_0 \mid \|p(n)\| \leq \varepsilon(n)\},$$

where $p(n)$ is a real polynomial and $\varepsilon\colon \mathbb{N}_0 \to \mathbb{R}_{>0}$ is a slowly decaying function, and we recall that $\|t\| = \min_{n\in\mathbb{Z}} |t-n|$. Note that when $\varepsilon(n) = \varepsilon_0$ is constant, these sets are nil–Bohr sets, as explained in Section I.1.7.

Specifically, we ask when is A an basis for $\mathbb{N}_0$ of a given finite order? That is, when does the k-fold sumset $kA = \{a_1 + \cdots + a_k \mid a_i \in A\}$ contain all sufficiently large integers? We will also be interested in when A is an *almost* basis of order k, by which we mean that kA has asymptotic density equal to 1. In fact, in all cases when we can prove that $d(kA) = 1$, our argument in fact gives a stronger conclusion $d^*(\mathbb{N}_0 \setminus kA) = 0$, where $d^*(B)$ denotes the Banach density of the set $B \subset \mathbb{N}_0$.

We consider two types of behaviour of $\varepsilon(n)$: we either demand that $\varepsilon(n) \to 0$, or that $\varepsilon(n)$ is bounded pointwise by a suitably small constant ε_0 (in which case we may equally well assume that $\varepsilon(n) = \varepsilon_0$). This technical issue will appear at various points in this chapter.

In the case when $\deg p = 1$, the problem is rather straightforward. We are essentially dealing with Bohr sets, which are simple and well studied objects (see e.g. [104, Chapter 4.4] or Section I.1.7). We expect that the sets kA should not be significantly larger than A, and hence that A should not be a basis of any order for sufficiently small ε. This expectation is formalised in the following proposition.

Proposition II.1.1. *Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, $\beta \in \mathbb{R}$ and $k \in \mathbb{N}$. Suppose $\varepsilon(n)$ is a function such that either $\varepsilon(n) \to 0$ or $\varepsilon(n) \leq \varepsilon_0 < \frac{1}{2k}$ for each n. Then the set*

$$A = \{n \in \mathbb{N}_0 \mid \|\alpha n + \beta\| \leq \varepsilon(n)\}$$

is not a basis of order k.

Proof. Our aim is to prove that there exists arbitrarily large integers N which do not admit a decomposition $N = \sum_{j=1}^{k} n_j$ with $n_j \in A$. Since α is irrational, for any $\gamma \in \mathbb{T}$ and $\delta > 0$ we may find arbitrarily large $N \in \mathbb{N}_0$ such that $\|N\alpha - \gamma\| \leq \delta$.

In the situation when $\varepsilon(n) \leq \varepsilon_0$ is bounded pointwise, take $\gamma = k\beta + \frac{1}{2}$ and $\delta < \frac{1}{2} - k\varepsilon_0$. If for some N with $\|N\alpha + \gamma\| \leq \delta$ we had a decomposition $N = \sum_{j=1}^{k} n_j$ with $n_j \in A$, then it would follow that

$$\frac{1}{2} - \delta \leq \|N\alpha + k\beta\| \leq \sum_{j=1}^{k} \|n_j\alpha + \beta\| \leq k\varepsilon_0,$$

which contradicts the choice of δ. Hence, $N \notin kA$, and consequently A is not a basis of order k.

In the situation when $\varepsilon(n) \to 0$, take $\gamma = k\beta + \frac{1}{2} + \gamma'$, where γ' is linearly independent from $\alpha, \beta, 1$ over $\mathbb{Q}$, and let $N_i \to \infty$ be such that $\|N_i\alpha - \gamma\| \to 0$ as $i \to \infty$. Suppose for the sake of contradiction that we have for each i a decomposition $N_i = \sum_{j=1}^{k} n_{i,j}$ with $n_{i,j} \in A$. We may assume that (possibly after rearranging the summands and passing to a subsequence) there is some l such that $n_{i,j} = n_j$ does not depend on i for $1 \leq i \leq l$ and $n_{i,j} \to \infty$ for $l < i \leq k$. Since for $l < i \leq k$ we have $\|n_{i,j}\alpha + \beta\| \to 0$, it follows that

$$\frac{1}{2} = \lim_{i\to\infty} \|N_i\alpha + k\beta + \gamma'\| \leq \left\| \sum_{j=1}^{l} (n_j\alpha + \beta) + \gamma' \right\|.$$

However, this would imply $\sum_{j=1}^{l}(n_j\alpha + \beta) + \gamma' \equiv \frac{1}{2} \pmod{\mathbb{Z}}$, which is impossible because of the choice of γ'. □

The problem is most interesting when $\deg p = 2$. One might expect A to behave roughly as a random set such that $n \in A$ with probability $\varepsilon(n)$, and hence to be a basis of finite order if $\varepsilon(n)$ decays reasonably slowly. A particular case of this problem was considered by Erdős, who asked the following[1].

Question II.1.2. *Is the set $A = \left\{ n \in \mathbb{N}_0 \;\middle|\; \left\|\sqrt{2}n^2\right\| \leq \frac{1}{\log n} \right\}$ a basis of order 2?*

[1]Personal communication from Ben Green; no written reference could be located.

Somewhat unexpectedly, the answer to this question is negative. One can even produce an explicit sequence $N_i = \frac{(3+2\sqrt{2})^{2i+1}-(3-2\sqrt{2})^{2i+1}}{2\sqrt{2}}$ such that $N_i \notin 2A$ for all sufficiently large i, where the important point is that $(3+2\sqrt{2})$ is a unit in $\mathbb{Q}[\sqrt{2}]$.

Several other constructions of this type are possible, each leading to a sequence $N_i \notin 2A$ with N_i growing exponentially with i. Hence, one may hope that the following weaker variant should have a positive answer. Recall that we call A an almost basis of order 2 if $d(2A) = 1$.

Question II.1.3. *Is the set* $A = \left\{ n \in \mathbb{N}_0 \;\middle|\; \left\|\sqrt{2}n^2\right\| \leq \frac{1}{\log n} \right\}$ *an almost basis of order 2?*

This is indeed the case. In fact, we can prove a stronger statement concerning the size of the complement $(2A)^c = \mathbb{N}_0 \setminus 2A$, namely that as $T \to \infty$ we have $|[T] \setminus 2A| \ll \log^C T$, where C is a constant.

In larger generality, we have the following collection of results.

Theorem II.1.4. *For* $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ *and function* $\varepsilon \colon \mathbb{N}_0 \to \mathbb{R}_{>0}$*, define*

$$A = \left\{ n \in \mathbb{N}_0 \;\middle|\; \left\|\alpha n^2\right\| \leq \varepsilon(n) \right\}.$$

Then the following statements hold.

(i) For any $\alpha \in \mathbb{R} \setminus \mathbb{Q}$*,* A *is an almost basis of order* 2*, provided that* $\varepsilon(n)$ *decays slowly enough.*

(ii) Moreover, for uncountably many exceptional values of α*,* A *is a basis of order* 2*, provided that* $\varepsilon(n)$ *decays slowly enough.*

(iii) In particular, for any $\alpha \in \mathbb{R} \setminus \mathbb{Q}$*,* A *is a basis of order* 3*, provided that* $\varepsilon(n)$ *decays slowly enough.*

(iv) There exists a constant $\delta > 0$ *such that for almost all* $\alpha \in \mathbb{R}$*,* A *is a basis of order* 3*, provided that* $\varepsilon(n) \gg 1/n^{\delta}$*.*

(v) However, for almost all α*,* A *is not a basis of order* 2*, as long as* $\varepsilon(n) \to 0$ *or* $\varepsilon(n) \leq \varepsilon_0$*, where* $\varepsilon_0 < \frac{1}{4}$ *is a constant.*

Above, the phrase "provided that $\varepsilon(n)$ decays slowly enough" may be expanded into "there exists $\varepsilon_0(n) \to 0$ such that if $\varepsilon(n) \geq \varepsilon_0(n)$ for all n, then the statement holds"; choice of ε_0 is allowed to depend on α. Likewise, "almost all" means "all except for a set of Lebesgue measure 0". We state the results in a more precise

manner when we approach the proofs. In fact, we will prove somewhat more, but at this point we include more succinct if slightly weaker statements.

We first address item II.1.4.v, which was the original motivation for this line of investigation. Because of II.1.4.ii, we cannot hope to obtain a result for all α, but we are able to cover a number of interesting cases, including Lebesgue-almost all reals (as quoted in the theorem), as well as all quadratic surds. This is done in Section II.2. Interestingly, for quadratic surds one cannot take ε_0 close to $\frac{1}{4}$, but we can explicitly compute upper bound for the largest possible value of ε_0; this is done is Section II.4.

Items II.1.4.i and II.1.4.ii are proved in Section II.3. Our key idea is to translate information about the complement of $2A$ into information about good rational appoximations of α. We are then able to use known equidistribution results as a black box, in order to show that if $(2A)^c$ had positive (upper asymptotic) density, then α would have too many good rational approximations. Item II.1.4.ii is proved by an explicit construction using continued fractions.

Item II.1.4.iii is an immediate consequence of II.1.4.i (or, strictly speaking, the proof thereof). In fact, our argument implies that $2A + B$ contains all sufficiently large integers for any set B with at least 2 elements; see remarks after the proof of Theorem II.3.4 for details.

Finally, item II.1.4.iv is obtained by means of the circle method in Section II.5. We prove a slightly more precise statement which holds for all sufficiently badly approximable α's.

For polynomials of higher degrees $\deg p \geq 3$, the situation becomes much simpler. The heuristic expectation that A should be a basis of order 2 is accurate in this case, as long as we impose the suitable genericity assumptions. Below we give a special case of our main result for polynomials of degree ≥ 3.

Theorem II.1.5. *Fix $d \geq 3$. For $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and function $\varepsilon\colon \mathbb{N}_0 \to \mathbb{R}_{>0}$, define*

$$A = \left\{ n \in \mathbb{N}_0 \;\middle|\; \left\| \alpha n^d \right\| \leq \varepsilon(n) \right\}.$$

Then the following statements hold.

(i) For almost all α, A is a basis of order 2, provided that $\varepsilon(n)$ decays slowly enough.

(ii) Nevertheless, for uncountably many α, A is not a basis of order 2, provided that $\varepsilon(n) \leq \varepsilon_0$ where $\varepsilon_0 < \frac{1}{4}$ is a constant.

In Section II.6 we will establish a more general result in which we consider sets $\{n \in \mathbb{N}_0 \mid \|p(n)\| \leq \varepsilon(n)\}$, and the polynomial p varies in a linear family. The bulk of the difficulty lies in proving II.1.5.i. We rely on similar ideas as for II.1.4.i, and relate each element in the complement of $2A$ to the lack of equidistribution of a certain polynomial sequence. Using known result about distribution of polynomial sequences, we then connect lack of equidistribution with a system of approximate rational dependencies, which generically turn out not to be satisfiable.

For II.1.5.ii, it suffices to take α sufficiently well approximable by rationals, and we can construct such α explicitly.

II.2 Non-bases of order 2

Our goal in this section is to prove that the sets

$$A_\varepsilon^\alpha := \left\{n \in \mathbb{N}_0 \mid \left\|\alpha n^2\right\| < \varepsilon(n)\right\} \tag{II.1}$$

are "usually" not bases of order 2, even when $\varepsilon(n) = \varepsilon_0$ is constant.

Theorem (II.1.4.v, reiterated)**.** *There exists a set $Z \subset \mathbb{R}$ of Lebesgue measure 0 such that the following is true. Let $\alpha \in \mathbb{R} \setminus Z$ and let $\varepsilon\colon \mathbb{N}_0 \to \mathbb{R}_{>0}$ be such that either $\varepsilon(n) \to 0$ or $\varepsilon(n) \leq \varepsilon_0$ for some $\varepsilon_0 < \frac{1}{4}$. Then, the set A_ε^α defined in* (II.1) *is not a basis of order 2.*

The same statement is true for any badly approximable $\alpha \in \mathbb{R}$, provided that $\varepsilon(n) \to 0$, and also for any quadratic surd $\alpha \in \mathbb{Q}[\sqrt{d}]$, provided that $\varepsilon(n) \leq \varepsilon_0(\alpha)$ for some constant $\varepsilon_0(\alpha)$.

This result is somewhat surprising, because a random (unstructured) set of similar size should be a basis of order 2.

Observation. Let $\varepsilon > 0$ be a constant and let $A \subset \mathbb{N}_0$ be a random set, with $\mathbb{P}(n \in A) = \varepsilon$, independently for each n. Then, for each n,

$$\mathbb{P}(\exists N > n \ : \ N \notin 2A) \ll \frac{(1-\varepsilon^2/2)^n}{\varepsilon^2}.$$

In particular, A is a basis of order 2 with probability 1.

Proof. It suffices to notice that each N has $\geq N/2 - 1$ independent representations $N = n_1 + n_2$, each appearing with probability ε^2, and use the union bound. □

We will prove a variety of partial results, with different restrictions on α and $\varepsilon(n)$, not all of which are included in Theorem II.1.4.v as stated above. For α, we separately address the "structured" case when α is a quadratic surd or, more generally, is badly approximable, and the "generic" case when α is selected from a suitable set of full measure. For $\varepsilon(n)$, we either assume that $\varepsilon(n) \to 0$ as $n \to \infty$ or that $\varepsilon(n) \leq \varepsilon_0$ is bounded by a constant (which is allowed to depend on α). Theorem II.1.4.v will follow from combining Propositions II.2.4 ($\alpha \in \mathbb{Q}[\sqrt{d}]$), II.2.5 ($\alpha$ badly approximable) II.2.6 (generic α). Definition of badly approximable numbers will be recalled in Section II.2.3.

II.2.1 General strategy

We begin by introducing a somewhat technical tool which will allow us to detect large integers N in the complement of $2A_\varepsilon^\alpha$. Importantly, we are able to reduce the task of proving that $N \notin 2A_\varepsilon^\alpha$ to the task of verifying a simple Diophantine inequality.

The basic idea is quite simple. Suppose that we allowed α to take rational values, and take for instance $\alpha = \frac{1}{2}$. Assuming that $\varepsilon(n) < \frac{1}{2}$ for all n, the set A_ε^α is far from being a basis of order 2. Indeed, we then have $A_\varepsilon^\alpha = 2\mathbb{N}_0$, which is not a basis of any order.

The following lemma makes this observation quantitative. We will use it multiple times.

Lemma II.2.1. *Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Suppose that for an odd integer N, there are integers k, m, with k even and m odd, and a real parameter $\delta > 0$, such that we have*

$$\left\| N\alpha - \frac{m}{k} \right\| < \frac{1-\delta}{kN}. \tag{II.2}$$

Then $N \notin 2A_\varepsilon^\alpha$ for any pointwise bounded $\varepsilon(n) \leq \varepsilon_0$, where $\varepsilon_0 = \frac{\delta}{2k}$.

Proof. Let us take γ with $|\gamma| < 1 - \delta$ so that $N\alpha \equiv \frac{m}{k} + \frac{\gamma}{kN} \pmod 1$. Consider any decomposition $N = n_1 + n_2$ with $n_1, n_2 \in \mathbb{N}_0$. We can then compute that

$$\begin{aligned} \left\| n_1^2\alpha - n_2^2\alpha \right\| &= \|(n_1 - n_2)N\alpha\| \\ &= \left\| \frac{(n_1 - n_2)m}{k} + \frac{n_1 - n_2}{N}\frac{\gamma}{k} \right\| \geq \frac{1}{k} - \frac{|\gamma|}{k} > \frac{\delta}{k} = 2\varepsilon_0. \end{aligned}$$

It follows that $n_1^2\alpha$ and $n_2^2\alpha \bmod 1$ cannot both lie in $(-\varepsilon_0, \varepsilon_0) \bmod 1$. Hence at least one of n_1, n_2 fails to belong to A_ε^α and consequently $N \notin 2A_\varepsilon^\alpha$. □

Our next result is in similar spirit, with the difference that instead of pointwise bound $\varepsilon(n) \leq \varepsilon_0$, we work with the condition $\varepsilon(n) \to 0$.

Remark. It might seem that a set A_ε^α with $\varepsilon(n) \to 0$ must necessarily be "smaller" than one with constant $\varepsilon(n) = \varepsilon_0$, and hence that Lemma II.2.2 below is strictly weaker than Lemma II.2.1. However, we wish to emphasise that for variable $\varepsilon(n)$ we allow the value $\varepsilon(n)$ to be large when n is small.

Because we expect the complement of $2A_\varepsilon^\alpha$ to have density 0, we cannot rule out a priori that small values of n play a role. In fact, for any $\varepsilon_0 > 0$, one can construct $\varepsilon(n) \to \varepsilon_0$ such that A_ε^α is basis of order 2, simply by exploiting the fact that $A_{\varepsilon_0}^\alpha$ is syndetic. Hence, it is not the case that small values of n can be altogether ignored.

Here and elsewhere, by a slight abuse of notation, we write $A_{\varepsilon_0}^\alpha$, allowing the symbol ε_0 to also denote the constant function $n \mapsto \varepsilon_0$.

Lemma II.2.2. *Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, $\varepsilon(n) \to 0$, and let $(N_i)_{i=1}^\infty$ be an increasing sequence of odd integers. Suppose that for each i, we have*

$$N_i\alpha = \frac{m_i}{k} + \frac{\gamma_i}{kN_i},$$

where m_i, k are integers, k is even and m_i is odd. Assume further that γ with $|\gamma| < 1$ is an accumulation point of γ_i. Then $N_i \notin 2A_\varepsilon^\alpha$ for infinitely many i, unless $\gamma + kn^2\alpha \in \mathbb{Z}$ for some integer n.

Proof. Passing to a subsequence, we may assume that $\gamma_i \to \gamma$ as $i \to \infty$.

Let ε_0 be such that $\frac{1-|\gamma|}{k} > 2\varepsilon_0 > 0$. Then from previous Proposition II.2.1 it follows that $N_i \notin 2A_{\varepsilon_0}^\alpha$ for sufficiently large i. Hence, if $N_i \in 2A_\varepsilon^\alpha$ for some i, then N_i needs to have a representation as $n_1 + n_2$ with $n_1 \in A_\varepsilon^\alpha \setminus A_{\varepsilon_0}^\alpha$, $n_2 \in A_\varepsilon^\alpha$. Note that the set $A_\varepsilon^\alpha \setminus A_{\varepsilon_0}^\alpha$ is finite, so passing to a subsequence again we may assume that there exists a single n_1 such that for each i we have $n_{2,i} := N_i - n_1 \in A_\varepsilon^\alpha$.

Directly from the membership condition for A_ε^α, we conclude that

$$\varepsilon(n_{2,i}) > \left\| (N_i - 2n_1)\left(\frac{m_i}{k} + \frac{\gamma_i}{kN_i}\right) + n_1^2\alpha \right\|.$$

We have $\varepsilon(n_{2,i}) \to 0$ and $\frac{N_i - 2n_1}{N_i}\frac{\gamma_i}{k} \to \frac{\gamma}{k}$ as $i \to \infty$. It follows that

$$\left\| \frac{m_i'}{k} + \frac{\gamma}{k} + n_1^2\alpha \right\| \to 0,$$

where $m_i' := (N - 2n_1)m_i \mod k$. Note that m_i' is odd and takes only finitely many values. Restricting to a subsequence, we may assume that $m_i' = m'$ is constant. Now, the expression in the limit above is independent of i, and hence

$$\left\| \frac{m'}{k} + \frac{\gamma}{k} + n_1^2\alpha \right\| = 0.$$

In particular, we have $\gamma + kn_1^2\alpha \in \mathbb{Z}$, contradicting the irrationality assumption. □

II.2.2 Quadratic irrationals

We will now prove Theorem II.1.4.v in the special case when $\alpha = \sqrt{2}$. The argument generalises to $\alpha \in \mathbb{Q}[\sqrt{d}]\backslash\mathbb{Q}$ without any new ideas. This case is already representative for some of our methods. Our immediate goal is the following result.

Proposition II.2.3. *Let $\varepsilon_1 := \frac{1}{4}(1 - \frac{1}{4\sqrt{2}})$. Suppose that either we have the pointwise bound $\varepsilon(n) \leq \varepsilon_0 < \varepsilon_1$ or that $\varepsilon(n) \to 0$. Then $A_\varepsilon^{\sqrt{2}}$ is not a basis of order 2. In the case when ε is pointwise bounded, we additionally have the quantitative bound:*

$$\left|[T] \setminus 2A_\varepsilon^{\sqrt{2}}\right| \gg \log T,$$

where the implicit constant depends at most on $\varepsilon_0, \varepsilon_1$.

Proof. Any positive integer solution (x, y) to the Pell equation

$$X^2 - 2Y^2 = 1, \tag{II.3}$$

gives rise to the rational approximation $\frac{x}{y}$ of α with $\frac{x}{y} - \sqrt{2} = \frac{1}{y(x+\sqrt{2}y)}$.

The fundamental solution to (II.3) is $(x, y) = (3, 2)$. If we let $\phi := 3 + 2\sqrt{2}$ and $\hat{\phi} := 3 - 2\sqrt{2}$, and define integer sequences a_i, b_i by $\phi^i = a_i + b_i\sqrt{2}$, then all solutions to (II.3) are of the form $(x, y) = (a_i, b_i)$ We note that a_i, b_i have the explicit formulas:

$$a_i := \frac{\phi^i + \hat{\phi}^i}{2}, \quad b_i := \frac{\phi^i - \hat{\phi}^i}{2\sqrt{2}}, \tag{II.4}$$

as well as recursive relations:

$$\begin{aligned} a_{i+2} &= 6a_{i+1} - a_i, & a_0 &= 1,\ a_1 = 3, \\ b_{i+2} &= 6b_{i+1} - b_i, & b_0 &= 0,\ b_1 = 2. \end{aligned}$$

It will be convenient to take $N_i = b_i/2$. For any i, N_i is an integer, and if i is odd, then N_i is odd. We then have

$$N_i\sqrt{2} = \frac{a_i}{2} + \frac{\gamma_i}{2N_i}, \qquad \text{where } \gamma_i = \hat{\phi}^i N_i = \frac{-1}{4\sqrt{2}} + O\left(\frac{1}{N_i^2}\right). \tag{II.5}$$

To prove the statement in the case when $\varepsilon(n) \leq \varepsilon_0 < \varepsilon_1$ is pointwise bounded, we apply Lemma II.2.1 to N_i, assuming that i is large enough and odd. It follows that $N_i \notin 2A_\varepsilon^{\sqrt{2}}$, and hence $A_\varepsilon^{\sqrt{2}}$ is not a basis of order 2. The quantitative estimate follows from the fact that $N_i = \Theta(\phi^i)$, and for any T there are $\Theta(\log T)$ values of i with $N_i < T$.

To prove the statement in the case when $\varepsilon(n) \to 0$, we similarly apply Lemma II.2.2 to the sequence N_i restricted to odd i, with $\gamma = \frac{-1}{4\sqrt{2}}$. The claim follows, unless there exists n such that $\gamma + 2n^2\sqrt{2} \in \mathbb{Z}$. Since $\sqrt{2}$ is irrational, that would imply that $2n^2 = \frac{1}{8}$, which is absurd. □

The result for general quadratic irrationals $\alpha \in \mathbb{Q}[\sqrt{d}]$ can be obtained with essentially the same argument. As usual, for $\phi = x + y\sqrt{d} \in \mathbb{Q}[\sqrt{d}]$ with $d \in \mathbb{N}$ not a square, we denote by $\hat{\phi}$ the conjugate: $\hat{\phi} = x - y\sqrt{d}$.

Proposition II.2.4. *For any $\alpha \in \mathbb{Q}[\sqrt{d}] \setminus \mathbb{Q}$ there exists $\varepsilon_1 = \varepsilon_1(\alpha)$ such that the following is true. Suppose that either $\varepsilon(n) \leq \varepsilon_0 < \varepsilon_1$ or that $\varepsilon(n) \to 0$. Then A_ε^α is not a basis of order 2.*

Proof. We may write $\alpha = \frac{a+b\sqrt{d}}{c}$, where a, b, c are integers. Let $\phi = x + y\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$ be a unit, and let $\mu := \nu_2(y)$, the largest power of 2 dividing y. Replacing ϕ by ϕ^{2^n} for large n, we may assume that μ is sufficiently large, or more concretely that $\mu > \nu_2(b)$ and $2^\mu > bc$.

Like before, we consider the integer valued sequences:

$$a_i = \frac{\alpha\phi^i - \hat{\alpha}\hat{\phi}^i}{2\sqrt{d}}c, \quad b_i = \frac{\phi^i - \hat{\phi}^i}{2\sqrt{d}}c. \tag{II.6}$$

Using $\mu > \nu_2(b)$ and $\nu_2(x) = 0$, we derive the following relations:

$$\nu_2(a_1) = \nu_2(ay + bx) = \nu_2(b), \quad \nu_2(b_1) = \nu_2(cy) = \nu_2(c) + \mu. \tag{II.7}$$

Because the sequences a_i and b_i are periodic modulo any power of 2, there exists some L such that for all $i \equiv 1 \pmod{L}$ we have $\nu_2(a_i) = \nu_2(a_1)$ and $\nu_2(b_i) = \nu_2(b_1)$. For any such i we define

$$N_i := \frac{b_i}{2^{\nu_2(c)+\mu}}, \quad m_i := \frac{a_i}{2^{\nu_2(b)}}, \quad k := 2^{\nu_2(c)-\nu_2(b)+\mu}. \tag{II.8}$$

It is straightforward, if mundane, to check that these quantities are integers, and that we have the relation

$$N_i\alpha = \frac{m_i}{k} + \frac{\gamma_i}{kN_i},$$

where $\gcd(m_i, k) = 1$ and γ_i are given by

$$\gamma_i = \frac{b\hat{\phi}^i}{2^{\mu+\nu_2(c)}} = \frac{\pm bc}{2^{\mu+\nu_2(bc)+1}\sqrt{d}} + o(1).$$

Here, $o(1)$ denotes an error term which goes to 0 as $i \to \infty$. The choice of μ guarantees that $|\gamma_i| < \frac{1}{2}$ for large i. In the case $\varepsilon(n) \leq \varepsilon_0$, it follows from Lemma II.2.1 that $N_i \notin 2A_\varepsilon^\alpha$, provided that we take $\varepsilon_1 \leq \frac{1}{4k}$.

To deal with the case $\varepsilon(n) \to 0$, we notice that $\gamma_i \to \gamma := \frac{\pm bc\sqrt{d}}{2^{\mu+\nu_2(bc)+1}d}$. By Lemma II.2.2, we have $N_i \notin 2A_\varepsilon^\alpha$ for sufficiently large i, unless $\gamma + kn^2\alpha \in \mathbb{Z}$ for some n. If it was the case that $\gamma + kn^2\alpha \in \mathbb{Z}$ for some n, then it would follow that $k2^{\mu+\nu_2(bc)+1}d \mid bc^2$. However, this is impossible, since

$$\nu_2(k2^{\mu+\nu_2(bc)+1}d) \geq \mu + \nu_2(bc) + 1 > \nu_2(bc^2). \quad \square$$

II.2.3 Badly approximable reals

We now turn to the proof of a variant of Theorem II.1.4.v for badly approximable values of α.

We say that α is *badly approximable* if for any $p, q \in \mathbb{Z}$, $q > 0$, we have

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{c(\alpha)}{q^2},$$

where $c(\alpha) > 0$ is a constant dependent only on α. The most well known example of such numbers are quadratic irrationals.

This is a more general situation than $\alpha \in \mathbb{Q}[\sqrt{d}]$, but still rather specific. In particular almost all $\alpha \in \mathbb{R}$ are *not* badly approximable. However, badly approximable α provide non-trivial and fairly explicit classes of examples when the conclusion of Theorem II.1.4.v holds (as opposed to an "almost surely" type of statement).

A useful characterisation of badly approximable reals is that these are precisely the ones whose continued fraction expansion has bounded entries, see A.3.4. A specific class of badly approximable real numbers which has attracted some attention are those whose entries are produced by a finite automata. For instance, it has been shown that such numbers are transcendental, unless their continued fraction expansion is periodic, see [22].

The main result in this section shows that the sets A_ε^α are not bases of order 2 for badly approximable α and sufficiently small ε.

Proposition II.2.5. *If α is badly approximable then there is $\varepsilon_1 = \varepsilon_1(\alpha)$ such that if $\varepsilon(n) \leq \varepsilon_0 < \varepsilon_1$ then A_ε^α is not a basis of order 2.*

Moreover, we have $|[T] \setminus 2A_\varepsilon^\alpha| \gg \log T$, where the implicit constant depends only on α.

We will make extensive use of the continued fraction expansion of 2α. The crucial role played by the continued fraction expansion explains why we were able to give rather elementary proofs for $\alpha = \sqrt{2}$ and $\alpha \in \mathbb{Q}[\sqrt{d}]$, whose expansion is particularly simple.

Continued fractions are a classical topic, and we assume some familiarity with the basic notions and theorems. For an accessible introduction, see e.g. [70], or the more analytic approach in [71]. For the perspective inspired by measurable dynamics, see [35, Chpt. 3]. We delegate a complete list of used facts to the Appendix A. Here, we just review several basic properties and introduce notation, which we will also use in subsequent sections. We will write:

$$2\alpha = [a_0; a_1, a_2, \dots] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}} \tag{II.9}$$

and a_i will denote the coefficients of 2α throughout this section (note that for technical reasons we consider 2α rather than α). Using obvious translation invariance, we may assume without loss of generality that $a_0 = 0$. By p_i/q_i we denote the partial approximations given by

$$\frac{p_i}{q_i} = [0; a_1, a_2, \dots, a_i] = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{\cdots}{a_i}}}. \tag{II.10}$$

These are essentially the best possible rational approximations of α (see A.3.3), and we have the error term of the form:

$$2\alpha = \frac{p_i}{q_i} + \frac{\delta_i}{q_i^2}, \tag{II.11}$$

where δ_i can be explicitly described by

$$|\delta_i| = q_i^2 \left(\frac{1}{q_i q_{i+1}} - \frac{1}{q_{i+1} q_{i+2}} + \frac{1}{q_{i+2} q_{i+3}} - \dots \right). \tag{II.12}$$

In particular, we have $|\delta_i| < \frac{q_i}{q_{i+1}} < \frac{1}{a_{i+1}} \leq 1$.

Proof of Proposition II.2.5.

Because α is badly approximable, so is 2α, and by A.3.4 the coefficients a_i are bounded, say $a_i \leq a_{\max}$. Let κ be large enough that $2^\kappa \nmid a_i$ for all i.

We claim that among any 4 consecutive indices $\{j, j+1, j+2, j+3\}$ we can find an index i such that p_i is odd and $\nu_2(q_i) < \kappa$.

For each i we have (e.g. by A.1.3) that $\gcd(p_i, p_{i+1}) = \gcd(q_i, q_{i+1}) = 1$, and in particular neither p_i, p_{i+1} nor q_i, q_{i+1} can both be even. If for some $j \leq i \leq j+2$ we have that both p_i and p_{i+1} are odd then either q_i or q_{i+1} is odd, and the claim holds. Otherwise, since p_i, p_{i+1} can never both be even, the parity of p_i alternates. It follows that for one of $i \in \{j, j+1\}$, both p_i, p_{i+2} are odd, while p_{i+1} is even. Then $q_{i+2} - q_i = a_{i+2}q_{i+1}$ is not divisible by 2^κ, so one of q_{i+2}, q_i is not divisible by 2^κ. Either i or $i+2$ is the sought index.

Suppose now that i is an index such that p_i is odd and $\nu_2(q_i) < \kappa$, whose existence we have just proved. Let us take

$$N_i := \frac{q_i}{k_i/2}, \qquad k_i := 2^{\nu_2(q_i)+1}, \tag{II.13}$$

where as usual $\nu_2(q_i)$ is the largest power of 2 dividing q_i. (The definition makes sense for arbitrary i, but we only apply it to i as above.)

Note that k_i is guaranteed to be an even integer, and N_i — an odd integer. More precisely, $k_i \leq 2^\kappa$ is a power of 2. It is slightly inconvenient that N_i do not need to be distinct for distinct i, but this will not lead to problems since N_i take any given value at most κ times.

Finally, we introduce $\gamma_i := 2\delta_i/k_i$, so that we have the relation

$$N_i\alpha = \frac{p_i}{k_i} + \frac{\gamma_i}{k_i N_i}.$$

Note that we have the bounds:

$$|\gamma_i| \leq |\delta_i| \leq \frac{q_i}{q_{i+1}} = \frac{q_i}{q_i + q_{i-1}} \leq 1 - \frac{1}{1 + a_{\max}} < 1.$$

We are now in position to apply Lemma II.2.1 (with $\delta = \frac{1}{1+a_{\max}}$). If follows that for $\varepsilon_0 < \frac{1}{2^{\kappa+1}(1+a_{\max})}$, if $\varepsilon(n) \leq \varepsilon_0$ for all n, then $N_i \notin 2A_\varepsilon^\alpha$ for all i as described above. In particular, the complement of $2A_\varepsilon^\alpha$ is infinite, proving the first part of the proposition.

For the quantitative bound, we begin by noticing that $\log N_i = \Theta(i)$. Hence, given T, we have $N_i \in [T]$ for $i \leq i_0(T)$, with $i_0(T) = \Theta(\log T)$. For any 4 consecutive values of i, sufficiently large, for at least one of them we have $N_i \notin 2A_\varepsilon^\alpha$. Thus,

$$|[T] \setminus 2A_\varepsilon^\alpha| \gg \frac{i_0(T)}{4\kappa} \gg \log T. \quad \square$$

Remark. In the above result we deal exclusively with pointwise bounded $\varepsilon(n)$. As noted earlier, it does not quite follow that analogous claim holds when $\varepsilon(n) \to 0$, since large values of $\varepsilon(n)$ for small n can lead to problems. The main difficulty which stops us from extending our results is establishing the irrationality condition in Lemma II.2.2. This can be done for specific values of α, but we do not give a general result.

Remark. We believe that our methods should extend to numbers such as

$$e^{\frac{1}{n}} = [1; n-1, 1, 1, 3n-1, 1, 1, 5n-1, 1, ...],$$
$$\tanh(1/n) = [0; n, 3n, 5n, 7n, \dots],$$

whose continued fraction expansions are well understood (see e.g. [71, Chapter II]). It is straightforward to adopt our argument to these situations, and the only reason why we do not pursue this further is that we doubt if any particular one of those results would be of much interest.

II.2.4 Generic reals

Finally, we consider "generic" values of α. We prove a version of II.1.4.v which is valid for α outside of a set of measure 0. Conveniently, in this case we can make the dependence on ε rather explicit.

Proposition II.2.6. *For all $\alpha \in \mathbb{R}$ except for a set of measure 0, the set A_ε^α fails to be a basis of order 2 if either $\varepsilon(n) \leq \varepsilon_0 < \frac{1}{4}$ for all n, or if $\varepsilon(n) \to 0$.*

We retain definitions and conventions from the previous section. Namely, we assume that $2\alpha \in (0,1)$ has expansion $2\alpha = [0; a_1, a_2, \dots]$ and $\frac{p_i}{q_i} = [0; a_1, a_2, \dots, a_i]$ and are the convergents.

The following description of the continued fraction expansion comes as no surprise. It can be construed as a continued fractions analogue of the fact that almost all numbers are normal.

Proposition II.2.7. *There exists a set Z of zero measure such that the following is true for $\alpha \notin Z$.*

Let $b = (b_i)_{i=1}^l \in \mathbb{N}^l$ be a finite string of integers. Let J be the set of indices where b occurs in the expansion $(a_i)_{i=1}^\infty$, i.e. the set of those $j \in \mathbb{N}$ for which $a_{j+t} = b_t$ for all $1 \leq t \leq l$. Then the asymptotic density $d(J) = \lim_{n\to\infty} \frac{1}{n} |J \cap [n]|$ of the set J exists and is positive.

Proof. Let $T\colon [0,1] \to [0,1]$ be the continued fraction map $T(x) = \left\{\frac{1}{x}\right\}$, where $\{\cdot\}$ denotes the fractional part, and let μ be the Gauss measure on $[0,1]$, $\mu(E) = \frac{1}{\log 2} \int_E \frac{dx}{x+1}$.

It is known that $([0,1], T, \mathcal{B}, \mu)$ is an ergodic measure preserving system, and that T acts on continued fraction expansions by a shift: $T([0; c_1, c_2, \dots]) = [0; c_2, c_3, \dots]$ (for details, see [35, Chpt. 3], and Appendix A).

Define $B \subset [0,1]$ to be the set of those $\beta \in [0,1]$ whose expansion is of the form $\beta = [0; b_1, b_2, \dots, b_l, *, *, \dots]$. In simpler terms, B is an interval with endpoints $[0; b_1, b_2, \dots, b_l]$ and $[0; b_1, b_2, \dots, b_l + 1]$. Clearly, $\mu(B) > 0$.

By the poitwise ergodic theorem, for all α but a set of zero measure we have

$$d(J) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} 1_B(T^n(2\alpha)) = \int 1_B d\mu = \mu(B). \quad \square$$

Proof of Proposition II.2.6, case $\varepsilon(n) \le \varepsilon_0 < \frac{1}{4}$. Let C be a large, odd integer, to be specified in the course of the proof. For almost all choices of α, the sequence (C, C, C) appears infinitely often in $(a_i)_{i=1}^{\infty}$, i.e. there exists an infinite set J such that for each $j \in J$ we have $a_{j+1} = a_{j+2} = a_{j+3} = C$.

For each $j \in J$ we claim that we can find $i = i(j) \in \{j, j+1, j+2\}$ such that p_i, q_i are both odd.

If p_j, q_j are odd, we are done. Else, because $\gcd(p_j, q_j) = 1$, precisely one of p_j, q_j is even; suppose for concreteness that p_j is even. If p_{j+1}, q_{j+1} are both odd we are done. Otherwise, q_{j+1} is odd, because p_j, p_{j+1} cannot both be even. We have the recursive relation $p_{j+2} = a_{j+2}p_{j+1} + p_j = Cp_{j+1} + p_j$, so p_{j+2} is odd. By similar argument, q_{j+2} is odd, so we are done.

For any $j \in J$, take $N_j = q_{i(j)}$ and $\gamma_j = \delta_{i(j)}$. By construction, N_j is odd and

$$N_j\alpha = \frac{p_{i(j)}}{2} + \frac{\gamma_j}{2N_j}.$$

We are now in position to apply Lemma II.2.1. It follows that $N_j \notin A_\varepsilon^\alpha$, provided that $\varepsilon(n) \le \varepsilon_1$ for all n, where $\varepsilon_1 < \frac{1}{4} - \frac{1}{4}|\gamma_j|$ for all j. We know that $|\gamma_j| < \frac{1}{a_{i(j)+1}} = \frac{1}{C}$, so it will suffice to ensure that $\varepsilon_0 < \frac{1}{4} - \frac{1}{4C}$, which can be accomplished by choosing sufficiently large C. $\square$

We will next deal with the situation when $\varepsilon(n) \to 0$. Surprisingly, this is more difficult, because care is needed to ensure that the irrationality condition in Lemma II.2.2 is satisfied.

We need a preliminary lemma about estimation of the error term δ_i based on the knowledge of a limited number of continued fraction coefficients. Recall that δ_i is related to α by $2\alpha = \frac{p_i}{q_i} + \frac{\delta_i}{q_i^2}$.

Lemma II.2.8. *Let l be a positive integer. There exists a function $\tilde{\delta}_l \colon \mathbb{N}^{2l+1} \to \mathbb{R}$ such that for any $n > l$ we have*

$$\left|\delta_n - (-1)^n \tilde{\delta}_l\left((a_i)_{i=n-l}^{n+l}\right)\right| \ll 2^{-l/2}, \tag{II.14}$$

where the implicit constant is absolute.

Proof. Recall that $\delta_n = q_n^2(2\alpha - \frac{p_n}{q_n})$. Using standard facts about continued fractions, putting $\rho_n = [a_n; a_{n+1}, \dots]$ we may write:

$$\begin{aligned}\delta_n &= q_n^2 \left(\frac{p_{n-1}\rho_n + p_{n-2}}{q_{n-1}\rho_n + q_{n-2}} - \frac{p_{n-1}a_n + p_{n-2}}{q_{n-1}a_n + q_{n-2}} \right) \\ &= \frac{(-1)^n \left(q_{n-1}a_n + q_{n-2}\right)^2 (\rho_n - a_n)}{\left(q_{n-1}a_n + q_{n-2}\right)\left(q_{n-1}\rho_n + q_{n-2}\right)}\end{aligned}$$

Recalling that $\frac{q_{n-2}}{q_{n-1}} = [0; a_{n-1}, a_{n-2}, \dots] := \lambda_n$ we may simplify the above formula to

$$\delta_n = (-1)^n(\rho_n - a_n)\frac{a_n + \lambda_n}{\rho_n + \lambda_n}.$$

Putting $\tilde{\rho} = \tilde{\rho}\left((a_i)_{i=n-l}^{n+l}\right) := [a_n; a_{n+1}, \dots, a_{n+l}]$ and $\tilde{\lambda} = \tilde{\lambda}\left((a_i)_{i=n-l}^{n+l}\right) := [0; a_{n-1}, a_{n-2}, \dots, a_{n-l}]$ we have $\rho_n = \tilde{\rho} + O(2^{-n/2})$ and $\lambda_n = \tilde{\rho} + O(2^{-n/2})$. It remains to put

$$\tilde{\delta}_l = \tilde{\delta}_l\left((a_i)_{i=n-l}^{n+l}\right) := (\tilde{\rho} - a_n)\frac{a_n + \tilde{\lambda}}{\tilde{\rho} + \tilde{\lambda}}. \quad \square$$

Proof of Proposition II.2.6, case $\varepsilon(n) \to 0$. Using Proposition II.2.7, for almost all choices of α, we may find arbitrarily long strings of 1's in the expansion $(a_i)_{i=1}^{\infty}$. More precisely, there exists an infinite set J and a sequence $l(j)$, $j \in J$ with $l(j) \to \infty$ as $J \ni j \to \infty$, such that for each $j \in J$ and for each $|t| \le l(j)$ we have $a_{j+t} = 1$.

Repeating the argument from the proof of the same proposition in the pointwise bounded case, we may find for each $j \in J$ and index $i = i(j) \in \{j, j+1, j+2\}$ such that p_i, q_i are both odd. Without loss of generality we may assume that $i(j) = j$, i.e. that p_j, q_j are both odd for $j \in J$.

Let us put $N_j := q_j$ and $\gamma_j := \delta_j$, so that

$$N_j\alpha \equiv \frac{p_{i(j)}}{2} + \frac{\gamma_j}{2N_j} \pmod 1.$$

Applying Lemma II.2.2, we conclude that either $N_j \notin A_\varepsilon^\alpha$ for infinitely many j, or for each limit point γ of γ_j there exists n such that $\gamma + 2n^2\alpha \in \mathbb{Z}$.

Passing to a subsequence, we may assume without loss of generality that γ_j converges. Using Lemma II.2.8 we may identify $\gamma := \lim_{j\to\infty} \gamma_j$:

$$\gamma = \pm\frac{1}{\varphi} \cdot \frac{1 + \frac{1}{\varphi}}{\varphi + \frac{1}{\varphi}} = \pm\frac{1}{\sqrt{5}},$$

where $\varphi = \frac{1+\sqrt{5}}{2} = [1; 1, 1, \dots]$.

There are two cases to consider, depending on whether α and γ are affinely independent over $\mathbb{Z}$. If they are, then we are done by Lemma II.2.2. Otherwise, $\alpha \in \mathbb{Q}[\sqrt{5}]$. However, we can exclude this case, since $\mathbb{Q}[\sqrt{5}]$ has measure 0 (alternatively, we can apply Proposition II.2.3). $\square$

Remark. It is tempting to try to repeat the argument for the case $\varepsilon(n) \leq \varepsilon_0$ in the case $\varepsilon(n) \to 0$. Arguing along these lines, one can find a sequence of odd integers N_j such that $N_j\alpha \equiv \frac{p_{i(j)}}{2} + \frac{\gamma_j}{2N_j} \pmod 1$ with $p_{i(j)}$ odd and $\gamma_j \to 0$. Lemma II.2.2 would be applicable with $\gamma = 0$. We may conclude (inspecting the proof of Lemma II.2.2) that for sufficiently large j, the only possible representation of N_j as a member of $2A_\varepsilon^\alpha$ is $N_j = N_j + 0$. However, $0 \in A_\varepsilon^\alpha$, and we cannot exclude the possibility that $N_j \in A_\varepsilon^\alpha$, hence the need for a more involved argument.

II.3 Bases and almost bases of order 2

In the previous Section II.2 we have seen that the sets A_ε^α (as defined in II.1) usually are not bases of order 2. Our goal in this section is to show that the sets $2A_\varepsilon^\alpha$ nevertheless tend to be quite sizeable. For the convenience of the reader we recall the content of our main theorem, stated in the introduction. Our first result deals with density, and applies in a fairly general situation.

Theorem (II.1.4.i, reiterated)**.** *Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Then, there exists a decreasing sequence $\varepsilon_\alpha(n) \to 0$ such that A_ε^α is an almost basis of order 2, provided that $\varepsilon(n) \geq \varepsilon_\alpha(n)$ for all n.*

We will also prove a more surprising result, which shows that results from Section II.2 cannot be generalised to all α.

Theorem (II.1.4.ii, reiterated)**.** *There exist an uncountable set $E \subset \mathbb{R}$ such that for any $\alpha \in E$, there exists a decreasing sequence $\varepsilon_\alpha(n) \to 0$ such that A_ε^α is a basis of order 2, provided that $\varepsilon(n) \geq \varepsilon_\alpha(n)$ for all n.*

As the reader will have noticed, because of the monotonicity of the family A_ε^α with respect to ε, in both theorems there is no loss of generality in assuming that $\varepsilon(n) = \varepsilon_\alpha(n)$ for all n.

II.3.1 Equidistribution and quantitative rationality

In Section II.2, specifically in Lemmas II.2.1 and II.2.2, we have identified a class of obstructions to A_ε^α being a basis of order 2. Namely, we found sufficient conditions for a large integer N to fail to belong to $2A_\varepsilon^\alpha$.

Here, our first goal is to prove that these obstructions are essentially the only possible ones. We obtain two subtly different results, which can be construed as partial converses to Lemmas II.2.1 and II.2.2. Because $\|\cdot\|$ is always at most $\frac{1}{2}$, we implicitly assume that $\varepsilon_0 \leq \frac{1}{2}$ in what follows.

Lemma II.3.1. *There exists a constant C such that the following is true. Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, let $\varepsilon(n)$ be pointwise bounded as $\varepsilon(n) \geq \varepsilon_0$, and suppose that $N \notin 2A_\varepsilon^\alpha$. Then there exists $0 < k \leq 1/\varepsilon_0^C$, such that*

$$\|kN\alpha\| \leq \frac{1}{N\varepsilon_0^C}. \tag{II.15}$$

Lemma II.3.2. *Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and $\varepsilon_1 > 0$. For any $\varepsilon_0 > \varepsilon_1$ there exists $N_0 = N_0(\alpha, \varepsilon_1, \varepsilon_0)$ such that following is true for $N \geq N_0$. Suppose that $\varepsilon(n) \geq \varepsilon_0$ for all n and that $N \notin 2A_\varepsilon^\alpha$. Then N is odd and there exist $k \in \mathbb{N}$, $m \in \mathbb{Z}$ with $2 \mid k$, $\gcd(m,k) = 1$, and $\gamma \in \mathbb{R}$ such that*

$$N\alpha = \frac{m}{k} + \frac{\gamma}{kN}, \qquad \frac{1-|\gamma|}{2k} > \varepsilon_1. \tag{II.16}$$

We pause to describe the difference between these two results. Both state that if $N \notin 2A_\varepsilon^\alpha$, then $N\alpha$ is well approximated by a rational with small denominator. In II.3.1, the quality of approximation is worse, but no additional assumptions are imposed on N. On the other hand, in II.3.2 we obtain detailed information, but we need to restrict to sufficiently large N. In particular, II.3.1 is non-vacuous in the regime $\varepsilon \sim 1/N^\delta$ with δ sufficiently small.

The first step in order to prove Lemmas II.3.1 and II.3.1 is to reduce the problem of representing N as an element of $2A_\varepsilon^\alpha$ to an equidistribution statement about an orbit on the torus.

Observation. Fix $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, let $\varepsilon(n) = \varepsilon_0$ be constant, and let $N \in \mathbb{N}$. Then $N \in 2A_\varepsilon^\alpha$ if and only if the quadratic orbit

$$x_n := \big(n^2\alpha, (N-n)^2\alpha\big) \in \mathbb{T}^2$$

enters the set $(-\varepsilon_0, \varepsilon_0)^2 \subset \mathbb{T}^2$ at time up to N, i.e. if there exists $0 \leq n < N$ such that $x_n \in (-\varepsilon_0, \varepsilon_0)^2$.

We are now ready to prove the main results in this section.

Proof of Lemma II.3.1. Since $N \notin 2A_\varepsilon^\alpha$, the orbit $(n^2\alpha, (N-n)^2\alpha)$ misses $(-\varepsilon_0, \varepsilon_0)^2$ up to time N. It follows that $(n^2\alpha, 2Nn\alpha)$ fails to be $(c\varepsilon_0^2, N)$-equidistributed with $c > 0$.

By the characterisation of equidistribution in Theorem I.1.25, it follows that there is a universal constant C such that we can find k_1, k_2 with $|k_i| \ll \frac{1}{\varepsilon_0^C}$, $(k_1, k_2) \neq (0,0)$, such that $\|k_1\alpha\| \ll \frac{1}{N^2\varepsilon_0^C}$ and $\|k_2 N\alpha\| \ll \frac{1}{N\varepsilon_0^C}$. Hence, for $i = 1, 2$ we have $\|k_i N\alpha\| \ll \frac{1}{N\varepsilon_0^C}$, and since both of k_i cannot be 0, the claim follows. □

Proof of Lemma II.3.2. It follows from the above Lemma II.3.1 that there exists $K = O_{\varepsilon_0}(1)$ such that $\|kN\alpha\| = O_{\varepsilon_0}(1/N)$ for some $1 \le k \le K$. Possibly replacing k by one of its divisors, we can therefore write:

$$N\alpha = \frac{m}{k} + \frac{\gamma}{kN},$$

where $\gcd(m,k) = 1$ is an integer, $|\gamma| \le G$ and $G = O_{\varepsilon_0}(1)$ is a constant.

Let $\delta > 0$ be a small number to be determined later and let $M = M(\delta, \alpha)$ be such that $(n^2\alpha \bmod 1)$ intersects any interval of length δ as n ranges over any progression $P = n_0 + l[M']$ with length $M' \ge M$ and step $l \le K$. We know that such M exists, for instance by Theorem I.1.25.

Note that for any $n \in \mathbb{N}_0$ we have

$$\left(n^2\alpha, (N-n)^2\alpha\right) = \left(n^2\alpha, n^2\alpha + \frac{(N-2n)m}{k} + \frac{N-2n}{N}\frac{\gamma}{k}\right).$$

Thus, the values $(n^2\alpha \bmod 1)$ and $((N-n)^2\alpha \bmod 1)$ depend only on $(n^2\alpha \bmod 1)$, $(N-2n \bmod k)$ and $\frac{N-2n}{N}$. If k is even then for any choice of n, the number $(N - 2n \bmod k)$ has the same parity as N, and obviously $\frac{N-2n}{N} \in [-1,+1]$. These turn out to be essentially the only restrictions.

Observation II.3.3. Let $\tau \in \mathbb{T},\ 1 \le b \le k,\ x \in [-1,1]$, and if k is even, assume additionally that $b \equiv N \pmod 2$. Then there exists some $0 \le n < N$ such that $\|n^2\alpha - \tau\| \le \delta$, $(N-2n)m \bmod k = b$ and $\left|\frac{N-2n}{N} - x\right| \le \frac{2KM}{N}$, provided that $N > 2KM$.

Proof. We can pick n_0 such that $\left|\frac{N-2n_0}{N} - x\right| \le \frac{2}{N}$. Next, we can pick n_1 with $|n_0 - n_1| \le k$ such that $(N-2n_1)m \equiv b \pmod k$ and $\left|\frac{N-2n_1}{N} - x\right| \le \frac{2K}{N}$.

Let $P = n_2 + k[M]$ be an progression of length M, step k, containing n_1, and contained in $[N]$. For $n \in P$ we have $\left|\frac{N-2n}{N} - x\right| \le \frac{2KM}{N}$ and $(N-2n)m \equiv b \pmod k$. For at least one of these values, we have $\|n^2\alpha - \tau\| \le \delta$. □

If N is even or k is odd, then taking $\tau = 0,\ b = 0,\ x = 0$ and setting $\delta = \varepsilon_0/2$ we find some $0 \le n < N$ that $\|n^2\alpha\| \le \varepsilon_0/2$ and $\|(N-n)^2\alpha\| \le \varepsilon_0/2 + \frac{2MKG}{N}$, unless $N \le 2KM$. Since $N \notin 2A_\varepsilon^\alpha$, this situation is only possible if $N \ll MKG/\varepsilon_0$. Hence, we may assume that N is odd, k is even.

If $|\gamma| \ge 1$, then taking $\tau = 0,\ b = 1,\ x = \frac{1}{\gamma},\ \delta = \varepsilon_0/2$ we again find some $n \le N$ that $\|n^2\alpha\| \le \varepsilon_0/2$ and $\|(N-n)^2\alpha\| \le \varepsilon_0/2 + 2MKG/N$, which leads to a contradiction, unless $N \ll MKG/\varepsilon_0$. Hence, we may assume this is not the case.

Finally, let us take $b = 1,\ x = -\operatorname{sgn}\gamma$ and $\tau = -\frac{1}{2k}(1 - |\gamma|)$ and $\delta = (\varepsilon_0 - \varepsilon_1)/2$. Then for some $n \le N$ (assuming $N \ge KG$) we have

$$\begin{aligned}\varepsilon_0 &\le \max\left(\left\|n^2\alpha\right\|, \left\|(N-n)^2\alpha\right\|\right)\\ &\le \frac{1-|\gamma|}{2k} + \frac{\varepsilon_0 - \varepsilon_1}{2} + 2\frac{MKG}{N}\end{aligned}$$

If it holds that $\frac{1-|\gamma|}{2k} \le \varepsilon_1$, then the above implies that $N \ll MKG/\varepsilon_0$. Otherwise, the decomposition $N\alpha = \frac{m}{k} + \frac{\gamma}{kN}$ obtained earlier satisfies the condition $\frac{1-|\gamma|}{2k} > \varepsilon_1$, and we have that k is even and N is odd from previous considerations. □

II.3.2 Almost bases of order 2

With tools introduced in II.3.1, we are ready to prove the first of the two main result of this section, of which Theorem II.1.4.i is a special case.

To formulate the theorem, we need an additional a piece of notation. For real α, the *irrationality measure of* α, denoted $\mu(\alpha)$, is the smallest value μ such that for any $\delta > 0$ we have

$$\left|\alpha - \frac{p}{q}\right| \ge \frac{c}{q^{\mu+\delta}}$$

for any p, q with $\alpha \ne \frac{p}{q}$, where $c = c(\alpha, \delta, \mu) > 0$ is a constant independent of p and q. If no such μ exists, then $\mu(\alpha) = \infty$. We also recall that α is said to be *badly approximable*, if it holds that $\left|\alpha - \frac{p}{q}\right| \ge \frac{c}{q^2}$ for any integers p, q, where $c = c(\alpha) > 0$.

For $\alpha \in \mathbb{Q}$ we have (somewhat artificially) $\mu(\alpha) = 1$, and for any other α, $\mu(\alpha) \ge 2$. For almost all (with respect to Lebesgue measure) α, we have $\mu(\alpha) = 2$. Specifically, this holds for algebraic numbers, which is a celebrated result due to Roth [98].

Theorem II.3.4. *Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Then, there exists a decreasing sequence $\varepsilon_\alpha(n) \to 0$ such that for any ε with $\varepsilon(n) \ge \varepsilon_\alpha(n)$, the set A_ε^α is an almost basis of order 2.*

Moreover, if $\mu(\alpha) < \infty$, then the assumption $\varepsilon(n) \le \varepsilon_\alpha(n)$ can be replaced with the assumption that $\frac{\log(1/\varepsilon(n))}{\log n} \to 0$. In this case, it additionally holds that

$$|[T] \setminus 2A_\varepsilon^\alpha| \ll T^{1-c},$$

where the constant $c > 0$ depends only of α.

Finally, if α is badly approximable, and $\varepsilon(n) \ge \varepsilon_0 > 0$ for all n, we have a sharper estimate

$$|[T] \setminus 2A_\varepsilon^\alpha| \ll \log T,$$

where the implicit constant depends only on α and ε_0.

We begin by proving a technical proposition which describes local sparsity of the complement of $2A_\varepsilon^\alpha$. We wish to point out that this is a slightly stronger type of statement than Theorem II.3.4, since even sets with extremely slow asymptotic growth can contain many consecutive elements.

Proposition II.3.5. *There exists a constant C such that the following is true. Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and suppose that $\varepsilon(n) \geq \varepsilon_0 > 0$ is pointwise bounded from below. Suppose that $N, N' \notin 2A_\varepsilon^\alpha$ and $N' > N$. Then the following statements hold.*

(i) If α is badly approximable, then $N' - N \gg N\varepsilon_0^C$.

(ii) If $\mu(\alpha) < \infty$ and $\tau > \frac{\mu(\alpha)-2}{\mu(a)-1}$, then $N' - N \gg_\tau N^{1-\tau}\varepsilon_0^C$.

(iii) If $\mu(\alpha) = \infty$ then $N' - N \gg \varepsilon_0^C w(N\varepsilon_0^C)$, where $w(t) \to \infty$ as $t \to \infty$.

Proof. Since $N, N' \notin 2A_\varepsilon^\alpha$, it follows from Lemma II.3.1 that, there are $k, k' \leq 1/\varepsilon_0^C$ such that $\|kN\alpha\|, \|k'N'\alpha\| \leq \frac{1}{N\varepsilon_0^C}$, where C is a universal constant.

Let $Q_0(\delta)$ denote the least positive integer such that $\|Q_0(\delta)\alpha\| \leq \delta$. For any irrational α we have $Q_0(\delta) \to \infty$ as $\delta \to 0$. Moreover, if $\mu > \mu(\alpha)$ and $\frac{1}{1-\tau} = \mu - 1$, (resp. if $\mu = 2$ and $\tau = 0$ if α is badly approximable), then $Q_0(\delta) \gg 1/\delta^{1-\tau}$.

Let $L := N - N'$ and let $m := kk' \leq 1/\varepsilon_0^{2C}$. We have

$$\|mL\alpha\| \leq k\|k'N'\alpha\| + k'\|kN\alpha\| \leq \frac{2}{N\varepsilon_0^{2C}},$$

and as a consequence we have

$$\frac{L}{\varepsilon_0^{2C}} \geq Q_0\left(\frac{2}{N\varepsilon_0^{2C}}\right) \gg \left(N\varepsilon_0^{2C}\right)^{1-\tau}.$$

In each case, this easily leads to the sought bound. □

Proof of Theorem II.3.4. We may assume without loss of generality that $\varepsilon(2n) \geq \frac{1}{2}\varepsilon(n)$, and that $\varepsilon(n)$ is non-increasing. Let us denote $Z := \mathbb{N}_0 \setminus 2A_\varepsilon^\alpha$ and enumerate $Z = \{N_i\}_{i=1}^\infty$ so that $N_{i+1} > N_i$.

Our first aim is to show that the sequence N_i increases rapidly enough. Let us take any i. If $N_{i+1} \geq 2N_i$, then we have sufficiently good lower bound for N_{i+1}, so suppose that this is not the case. Since $N_{i+1}, N_i \notin 2A_{\varepsilon(N_{i+1})}^\alpha$, we may apply Proposition II.3.5 to conclude that

$$N_{i+1} - N_i \gg \begin{cases} N_i\varepsilon(N_i)^C, & \text{if } \alpha \text{ b. approx., with } C > 0, \\ N_i^c\varepsilon(N_i)^C, & \text{if } \mu(\alpha) < \infty, \text{ with } C, c > 0, \\ \varepsilon(N_i)^C w(N_i\varepsilon(N_i)^C), & \text{else, with } w(t) \to \infty,\ C > 0. \end{cases} \tag{II.17}$$

In the general case, we have $N_{i+1} - N_i \to \infty$, as $i \to \infty$, provided that

$$\varepsilon(n)^C w(n\varepsilon(n)^C) \to 0,$$

as $n \to \infty$. The latter condition is satisfied for constant $\varepsilon(n)$, and hence also for some slowly decaying $\varepsilon_\alpha(n) \to 0$. Lack of control on w makes it impossible to say anything more explicit about ε_α.

In the case when $\mu(\alpha) < \infty$, let us assume that $\varepsilon(n) \gg 1/N^\delta$, where δ is small enough. We have $N_{i+1} \geq N_i + c_2 N_i^{c_1}$ with $c_1, c_2 > 0$. A simple inductive argument shows that in this case we have $N_i \gg i^{1+c}$ for some $c > 0$. Hence, $|[T] \cap Z| \ll T^{\frac{1}{1+c}}$, proving the sought bound.

In the case when α is badly approximable and $\varepsilon(n) \geq \varepsilon_0$ is bounded pointwise, we have $N_{i+1} \geq N_i(1 + \varepsilon_0^C)$ with some constant C. It follows by a simple inductive argument that $\log N_i \gg i$. In particular, $|[T] \cap Z| \ll \log T$. □

Remark. The argument gives a slightly stronger conclusion than stated. Namely, under the assumptions in Theorem II.3.4, the distance between consecutive elements of $\mathbb{N}_0 \setminus 2A_\varepsilon^\alpha$ tends to ∞. In particular, this immediately implies II.1.4.iii.

Remark. Essentially the same argument leads to a result in higher dimension. More precisely, if $\alpha \in \mathbb{R}^r$ and we define

$$A = \left\{ n \in \mathbb{N} \;\middle|\; \left\| n^2 \alpha_i \right\| \leq \varepsilon \text{ for } i = 1, \ldots, r \right\},$$

where $\varepsilon > 0$ is constant, then one can show that $2A$ has density 1, provided that $k \cdot \alpha = \sum_i k_i \alpha_i$ is irrational for all $k \in \mathbb{Z}^r \setminus \{0\}$.

II.3.3 Exceptional values of α

We have seen in Section II.2 that the sets A_ε^α tend not to be bases of order 2. The main result of this section shows that such statements do not generalise to *all* values of α: we can find values of α such that the set A_ε^α is a basis of order 2 as soon as $\varepsilon(n) \geq \varepsilon_0 > 0$.

For such values of α we also have that A_ε^α is a basis of order 2 for some $\varepsilon(n) \to 0$. However, we have little control over the rate of convergence, so we do not pursue this issue further.

Our approach amounts to carefully preventing the conditions (II.16) in Lemma II.3.2 from being satisfied. The crucial step is establishing some control over all good approximations of α.

Throughout this section, we work with $\alpha \in (0,1) \setminus \mathbb{Q}$, we let a_i denote the digits in the continued fraction expansion of α:

$$\alpha = [a_0; a_1, a_2, \dots] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}} \tag{II.18}$$

(note difference with usage in Section II.2), and $\frac{p_i}{q_i}$ denote the rational approximations of α arising from the truncated continued fractions: $\frac{p_i}{q_i} = [a_0; a_1, a_2, \dots, a_i]$.

Recall that $\nu_p(a)$ denotes the largest power of the prime p dividing a. We will be interested in α satisfying the following conditions:

$$\nu_2(q_i) \to \infty \qquad \text{as } i \to \infty \text{ through even numbers,} \tag{II.19}$$

$$\nu_p(q_i) \to \infty \qquad \text{as } i \to \infty \text{ through odd numbers, for } p \text{ odd prime.} \tag{II.20}$$

We observe that if a_i obey the conditions (II.19, II.20), then they also obey the following conditions:

$$\nu_2(a_i) \to \infty \qquad \text{as } i \to \infty \text{ through even numbers,} \tag{II.21}$$

$$\nu_p(a_i) \to \infty \qquad \text{as } i \to \infty \text{ through odd numbers, for } p \text{ odd prime.} \tag{II.22}$$

This is an easy consequence the facts that $a_i = \frac{q_i - q_{i-2}}{q_{i-1}}$ (A.1.2) and $\gcd(q_i, q_{i-1}) = 1$ (A.1.3).

The following observation ensures that our considerations are not vacuous. It is not difficult, but the proof is slightly mundane.

Lemma II.3.6. *There exist uncountably many α such that the conditions (II.19, II.20) (and hence also (II.21, II.22)) are satisfied. Moreover, for any $h_i \in \mathbb{N}$ with $h_i \to \infty$ as $i \to \infty$, we can additionally require that $a_i \leq h_i$ for all i.*

Proof. Let p be a prime and $k_i^{(p)}$ a sequence of integers with $k_i^{(p)} \to \infty$. We will construct a sequence $a_i^{(p)}$ such that $a_i^{(p)} \leq p^{k_i^{(p)}}$ and an associated sequence $q_i^{(p)}$ (related to $a_i^{(p)}$ by $r_i^{(p)}/q_i^{(p)} = [a_0^{(p)}; a_1^{(p)}, a_2^{(p)}, \dots, a_i^{(p)}]$ for some $r_i^{(p)}$), in such a way that $\nu_p(q_i^{(p)}) \geq k_i^{(p)}$ for sufficiently large i. Once this is done, we choose some sequences $k_i^{(p)}$ with $k_i^{(p)} \to \infty$ such that $\prod_p p^{k_i^{(p)}} \leq h_i$, and define $a_i \leq \prod_p p^{k_i^{(p)}}$ by requiring that $a_i \equiv a_i^{(p)} \pmod{p^{k_i^{(p)}}}$. It is clear that thus defined sequence a_i satisfies (II.19, II.20).

To construct $a_i^{(p)}$ we proceed by induction. We restrict to the case when $p \neq 2$, the case $p = 2$ is fully analogous. We may assign arbitrary values $a_i^{(p)}$ for a number of small values of i. In particular, in the construction we may assume that i is large enough that $k_i^{(p)} \geq 1$. Suppose that $a_1^{(p)}, \dots, a_i^{(p)}$ have been constructed for some $i \equiv 1$

(mod 2). Since at most one of $q_i^{(p)}, q_{i+1}^{(p)}$ is divisible by p, we can choose $a_{i+1}^{(p)} \leq p \leq p^{k_{i+1}^{(p)}}$ so that that $q_{i+1}^{(p)} = a_{i+1}^{(p)} q_i^{(p)} + q_{i-1}^{(p)} \not\equiv 0 \pmod{p}$. Next, since $p \nmid q_{i+1}^{(p)}$, we may choose $a_{i+2}^{(p)} \leq p^{k_{i+2}^{(p)}}$ so that $q_{i+2}^{(p)} = a_{i+2}^{(p)} q_{i+1}^{(p)} + q_i^{(p)} \equiv 0 \pmod{p^{k_{i+2}^{(p)}}}$. This finishes the inductive step. It follows from the construction that $\nu_p(q_i) \geq p^{k_i^{(p)}}$ for all but finitely many i, so the sequence satisfies the required conditions.

To show that the number of possible choices of α is uncountable, we notice that different choices of the sequences $k_i^{(p)}$ produce different α. Since there are uncountably many choices for $k_i^{(p)}$, there are also uncountably many choices of α. □

Theorem II.1.4.ii is an immediate consequence of the following slightly more technical result, paired with Lemma II.3.6.

Proposition II.3.7. *Suppose that for some α, the conditions (II.19, II.20) and (II.21, II.22) are satisfied, and that $\frac{\log a_i}{i} \to 0$. Then A_ε^α is a base of order 2 provided that there exists $\varepsilon_0 > 0$ such that $\varepsilon(n) \geq \varepsilon_0$ for all n.*

Proof. For a proof by contradiction, suppose that there is some $\varepsilon_0 > 0$ such that $A_{\varepsilon_0}^\alpha$ is not a base of order 2.

By Lemma II.3.2, for infinitely many odd N there exists k, m, γ with k even, $\gcd(m, k) = 1$, such that

$$N\alpha = \frac{m}{k} + \frac{\gamma}{kN}, \qquad \frac{1}{2k}(1 - |\gamma|) > \frac{\varepsilon_0}{2}.$$

Because k is automatically bounded by $k \leq \frac{1}{\varepsilon_0}$, we may assume that k does not depend on N. Moreover, we may assume that m and N are coprime, because a similar relation is satisfied for $N' = \frac{N}{\gcd(N,m)}$, $m' = \frac{m}{\gcd(N,m)}$ and $\gamma' = \frac{\gamma}{\gcd(N,m)^2}$.

Let us fix N, but we reserve the right to assume that N is sufficiently large in terms of ε_0 and k, and take m, γ as above. Put $q = kN$ and $p = m$, and let i be the largest index such that $\frac{p}{q}$, lies between α and $\frac{p_i}{q_i}$, where $q := kN$. We may without loss of generality assume that i is odd so that $\alpha < \frac{p_{i+2}}{q_{i+2}} < \frac{p}{q} \leq \frac{p_i}{q_i}$. The other case, when $\alpha > \frac{p_{i+2}}{q_{i+2}} > \frac{p}{q} \geq \frac{p_i}{q_i}$ is fully analogous.

The case when $\frac{p}{q} = \frac{p_i}{q_i}$ is particularly simple. Because k is even, $q_i = q = kN$ is even, and in particular i is even (because of assumption (II.19)). Since N is odd, $\nu_2(q_i) = \nu_2(k)$ is bounded. However, this contradicts condition II.19, provided that N (and hence i) is sufficiently large. Hence, we may assume that $\frac{p}{q} < \frac{p_i}{q_i}$

We now deal with the general $\frac{p}{q}$. We can write $\frac{p}{q} = \frac{ap_{i+1} + bp_i}{aq_{i+1} + bq_i}$ for some coprime $a, b \in \mathbb{N}$, simply because $\frac{p_{i+1}}{q_{i+1}} < \frac{p}{q} < \frac{p_i}{q_i}$. A straightforward computation using A.1.3 shows that

$$\frac{p}{q} - \frac{p_{i+2}}{q_{i+2}} = \frac{(a_{i+2}b - a)(p_{i+1}q_i - p_i q_{i+2})}{qq_{i+2}} = \frac{\Delta}{qq_{i+2}},$$

where $\Delta := a_{i+2}b - a \geq 1$. It follows that

$$\frac{k\,|\gamma|}{q^2} = \frac{|\gamma|}{kN^2} = \left|\frac{p}{q} - \alpha\right| > \left|\frac{p}{q} - \frac{p_{i+2}}{q_{i+2}}\right| = \frac{\Delta}{qq_{i+2}}.$$

Thus, we may have derive the bound

$$\gamma \geq \frac{\Delta}{k}\frac{q}{q_{i+2}} = \frac{\Delta}{k}\left(b - \frac{\Delta}{q_{i+2}}\right). \tag{II.23}$$

To have some rather crude control on the size of Δ, we note that $q \geq q_i$ (A.3.3) so

$$1 \geq |\gamma| \geq \frac{\Delta}{k}\frac{q}{q_{i+2}} \geq \frac{\Delta}{4ka_{i+2}a_{i+1}},$$

which leads to $\Delta \leq 4ka_{i+2}a_{i+1} \ll \left(1 + \frac{1}{10}\right)^i$. On the other hand, because of A.1.2 we have $q_i \gg 2^{i/2}$, so if N (and hence also i) is sufficiently large, then we have $\frac{\Delta}{q_{i+2}} < \varepsilon_0$. Combining this with previous bounds, we find that

$$1 - 2k\varepsilon_0 \geq |\gamma| > \frac{\Delta}{k}\left(b - \varepsilon_0\right),$$

which in particular implies that $1 > \frac{\Delta b}{k}$, provided that ε_0 is small enough.

Let us write $k = k_0k_1$ as a product of a power of 2 and an odd integer. Recall that we have $k_0k_1 = \frac{q}{N} \mid aq_{i+1} + bq_i$. Assuming that N (and hence i) is sufficiently large, and possibly exchanging the order of k_0, k_1, we conclude from conditions (II.19, II.20) that $k_0 \mid q_i,\ \gcd(k_0, q_{i+1}) = 1$ and $k_1 \mid q_{i+1},\ \gcd(k_1, q_i) = 1$. The divisibility condition $k_0 \mid aq_{i+1} + bq_i$ reduces to $k_0 \mid a$ and likewise $k_1 \mid aq_{i+1} + bq_i$ reduces to $k_1 \mid b$.

Clearly, $k_1 \mid b$ implies that $b \geq k_1$. From $k_0 \mid a$, we have $k_0 \mid \Delta = a_{i+2}b - a$ because of conditions (II.21, II.22). Consequently, $\Delta \geq k_0$. Thus, we find $\Delta b \geq k_0k_1 = k > \Delta b$. This is the sought contradiction, which finishes the proof. □

II.4 Threshold for being a basis of order 2

We have shown in Section II.2 that for a variety of values of α, the set A_ε^α defined by (II.1) fails to be a basis of order 2 when $\varepsilon(n) < \varepsilon_1(\alpha)$ is sufficiently small. On the other hand, if ε is sufficiently large, then A_ε^α is large as well. In the extreme case, if $\varepsilon(n) \geq 1/2$, then $A_\varepsilon^\alpha = \mathbb{N}_0$ which trivially is a basis of order 2. The set A_ε^α is increasing with respect to ε, in the sense that $A_\varepsilon^\alpha \subset A_{\varepsilon'}^\alpha$ if $\varepsilon(n) \leq \varepsilon'(n)$. Hence, there exists a threshold value of $\varepsilon_{\text{thr}} = \varepsilon_{\text{thr}}(\alpha)$ where A_ε^α becomes a basis of order 2.

Definition II.4.1. Given $\alpha \in \mathbb{R}$, the threshold $\varepsilon_{\text{thr}} = \varepsilon_{\text{thr}}(\alpha)$ is the unique real number with the property that for any $\varepsilon_1 < \varepsilon_{\text{thr}} < \varepsilon_2$ the following hold:

(i) the set A_ε^α is not a basis of order 2 if $\varepsilon(n) \leq \varepsilon_1$;

(ii) the set A_ε^α is a basis of order 2 if $\varepsilon(n) \geq \varepsilon_2$.

Our goal in this section is to understand the behaviour of $\varepsilon_{\text{thr}}(\alpha)$ as α varies. Note that in Definition II.4.1 it is no loss of generality to assume that ε is a constant function. Hence, from now on we will restrict our attention to the case when ε is constant, and will denote this constant value again by ε. Under this restriction, in conditions II.4.1.i and II.4.1.ii above, we may assume that $\varepsilon = \varepsilon_1$ and $\varepsilon = \varepsilon_2$ respectively.

We begin with the following trivial upper bound of $\varepsilon_{\text{thr}}(\alpha)$.

Observation II.4.2. For any $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ we have $\varepsilon_{\text{thr}}(\alpha) \leq \frac{1}{4}$.

Proof. This follows immediately from Lemma II.3.2. □

This is also what one might (at least heuristically) expect on grounds of density and results such as Schnirelmann's theorem. The assumption of irrationality is crucial for the bound above, as the following example shows.

Example II.4.3. For $\alpha = \frac{1}{3}$ we have $\varepsilon_{\text{thr}}(\alpha) = \frac{1}{3} > \frac{1}{4}$.

Proof. Since $n^2\alpha \bmod 1 \in \{0, \frac{1}{3}, -\frac{1}{3}\}$, we have $A_\varepsilon^\alpha = \mathbb{N}_0$ if $\varepsilon > \frac{1}{3}$, and A_ε^α is trivially a basis of order 2. Conversely, if $\varepsilon < \frac{1}{3}$ then $A_\varepsilon^\alpha = 3\mathbb{N}_0$, which is a not a basis of any order. □

By a similar argument, we have $\varepsilon_{\text{thr}}(\alpha) > \frac{1}{4}$ for $\alpha = \frac{2}{5}, \frac{1}{7}, \frac{3}{17}$, and so on. From this point, we will restrict to the case when α is irrational. For a generic choice of α, we can compute the value of $\varepsilon_{\text{thr}}(\alpha)$ explicitly.

Proposition II.4.4. *For all $\alpha \in \mathbb{R}$, except for a set of measure 0, we have $\varepsilon_{\text{thr}}(\alpha) = \frac{1}{4}$.*

Proof. We have $\varepsilon_{\text{thr}}(\alpha) \leq \frac{1}{4}$ for almost all α by above Observation II.4.2. Conversely, suppose that $\varepsilon < \frac{1}{4}$. Then A_ε^α fails to be a basis of order 2 for almost all α by Proposition II.2.6. □

II.4.1 Quadratic irrationals

The problem of determining $\varepsilon_{\text{thr}}(\alpha)$ is most interesting for $\alpha \in \mathbb{Q}[\sqrt{d}]$, where $d > 1$ is squarefree. In this case, we are able to determine the exact value of $\varepsilon_{\text{thr}}(\alpha)$, which usually is not equal to $\frac{1}{4}$. Our argument relies on a careful study of the possible ways in which conditions of Lemma II.2.1 could be satisfied. This involves considering a significant number of cases with little apparent pattern. Consequently, we construct an algorithm which computes $\varepsilon_{\text{thr}}(\alpha)$, as opposed to an explicit formula. Throughout this section, for $\xi = x + y\sqrt{d} \in \mathbb{Q}[\sqrt{d}]$, with d not a square, we denote $\hat{\xi} = x - y\sqrt{d}$, $\mathrm{Tr}(\xi) = \xi + \hat{\xi}$ and $\mathrm{N}(\xi) = \xi\hat{\xi}$.

We will need the following lemma, which gives a description of all good approximations of quadratic surds.

Lemma II.4.5. *Fix $\alpha \in \mathbb{Q}[\sqrt{d}] \setminus \mathbb{Q}$, with $d > 1$, squarefree, and let ϵ denote the fundamental unit in $\mathbb{Q}[\sqrt{d}]$ with $|\epsilon| > 1$. For any $C > 0$ there exists $C' > 0$ such that the following is true. Let p, q be integers such that $\left|\alpha - \frac{p}{q}\right| \leq \frac{C}{q^2}$. Then there exists an index i and $\tau = \frac{u+v\sqrt{d}}{w} \in \mathbb{Q}[\sqrt{d}]$ with $u, v, v \in \mathbb{Z}$, $|u|, |v|, |w| \leq C'$ such that $q = \mathrm{Tr}(\tau\epsilon^i)$.*

Here and elsewhere, if $\tau = \frac{u+v\sqrt{d}}{w}$ and $|u|, |v|, |w|$ are bounded, we shall informally say that τ has bounded complexity. Since we will need to deal with continued fraction approximations, we take as usual a_i to be the digits of α, so that $\alpha = [a_0; a_1, a_2, \dots]$, and we denote the partial approximations as $p_i/q_i = [a_0; a_1, a_2, \dots, a_i]$. From this point, we consider α to be fixed, and allow all implied constants to depend on α.

Proof. We may find an index i such that $\frac{p}{q}$ lies between $\frac{p_i}{q_i}$ and $\frac{p_{i+2}}{q_{i+2}}$; see A.1.5 (here, we allow $i = -1$ with $\frac{p_{-1}}{q_{-1}} := \frac{1}{0} := +\infty$). Following the reasoning in Proposition II.3.7, we may find coprime $a, b \in \mathbb{N}_0$ with $\frac{p}{q} = \frac{ap_{i+1}+bp_i}{aq_{i+1}+bq_i}$. A standard computation shows that $q \leq Cq_{i+2}$, which combined with A.3.5 implies that $a, b \ll C$.

Hence, it will suffice to show that for each j, there are some i and τ with with bounded complexity, such that $q_j = \mathrm{Tr}(\tau\epsilon^i)$. Moreover, it will suffice to prove this claim for $j \in J$ for a syndetic set $J \subset \mathbb{N}$.

Let us write $\alpha = \frac{a+b\sqrt{d}}{c}$. Note that $\|c\,\mathrm{Tr}\,(\epsilon^i)\alpha\| = O\left(1/|\epsilon|^i\right)$. Hence, repeating the argument above, we see that for any i we have $c\,\mathrm{Tr}(\epsilon^i) = a(i)q_{j(i)+1} + b(i)q_{j(i)}$, for some index $j(i)$ and for some bounded $a(i), b(i)$. Moreover, since q_{j+1}/q_j is bounded, the sequence $j(i+1) - j(i)$ is bounded, and hence $J := \{j(i) \mid i \in \mathbb{N}\}$ is syndetic.

By the same argument, we have $c\,\mathrm{Tr}(\epsilon^i\sqrt{d}) = a'(i)q_{j(i)+1} + b'(i)q_{j(i)}$. Note that we can assume that the same function $j(\cdot)$ appears here as above. Finally, observe that the matrix $\begin{bmatrix} a(i) & b(i) \\ a'(i) & b'(i) \end{bmatrix}$ is invertible. For if it were not, then it would hold that

$$\frac{\mathrm{Tr}(\epsilon^i)}{\mathrm{Tr}(\epsilon^i\sqrt{d})} = \frac{a(i)}{a'(i)} = \frac{b(i)}{b'(i)},$$

which is impossible if i is large enough, since $a(i), a'(i)$ and $\gcd\Big(\mathrm{Tr}(\epsilon^i),\ \mathrm{Tr}(\epsilon^i\sqrt{d})\Big)$ are bounded while $\mathrm{Tr}(\epsilon^i)$, $\mathrm{Tr}(\epsilon^i\sqrt{d})$ are not. It follows that:

$$q_{j(i)} = x(i)\,\mathrm{Tr}(\epsilon^i) + y(i)\,\mathrm{Tr}(\epsilon^i\sqrt{d}) = \mathrm{Tr}(\tau(i)\epsilon^i),$$

where $x(i), y(i)$ are rational of bounded height and $\tau(i) = x(i) + y(i)\sqrt{d}$ has bounded complexity. □

We will now show that if the set A_ε^α fails to be a basis of order 2, then one can find a certificate of a rather specific type for this.

Proposition II.4.6. *Fix $\alpha \in \mathbb{Q}[\sqrt{d}] \setminus \mathbb{Q}$, with $d > 1$, squarefree. Also, let $\varepsilon_0 > 0$ be fixed. Then for sufficiently large C, the following conditions are equivalent:*

(i) The set A_ε^α is not a basis of order 2 for any $\varepsilon < \varepsilon_0$.

(ii) There exist coprime $k, m \geq 1$, with k even and m, and a sequence of odd integers $(N_i)_{i=1}^\infty$ and $\gamma \in \mathbb{R}$ such that as $i \to \infty$ we have

$$N_i\alpha \equiv \frac{m}{k} + \frac{\gamma + o(1)}{kN_i} \pmod 1, \qquad \frac{1-|\gamma|}{2k} \geq \varepsilon_0. \tag{II.24}$$

(iii) There exists $k \in 2\mathbb{N}$, $\tau = \frac{u+v\sqrt{d}}{w} \in \mathbb{Q}[\sqrt{d}]$ with $|u|, |v|, |w| \leq C$ and a unit $\phi \in \mathbb{Q}[\sqrt{d}]$, $N(\phi) = +1$, such that the condition (ii) is satisfied for $N_i := \mathrm{Tr}(\phi^i\tau)/k$, $m := \mathrm{Tr}(\tau\alpha)$ and $\gamma := 2\frac{b\sqrt{d}\,\mathrm{N}(\tau)}{ck}$.

(iv) There exists $k \in 2\mathbb{N}$ and $\tau = \frac{u+v\sqrt{d}}{w} \in \mathbb{Q}[\sqrt{d}]$ with $|u|, |v|, |w| \leq C$, such that $N := \mathrm{Tr}(\tau)/k \in 2\mathbb{Z}+1$, $m := \mathrm{Tr}(\tau\alpha) \in 2\mathbb{Z}+1$, and $\gamma := 2\frac{b\sqrt{d}\,\mathrm{N}(\tau)}{ck}$ satisfies $\frac{1}{2k}(1-|\gamma|) \geq \varepsilon_0$.

Proof. It is easy to check that (iii) implies both (ii) and (iv).

If (ii) is satisfied, then $N_i \notin 2A_\varepsilon^\alpha$ if $\varepsilon < \frac{1}{2k}(1 - |\gamma| - o(1))$ by Lemma II.2.1. This condition is clearly satisfied if $\varepsilon < \varepsilon_0$ and i is sufficiently large. Thus, the complement

of $2A_\varepsilon^\alpha$ contains all but finitely many values of N_i. Hence, A_ε^α is not a basis of order 2, proving (i).

We will next show that (iv) implies (iii). Suppose that (iv) holds for some k and $\tau = \frac{u+v\sqrt{d}}{w}$, and let m, γ be defined accordingly. If $\gcd(m,k) > 1$, we may replace k by $k/\gcd(k,m)$ and τ by $\tau/\gcd(k,m)$. Consequently, we may assume without loss of generality that $\gcd(k,m) = 1$.

To begin with, note that for any $\phi \in \mathbb{Q}[\sqrt{d}]$ we have

$$\mathrm{Tr}(\tau\phi^i)\alpha = \mathrm{Tr}(\alpha\tau\phi^i) + (\alpha - \hat{\alpha})\hat{\tau}\hat{\phi}^i = \mathrm{Tr}(\alpha\tau\phi^i) + 2\frac{b\sqrt{d}}{c} \cdot \frac{\mathrm{N}(\tau)\,\mathrm{N}(\phi)^i}{\tau\phi^i} \tag{II.25}$$

Let $K := \mathrm{lcm}(2k, c)w$. The ideal generated by $2K$ inside the ring of integers of $\mathbb{Q}[\sqrt{d}]$ clearly has finite index, so there exists a unit $\phi \in 1 + K\mathbb{Z}[\sqrt{d}]$ with $|\phi| > 1$. Replacing ϕ by ϕ^2, we may assume that $\mathrm{N}(\phi) = +1$.

Let $N_i := \mathrm{Tr}(\tau\phi^i)/k$ and $m_i := \mathrm{Tr}(\alpha\tau\phi^i)$. The choice of ϕ guarantees that $N_i,\ m_i$ are odd integers, and $m_i \equiv m \pmod{k}$.

Inserting definitions of N_i and m_i into (II.25) and using the fact that $\tau\phi^i = \mathrm{Tr}(\tau\phi^i) + o(1)$ as $i \to \infty$, we find that

$$N_i\alpha = \frac{m}{k} + \frac{\gamma + o(1)}{kN_i}, \qquad \gamma := 2\frac{b\sqrt{d}\,\mathrm{N}(\tau)}{ck},$$

as required in (iii).

We will next show that (i) implies (iv). Assume that (i) is satisfied and take $\varepsilon < \varepsilon_0$. (We reserve the right to assume that ε is sufficiently close to ε_0, in terms of ε_0 and α.) Since A_ε^α is assumed not to be a basis of order 2, we may select an infinite, increasing sequence $(N_i)_i$ with $N_i \notin 2A_\varepsilon^\alpha$. (Again, we reserve the right to restrict to a subsequence of $(N_i)_i$.)

It follows from Propsition II.3.2 that N_i are odd and we can select some coprime m_i, k_i with k_i even, such that

$$N_i\alpha \equiv \frac{m_i}{k_i} + \frac{\gamma_i}{N_i} \pmod{1}, \qquad \frac{1}{2k_i}\left(1 - |\gamma_i|\right) > \varepsilon. \tag{II.26}$$

Because $m_i, k_i \le \frac{4}{\varepsilon} \le \frac{8}{\varepsilon_0}$ (assuming that $\varepsilon > \frac{\varepsilon_0}{2}$), we may assume (after passing to subsequence) that all $m_i,\ k_i$ do not depend on i: $m_i = m,\ k_i = k$.

Since $\|kN_i\alpha\| \le \frac{1/\varepsilon}{N_i} \le \frac{2/\varepsilon_0}{N_i}$, it follows from Lemma II.4.5 that for each i we can find τ_i and n_i such that $kN_i = \mathrm{Tr}\left(\tau_i\epsilon^{n_i}\right)$, where ϵ denotes the fundamental unit of $\mathbb{Q}[\sqrt{d}]$ and τ_i have bounded complexity. Passing to subsequence, we may assume that $\tau_i = \tau = \frac{u+v\sqrt{d}}{w}$ does not depend on i.

Up to this point, ε took an unspecified value smaller than ε_0, and the values m, k, τ implicitly depended on ε. However, since the number of choices of m, k, τ is bounded, we may assume that they are independent of ε.

Using the relation (II.25) (with ϵ in place of ϕ), for any i we may assume that $m_i = \mathrm{Tr}\,(\alpha\tau\epsilon^{n_i})$, and

$$\gamma_i = \pm\gamma + o(1), \qquad \gamma = \frac{b\sqrt{d}}{c} \cdot \frac{\mathrm{N}(\tau)}{k} = \frac{b(u^2 - dv^2)}{ckw^2}.$$

Passing to the limit $i \to \infty$ in (II.26), and using that $\varepsilon < \varepsilon_0$ was arbitrary we conclude that $\frac{1}{2k}\,(1 - |\gamma|) \geq \varepsilon_0$.

Let J denote the set of j such that $N_i' := \mathrm{Tr}\,(\tau\epsilon^i)\,/k$ and $m_i' := \mathrm{Tr}\,(\alpha\tau\epsilon^i)$ are both odd integers. Above considerations show that $J \neq \emptyset$, and J is a union of arithmetic progressions of bounded step directly from definition. Let n_0 be the smallest positive element of J, and put $\tau' := \tau\epsilon^{n_0}$, $N' := \mathrm{Tr}\,(\tau')\,/k$, and $m' := \mathrm{Tr}\,(\alpha\tau')$. We conclude that (iv) is satisfied for N', m', τ' and $\gamma' = \pm\gamma$. □

Corollary II.4.7. *For $\alpha \in \mathbb{Q}[\sqrt{d}]$, the threshold $\varepsilon_{\mathrm{thr}}(\alpha)$ is given by the formula:*

$$\varepsilon_{\mathrm{thr}}(\alpha) = \max\left\{ \frac{1 - |\gamma(\alpha, \tau, k)|}{2k} \;\middle|\; (k, \tau) \in \mathcal{S} \right\}, \tag{II.27}$$

where $\mathcal{S}$ is the set of all admissible pairs (k, τ):

$$\mathcal{S} = \left\{ (k, \tau) \;\middle|\; \tau \in \mathbb{Q}[\sqrt{d}], k \in 2\mathbb{N}, \mathrm{Tr}(\tau)/k \in 2\mathbb{Z} + 1, \mathrm{Tr}(\alpha\tau) \in 2\mathbb{Z} + 1 \right\}, \tag{II.28}$$

and γ is given by $\gamma(\alpha, \tau, k) := (\alpha - \hat{\alpha})\,\frac{N(\tau)}{k}$.

II.4.2 The algorithmic approach

We will now derive from II.4.7 an effective algorithm, which can in principle compute $\varepsilon_{\mathrm{thr}}(\alpha)$ for any $\alpha \in \mathbb{Q}[\sqrt{d}]$. In order to achieve this, we need an efficient way to iterate over $\mathcal{S}$ from (II.28), and a stopping criterion which will ensure that no more elements of $\mathcal{S}$ need to be considered. We begin with a somewhat more basic description of $\mathcal{S}$.

Observation II.4.8. The pairs $(k, \tau) \in \mathcal{S}$ are precisely the ones where τ takes the form:

$$\tau = (x + 1/2)\,k + \frac{\sqrt{d}}{db}\left((y + 1/2)\,c - (x + 1/2)\,ak\right),$$

with $x, y \in \mathbb{Z}$. In particular, $\tau \in \frac{1}{2b\sqrt{d}}\mathbb{Z}[\sqrt{d}]$.

Proof. The condition $\mathrm{Tr}(\tau)/k \in 2\mathbb{Z}+1$ holds if and only if τ is of the form $2\tau = ku_1 + \frac{v}{w}\sqrt{d}$ with $u_1, v, w \in \mathbb{Z}$, u_1 odd, v and w coprime. Next, the condition $\mathrm{Tr}(\alpha\tau) \in 2\mathbb{Z}+1$ holds if and only if $aku_1 + bd\frac{v}{w}$ is an odd multiple of c. This implies, in particular, that $w \mid bd$, and thus we may write $\frac{v}{w} = \frac{v_1}{db}$ for some v_1. Hence, the condition above simplifies to the requirement that $v_1 + aku_1$ should be an odd multiple of c. Writing $u_1 = 2x + 1$ and $v_1 + aku_1 = (2y + 1)c$ we find the required form for τ. □

Recall that $\varepsilon_{\mathrm{thr}}(\alpha) = \max\{\varepsilon(\alpha, \tau, \kappa) \mid (k, \tau) \in \mathcal{S}\}$, where $\varepsilon(\alpha, \tau, \kappa) := \frac{1-|\gamma(\alpha,\tau,k)|}{2k}$, and $\gamma(\alpha, \tau, k) := (\alpha - \hat{\alpha})\frac{N(\tau)}{k}$. Because of the trivial bound $\varepsilon(\alpha, \tau, \kappa) \leq \frac{1}{2k}$, it is reasonable to consider the elements of $\mathcal{S}$ in the order of increasing k. Note that as soon as a bound $\varepsilon_{\mathrm{thr}}(\alpha) \geq \varepsilon_0 > 0$ is available, we may stop the iteration at $k_0 = \lceil 1/\varepsilon_0 \rceil$.

Note also that, given k and α, the value of $\varepsilon(\alpha, \tau, \kappa)$ depends only on $\mathrm{N}(\tau)$, and is decreasing in $|\mathrm{N}(\tau)|$. Hence, for a fixed value of k, we only need to find the pair $(k, \tau) \in \mathcal{S}$ with the smallest possible norm of τ (if any such pair exists).

It is computationally feasible to enumerate all elements of $\mathbb{Z}[\sqrt{d}]$ (or $\frac{1}{2b\sqrt{d}}\mathbb{Z}[\sqrt{d}]$, recall Corollary II.4.7) of a given norm, up to multiplication by the fundamental unit, see [29, Chpt. 5].[2] Let ϵ denote the fundamental unit (which is also feasible to compute, see [29]), and let $\tau_1, \tau_2, \ldots$ be the list of elements of $\frac{1}{2b\sqrt{d}}\mathbb{Z}[\sqrt{d}]$, in the order of increasing norm (in absolute value), such that any $\tau \in \frac{1}{2b\sqrt{d}}\mathbb{Z}[\sqrt{d}]$ takes the form $\tau = \tau_{j,i} = \tau_j\epsilon^i$ for some i, j. For each i, j, it is straightforward to check if $(k, \tau_{j,i})$ is an element of $\mathcal{S}$, and any (k, τ) takes this form. Finally, note that for each j, the sequences $\mathrm{Tr}(\tau_{j,i}) \bmod 2k$ and $\mathrm{Tr}(\alpha\tau_{j,i}) \bmod 2$ are periodic[3], so for each j, there are only finitely many values of i which need to be checked in order to determine if $(k, \tau_{j,i}) \in \mathcal{S}$ for some i. We are now ready to give the promised algorithm, written in pseudocode as Algorithm 1.

Example II.4.9. We can compute[4] several values of $\varepsilon_{\mathrm{thr}}(\alpha)$:

$$\varepsilon_{\mathrm{thr}}\left(\sqrt{2}\right) = \frac{1}{4} - \frac{1}{8\sqrt{2}}, \qquad \varepsilon_{\mathrm{thr}}\left(\frac{1}{\sqrt{2}}\right) = \frac{1}{8} - \frac{1}{16\sqrt{2}},$$
$$\varepsilon_{\mathrm{thr}}\left(\sqrt{3}\right) = \frac{1}{8} - \frac{1}{32\sqrt{3}}, \qquad \varepsilon_{\mathrm{thr}}\left(\frac{1}{\sqrt{3}}\right) = \frac{1}{4} - \frac{1}{8\sqrt{3}}.$$

[2]Strictly speaking, one should work with the ring of integers $\mathcal{O}_{\mathbb{Z}[\sqrt{d}]}$ rather than $\mathbb{Z}[\sqrt{d}]$. Recall that $\mathbb{Z}[\sqrt{d}] \subset \mathcal{O}_{\mathbb{Z}[\sqrt{d}]}$, with strict inclusion when $d \equiv 1 \pmod 4$. However, this technical point will not make significant difference.

[3]Here, we allow $x \bmod m$ for $x \in \mathbb{Q}$ to denote the unique $y \in [0, m)$ with $x - y \in \mathbb{Z}m$; x need not be an integer.

[4]The computation was performed using Wolfram Mathematica; code is available upon request.

Input : The quadratic irrational $\alpha = \frac{a+b\sqrt{d}}{c}$.
Output: The threshold value $\varepsilon_{\text{thr}}(\alpha)$.

$\epsilon :=$ FundamentalUnit$(\sqrt{d})$;
$\varepsilon_{\text{best}} := 0$; // the best known bound for ε_{thr}
for $k := 2, 4, \ldots$ **do**
  if $\frac{1}{2k} \leq \varepsilon_{\text{best}}$ **then**
    break; // no better bounds for ε_{thr} can be found
  end
  for $n = 1, 2, \ldots$ **do**
    $\varepsilon_0 := \frac{1}{2k}\left(1 - 2\frac{b\sqrt{d}}{c} \cdot \frac{n}{k}\right)$;
    if $\varepsilon_0 \leq \varepsilon_{\text{best}}$ **then**
      break; // no better bounds for ε_{thr} can be found
    end
    $S :=$ Representatives$(\sqrt{d}, n)$ // list of all alg. integers in $\mathbb{Q}[\sqrt{d}]$ with norm $\pm n$, up to powers of ϵ
    for $\sigma \in S$ **do**
      for $i = 1, 2, \ldots$ **do**
        $\tau := \frac{1}{2b\sqrt{d}}\sigma\varepsilon^i$;
        if $\mathrm{Tr}(\tau)/k \in 2\mathbb{Z}+1$ *and* $\mathrm{Tr}(\alpha\tau) \in 2\mathbb{Z}+1$ **then**
          $\varepsilon_{\text{best}} := \varepsilon_0$; // new best bound for ε_{thr}
          next k; // pass to the next iteration of the outermost loop (over k)
        end
        if $\epsilon^i \equiv 1 \pmod{4kbc\sqrt{d}}$ **then**
          break; // period has been reached
        end
      end
    end
  end
end
return $\varepsilon_{\text{best}}$;

Algorithm 1: Algorithm computing $\varepsilon_{\text{thr}}(\alpha)$ for $\alpha \in \mathbb{Q}[\sqrt{d}]$.

As a direct consequence, we can derive a non-trivial bound for $\varepsilon_{\mathrm{thr}}(\alpha)$.

Corollary II.4.10. *If* $\alpha = \frac{a+b\sqrt{d}}{c} \in \mathbb{Q}[\sqrt{d}]$ *then* $\varepsilon_{\mathrm{thr}}(\alpha) \leq \frac{1}{4} - \frac{1}{4bc\sqrt{d}}$.

Proof. It follows from II.4.7 and II.4.8 that $\varepsilon_{\mathrm{thr}}(\alpha) = \frac{1}{2k}(1 - |\gamma|)$ with $\gamma = \frac{n}{bc\sqrt{d}}$ for some $n \in \mathbb{Z} \setminus \{0\}$ and $k \in 2\mathbb{N}$. □

II.5 Bases of order 3

From results in Sections II.2 and II.3 it follows that for $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and slowly decaying $\varepsilon(n)$, the sets A_ε^α are generally not bases of order 2, even though $2A_\varepsilon^\alpha$ have density 1. One thus expects that for a sufficiently large number N, there should be multiple ways to represent N as $n_1+n_2+n_3$ with $n_i \in A_\varepsilon^\alpha$. In this section, we prove that (under some modest assumptions) this is indeed the case, and in particular A_ε^α are bases of order 3. In fact, we are able to show this under considerably weaker assumption on $\varepsilon(n)$ than in Section II.3, and we give asymptotics for the number of such representations.

Our main tool is the Hardy-Littlewood circle method. For a fairly exhaustive introduction, see [107]. The argument is standard, and we present it here mainly for expository purposes. For this reason, we do not put much effort into optimising various constants.

Throughout this section, we assume that $\varepsilon(n)$ is non-increasing with $0 < \varepsilon(n) \leq \frac{1}{2}$, and that $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ with irrationality measure $\mu := \mu(\alpha) < \infty$. For a large integer N, we denote:

$$A_{\varepsilon,N}^\alpha := A_{\varepsilon(N)}^\alpha \cap [N] = \left\{0 \leq n < N \;\middle|\; \left\|\alpha n^2\right\| \leq \varepsilon(N)\right\}. \tag{II.29}$$

Note that monotonicity of $\varepsilon(n)$ implies that $A_{\varepsilon,N}^\alpha \subset A_\varepsilon^\alpha$. We further denote the number of representations of N in $3A_{\varepsilon,N}^\alpha$ by $R_\varepsilon^\alpha(N)$, so

$$R_\varepsilon^\alpha(N) = \left|\left\{(n_1,n_2,n_3) \in \left(A_{\varepsilon,N}^\alpha\right)^3 \;\middle|\; n_1+n_2+n_3 = N\right\}\right|. \tag{II.30}$$

Our aim is to estimate $R_\varepsilon^\alpha(N)$. We devote the remainder of this section to proving the following result, whose special case $\mu = 2$ implies Theorem II.1.4.iv.

Theorem II.5.1. *For any* $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ *with* $\mu = \mu(\alpha) < \infty$, *there exists* $D = D(\mu)$ *such that for* $N \geq N_0$, *then* $R_\varepsilon^\alpha(N)$ *obeys the asymptotic formula:*

$$R_\varepsilon^\alpha(N) = \frac{1}{2}N^2\varepsilon(N)^3\big(1 + O_\alpha(\varepsilon(N))\big), \tag{II.31}$$

provided that $\varepsilon(N) \geq 1/N^{1/D}$ *(the implicit constant depends only on* α*).*

In particular, if $\limsup_{n\to\infty} \frac{\log 1/\varepsilon(n)}{\log n} = 0$, *then* A_ε^α *is a basis of order* 3 *and* $R_\varepsilon^\alpha(N) \sim \frac{1}{2}N^2\varepsilon(N)^3$.

From this point, we assume the notation as in the above statement. Hence, from this point $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ with $\mu = \mu(\alpha)$ is fixed, and N denotes an integer about which we may assume that it is large in terms of α.

II.5.1 Set-up

The statement of Theorem II.5.1 is unaffected if we replace ε with the constant function taking the value $\varepsilon(N)$. Hence, without loss of generality we assume that ε is indeed constant, and to simplify notation we denote this constant value again by ε.

We will work with the Fourier transform $f = f_{\alpha,\varepsilon,N}$ of the characteristic function of the set $A^{\alpha}_{\varepsilon,N}$:

$$f(x) := \sum_{n \in A^{\alpha}_{\varepsilon,N}} e(nx). \tag{II.32}$$

A standard Fourier inversion argument gives

$$R^{\alpha}_{\varepsilon}(N) = \int_{\mathbb{T}} e(-Nx) f(x)^3 dx. \tag{II.33}$$

It is convenient introduce an integer parameter Q with

$$1/\varepsilon^{D_2} < Q < N^{1/D_1},$$

where the constants $D_1 = D_1(\mu),\ D_2 = D_2(\mu)$ are yet to be specified. Such choice is possible as long as $D_1 D_2 < D$.

Given a choice of N and Q as above, we define the major arcs:

$$\mathcal{M} := \bigcup_{p,q \le Q} \mathcal{M}_{p,q}, \qquad \mathcal{M}_{p,q} = \left\{ \alpha \in \mathbb{T} \ : \ \left\| \alpha - \frac{p}{q} \right\| \le \frac{Q}{N} \right\}.$$

We also define the minor arc $\mathfrak{m}$, which is given simply by: $\mathfrak{m} := \mathbb{T} \setminus \mathcal{M}$. The rationale is that f should be easy to understand on $\mathcal{M}_{p,q}$, and negligibly small on $\mathfrak{m}$.

II.5.2 Minor arcs

We begin by considering f on the minor arc $\mathfrak{m}$. We will use the value of D_2 provided by the following proposition.

Claim II.5.2. There exists a choice of $D_2 = D_2(\mu)$ and $N_0 = N_0(\alpha)$ such that for $N \ge N_0$, have the bound

$$\sup_{x \in \mathfrak{m}} |f(x)| \le N\varepsilon^3. \tag{II.34}$$

Proof. To begin with, note that for any constant C we choose, we may without loss of generality assume that $n^2\alpha \bmod 1$ is $1/\varepsilon^C$-equidistributed up to time N for all $N \geq N_0(\alpha)$. Indeed, if this was not the case, then using Theorem I.1.25 we could find $k \neq 0$ with $k \ll 1/\varepsilon^{O(1)}$ and $\|k\alpha\| \ll 1/(\varepsilon^{O(1)}N^2)$. On the other hand, $\|k\alpha\| \gg_\alpha 1/k^\mu$ (by definition of μ; we could put any constant $> \mu - 1$ in place of μ). Combining these bounds, we conclude that $N \ll_\alpha 1/\varepsilon^{O(\mu)}$. However, if N is sufficiently large in terms of α and D_1 is sufficiently large in terms of μ, this leads to a contradiction.

Suppose now that $x \in \mathbb{T}$ is such that $|f(x)| \geq N\varepsilon^3$; we will show that $x \in \mathcal{M}$. We may rewrite the above assumption as

$$\left| \mathop{\mathbb{E}}_{n<N} 1_I(n^2\alpha)e(nx) \right| \geq \varepsilon^3,$$

where 1_I denotes the characteristic function of the interval $I = (-\varepsilon, \varepsilon) \bmod 1$. Let ψ be a continuous function approximating 1_I so that $\|\psi\|_{\mathrm{Lip}} \ll 1/\varepsilon^{O(1)}$ and

$$\mathop{\mathbb{E}}_{n<N} \psi(n^2\alpha)e(nx) \geq \varepsilon^3/2. \tag{II.35}$$

(Knowing that $n^2\alpha \bmod 1$ is sufficiently equidistributed, we may take ψ to be a piecewise-linear approximation of 1_I.) Define $\Psi(\alpha, \beta) := \psi(\alpha)e(\beta)$. This is a continuous function on $\mathbb{T}^2$, with $\|\Psi\|_{\mathrm{Lip}} \ll 1/\varepsilon^{O(1)}$ and $\int \Psi = 0$. Hence, inequality (II.35) above implies that $(n^2\alpha, nx)$ fails to be $\varepsilon^{O(1)}$-equidistributed up to time N.

From the Quantitative Weyl Theorem I.1.25 it follows that there are $k_1, k_2 \ll 1/\varepsilon^{O(1)}$, not both 0, such that we have $\|k_1 x\| \ll 1/\left(\varepsilon^{O(1)}N\right)$ and $\|k_2\alpha\| \ll 1/\left(\varepsilon^{O(1)}N^2\right)$. Because of our initial remark, we have $k_2 = 0$ and hence $k_1 \neq 0$. Let $q = k_1$ and let p be such that p/q is as close to x as possible. Then $\|p/q - x\| \ll 1/\left(\varepsilon^{O(1)}N\right)$, which implies that $x \in \mathcal{M}$, provided that D_2 exceeds some absolute constant dependent on μ and N is large enough in terms of D_2 and implicit constants present above. □

II.5.3 Major arcs

We now proceed to the major arcs. First, we consider the arcs $\mathcal{M}_{p,q} \neq \mathcal{M}_{0,1}$. As before, we will use the value of D_1 provided by the following bound.

Claim II.5.3. There exists a choice of $D_1 = D_1(\mu)$, $c_1 = c_1(\mu)$ and $N_0 = N_0(\alpha)$ such that for $N \geq N_0$, for any coprime $p, q \leq Q$, $p \neq 0$ we have the bound

$$\sup_{x \in \mathcal{M}_{p,q}} |f(x)| \leq N^{1-c_1}.$$

Proof. Throughout the proof, we assume that N is sufficiently large in terms of α. Pick $M := \lfloor N^{1/2} \rfloor$, and split $[N]$ into $N/M + O(1)$ intervals of length $M + O(1)$. For each such interval J, and for any τ with $|\tau| < Q/N$, we will prove that

$$\sum_{m \in J \cap A^{\alpha}_{\varepsilon,N}} e\left(m\left(p/q + \tau\right)\right) = O\left(M^{1-c}\right)$$

for a constant $c > 0$ (dependent only μ). Once this is achieved, summing all intervals yields the claim.

Let us first consider the case when $\tau = 0$. We split J into congruence classes $J_r := J \cap (q\mathbb{Z} + r)$, so that $e(mp/q) = e(rp/q)$ for $m \in J_r$. Hence,

$$\sum_{m \in J \cap A^{\alpha}_{\varepsilon,N}} e\left(mp/q\right) = \sum_{r \bmod q} \left|J_r \cap A^{\alpha}_{\varepsilon,N}\right| e\left(rp/q\right)$$

We begin by showing that if $N \geq N_0(\alpha)$ is sufficiently large, then

$$\left|J_r \cap A^{\alpha}_{\varepsilon,N}\right| = 2\varepsilon \frac{M}{q} + O(M^{1-c}), \tag{II.36}$$

for a small constant $c > 0$ to be determined in the course of the proof, where the implicit constant is actually bounded by 1. Put $\delta = M^{-c}$, fix a choice of r, and suppose for the sake of contradiction that

$$\left| \left|J_r \cap A^{\alpha}_{\varepsilon,N}\right| - 2\varepsilon \frac{M}{q} \right| > \delta M.$$

Let $M' = |J_r|$; clearly, $M' = M/q + O(1)$. Dividing by M', we can conclude from the above inequality that

$$\left| \mathop{\mathbb{E}}_{n \in J_r} 1_{A^{\alpha}_{\varepsilon,N}}(n) - 2\varepsilon \right| > \delta, \tag{II.37}$$

where the omitted error terms are negligible, provided that D_1 is sufficiently large in terms of c.

Let $I := (-\varepsilon, \varepsilon) \subset \mathbb{T}$, and denote $n_0 = \min J_r$ so that $J_r = n_0 + q[M']$. Expanding definitions, we conclude that

$$\left| \mathop{\mathbb{E}}_{m < M'} 1_I\left((mq + n_0)^2 \alpha\right) - \int_{\mathbb{T}} 1_I(x) dx \right| > \delta.$$

Suppose without loss of generality that $\mathbb{E}_{m<M'} 1_I((mq+r)^2\alpha) > \int 1_I$; the other case being analogous. Let $\psi \colon \mathbb{T} \to \mathbb{R}$ be an approximation of 1_I such that $\int |\psi - 1_I| \ll \delta^2$, $\|\psi\|_{\mathrm{Lip}} \ll 1/\delta^2$ and $\psi > 1_I$. Then it remains true that

$$\left| \mathop{\mathbb{E}}_{m < M'} \psi((mq + n_0)^2 \alpha) - \int_{\mathbb{T}} \psi(x) dx \right| \gg \delta \gg \delta^3 \left\|\psi\right\|_{\mathrm{Lip}}.$$

It follows that $(mq+n_0)^2\alpha$ fails to be $O(\delta^3)$-equidistributed up to time M'. Thus, there exists $k \ll 1/\delta^{O(1)}$ such that $\|kq^2\alpha\| \leq 1/\left(M'^2\delta^{O(1)}\right)$. On the other hand, $\|kq^2\alpha\| \gg_\alpha (kQ^2)^{-\mu}$. Combining these bounds leads to $Q^\mu \gg_\alpha N^{1/2-O(c\mu)}$. Choosing c small enough in terms of μ, and assuming that N is large enough in terms of α, we may ensure that the last inequality implies that $Q^\mu > N^{1/4}$. If D_1 is chosen sufficiently large in terms of μ, this is a contradiction, which finishes the proof of (II.36)

Applying (II.36), we can now compute:

$$\begin{aligned}\sum_{m\in J\cap A_\varepsilon^\alpha} e(mp/q) &= \sum_{r=0}^{q-1} e(rp/q)\left(2\varepsilon\frac{M}{q}+O(M^{1-c})\right)\\ &= 2\varepsilon\frac{M}{q}\sum_{r=0}^{q-1} e(rp/q) + O\left(QM^{1-c}\right) = O\left(M^{1-c/2}\right),\end{aligned}$$

assuming that D_1 is sufficiently large in terms of c.

Finally, we deal with $\tau \neq 0$. For any $0 \leq r < q$ and any $m \in J$ we have $e(m(p/q+\tau)) = e(\tau m_0)e(rp/q) + O(M\tau)$, where $m_0 := \min J$. It follows that

$$\begin{aligned}\sum_{m\in J\cap A_\varepsilon^\alpha} e(m(p/q+\tau)) &= e(\tau m_0)\sum_{m\in J\cap A_\varepsilon^\alpha}\left(e\left(\frac{mp}{q}\right)+O(MQ/N)\right)\\ &= O(M^{1-c/2}) + O(M^2Q/N) = O(M^{1-c/2}).\end{aligned}$$

□

II.5.4 Main contribution

Finally, we account for the arc $\mathcal{M}_{0,1}$. This is the only component that will significantly contribute to $R_\varepsilon(N)$. To simplify the argument, let us introduce an auxiliary function

$$g(x) = 2\varepsilon\sum_{n<N} e(nx). \tag{II.38}$$

By a standard argument, we have

$$g(x)^3 = (2\varepsilon)^3\sum_n S(n)e(nx),$$

where $S(n)$ denotes the number of representations of n as $n_1+n_2+n_3$ with $0 \leq n_i < N$. Of course, $S(N) = \frac{1}{2}N^2 + O(N)$.

Claim II.5.4. There exists a choice of $D_1 = D_1(\mu)$, $D_2 = D_2(\mu)$, and $N_0 = N_0(\alpha)$, such that for $N \geq N_0$, we have the bound

$$\sum_{x\in\mathbb{T}} |f(x) - g(x)| \ll N\varepsilon^3. \tag{II.39}$$

Proof. Assume throughout that N is sufficiently large in terms of α. If $x \notin \mathcal{M}_{0,1}$, then $|g(x)| \ll N\varepsilon^3$. This can be proved for instance by a very simplified version of our earlier arguments bounding $|f(x)|$. Hence, bound (II.39) follows immediately from Claim II.5.2 (for $x \in \mathfrak{m}$) or Claim II.5.3 (for $x \in \mathcal{M}_{p,q}$, $p \neq 0$). It remains to deal with $x \in \mathcal{M}_{0,1}$.

Like before in Claim II.5.3, we split $[N]$ into $N/M+O(1)$ pieces of length $M+O(1)$, where $M = \lfloor N^{1/2} \rfloor$. Let J be one of those intervals, and take $|\tau| < Q/N$. We have $e(m\tau) = e(m'\tau)+O(MQ/N)$ for any $m, m' \in J$. It follows that, for sufficiently small constant $c = c(\mu)$, we have

$$\begin{aligned}\sum_{m\in J\cap A_\varepsilon^\alpha} e(m\tau) &= |J \cap A_\varepsilon^\alpha| \left(\mathop{\mathbb{E}}_{n\in J} e(n\tau) + O(MQ/N) \right) \\ &= \left(2\varepsilon M + O(M^{1-c})\right) \left(\mathop{\mathbb{E}}_{n\in J} e(n\tau) + O(MQ/N) \right) \\ &= 2\varepsilon \sum_{m\in J} e(m\tau) + O(M^{1-c}).\end{aligned}$$

Summing over all intervals gives (II.39). □

Our main theorem now follows as a simple corollary.

Proof of Theorem II.5.1. Fix the choice of D_1, D_2 and N_0 from Claim II.5.4, and pick some $D > D_1D_2$. By picking the implicit constant in (II.31) to be sufficiently large, we may assume that Theorem II.5.1 holds for $N \leq N_0$, so we can assume that $N > N_0$.

Using Claim II.5.4 and equation II.33, we may now estimate

$$\begin{aligned}\left|R_\varepsilon^\alpha(N) - (2\varepsilon)^3 S(N)\right| &= \left| \int_{\mathbb{T}} \left(f^3(x) - g^3(x)\right) e(-Nx) dx \right| \\ &\ll \sup_{x\in\mathbb{T}} |f(x) - g(x)| \int_{\mathbb{T}} \left(|f(x)|^2 + |g(x)|^2\right) dx \\ &\ll N\varepsilon^3 \left(\varepsilon N + \varepsilon^2 N\right) \ll N^2\varepsilon^4.\end{aligned}$$

After a simple rearrangement, this gives (II.31). □

II.6 Higher degrees

In this section we deal with sets

$$A_\varepsilon^p := \{n \in \mathbb{N} \mid \|p(n)\| \leq \varepsilon(n)\}, \tag{II.40}$$

where $p\colon \mathbb{Z} \to \mathbb{R}$ is a polynomial, generally of degree higher than 2, and $\varepsilon(n)$ is a slowly decaying function. Our main goal is to prove a slight strengthening of Theorem II.1.5.i.

Theorem (II.1.5.i, reiterated)**.** *There exists a set $Z \subset \mathbb{R}$ of measure 0 such that for any $\varepsilon(n) > 0$ with $\lim \frac{\log 1/\varepsilon(n)}{\log n} \to 0$ and any $\alpha \in \mathbb{R} \setminus Z$, the set A^p_ε defined in* (II.40) *with $p(n) = \alpha n^d$ is a basis of order 2.*

Note that the restriction $\lim \frac{\log 1/\varepsilon(n)}{\log n} \to 0$ is just another way of saying that $\varepsilon(n) = n^{-o(1)}$. In particular, any function of the form $\varepsilon(n) = \log^{-C} n$ will be suitable.

We will also give a simple argument for II.1.5.ii.

Theorem (II.1.5.ii, reiterated)**.** *There exists a closed uncountable set $E \subset \mathbb{R}$ such that for any $\varepsilon_0 < \frac{1}{4}$, any $\varepsilon(n)$ with $\varepsilon(n) \leq \varepsilon_0$ and any $\alpha \in E$, the set A^p_ε defined in* (II.40) *is not a basis of order 2.*

II.6.1 Bases of order 2

In degree at least 3, the generic behaviour is that A^p_ε is a basis of order 2. We can prove a result for p varying over an affine subspace $\mathcal{P}$ of the $\mathbb{R}$-vector space $\mathbb{R}[x]$. For brevity, we refer to such $\mathcal{P}$ as an affine family of polynomials. Note that $\mathcal{P}$ has a canonical Haar measure (defined up to a constant factor), and hence we have a notion of zero measure sets.

Theorem II.6.1. *Let $\mathcal{P} \subset \mathbb{R}[x]$ be an affine family of polynomials, and let $\varepsilon(n) > 0$ be such that $\frac{\log 1/\varepsilon(n)}{\log n} \to 0$. Then at least of the following holds:*

(i) For all $p \in \mathcal{P}$ we have $\deg p \leq 2$.

(ii) There is $p \in \mathcal{P}$ such that for all $q \in \mathcal{P}$ we have $\deg p > \deg(p - q)$.

(iii) For all $p \in \mathcal{P}$ except for a set of measure 0, the set A^p_ε is a basis of order 2.

Proof of Theorem II.1.5.i assuming Theorem II.6.1. Apply Theorem II.6.1 to the linear family of polynomials $\mathcal{P} = \{\alpha x^d \mid \alpha \in \mathbb{R}\}$, with $d \geq 3$. It is clear that neither of the conditions (i) and (ii) holds for $\mathcal{P}$. Hence, we have condition (iii), which is precisely the claim of II.1.5.i. □

Perhaps a more useful restatement of the above theorem is that if $\mathcal{P}$ is an affine family of polynomials not satisfying (i) and (ii), then $\mathcal{P}$ must satisfy (iii). We clearly need to include condition (i), because the behaviour for polynomials of degree 2 is

different. Condition (ii) is meant to exclude the possibility that the behaviour A_ε^p is controlled by a highest degree term which is constant in p.

In the above theorem, we cannot replace "almost all $p \in \mathcal{P}$" with "all $p \in \mathcal{P}$", because A_ε^p need not be a basis of order 2, for example, when p is rational. We also believe there exist $p \in \mathbb{R}[x]$ with $\deg p \geq 3$ and highly irrational leading coefficients such that A_ε^p is not a basis of order 2.

Before we prove Theorem II.6.1, we will need a simple geometric lemma.

Lemma II.6.2. *Let $\mathcal{P}$ be an affine space equipped with a volume form. Let $\alpha, \beta \colon \mathcal{P} \to \mathbb{R}$ be affine forms, let $B \subset \mathcal{P}$ be an open convex set, and let $k, l \in \mathbb{Z}$ be integers such that $k\alpha + l\beta$ is non-constant. Then there exists $r_0 = r_0(\alpha, \beta, B)$ such that for $r \geq r_0$ and arbitrary $\delta > 0$ we have*

$$\mathbb{P}_{v \in rB} \left(\|k\alpha(v) + l\beta(v)\| \leq \delta \right) = 2\delta \left(1 + o(1)\right), \tag{II.41}$$

as $r \to \infty$, where the error term is bounded uniformly in δ and k, l (but may depend on α, β and B).

Proof. If α and β are affinely dependent, then the problem becomes simpler, and can be solved by an argument similar to the one presented below. Let us suppose that α, β are not affinely dependent.

It is easy to construct a parallelepiped K such that $(\alpha(v), \beta(v))$ are uniformly distributed in $\mathbb{T}^2$ for $v \in K$. If K is such parallelepiped then

$$\mathbb{P}_{v \in K} \left(\|k\alpha(v) + l\beta(v)\| \leq \delta \right) = 2\delta.$$

It is elementary that for each r, there exist collections $C_+(r)$, $C_-(r)$ of such parallelepipeds with $C_+(r) \supset rB \supset C_-(r)$ and $\frac{\mathrm{vol}(C_\pm(r))}{\mathrm{vol}(rB)} \to 1$ as $r \to \infty$. The proof now follows by a sandwiching argument. □

Proof of Theorem II.6.1. We shall assume that neither of the conditions (i), (ii) holds, and derive condition (iii). We may assume that $\varepsilon(n) \gg \frac{1}{n^\delta}$, where δ is a small positive constant yet to be determined, and that $\varepsilon(n)$ is decreasing.

Given $p \in \mathcal{P}$, we define $Z^{(p)}$ to be the set of N such that $N \notin 2A^p_{\varepsilon(N)}$. Since $Z^{(p)} \supset \mathbb{N}_0 \setminus 2A_\varepsilon^p$, it will suffice to show that $Z^{(p)}$ is almost surely finite. For this, it is enough to prove that $\mathbb{E}_{p \in rB} \left| Z^{(p)} \right| < \infty$, where $B \subset \mathcal{P}$ denotes a unit ball with respect to some norm on $\mathcal{P}$.

Take any $N \in Z^{(p)}$. Following the argument in Lemma II.3.1, the orbit $(p(n), p(N-n))$ is not $1/\varepsilon(N)^{O(1)}$-equidistributed for $0 \le n < N$. Hence, by Theorem I.1.25, there exist $(k,l) \in \mathbb{Z}^2 \setminus \{(0,0)\}$ such that

$$\|kp(n) + lp(N-n)\|_{\mathcal{C}^\infty[N]} \ll \frac{1}{\varepsilon(N)^{O(1)}} \ll N^{O(\delta)}, \quad |k|, |l| \ll N^{O(\delta)}, \tag{II.42}$$

where $\|\sum_i \alpha_i n^i\|_{\mathcal{C}^\infty[N]} := \max_i N^i \|\alpha_i\|$.

Given k, l and p, let $Z_{k,l}^{(p)}$ denote the set of all N satisfying the above bound (II.42) (for some choice of the implicit constants). We have

$$Z^{(p)} \subset \bigcup_{(k,l)\neq(0,0)} Z_{k,l}^{(p)}.$$

We will allow for a certain finite set $Z_* \subset \mathbb{N}$ of N which may belong to particularly many sets $Z_{k,l}^{(p)}$. It will suffice if (for suitable choice of Z_*) we prove the following bound:

$$\sum_{(k,l)\neq(0,0)} \mathop{\mathbb{E}}_{p\in rB} \left| Z_{k,l}^{(p)} \setminus Z_* \right| < \infty. \tag{II.43}$$

We can write $p(x) = \sum_{i=0}^d \alpha_i^{(p)} x^i$, where $\alpha_i^{(p)}$ are affine functions of p. Because conditions (ii) and (iii) do not hold, $\alpha_d^{(p)}$ is not a constant function of p, and $d \ge 3$.

A straightforward manipulation of (II.42) shows that if $N \in Z_{k,l}^{(p)}$ then

$$\left\| k\alpha_d^{(p)} + (-1)^d l\alpha_d^{(p)} \right\| \ll \frac{1}{N^{d-O(\delta)}}, \tag{II.44}$$

$$\left\| k\alpha_{d-1}^{(p)} - (-1)^d l\alpha_{d-1}^{(p)} - (-1)^d l\alpha_d^{(p)} N \right\| \ll \frac{1}{N^{d-1-O(\delta)}}. \tag{II.45}$$

If $k + (-1)^d l \neq 0$ then the first bound (II.44) together with Lemma II.6.2 implies for $r \ge r_0 = r_0(\alpha_d, \alpha_{d-1}, B)$ that

$$\mathbb{P}_{p\in rB}\left(N \in Z_{k,l}^{(p)}\right) \ll \frac{1}{N^{d-O(\delta)}}. \tag{II.46}$$

Else, if $k + (-1)^d l = 0$, then likewise the second bound (II.45) together with Lemma II.6.2 implies for $r \ge r_0$ that

$$\mathbb{P}_{p\in rB}\left(N \in Z_{k,l}^{(p)}\right) \ll \frac{1}{N^{d-1-O(\delta)}}, \tag{II.47}$$

unless $2\alpha_{d-1}^{(p)} + N\alpha_d^{(p)}$ is constant in p. The latter condition can only hold for a single value of N, independent of k and l. Letting Z_* consist of this specific N (or $Z_* = \emptyset$ if no such N exists), we conclude that for any N we have the bound

$$\mathbb{P}_{p\in rB}\left(N \in Z_{k,l}^{(p)} \setminus Z_*\right) \ll \frac{1}{N^{d-1-O(\delta)}}. \tag{II.48}$$

Because for $N \in Z_{k,l}^{(p)}$ we have $|k|, |l| \ll N^{O(\delta)}$, at the cost of worsening implicit constants, we may rewrite (II.47) as

$$\mathbb{P}_{p\in rB}\left(N \in Z_{k,l}^{(p)} \setminus Z_*\right) \ll \frac{1}{(k^4+l^4)N^{d-1-O(\delta)}} \tag{II.49}$$

Taking δ sufficiently small, we can now derive

$$\sum_{k,l} \mathop{\mathbb{E}}_{p\in rB} \left|Z_{k,l}^p \setminus Z_*\right| \ll \sum_{k,l} \frac{1}{k^4+l^4} \sum_N \frac{1}{N^{2.1}} < \infty. \tag{II.50}$$

This finishes the proof. □

Remark. The same ideas can be applied to higher dimensions. One then defines

$$A_\varepsilon^p := \{n \in \mathbb{N} \mid \|p_i(n)\| \leq \varepsilon(n),\ 1 \leq i \leq r\},$$

for a polynomial map $p(n) = (p_i(n))_{i=1}^r$. For $\varepsilon(n) \geq \varepsilon_0 > 0$, these sets will generically be bases of order 2.

Because the general version of the equidistribution Theorem I.1.25 holds for arbitrary nilmanifolds, similar arguments can be applied to "generic" nil–Bohr sets (see Section I.1.7 for more details). Unfortunately, this genericity assumption cannot be removed, which makes the resulting theorem much less appealing.

II.6.2 Non-bases of order 2

We close this section considering situations when the sets A_ε^p fail to be bases of order 2. We show that for higher degrees of polynomials, it is still possible for A_ε^p to fail to be a basis of order 2. For the sake of concreteness, we work with the polynomials of the specific form $p(n) = \alpha n^d$.

Proof of Theorem II.1.5.ii. Let d be fixed. We first claim, in analogy to Lemma II.2.2, that for any $\varepsilon_0 < \frac{1}{4}$ there is some $N_0 = N_0(d, \varepsilon_0)$ such that if $N > N_0$ is odd and $\varepsilon(n) \leq \varepsilon_0$ for all n, and if N and α satisfy

$$\left\|N\alpha - \frac{1}{2}\right\| \leq \frac{1}{N^d},$$

then $N \notin 2A_\varepsilon^p$. Indeed, if $0 \leq n_1, n_2 \leq N$ are such that $n_1 + n_2 = N$ then

$$\left\|n_1^d\alpha - (-1)^d n_2^d\alpha\right\| = \left\|\sum_{j=0}^{d-1}(-1)^j n_1^{d-1-j} n_2^j N\alpha\right\| = \frac{1}{2} + O\left(1/N\right),$$

where the implicit constant in the error term depends only on d. Thus it is impossible that $n_1, n_2 \in A_\varepsilon^p$ if $\varepsilon(n) \le \varepsilon_0 < \frac{1}{4}$ and N is sufficiently large in terms of ε_0.

Let N_i be a rapidly increasing sequence of odd integers, and set

$$\Gamma := \bigcap_{i=1}^{\infty} \Gamma_i, \quad \Gamma_i := \left\{ \alpha \in \mathbb{T} \;\middle|\; \|N_i \alpha - 1/2\| \le \frac{1}{N_i^d} \right\}.$$

For any $\alpha \in \Gamma$ and any $\varepsilon_0 < \frac{1}{4}$, we have by the above observation that $N_i \notin 2A_\varepsilon^p$ for all but finitely many i, provided that $\varepsilon(n) \le \varepsilon_0$. It remains to show that Γ is uncountable.

Note that for each i, Γ_i is a union of N_i closed intervals of length $\frac{2}{N_i^{d+1}}$, equally spaced in $\mathbb{T}$. Assuming N_i are increasing rapidly enough, each set $\bigcap_{j<i} \Gamma_j$ is a union of closed intervals, and each of these intervals intersects at least two different intervals in Γ_i.

It now follows easily that Γ contains a homeomorphic copy of the Cantor set, and hence is uncountable. The set E in II.1.5.ii can be taken to be $\Gamma + \mathbb{Z}$. $\square$

Chapter III

Weakly mixing sets and polynomial equations

III.1 Introduction

In this chapter, we investigate patterns appearing in *weakly mixing* sets of integers, which were first proposed by Furstenberg, and then studied by Fish [41], [40]. Motivated by Furstenberg correspondence principle I.1.8, in Section III.2 we define weakly mixing sets as sets of recurrence times in weakly mixing systems. For now, suffice it to say that weakly mixing sets are *pseudorandom*, in the sense that, while being deterministic, they enjoy many of the properties of random sets. The original motivation for this line of study was an attempt to provide a counterpart to results from Chapter II in a pseudorandom setting, although the project eventually progressed in a somewhat different direction.

To motivate introduction of the weakly mixing sets sets, let us first recall that subsets of integers with positive density are guaranteed to contain various configurations. We recall one of the most general results among the progeny of Szemerédi's theorem.

Theorem (I.1.10)**.** *Let $A \subset \mathbb{N}_0$ be a set with positive Banach density, and let $p_i \in \mathbb{Z}[x]$ for $1 \leq i \leq r$ be an intersective family of polynomials with $p_i(n) \to \infty$ as $n \to \infty$. Then, there exists $m, n \in \mathbb{N}_0$ such that*

$$m, m + p_1(n), \ldots, m + p_r(n) \in A.$$

Note that the conclusion of the above theorem fails if p_i are not intersective. Moreover, the offending set A can be very structured: indeed, an (infinite) arithmetic progression will do.

On the other hand, one expects that more can be proved if A is forced to be *unstructured*. In the extreme case, when A is a random set, constructed by declaring $n \in A$ with a certain probability $p > 0$, independently for all n, then with probability 1, A will contain many occurrences of the pattern, $m, m + p_1(n), m + p_2(n), \dots, m + p_r(n)$ for any polynomials $p_1, \dots, p_r$ (or, indeed, the pattern $m, m + a_1, \dots, m + a_r$ for any $a_i \in \mathbb{Z}$). Thus, it is of interest to see which notions of pseudo-randomness guarantee existence of various patterns. Since the proof of Theorem I.1.10 is ergodic, it is natural to consider notions of pseudo-randomness originating from measurable dynamics.

The notion of pseudo-randomness which we investigate here is that of being weakly mixing. We point out that that any *normal* set is weakly mixing, where we recall that a set A is normal if in the characteristic sequence $(1_A(n))_{n=1}^{\infty}$, any pattern $\epsilon = \epsilon_1\epsilon_2 \dots \epsilon_r$ of 0's and 1's appears with the same limiting frequency as for a genuinely random set, that is $p^{|\epsilon|}(1-p)^{r-|\epsilon|}$, where $|\epsilon| = \sum_i \epsilon_i$. On the other hand, one can show that weakly mixing sets are Gowers-uniform of all orders, in the sense that for each s, $\|1_A\|_{U^s[N]} = o(1)$ as $N \to \infty$ (for definition, see Section I.1.11).

In [41], Fish characterised all the linear patterns which are guaranteed to appear in a weakly mixing set. We give a special (yet representative) case of this characterisation. Alternatively, this can be seen as a consequence of the Generalised von Neumann Theorem.

Theorem III.1.1 (Fish [41]). *Let A be a weakly mixing set. Suppose that $a_i, b_i \in \mathbb{N}$ and $c_i \in \mathbb{Z}$ for $1 \le i \le r$ are such that for all $i \neq j$ we have* $\det \begin{bmatrix} a_i & a_j \\ b_i & b_j \end{bmatrix} \neq 0$. *Then, there exist $n, m \in \mathbb{N}$ such that $a_i n + b_i m + c_i \in A$ for all $1 \le i \le r$.*

For example, a weakly mixing set A will contain the pattern $n, m, n + m, n + 2m, 2n + m$, which is not guaranteed to appear on the grounds of density alone. However, unlike in the case of a normal set, a weakly-mixing set is not guaranteed to contain two consecutive elements $n, n+1$. When it comes to polynomial patterns, one has a result of a somewhat different flavour, which bears resemblance to Theorem I.1.10.

Theorem III.1.2 (Fish [40]). *Let $A \subset \mathbb{N}_0$ be a weakly-mixing set, and let $B \subset \mathbb{N}_0$ be a set with positive density. Let $p_i(x) \in \mathbb{Z}[x]$ for $1 \le i \le r$ be polynomials of equal degree, such that for all $i \neq j$ we have* $\deg(p_i - p_j) > 0$, *and $p_i(n) \to \infty$ as $n \to \infty$.*

Then, for all $n \in \mathbb{N}_0$ except[1] *for a set of density* 0*, there exist $m \in B$ such that $p_1(n) - m, p_2(n) - m, \ldots, p_r(n) - m \in A$.*

Returning to our original motivation, one of the questions investigated in Chapter II was whether the sets

$$A = \{n \in \mathbb{N}_0 \mid \|q(n)\| \leq \varepsilon\} \tag{III.1}$$

are bases for $\mathbb{N}_0$, where $q(x) \in \mathbb{R}[x]$ and $\varepsilon > 0$. A special, yet representative, case, we may take B to be the Bohr set

$$B = \{n \in \mathbb{N}_0 \mid \|n\alpha\| \leq \varepsilon\}, \tag{III.2}$$

where $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and $\varepsilon > 0$, and ask the following question.

Question III.1.3. *Let $p(x) \in \mathbb{Z}[x]$ be a polynomial with integer coefficients, and let B be given by* (III.2) *for some $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and $\varepsilon > 0$. Is it true that any sufficiently large integer N can be represented as $N = n_1 + n_2$, where $p(n_1), p(n_2) \in B$?*

In this chapter, we investigate what happens when the highly structured Bohr set B in Question III.1.3 is replaced with a weakly mixing set, which is highly pseudo-random.

When $\deg p = 1$, then it is not a significant loss of generality to assume that $p(x) = x$. The question then becomes: Is any weakly-mixing set A a basis of order 2? The answer to this is negative, but A is *almost* a basis of order 2, in the sense $A + A$ has density 1 (see [41]). In the case when $\deg p \geq 2$, one cannot expect to guarantee that a weakly-mixing set A contains *any* elements from $\{p(n) \mid n \in \mathbb{N}_0\}$. Indeed, if A is weakly-mixing, then so is $A \setminus Z$ for any Z of density 0, and thus in particular $A \setminus \{p(n) \mid n \in \mathbb{N}_0\}$ is weakly mixing. However, we are able to prove the following.

Theorem III.1.4. *Let $A \subset \mathbb{N}_0$ be a weakly-mixing set, and let $p(x) \in \mathbb{Z}[x]$ be a non-constant polynomial. Then, all $N \in \mathbb{N}_0$ except for a set of* 0 *density can be represented as $N = n_1 + n_2$, where $n_1, n_2 \in \mathbb{N}_0$ are such that $p(n_1) + m \in A$ and $p(n_2) + m \in A$ for some $m \in A$. Moreover, if* $\deg p$ *is odd then the same conclusion holds for all but finitely many N.*

The above theorem is a direct consequence of two more technical results, which may be of independent interest. To formulate them, we need to introduce some

[1] Here and elsewhere, when a statement is said to hold for "all N except for a set of 0 density", we mean that there exists a set $S \subset \mathbb{N}_0$ with density 1 such that the statement holds for all $N \in S$. The meaning of the phrase "all but finitely many" is analogous.

terminology. For a polynomial p, we denote by $\deg p$ and $\operatorname{lc} p$ the degree and leading coefficient of p, respectively, so that the leading term of p is $x^{\deg p} \operatorname{lc} p$. We shall say that a family of polynomials $p_i^{(N)}(x) \in \mathbb{Z}[x]$ for $1 \leq i \leq r,\ N \in \mathbb{N}_0$, is *uniformly admissible* if the following conditions hold:

(i) For each i and N, $\deg p_i^{(N)} > 0$, and $\operatorname{lc} p_i^{(N)},\ \deg p_i^{(N)}$ do not depend on N.

(ii) For each $i \neq j$, $\deg(p_i^{(N)} - p_j^{(N)}) > 0$, and $\operatorname{lc}(p_i^{(N)} - p_j^{(N)}),\ \deg(p_i^{(N)} - p_j^{(N)})$ do not depend on N.

For instance, the pair $p_1^{(N)}(n) = n^3,\ p_2^{(N)}(n) = (N-n)^3$ is uniformly admissible, but the pair $p_1^{(N)}(n) = n^2,\ p_2^{(N)}(n) = (N-n)^2$ is not.

Theorem III.1.5. *Let $A \subset \mathbb{N}_0$ be a weakly-mixing set, and let $B \subset \mathbb{N}_0$ be a set with positive density. Let $p_i^{(N)}(x) \in \mathbb{Z}[x]$ for $1 \leq i \leq r,\ N \in \mathbb{N}_0$, be a family of polynomials which is* uniformly admissible. *Then, there exists N_0 such that for any $N > N_0$, there are $0 \leq n < N$ and $m \in B$ such that*

$$p_1^{(N)}(n) + m, \ldots, p_r^{(N)}(n) + m \in A$$

Rather than fully general families of polynomials, we are interested specifically in those which are themselves given by polynomial formulas. In other words, we are concerned with a sequence $p_i^{(N)}(x) \in \mathbb{Z}[x, N],\ 1 \leq i \leq r$. We will say that such sequence is *admissible* if the following holds:

(i) For each i, $\deg_x p_i^{(N)}(x) > 0$ (as polynomial in two variables).

(ii) For each $i \neq j$, $\deg_x(p_i^{(N)}(x) - p_j^{(N)}(x)) > 0$.

For instance, the pair $p_1^{(N)}(n) = n^k,\ p_2^{(N)}(n) = (N-n)^k$ is admissible for each $k \in \mathbb{N}$.

Theorem III.1.6. *Let $A \subset \mathbb{N}_0$ be a weakly-mixing set, and let $B \subset \mathbb{N}_0$ be a set with positive density. Let $p_i^{(N)}(x) \in \mathbb{Z}[x, N]$ for $1 \leq i \leq r$ be a polynomial family of polynomials, which is* admissible. *Then, there exists a set $S \subset \mathbb{N}_0$ with density $d(S) = 1$, such that for all $N \in S$, there are $0 \leq n < N$, $m \in B$ such that*

$$p_1^{(N)}(n) + m, \ldots, p_r^{(N)}(n) + m \in A.$$

Proof of Theorem III.1.4 assuming III.1.5 and III.1.6. In the case when $\deg p$ is odd, the pair $p_1^{(N)}(x) = p(x),\ p_2^{(N)}(x) = p(N-x)$ is uniformly admissible, hence by Theorem III.1.5 applied with $B = A$, for all sufficiently large N there exist $n, m \in A$ such that $p(n) + m \in A,\ p(N-n) + m \in A$. It remains to put $n_1 = n,\ n_2 = N - n$.

In the case when $\deg p$ is even, we apply Theorem III.1.6, and use the fact that the pair $p_1^{(N)}(x), p_2^{(N)}(x)$ just defined (but now viewed as an element of $\mathbb{Z}[x, N]$) is admissible. The remainder of the argument is fully analogous. □

Remark. Our proof of Theorem III.1.4 depends on the parity of $\deg p$ in a crucial way. However, it is not a priori clear that the conclusion of this theorem should depend on $\deg p$. In fact, the author believes that the stronger conclusion (the set of exceptional N being finite) holds also when $\deg p$ is even, but it does not appear to be possible to obtain this result with our methods.

This chapter is organised as follows. In Section III.2 we give basic definitions, specifically we define the weakly-mixing sets. In Section III.3 we reduce Theorem III.1.5 to a uniform convergence statement in ergodic theory, and prove a special case of it. In Section III.4 we introduce the PET induction and finish the proof of Theorem III.1.5. In Section III.5 we again reduce Theorem III.1.6 to a statement about certain ergodic averages, and then prove this statement.

Methods in this chapter draw heavily on the work of Bergelson [10] and Fish [41], [40], and may be primarily of interest as expository material. Many of the ideas we use can be traced back to their, or earlier, work.

III.2 Definitions

It will be convenient to view a set $A \subset \mathbb{N}_0$ of positive upper density as arising from dynamics. Let $\Omega := \{0,1\}^{\mathbb{N}_0}$ denote the *shift space*, taken with the natural product topology and the Borel σ-algebra. On Ω, we may define the *shift map* given by $(Sx)(i) := x(i+1)$. To $A \subset \mathbb{N}_0$ we can always associate its characteristic function $1_A \in \Omega$, which gives rise the *subshift* $X_A := \mathrm{cl}\,\{S^n 1_A \mid n \in \mathbb{N}_0\}$, which is evidently a closed and S-invariant subspace of Ω.

Definition III.2.1. A set $A \subset \mathbb{N}_0$ with positive upper density is *weakly-mixing* if and only if the point $1_A \in X_A$ is generic for some ergodic S-invariant probability

measure μ_A (which is necessarily unique), such that the resulting measure preserving system denoted $\mathcal{X}_A = (X_A, S, \mathcal{B}(X_A), \mu_A)$ is weakly mixing.

We stress that a weakly mixing set A is in particular required to have positive upper density, and it has a density since

$$\mathop{\mathbb{E}}_{n\leq N} 1_A(n) \xrightarrow[N\to\infty]{} \mu(\{x \in \Omega \mid x(0) = 1)\}\,.$$

A seemingly more general definition of weakly mixing systems is possible. Note that any weakly mixing measure preserving system gives rise to multiple examples of weakly mixing sets of integers via the following construction.

Lemma III.2.2. *A set $A \subset \mathbb{N}_0$ is weakly mixing if and only if it takes the form $A = \{n \in \mathbb{N}_0 \mid f(T^n x_0) = 1\}$, where $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ is a weakly mixing system, $f \in L^\infty(\mu)$ takes values 0 and 1, $\int f d\mu > 0$, and x_0 is f-*generic*. Here, a point x_0 is f-generic if for any g in the algebra generated by f, Tf, $T^2 f, \ldots$, we have the convergence of the averages:*

$$\mathop{\mathbb{E}}_{n<N} g(T^n x_0) \xrightarrow[N\to\infty]{} \int_X g d\mu.$$

Proof. Clearly, any weakly mixing system is of the aforementioned form, with $\mathcal{X} = \mathcal{X}_A$, so only one implication needs to be proved. Suppose that a set

$$A = \{n \in \mathbb{N}_0 \mid f(T^n x_0) = 1\}$$

is as in the latter definition. Define the measurable map $F\colon X \to \Omega$ given by $x \mapsto (f(T^n x))_{n\in\mathbb{N}_0}$. It is clear that $F \circ T = S \circ F$, and hence the pushforward $\nu := F_*\mu$ is a S-invariant measure on Ω. Since $\mathcal{Y} = (\Omega, S, \mathcal{B}(\Omega), \nu)$ is a factor of X, it is easy to check that $\mathcal{Y}$ is weakly mixing. Note also that $1_A = F(x_0)$.

It remains to check that 1_A is generic for thus defined ν. It will suffice to verify that for any cylinder $U = \{x \in \Omega \mid x(0) = \epsilon_0, \ldots, x(r) = \epsilon_r\}$ it is the case that $\mathbb{E}_{n<N} 1_U(S^n 1_A) \to \nu(U)$. But this is an easy consequence of f-genericity of x_0. Indeed, let $f_i(x) = f(x)$ if $\epsilon_i = 1$ and $f_i(x) = 1 - f(x)$ if $\epsilon_i = 0$; then

$$\mathop{\mathbb{E}}_{n<N} 1_U(S^n 1_A) = \mathop{\mathbb{E}}_{n<N} \prod_{i\leq r} f_i(T^{n+i} x_0) \xrightarrow[N\to\infty]{} \int_X \prod_{i\leq r} T^i f_i d\mu = \nu(U). \quad \square$$

We close with a remark on invertible extensions. Any of the systems $\mathcal{X}_A$ mentioned above has an invertible extension, such that any generic point lifts to an ergodic point.

Hence, for any weakly mixing set A, we may assume that it originates from an invertible weakly mixing system via the construction in Lemma III.2.2. In simpler terms, we may relate A to a two-sided shift rather than the one-sided one used in Definition III.2.2. Most of the time, we assume invertibility for the sake of convenience, but our results on ergodic averages, such as Theorem III.3.2 and III.5.1 remain true for non-invertible systems with minor modifications.

III.3 Uniform ergodic theorem

In this section, we will first show how Theorem III.1.5 can be deduced from a result in ergodic theory, namely Theorem III.3.2, concerning uniform convergence of certain averages for weakly mixing systems. We then go on to prove this statement in the second part of this section and Section III.4, thus completing the proof of Theorem III.1.5.

III.3.1 Outline and initial reductions

We begin with the aforementioned reduction to ergodic theory. Let us place ourselves in the situation of Theorem III.1.5, so in particular A is a weakly mixing set. Because A is already related to an m.p.s. $\mathcal{X}_A = (X_A, S, \mathcal{B}_A, \mu_A)$, it comes as no surprise that we will be interested in averages of functions for this system.

Fix a family of polynomials $p_i^{(N)}$, $1 \leq i \leq r$ as in Theorem III.1.5 or III.1.6, and assume for simplicity that $B = A$. For large integers M, N, let $\mathcal{N}(N, M)$ denote the number of solutions to

$$p_1^{(N)}(n) + m \in A, \quad \ldots, \quad p_r^{(N)}(n) + m \in A, \quad m \in A \tag{III.3}$$

with $0 \leq n < N$, $0 \leq m < M$. Let also $f_0 \in \mathcal{C}(X_A)$ be the function given by $f_0(x) = x(0)$. We may then approximate, at least heuristically, that

$$\begin{aligned}
\frac{1}{NM}\mathcal{N}(N,M) &= \mathop{\mathbb{E}}_{m<M}\mathop{\mathbb{E}}_{n\leq N} 1_A(m)\prod_{i=1}^{r} 1_A(p_i^{(N)}(n)+m) \\
&= \mathop{\mathbb{E}}_{m<M}\mathop{\mathbb{E}}_{n<N} S^m f_0(1_A)\prod_{i=1}^{r} S^{p_i^{(N)}(n)+m} f_0(1_A) \\
&\overset{(1)}{\approx} \int_{X_A} f_0 \cdot \mathop{\mathbb{E}}_{n<N}\prod_{i=1}^{r} S^{p_i^{(N)}(n)} f_0 \, \mathrm{d}\mu_A \\
&\overset{(2)}{\approx} \left(\int_{X_A} f_0(x)\, \mathrm{d}\mu_A\right)^{r+1} = d(A)^{r+1}.
\end{aligned}$$

The approximation labelled (1) is simply the ergodic theorem, and is valid as long as M is sufficiently large, with N fixed. The key difficulty lies in making precise and justifying step (2), which will involve understanding the convergence of averages such as the one under the integral.

Study of similar averages was pioneered by Bergelson in [10], but without the dependence of the polynomials on N. We shall call a sequence of $r \geq 1$ polynomials $(p_i)_{i=1}^r$ *admissible* if none of p_i and $p_i - p_j$ with $i \neq j$ are constant.

Theorem III.3.1 (Bergelson [10])**.** *Suppose that an m.p.s. $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ is weakly mixing and invertible. Let $(p_i)_{i=1}^r$ be an admissible sequence of polynomials. Let $f_i \in L^\infty(\mu)$ for $1 \leq i \leq r$. Then,*

$$\mathop{\mathbb{E}}_{n<N} \prod_{i=1}^{r} T^{p_i(n)} f_i \xrightarrow[N\to\infty]{L^2} \prod_{i=1}^{r} \int f_i d\mu. \tag{III.4}$$

Here, we need a slight variation of the above theorem, already mentioned in the introduction. Refining the notion of admissibility, we shall call a family of sequences of polynomials $(p_i^{(t)})_{i=1}^r$ *uniformly admissible* if for any i, $\deg(p_i^{(t)}) > 0$ and $\mathrm{lc}(\deg(p_i^{(t)})$ are independent of t, and if likewise for any $i \neq j$, $\deg(p_i^{(t)} - p_j^{(t)}) > 0$ and $\deg(p_i^{(t)} - p_j^{(t)})$ are independent of t. (Here, t runs over some unspecified index set I.) For example, the family $(x^2 + a_1^{(t)}, x^2 + x + a_2^{(t)})$ is uniformly admissible, but $(x^2 + b_1^{(t)} x + a_1^{(t)}, x^2 + x + a_2^{(t)})$ is, in general, not (unless $b_1^{(t)}$ is independent of t and $b_1^{(t)} \neq 1$).

Theorem III.3.2. *Suppose that an m.p.s. $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ is weakly mixing and invertible. Let $(p_i^{(t)})_{i=1}^r$ be a uniformly admissible family of sequences of polynomials, indexed by $t \in I$. Let $f_i \in L^\infty(\mu)$ for $1 \leq i \leq r$. Then,*

$$\mathop{\mathbb{E}}_{n<N} \prod_{i=1}^{r} T^{p_i^{(t)}(n)} f_i \xrightarrow[N\to\infty]{L^2} \prod_{i=1}^{r} \int f_i d\mu, \quad \textit{uniformly in } t. \tag{III.5}$$

Before embarking upon the proof of the above theorem, we explain how it completes the proof of the first of our main results.

Proof of Theorem III.1.5 assuming Theorem III.3.2. Recall that A takes the form

$$A = \{n \in \mathbb{N}_0 \mid f(T^n x) = 1\}$$

where $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ is a weakly mixing m.p.s. and x_0 is a generic point for the $\{0,1\}$-valued function $f \in L^\infty(\mu)$ with $\int f d\mu = d(A)$. Without loss of generality, we may assume that $\mathcal{X}$ is invertible.

Suppose, for the sake of contradiction, that the system

$$p_1^{(N)}(n) + m \in A, \quad \dots, \quad p_r^{(N)}(n) + m \in A, \quad m \in B \tag{III.6}$$

has no solution, and consider the quantity

$$\mathcal{L}(N) = \limsup_{M\to\infty} \left| \mathop{\mathbb{E}}_{m<M} \mathop{\mathbb{E}}_{n<N} 1_B(m) \left(\prod_{i=1}^{r} 1_A(m + p_i^{(N)}(n)) - d(A)^r \right) \right|,$$

which can be construed as the (normalised) deviation of the number of solutions to (III.6) from the expected value of $d(B)d(A)^r MN$. On one hand, since (III.6) lacks solutions, we have $\mathcal{L}(N) = d(B)d(A)^r$. On the other hand, we may approximate, with the use of Cauchy-Schwartz and the ergodic theorem, that

$$\begin{aligned}
\mathcal{L}(N) &\leq \lim_{M\to\infty} \left(\mathop{\mathbb{E}}_{m<M} \left(\mathop{\mathbb{E}}_{n<N} \prod_{i=1}^{r} 1_A(m + p_i^{(N)}(n)) - d(A)^r \right)^2 \right)^{1/2} \\
&= \left(\int_X \left(\mathop{\mathbb{E}}_{n<N} \prod_{i=1}^{r} T^{p_i^{(N)}(n)} f - d(A)^r \right)^2 d\mu \right)^{1/2} \\
&= \left\| \mathop{\mathbb{E}}_{n<N} \prod_{i=1}^{r} T^{p_i^{(N)}(n)} f - d(A)^r \right\|_{L^2(\mu)}.
\end{aligned}$$

An application of Theorem III.3.2 gives, in particular:

$$\mathop{\mathbb{E}}_{n<N} \prod_{i=1}^{r} T^{p_i^{(N)}(n)} f \xrightarrow[N\to\infty]{L^2} d(A)^r.$$

Thus, if N is large enough, then we conclude that $\mathcal{L}(N) < d(B)d(A)^r$, which is the sought for contradiction. □

Remark. In the case $A = B$, a similar reasoning gives the asymptotic formula for the number of solutions $\mathcal{N}(N, M)$ mentioned at the beginning of this section:

$$\mathcal{N}(N, M) = d(A)^{r+1} MN(1 + o_{N\to\infty}(1) + o_{N;M\to\infty}(1)).$$

III.3.2 Uniform convergence for linear polynomials

The remainder of this section is devoted to proving Theorem III.3.2 in the linear case; we delegate higher degrees to Section III.4. In our argument, we follow the approach of Bergelson rather closely, taking care to account for uniformity of convergence. We will need an uniform version of van der Corput Lemma, which is a slight variation on the usual statement. We include the proof, which is rather standard for the convenience of the reader.

Lemma III.3.3 (Uniform van der Corput)**.** *Suppose that $(u_n^{(t)})_{n=1}^{\infty}$ is a sequence of vectors in a Hilbert space $\mathcal{H}$ with $\left\|u_n^{(t)}\right\| \leq 1$, indexed by $t \in I$. Suppose further that for some $s_h^{(t)} \in \mathbb{R}_{\geq 0}$ we have*

$$\left| \mathop{\mathbb{E}}_{n<N} \left\langle u_n^{(t)},\ u_{n+h}^{(t)} \right\rangle \right| \leq s_h^{(t)} + o_{N\to\infty}(1), \tag{III.7}$$

where the error term is uniform in t. Suppose further that

$$\mathop{\mathbb{E}}_{h<H} s_h^{(t)} \xrightarrow[H\to\infty]{} 0, \quad \textit{uniformly in } t. \tag{III.8}$$

Then we have

$$\mathop{\mathbb{E}}_{n<N} u_n^{(t)} \xrightarrow[N\to\infty]{} 0, \quad \textit{uniformly in } t. \tag{III.9}$$

Proof. For fixed H we have

$$\left\| \mathop{\mathbb{E}}_{n<N} \mathop{\mathbb{E}}_{h<H} u_{n+h}^{(t)} \right\|^2 = \left\| \mathop{\mathbb{E}}_{n<N} u_n^{(t)} \right\|^2 + O\left(1/N\right).$$

By Cauchy-Schwartz inequality, we can bound:

$$\left\| \mathop{\mathbb{E}}_{n<N} \mathop{\mathbb{E}}_{h<H} u_{n+h}^{(t)} \right\|^2 \leq \mathop{\mathbb{E}}_{h,h'<H} \left| \mathop{\mathbb{E}}_{n<N} \left\langle u_{n+h}^{(t)},\ u_{n+h'}^{(t)} \right\rangle \right|.$$

For each summand above we have the bound:

$$\left| \mathop{\mathbb{E}}_{n<N} \left\langle u_{n+h}^{(t)},\ u_{n+h'}^{(t)} \right\rangle \right| = \left| \mathop{\mathbb{E}}_{n<N} \left\langle u_n^{(t)},\ u_{n+|h-h'|}^{(t)} \right\rangle \right| + o_{H;N\to\infty}\left(1\right) \leq s_h^{(t)} + o(1).$$

It follows that

$$\left\| \mathop{\mathbb{E}}_{n<N} u_n^{(t)} \right\|^2 \leq \mathop{\mathbb{E}}_{h,h'<H} s_{|h-h'|}^{(t)} + o_{H;N\to\infty}(1).$$

Summing by parts, we conclude that

$$\mathop{\mathbb{E}}_{h,h'<H} s_{|h-h'|}^{(t)} \leq \mathop{\mathbb{E}}_{h<H} \frac{h}{H} \mathop{\mathbb{E}}_{h'\leq h} s_h^{(t)} + o_{H\to\infty}(1).$$

Let $K \leq H$ be arbitrary. Splitting the above average into $h < K$ and $K \leq h < H$ gives:

$$\mathop{\mathbb{E}}_{h<H} \frac{h}{H} \mathop{\mathbb{E}}_{h'\leq h} s_h^{(t)} \leq \frac{K}{H} + \max_{K\leq h<H} s_h^{(t)} = o_{K;H\to\infty}(1) + o_{K\to\infty}(1).$$

It follows that

$$\left\| \mathop{\mathbb{E}}_{n<N} u_n^{(t)} \right\|^2 = o_{H;N\to\infty}(1) + o_{K;H\to\infty}(1) + o_{K\to\infty}(1),$$

which implies the sought convergence. □

Proof of Theorem III.3.2, linear case. We will now deal with the case of Theorem III.3.2 when $\deg p_i^{(t)} = 1$ for all i. In this case, $p_i^{(t)}$ are necessarily of the form $p_i^{(t)}(x) = a_i x + b_i^{(t)}$, where a_i are distinct integers which do not depend on t.

We may assume without loss of generality that for each i, we have $\int f_i d\mu = 0$. Indeed, if this is not the case, we may simply replace the original functions by $\tilde{f}_i := f_i - \int f_i d\mu$. Likewise, we may assume $\|f_i\|_\infty \leq 1$, else we may rescale.

The case $r = 1$, when there is only one polynomial, is simple. Indeed, then

$$\mathop{\mathbb{E}}_{n<N} T^{p_1^{(t)}(n)} f_1 = T^{b_i^{(t)}} \left(\mathop{\mathbb{E}}_{n<N} T^{a_1 n} f_1 \right).$$

The average in the brackets does not depend on t, and converges to 0 in L^2 which follows e.g. from Theorem III.3.1. Since $T^{b_i^{(t)}}$ preserves the L^2 norm, we have

$$\mathop{\mathbb{E}}_{n<N} T^{p_1^{(t)}(n)} f_1 \xrightarrow[N\to\infty]{L^2} 0, \text{ uniformly in } t.$$

For $r \geq 2$ we proceed by induction using van der Corput Lemma. Let us write $u_n := \prod_{i=1}^r T^{p_i^{(t)}(n)} f_i$. we have

$$\begin{aligned} \mathop{\mathbb{E}}_{n<N} \left\langle u_n^{(t)},\ u_{n+h}^{(t)} \right\rangle &= \int \prod_{i=1}^{r} T^{a_i n + b_i^{(t)}} \left(f_i \cdot T^{a_i h} f_i \right) d\mu \\ &= \int \tilde{f}_{r,h} \cdot \mathop{\mathbb{E}}_{n<N} \prod_{i=1}^{r-1} T^{\tilde{a}_i n + \tilde{b}_i^{(t)}} \tilde{f}_{i,h} d\mu \end{aligned}$$

where we define:

$$\tilde{a}_i := a_i - a_r, \quad \tilde{b}_i^{(t)} := b_i^{(t)} - b_r^{(t)}, \quad \tilde{f}_{i,h} = f_i \cdot T^{a_i h} f_i.$$

The sequence of polynomials $\tilde{p}_i^{(t)}(x) = \tilde{a}_i x + \tilde{b}_i^{(t)}$ for $1 \leq i \leq r-1$ is uniformly admissible, since $\tilde{p}_i^{(t)} - \tilde{p}_j^{(t)} = p_i^{(t)} - p_j^{(t)}$. Hence, we can invoke the inductive assumption to conclude that

$$\mathop{\mathbb{E}}_{n<N} \prod_{i=1}^{r-1} T^{\tilde{a}_i n + \tilde{b}_i^{(t)}} \tilde{f}_{i,h} = \prod_{i=1}^{r-1} \int \tilde{f}_{i,h} d\mu + o_{h;N\to\infty}(1),$$

with the error term independent of t. Letting $s_h = \left| \prod_{i=1}^r \int \tilde{f}_{i,h} d\mu \right|$ we have

$$\left| \mathop{\mathbb{E}}_{n<N} \left\langle u_n^{(t)},\ u_{n+h}^{(t)} \right\rangle \right| \leq s_h + o_{h;N\to\infty}(1).$$

By a standard argument, we have $\mathbb{E}_{h<H} s_h \to 0$ as $H \to \infty$, and convergence is automatically uniform in t, since s_h does not depend on t. We are now in position to apply Lemma III.3.3 to conclude that $\mathbb{E}_{n<N} u_n^{(t)} \to 0$ in L^2 as $N \to \infty$, uniformly in t. This finishes the inductive step, and thus the proof in the case $\deg p_i^{(t)} = 1$.

□

III.4 PET induction

To prove the general case of Theorem III.3.2 we will use PET induction. We will now introduce the key concepts in separation from the proof of this particular result.

III.4.1 Definitions and basic properties

Definition III.4.1 (Characteristic vector)**.** Let $p = (p_i)_{i=1}^r$ be the a sequence of polynomials (not necessarily admissible). For $k \in \mathbb{N}$, let F_k be the set of those p_i with $\deg p_i = k$. We define the *characteristic vector* of p, which we denote by $V(p)$, by declaring $V(p)_k$ to be equal to the number of different leading coefficients $\mathrm{lc}(p)$ for $p \in F_k$.

It does not matter much if we define characteristic vectors to have entries for all k or just for $k \leq \max_i \deg p_i$. For the sake of simplicity, we assume the former. We make the set of possible characteristic vectors (i.e. $\mathbb{N}_0$-valued sequences with finitely many non-zero entries) into an ordered set by introducing reverse-lexicographical order. Recall that $V > V'$ if for the largest k with $V_k \neq V'_k$ we have $V_k > V'_k$. It is well known fact that $\mathbb{N}_0^{\mathbb{N}}$ is well-ordeder by the reverse lexicographical order. Thus, any decreasing sequence $V > V' > V'' > \dots$ has to be finite.

For the inductive step, we shall need the following operation. Let $p = (p_i)_{i=1}^r$ be a sequence of polynomials. We may, without loss of generality, assume that $\deg p_1 \geq \deg p_2 \geq \cdots \geq \deg p_r$. We then define the sequence $\tilde{p}_h$ to be the concatenation of two sequences:

$$\tilde{p}_{0,h,i}(x) := p_i(x) - p_r(x), \qquad 1 \leq i \leq r-1, \tag{III.10}$$

$$\tilde{p}_{1,h,i}(x) := p_i(x+h) - p_r(x), \qquad 1 \leq i \leq r. \tag{III.11}$$

The following statements give base for the PET induction. We cite it here merely as a list of facts. For proofs, which are not difficult, we refer the reader to [10].

Fact III.4.2. Let $p = (p_i)_{i=1}^r$ be an admissible sequence of polynomials.

(i) If $V(p)_1 = 0$, then $\tilde{p}_h$ is admissible for all but finitely many h.

(ii) The characteristic vector $V(\tilde{p}_h)$ takes the same value for all but finitely many values of h.

(iii) We have $V(\tilde{p}_h) < V(p)$.

Essentially the same statement is true with admissible sequences of polynomials replaced by a uniformly admissible families of sequences.

Lemma III.4.3. *Let $p^{(t)} = (p_i^{(t)})_{i=1}^r$ be a uniformly admissible sequence of polynomials.*

(i) The characteristic vector $V(p^{(t)})$ does not depend on t.

(ii) For all but finitely many h, the degrees and leading coefficients of $\tilde{p}_{h,k}^{(t)}$ and $\tilde{p}_{h,k}^{(t)} - \tilde{p}_{h,l}^{(t)}$ do not depend on t.

(iii) For all but finitely many h, $V(\tilde{p}_h^{(t)})$ does not depend on h and t, and we have $V(\tilde{p}_h^{(t)}) < V(p^{(t)})$.

(iv) If $V(p^{(t)})_1 = 0$ then for all but finitely many h, $\tilde{p}_h^{(t)}$ is uniformly admissible.

(v) If $V(p^{(t)})_1 \geq 1$ then for all but finitely many h, we have for any $(\sigma, i) \neq (\rho, j)$ ($\sigma, \rho = 0$ or 1, $i, j \leq r$ or $\leq r-1$ accordingly) that $\deg(\tilde{p}_{\sigma,h,i}^{(t)} - \tilde{p}_{\rho,h,j}^{(t)}) \geq 1$, except when $i = j$ and $\deg p_i^{(t)} = 1$.

Proof. Item (i) is clear, since $V(p^{(t)})$ depends only on leading coefficients and degrees of polynomials in $p^{(t)}$, and these are independent of t.

For item (ii), we first deal with leading terms of polynomials in $\tilde{p}_h^{(t)}$. These are either of the form $p_i^{(t)}(x) - p_r^{(t)}(x)$, or of the form $p_i^{(t)}(x+h) - p_r^{(t)}(x)$. In the former case, the leading term does not depend on t by assumption. In the latter case, we reduce to the former unless $\deg p_i^{(t)} = \deg p_r^{(t)}$ and $\operatorname{lc} p_i^{(t)} = \operatorname{lc} p_r^{(t)}$. When this happens, we can write $p_i^{(t)}(x) = ax^d + b_1^{(t)} x^{d-1} + q_1^{(t)}(x)$ and $p_r^{(t)}(x) = ax^d + b_2^{(t)} x^{d-1} + q_2^{(t)}(x)$, where a and $b := b_1^{(t)} - b_2^{(t)}$ are independent of t. It follows that $\operatorname{lc}(p_i^{(t)}(x+h) - p_r^{(t)}(x)) = dah + b$, except for at most one value of h, when this is 0.

Secondly, we deal with leading terms of differences. They are of the form $p_i(x + \sigma h) - p_j(x + \rho h)$ with $\sigma, \rho \in \{0, 1\}$. In the case when $\rho = 0$, we use essentially the same argument as before. The case $\rho = 1,\ \sigma = 0$ follows by the same argument as the case $\rho = 0,\ \sigma = 1$. Finally, the case $\rho = 1,\ \sigma = 1$ follows from the case $\rho = 0,\ \sigma = 0$ by a change of variable.

For item (iii), we notice that the argument in (ii) shows that $V(\tilde{p}_h^{(t)})_k = V(p^{(t)})_k$ for all $k > \deg p_r$. It will suffice to check that for $k = \deg p_r$ we have $V(\tilde{p}_h^{(t)})_k < V(p^{(t)})_k$. This follows, because each $p_i^{(t)}$ with $\deg p_i^{(t)} = \deg p_r^{(t)}$ contributes 1 to $V(\tilde{p}_h^{(t)})_k$, except for $p_r^{(t)}$ itself.

For items (iv) and (v), we notice that the only condition that remains to be checked to verify that $\tilde{p}_h^{(t)}$ are uniformly admissible for almost all h is that the differences

$p^{(t)}_{\sigma,h,i}(x) - p^{(t)}_{\rho,h,j}(x)$ should be non-constant. The only possible degrees of such difference are $\deg(p^{(t)}_i - p^{(t)}_j)$ (if the leading terms differ or $\sigma = \rho$), or $\deg p^{(t)}_i - 1$ (otherwise). In the former case, we have $\deg(p^{(t)}_i - p^{(t)}_j) > 0$ by assumption. In the latter case, we have $\deg p^{(t)}_i - 1 > 0$, unless $\deg p^{(t)}_i = 1$. □

III.4.2 Uniform convergence in higher degrees

With this machinery, we are ready to complete the proof.

Proof of Theorem III.3.2, general case. We proceed by induction of $V = V(p^{(t)})$. Because the set of all characteristic vectors is well-ordered, we may assume that the claim of the theorem is true for any $p'^{(t)}$ with $V(p'^{(t)}) < V$. We have already dealt with all $V < (0, 1, 0, \cdots)$.

Take any $V \geq (0, 1, 0, \cdots)$, any uniformly admissible $p^{(t)} = (p^{(t)}_i)_{i=1}^r$ such that $V(p^{(t)}) = V$, and let $f_i \in L^\infty(\mu)$. As before, we may assume that $\int f_i d\mu = 0$ and $\|f_i\|_\infty \leq 1$ for all i. We need to show that

$$\mathop{\mathbb{E}}_{n<N} \prod_{i=1}^r T^{p^{(t)}_i(n)} f_i \xrightarrow[N\to\infty]{L^2} 0, \quad \text{uniformly in } t.$$

Let $u_n := \prod_{i=1}^r T^{p^{(t)}_i(n)} f_i$. Bearing in mind that we hope to apply van der Corput Lemma, we compute that

$$\mathop{\mathbb{E}}_{n<N} \langle u_n,\ u_{n+h}\rangle = \mathop{\mathbb{E}}_{n<N} \int \prod_{i=1}^r T^{p^{(t)}_i(n)} f_i \cdot \prod_{i=1}^r T^{p^{(t)}_i(n+h)} f_i d\mu \tag{III.12}$$

$$= \int f_r \cdot \mathop{\mathbb{E}}_{n<N} \left(\prod_{i=1}^{r-1} T^{\tilde{p}^{(t)}_{0,h,i}(n)} f_i \cdot \prod_{i=1}^r T^{\tilde{p}^{(t)}_{1,i,h}(n)} f_i \right) d\mu \tag{III.13}$$

with $\tilde{p}_{0,h,i}$ and $\tilde{p}_{1,h,i}$ defined as in (III.10) and (III.11). Except for finitely many values of h, we have that $V(\tilde{p}_h) < V$ does not depend on t.

We now need to branch out into two cases. Suppose first that $V(p^{(t)})_1 = 0$. Then, $\tilde{p}^{(t)}_h$ is uniformly admissible and $V(p^{(t)}_h) < V$, so by the inductive assumption we may write:

$$\mathop{\mathbb{E}}_{n<N} \left(\prod_{i=1}^{r-1} T^{\tilde{p}^{(t)}_{0,h,i}(n)} f_i \cdot \prod_{i=1}^r T^{\tilde{p}^{(t)}_{1,h,i}(n)} f_i \right) = o_{h;N\to\infty}(1). \tag{III.14}$$

The decay rate implicit in the o-notation is independent of t. Hence, the assumptions of van der Corput Lemma III.3.3 are satisfied with $s_h = 0$ for all but finitely many h. Application of the lemma gives precisely the sought convergence.

Secondly, suppose that $V(p^{(t)})_1 \neq 0$. In this case, let s denote the number of linear polynomials among $p_i^{(t)}$, and let $r' = r - s + 1$. We will adapt the argument from the linear case. If the linear polynomials in $p^{(t)}$ are given by $p_i^{(t)}(x) = a_i x + b_i^{(t)}$ then the expression under the integral in (III.13) becomes:

$$\tilde{f}_r \cdot \mathop{\mathbb{E}}_{n<N} \left(\prod_{i=1}^{r'} T^{\tilde{p}_{0,h,i}^{(t)}(n)} f_i \cdot \prod_{i=1}^{r'} T^{\tilde{p}_{1,h,i}^{(t)}(n)} f_i \right) \cdot \prod_{i=r'+1}^{r-1} T^{\tilde{a}_i n + \tilde{b}_i^{(t)}} \tilde{f}_{i,h},$$

where $\tilde{a}_i = a_i - a_r$, $b_i^{(t)} = b_i^{(t)} - b_r^{(t)}$ and $\tilde{f}_{i,h} = f_i \cdot T^{a_i h} f_i$. We may now apply the inductive assumption to the uniformly admissible sequence $\tilde{p}_h'^{(t)}$, which is the concatenation of $(\tilde{p}_{0,h,i}^{(t)}(x))_{i\in[r']}$, $(\tilde{p}_{1,h,i}^{(t)})_{i\in[r']}$ and $(\tilde{a}_i x + \tilde{b}_i^{(t)})_{i\in(r',r]}$. Note that $V(\tilde{p}_h'^{(t)}) < V(\tilde{p}_h^{(t)}) < V$ and that $r' > 0$ because $V > (0, 1, 0, \dots)$. Thus, we recover the bound from (III.14), and the rest of the argument proceeds in the same way. □

III.5 Doubly polynomial averages

We now deal with polynomial families of polynomials, such as the ones which appear in Theorem III.1.6. Our first step is again to translate the problem into a question about convergence of certain polynomial averages. Fortunately, we will be able to essentially reduce the problem to known results on L^2 convergence of polynomial averages along Følner sequences.

III.5.1 Initial reductions

In this section, rather than uniform convergence we dealt with in Section III.3, we will be interested only in convergence of averages such as $\mathbb{E}_{n\leq N(t)} \prod_{i=1}^{r} T^{p_i^{(t)}(n)} f_i$, as $t \to \infty$ for a specific sequence $N(t)$. We can afford to be quite flexible in choice of $N(t)$; the only conditions we need to impose are

$$N(t) \xrightarrow[t\to\infty]{} \infty, \qquad \frac{N(t) - N(t+h)}{N(t)} \xrightarrow[t\to\infty]{} 0, \text{ for any } h \in \mathbb{Z}. \tag{III.15}$$

As defined in the introduction, a polynomial family of polynomial sequences $p_i^{(t)}(x) \in \mathbb{Z}[x, t]$, $1 \leq i \leq r$, is *admissible* if $\deg_x p_i^{(t)}(x) > 0$ and $\deg_x(p_i^{(t)}(x) - p_j^{(t)}(x)) > 0$ for all $i \neq j$.

Theorem III.5.1. *Suppose that an m.p.s. $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ is weakly mixing and invertible, and that the sequence $N(t)$ obeys* (III.15)*. Let $(p_i^{(t)})_{i=1}^r$ be an admissible*

polynomial family of sequences of polynomials. Let $f_i \in L^\infty(\mu)$. Then there exists a set $S \subset \mathbb{N}$ with density $d(S) = 1$ such that

$$\mathop{\mathbb{E}}_{n \leq N(t)} \prod_{i=1}^{r} T^{p_i^{(t)}(n)} f_i \xrightarrow[t\to\infty, t\in S]{L^2} \prod_{i=1}^{r} \int f_i d\mu. \tag{III.16}$$

Proof of Theorem III.1.6 assuming Theorem III.5.1. Clearly, $N(t) = t$ satisfies (III.15). The argument follows by a direct repetition of the proof of Theorem III.1.5 in Section III.3. □

Remark. It is an immediate consequence of Bergelson's Theorem III.3.1 that the conclusion of Theorem III.5.1 also holds when the sequence $N(t)$, instead of obeying (III.15), is sufficiently steeply increasing. This leaves open an interesting gap. It is possible that Theorem III.5.1 holds when (III.15) is replaced with the weaker condition $N(t) \to \infty$ as $t \to \infty$. In Proposition III.6.1 we verify this for a single linear polynomial.

We devote most of the rest of this section to proving Theorem III.5.1. To begin with, we cite a simple lemma, which allows us to conveniently reformulate the problem.

Lemma III.5.2. *Let $(a_n)_{n=1}^{\infty}$ be a sequence with $a_n \in [0, 1]$. Then the following conditions are equivalent:*

(i) We have convergence $\mathop{\mathbb{E}}_{n<N} a_n \xrightarrow[N\to\infty]{} 0$.

(ii) There exists $J \subset \mathbb{N}$ with $d(J) = 1$ such that $a_n \xrightarrow[n\to\infty, n\in J]{} 0$.

Proof. See [35, Chpt 2.7] □

In the situation of Theorem III.5.1 we may always assume that $\int_X f_i d\mu = 0$ and $\|f_i\| \leq 1$ for each i. With this assumption, Theorem III.5.1 will follow by Lemma III.5.2 if we are able to show that

$$\mathop{\mathbb{E}}_{t \leq M} \left\| \mathop{\mathbb{E}}_{n<N(t)} \prod_{i=1}^{r} T^{p_i^{(t)}(n)} f_i \right\|_{L^2(\mu)}^2 \xrightarrow[M\to\infty]{} 0. \tag{III.17}$$

Expanding and using Cauchy-Schwartz, we conclude that (III.17) will follow from

$$\mathop{\mathbb{E}}_{(t,n,m)\in\Phi(M)} \prod_{i=1}^{r} T^{p_i^{(t)}(n)} f_i \cdot \prod_{i=1}^{r} T^{p_i^{(t)}(m)} f_i \xrightarrow[M\to\infty]{} 0, \tag{III.18}$$

where the average is being taken over the set

$$\Phi(M) = \left\{ (t, n, m) \in \mathbb{Z}^3 \;\middle|\; 1 \leq t \leq M,\ 0 \leq n < N(t),\ 0 \leq m \leq N(t) \right\}. \tag{III.19}$$

III.5.2 Polynomial Følner averages

Recall that a sequence $(\Psi(M))_{M=1}^{\infty}$ of finite subsets of an abelian group G is a *Følner sequence* if for any $h \in G$ it holds that

$$\frac{|(\Psi(M)+h)\triangle\Psi(M)|}{|\Psi(M)|} \xrightarrow[M\to\infty]{} 0,$$

where $\triangle$ denotes the symmetric difference.

Lemma III.5.3. *If the sequence $N(t)$ obeys condition* (III.15)*, then the sequence $\Phi(M)$ defined by* (III.19) *is Følner.*

Proof. Fix a choice of $(h,a,b) \in \mathbb{Z}^3$. It will be convenient to assume a probabilistic perspective: we choose $(t,n,m) \in \Phi(M)$ uniformly at random, and show that asymptotically almost surely as $M \to \infty$ (shortened to a.a.s.) we have $(t+h, n+a, m+b) \in \Phi(M)$.

It is easy to check that for any constants A, B it holds a.a.s. that $A \le t \le M - B$. In particular, a.a.s. $1 \le t+h \le M$, and also $1 \le n+a, m+b$. For any s, we may then estimate

$$\begin{aligned}\mathbb{P}\left(n+a > N(t+h)\middle|t=s\right) &\le \mathbb{P}\left(n+a > N(t)\middle|t=s\right)\\ &\quad + \mathbb{P}\left(N(t) \ge n+a > N(t+h)\middle|t=s\right)\\ &\le \frac{a}{N(s)} + \frac{N(s)-N(s+h)}{N(s)}\end{aligned}$$

For any choice of $\varepsilon > 0$, it is true a.a.s. that $a/N(t) < \varepsilon$ and $\frac{N(t)-N(t+h)}{N(s)} < \varepsilon$; hence a.a.s. $n+a \le N(t+h)$. By a symmetric argument, a.a.s. $m+b \le N(t+h)$. □

Averages such as (III.18), or more generally of the form

$$\mathop{\mathbb{E}}_{n\in\Psi(M)} \prod_{i=1}^{r} T^{q_i(n)} f_i, \tag{III.20}$$

where $\Psi(M)$ is a Følner sequence in $\mathbb{Z}^d$ and $q_i \colon \mathbb{Z}^d \to \mathbb{Z}$ are polynomials, are well studied. To prove convergence of these averages to $\prod_{i=1}^{r} \int f_i d\mu$, one can in principle apply a standard PET induction argument reminiscent of that in [10]. Unfortunately, this result is not recorded in the literature, but we can see it as a special case of a much stronger theorem.

In a larger generality, averages such as (III.20) have been studied by Leibman without the assumption that the system should be weakly mixing. Crucially, one can

show that they converge in L^2. Even though, unlike the case of the classical ergodic theorem of von Neumann, the limit function will not in general be T-invariant, we have a convenient description of the limit in terms of the Host-Kra-Ziegler factors, discussed in Section I.1.6.

Theorem III.5.4 (Leibman [82])**.** *Suppose that an m.p.s. $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ is invertible. Let $\Psi(M)$ be a Følner sequence, and let $q_i \colon \mathbb{Z}^d \to \mathbb{Z}$ be polynomials and $f_i \in L^\infty(\mu)$ for $1 \leq i \leq r$. Then, the averages*

$$\mathop{\mathbb{E}}_{\mathbf{n}\in\Psi(M)} \prod_{i=1}^{r} T^{q_i(\mathbf{n})} f_i \tag{III.21}$$

converge in $L^2(X)$ as $M \to \infty$. Moreover, supposing that $\deg q_i > 0$ and $\deg(q_i - q_j) > 0$ for each $i \neq j$, there exists an integer k, dependent only on r and the maximal degree $\max_{1\leq i\leq r} \deg q_i$, such that the Host-Kra-Ziegler factor $\mathcal{Z}_k$ of $\mathcal{X}$ is characteristic for convergence of the averages (III.21)*, in the sense that*

$$\lim_{M\to\infty} \mathop{\mathbb{E}}_{\mathbf{n}\in\Psi(M)} \prod_{i=1}^{r} T^{q_i(\mathbf{n})} f_i = \lim_{M\to\infty} \mathop{\mathbb{E}}_{\mathbf{n}\in\Psi(M)} \prod_{i=1}^{r} T^{q_i(\mathbf{n})} \mathbb{E}(f_i|\mathcal{Z}_k).$$

In fact, the convergence of (III.21) is established by verifying convergence under the additional assumption that $\mathcal{X}$ is a nilsystem [84] (or indeed an inverse limit of nilsystems), and then combining the fact that $\mathcal{Z}_k$ is characteristic for (III.21) with the fact that $\mathcal{Z}_k$ is an inverse limit of nilsystems [65]. We note that the second part of this theorem is not stated in this precise language in [82], but our restatement is an immediate consequence of Theorem 3 therein, by a standard telescoping argument.

For the purposes of this chapter, the key point about Host-Kra-Ziegler factors is that for weakly mixing systems, they are trivial.

Fact III.5.5. Suppose that the m.p.s. $\mathcal{X}$ is weakly mixing. Then, for any k, the Host-Kra-Ziegler factor $\mathcal{Z}_k$ of $\mathcal{X}$ is trivial. In particular $\mathbb{E}(f|\mathcal{Z}_k) = \int_X f d\mu$ for each $f \in L^\infty(\mu)$.

Proof. This is mentioned e.g. in [65]. With implicit definition of $\mathcal{Z}_k$ as above, one can prove by a simple induction that $|||f|||_k^{2^k} = \left(\int_X f d\mu\right)^{2^k}$ for each k, from which the claim easily follows. □

Corollary III.5.6. *Suppose that an m.p.s. $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ is invertible. Let $\Psi(M)$ be a Følner sequence, let $q_i \colon \mathbb{Z}^d \to \mathbb{Z}$ be polynomials with $\deg q_i > 0$ and $\deg(q_i - q_j) > 0$ for $i \neq j$, and $f_i \in L^\infty(\mu)$ for $1 \leq i \leq r$. Then,*

$$\mathop{\mathbb{E}}_{\mathbf{n}\in\Psi(M)} \prod_{i=1}^{r} T^{q_i(\mathbf{n})} f_i \xrightarrow[M\to\infty]{L^2} \prod_{i=1}^{r} \int_X f_i d\mu. \tag{III.22}$$

We are now ready to finish our argument.

Proof of Theorem III.5.1. Apply Corollary III.5.6 to the averages (III.18), together with previous discussion in this section. □

Remark. Using the same argument, one can prove Theorem III.5.1 under a weaker assumption. Indeed, the condition that A should be a weakly-mixing set can be replaced by the weaker requirement that A should give rise to an ergodic m.p.s. $\mathcal{X}_A$ via the construction in Section III.2, and that 1_A should be orthogonal to the k-th Host-Kra-Ziegler factor of $\mathcal{X}_A$, in the sense that $\mathbb{E}(1_A - d(A)1_{X_A}|\mathcal{Z}_k) = 0$, for some sufficiently large k.

III.6 Concluding remarks

We end with a slightly stronger version of Theorem III.5.1 for a single polynomial of degree 1, as suggested in an earlier remark.

Proposition III.6.1. *Suppose that an m.p.s. $\mathcal{X} = (X, T, \mathcal{B}, \mu)$ is weakly mixing and invertible, and let $p^{(t)}(x) \in \mathbb{Z}[t, x]$ be polynomial with $\deg_x p^{(t)}(x) = 1$, and $f \in L^\infty(\mu)$. Then there exists a set $S \subset \mathbb{N}$ with density $d(S) = 1$ and*

$$\mathop{\mathbb{E}}_{n<N} T^{p^{(t)}(n)} f \xrightarrow[N\to\infty]{L^2} \int f d\mu, \quad \text{uniformly in } t \in S. \tag{III.23}$$

Proof. Write $p^{(t)}(x) = a(t)x + b(t)$. Since $T^{b(t)}$ is an isometry, without loss of generality we may assume that $b(t) = 0$. As usual, we may assume that $\int f d\mu = 0$ and $\|f\|_\infty \leq 1$. If $a(t)$ is constant in t, we are done e.g. by Bergelson's theorem III.3.1, so assume this is not the case.

By the spectral theorem, there is a probability measure $\nu = \nu_f$ on $\mathbb{T}$ such that

$$\left\| \mathop{\mathbb{E}}_{n<N} T^{p^{(t)}(n)} f \right\|^2 = \left\| \mathop{\mathbb{E}}_{n\leq N} e\left(p(n)\alpha\right) \right\|^2_{L^2(\mathbb{T},\nu)} = \int_{\mathbb{T}} \left| \mathop{\mathbb{E}}_{n\leq N} e\left(a(t)n\alpha\right) \right|^2 d\nu(\alpha),$$

where as usual $e(\alpha) = e^{2\pi i\alpha}$. It is a well-known fact that since $\mathcal{X}$ is weakly mixing, the measure ν has no atoms.

We have the elementary inequality $|1 - e(\alpha)| \geq 4\|\alpha\|$. Thus, if for some $\delta > 0$ and $\alpha \in \mathbb{T}$ we have $\|a\alpha\| \geq \delta$ then

$$\left| \mathop{\mathbb{E}}_{n\leq N} e\left(an\alpha\right) \right| = \left| \frac{1 - e(aN\alpha)}{N(1 - e(a\alpha))} \right|^2 \ll \frac{1}{(N\delta)^2} \tag{III.24}$$

where the implied constant is absolute (and equal to $\frac{1}{16}$).

Let us denote by $\Gamma_{a,\delta}$ the set of $\alpha \in \mathbb{T}$ for which we have inequality $\|a\alpha\| < \delta$. Using (III.24) for $\alpha \notin \Gamma_{a,\delta}$ and the trivial bound for $\alpha \in \Gamma_{\alpha,\delta}$ we find that

$$\int_{\mathbb{T}} \left| \mathop{\mathbb{E}}_{n\leq N} e\left(an\alpha\right) \right|^2 d\nu(\alpha) \leq \nu\left(\Gamma_{a,\delta}\right) + \frac{O(1)}{(N\delta)^2}.$$

We claim that there exists a set $S \subset \mathbb{N}$ with $d(S) = 1$ such that

$$\nu\left(\Gamma_{a(t),\delta}\right) \xrightarrow[\delta\to 0]{} 0, \quad \text{uniformly in } t \in S. \tag{III.25}$$

Suppose that the claim has been established. Then,

$$\left\| \mathop{\mathbb{E}}_{n<N} T^{p^{(t)}(n)} f \right\|^2 = o_{\delta\to 0}(1) + o_{\delta;N\to\infty}(1) = o_{N\to\infty}(1),$$

with decay rates uniform in $t \in S$. This finishes the proof of the theorem. Hence, it remains to find S with (III.25). Our construction relies on the following observation.

Lemma III.6.2. *Given δ, there exists a set S_δ with density 1 such that*

$$\left|\nu\left(\Gamma_{a(t),\delta}\right) - \lambda\left(\Gamma_{a(t),\delta}\right)\right| \xrightarrow[t\to\infty,\ t\in S_\delta]{} 0, \tag{III.26}$$

where λ denotes that Lebesgue measure (so $\lambda\left(\Gamma_{a(t),\delta}\right) = 2\delta$ for almost all t).

Proof. Because of the Lemma III.5.2, it will suffice to prove that for fixed $\delta > 0$ we have

$$\mathop{\mathbb{E}}_{t\leq M} \left(\nu\left(\Gamma_{a(t),\delta}\right) - \lambda\left(\Gamma_{a(t),\delta}\right)\right)^2 \xrightarrow[M\to\infty]{} 0,$$

This will follow once we show that

$$\mathop{\mathbb{E}}_{t\leq M} \nu\left(\Gamma_{a(t),\delta}\right) \xrightarrow[M\to\infty]{} \lambda\left(\Gamma_{a(t),\delta}\right), \qquad \mathop{\mathbb{E}}_{t\leq M} \nu\left(\Gamma_{a(t),\delta}\right)^2 \xrightarrow[M\to\infty]{} \lambda\left(\Gamma_{a(t),\delta}\right)^2.$$

For the first limit, we can rewrite:

$$\begin{aligned}\lim_{M\to\infty} \mathop{\mathbb{E}}_{t\leq M} \nu\left(\Gamma_{a(t),\delta}\right) &= \lim_{M\to\infty} \int_{\mathbb{T}} \mathop{\mathbb{E}}_{t\leq M} \chi_\delta(a(t)\alpha) d\nu(\alpha) \\ &= \int_{\mathbb{T}} \lim_{M\to\infty} \mathop{\mathbb{E}}_{t\leq M} \chi_\delta(a(t)\alpha) d\nu(\alpha),\end{aligned}$$

where χ_δ denotes the characteristic function of the interval $(-\delta, \delta)$ mod 1, and the interchange of limit and the integral is justified by the dominated convergence theorem (assuming that the last limit exists).

For $\alpha \notin \mathbb{Q}$, since χ_δ is Riemann-integrable and $\alpha a(t)$ is polynomial with irrational coefficients, we have by a classical theorem of Weyl that

$$\lim_{M\to\infty} \mathop{\mathbb{E}}_{t\leq M} \chi_\delta(a(t)\alpha) = \int_{\mathbb{T}} \chi_\delta(\alpha) d\lambda = \lambda(\Gamma_{a(t),\delta}).$$

As for $\alpha \in \mathbb{Q}$, we know that ν has no atoms so $\nu(\mathbb{Q}) = 0$. It follows that

$$\lim_{M\to\infty} \mathop{\mathbb{E}}_{t\leq M} \nu\left(\Gamma_{a(t),\delta}\right) = \lambda(\Gamma_{a(t),\delta}).$$

For the second limit, we can proceed analogously. We first rewrite:

$$\begin{aligned}\lim_{M\to\infty} \mathop{\mathbb{E}}_{t\leq M} \nu\left(\Gamma_{a(t),\delta}\right)^2 &= \lim_{M\to\infty} \int_{\mathbb{T}\times\mathbb{T}} \mathop{\mathbb{E}}_{t\leq M} \chi_\delta(a(t)\alpha)\cdot\chi_\delta(a(t)\beta) d\nu(\alpha) d\nu(\beta) \\ &= \int_{\mathbb{T}\times\mathbb{T}} \lim_{M\to\infty} \mathop{\mathbb{E}}_{t\leq M} (\chi_\delta\times\chi_\delta)(a(t)\alpha,\ a(t)\beta) d(\nu\times\nu)(\alpha,\beta).\end{aligned}$$

The set of $(\alpha,\beta) \in \mathbb{T}\times\mathbb{T}$ which are linearly dependent over $\mathbb{Q}$ has $\nu\times\nu$ measure 0, since it is a union of countably many lines $\{(\alpha,\beta) \mid k\alpha + l\beta + m = 0\}$ $(k,l,m \in \mathbb{Z})$ whose measure is 0 by e.g. Fubini's theorem combined with ν having no atoms.

If (α,β) are $\mathbb{Q}$-linearly independent, then by Weyl's theorem, the sequence $(a(t)\alpha,\ a(t)\beta)$ is equidistributed. It follows that

$$\lim_{M\to\infty} \mathop{\mathbb{E}}_{t\leq M} (\chi_\delta\times\chi_\delta)(a(t)\alpha,\ a(t)\beta) = \int_{\mathbb{T}\times\mathbb{T}} (\chi_\delta\times\chi_\delta) d(\lambda\times\lambda).$$

As a consequence, we have the sought convergence:

$$\lim_{M\to\infty} \mathop{\mathbb{E}}_{t\leq M} \nu\left(\Gamma_{a(t),\delta}\right)^2 = \left(\Gamma_{a(t),\delta}\right)^2.$$

This finishes the proof of the lemma. □

Lemma III.6.3. *Let $\{S_n\}_{n=1}^\infty$ be a sequence of sets $S_n \subset \mathbb{N}$ with density 1. Then there exists a single set $S \subset \mathbb{N}$ with density 1 such that for each n, $S \setminus S_n$ is finite.*

Proof. We may assume without loss of generality that the family S_n is descending, else we may replace S_n by $\bigcap_{m\leq n} S_n$. We define S by declaring that $k \in S$ if and only if $k \in S_{n(t)}$ where $n(k)$ is an increasing function yet to be determined. If $n(k) \to \infty$ as $k \to \infty$, then clearly $S \setminus S_n$ is finite for any n. It remains to check that if $n(t)$ increases sufficiently slowly, then S has density 1. This is a simple consequence of the fact that for each n, S_n has density 1. □

We are now ready to finish the proof of the proposition. Let S_δ be the sets constructed in the above Lemma III.6.2. Let S be a set with $d(S) = 1$ and $S \setminus S_\delta$ finite for each $\delta \in \mathbb{Q}$. Then for any $t \in S$ it holds that

$$\nu\left(\Gamma_{a(t),\delta}\right) = \lambda\left(\Gamma_{a(t),\delta}\right) + o_{\delta;t\to\infty}(1).$$

We need to show that for any $\varepsilon > 0$ one can find $\delta(\varepsilon)$ such that for $\delta < \delta(\varepsilon)$ and any $t \in S$ we have

$$\nu\left(\Gamma_{a(t),\delta}\right) < \varepsilon. \tag{III.27}$$

Let δ_0 be such that $\lambda\left(\Gamma_{a(t),\delta_0}\right) = 2\delta_0 < \varepsilon/2$. We can then find t_0 such that for $t \geq t_0$, $t \in S$ we have $\nu\left(\Gamma_{a(t),\delta_0}\right) < \varepsilon/2 + \varepsilon/2 = \varepsilon$. Since $\nu\left(\Gamma_{a(t),\delta}\right)$ is decreasing in δ, the bound (III.27) holds for any $\delta < \delta_0$, and $t \geq t_0,\ t \in S$.

On the other hand, for each $1 \leq t < t_0$, because $\bigcap_{\delta>0} \Gamma_{a(t),\delta}$ is a finite set and ν is atomless, there is some $\delta_t > 0$ such that (III.27) holds for $\delta < \delta_t$.

Taking $\delta(\varepsilon) = \min\{\delta_t \mid 0 \leq t < t_0\}$, we find that (III.27) holds for all $\delta < \delta(\varepsilon)$. This finishes the proof of the claim. □

Chapter IV

Combinatorial characterisation of nil–Bohr sets of integers

IV.1 Introduction

Among the basic problems in additive combinatorics is the study of various notions of *largeness* which may apply to a set of integers. In this chapter we are specifically interested in one such notion, namely that of containing a nil-Bohr set, or a set of recurrence times for a d-step nilrotation. For background on nil–Bohr sets, see I.1.7.

The study of these sets was pioneered by Host and Kra [67], with later developments due to Huang, Shao and Ye [68], Tu [105], and Bergelson and Leibman [17]. In [67], it was realised that nil–Bohr sets bear a striking relation to a purely combinatorial class of SG_k^* sets which we will shortly define. Namely, it was shown that a SG_d^* set is (strongly) piecewise-Nil_d–Bohr. Here, we prove the reverse implication, although in a weaker form.

Even though a proper motivation for our results requires more context, we are able to express some of them in relatively basic terms. Our first result is the following.

Theorem IV.1.1. *Fix a polynomial $p \in \mathbb{R}[x]$ of degree d with $p(0) = 0$, and a sequence $(n_i)_{i=1}^{\infty}$ of positive integers. Then, for any $\varepsilon > 0$, there exists a finite, non-empty set $\alpha \subset \mathbb{N}$, with gaps bounded by d, such that $\left\| p\left(\sum_{i\in\alpha} n_i\right)\right\| \leq \varepsilon$.*

Above, $\|t\| = \min_{n\in\mathbb{Z}} |t - x|$ denotes the distance from the closest integer. From now on, let us denote by $\mathcal{F}$ the family of all finite non-empty subsets of $\mathbb{N}$. It is customary to denote elements of $\mathcal{F}$ by lowercase Greek letters $\alpha, \beta, \ldots$. For a set $\alpha = \{i_1, i_2, \ldots, i_r\}$, where $i_1 < i_2 < \cdots < i_r$, the *gaps* of α are the integers $i_2 - i_1, i_3 - i_2, \ldots, i_r - i_{r-1}$.

Bootstrapping (a slight modification of) the above result, we obtain a recurrence statement for nilrotations.

Theorem IV.1.2. *Let G be a d-step nilpotent Lie group and let $\Gamma < G$ be a cocompact, discrete subgroup. Fix $g \in G$, an open neighbourhood $e\Gamma \in U \subset G/\Gamma$, as well as a sequence $(n_i)_{i=1}^{\infty}$ of positive integers.*

Then, there exists $\alpha \in \mathcal{F}$ with gaps bounded by $d' = 4d$, such that $g^{\sum_{i\in\alpha} n_i}\Gamma \in U$.

The bound $d' = 4d$ results from an inductive argument. By a marginally more careful computation, this could be improved to $d' = 3d + \log_2 d + O(1)$, but we sacrifice this inconsequential improvement for the sake of readability. The optimal value is believed to be $d' = d$, but our argument does not yield this conclusion.

Finally, we point out that the recent paper of Bergelson and Leibman [17] proves highly relevant results. It follows as a special case of Theorem 0.3 in [17] that, in the situation of Theorem IV.1.1, there is some $r = r(d, \varepsilon)$ such that there exists $\alpha \in \mathcal{F}$ with $\max \alpha \leq r$ such that

$$\left\| p\left(\sum_{i\in\alpha} n_i \right) \right\| \leq \varepsilon.$$

Hence, trivially, we may ensure that α has gaps bounded in terms of d and ε; however, our result gives a good bound on the gaps, which also is uniform in ε. Similarly, a version of Theorem IV.1.2 can be read off Theorem 0.5 in [17].

Recall from Section I.1.5 that an IP set is one which contains the set of finite sums,

$$\mathrm{FS}(n_i) = \{ n_\alpha \mid \alpha \in \mathcal{F} \}, \tag{IV.1}$$

where $n_\alpha := \sum_{i\in\alpha} n_i$ for $\alpha \in \mathcal{F}$, for some sequence $(n_i)_{i=1}^{\infty}$ in $\mathbb{N}$.

In analogy to the IP sets $\mathrm{FS}(n_i)$, we define for $k \geq 0$ the sets $\mathrm{SG}_k(n_i)$, where the index sets are additionally required to have bounded gaps. Let $\mathcal{S}_k \subset \mathcal{F}$ denote the set of finite sets of integers $\alpha \in \mathcal{F}$, whose gaps are bounded by k; we might call such sets k-syndetic. In other words, we require that for any $i \in \alpha$, either $i = \max \alpha$, or there exists $j \in \alpha$ with $i < j \leq i + k$. We allow the degenerate case $k = 0$, where $\mathcal{S}_0 = \mathbb{N}$ (up to identification $i \mapsto \{i\}$). For a sequence of integers $(n_i)_{i=1}^{\infty}$ we put

$$\mathrm{SG}_k(n_i) = \left\{ \sum_{i\in\alpha} n_i \;\middle|\; \alpha \in \mathcal{S}_k \right\}. \tag{IV.2}$$

In analogy with IP sets, we define a set $A \subset \mathbb{N}$ to be a SG_k set if it contains a set of the form $\mathrm{SG}_k(n_i)$ for some sequence n_i. To the best of our knowledge, this definition

first appears in [67]. We have an obvious chain of inclusions $\mathrm{SG}_1(n_i) \subset \mathrm{SG}_2(n_i) \subset \cdots \subset \mathrm{FS}(n_i)$, whence $\mathrm{SG}_1 \supset \mathrm{SG}_2 \supset \cdots \supset \mathrm{IP}$.

The simplest non-degenerate example, though possibly a misleading one, is when $k = 1$. Then $\mathrm{SG}_k(n_i)$ consists precisely of the consecutive sums $\sum_{i=u}^{v} n_i$, and it has been noted that $\mathrm{SG}_1(n_i) = \Delta(S) := (S - S) \cap \mathbb{N}$ where $S = \{\sum_{i=1}^{v} n_i \mid v \in \mathbb{N}\}$. It is not difficult to see that conversely, any set of the form $\Delta(S)$ as above can be expressed as $\mathrm{SG}_1(n_i)$ for some sequence (n_i). Thus, SG_1 sets coincide with the well-studied class of Δ sets (see also [13]).

Using the notation introduced above we define the class SG_k^*, consisting of the sets $B \subset \mathbb{N}$ such that for any choice of integers $(n_i)_{i=1}^{\infty}$, there exists some $\alpha \in \mathcal{S}_k$ with $n_\alpha \in B$. We note the reversed chain of inclusions: $\mathrm{SG}_1^* \subset \mathrm{SG}_2^* \subset \cdots \subset \mathrm{IP}^*$.

As noted earlier, there is a somewhat unexpected connection between the a priori unrelated notions of SG_k^* sets and Nil_d–Bohr sets.

Following the usual convention, we say that a set A is piecewise Nil_d–Bohr_0 if there exists a thick set T (i.e. such that T contains arbitrarily long intervals) and a Nil_d–Bohr_0 set B such that $A = B \cap T$. It was proved in [67] that any SG_d^* set is piecewise Nil_d–Bohr_0. (In fact, the result proved there is stronger, with a more rigid notion of "strongly piecewise".) The following question arises naturally:

Question IV.1.3. *Is any* Nil_d–Bohr_0 *set a* SG_d^* *set?*

The main result in this chapter is an answer to a weaker variant of this question. The reader will have no problem checking that the following is merely a succinct restatement of Theorem IV.1.2.

Theorem (IV.1.2, short version)**.** *Any* Nil_d–Bohr_0 *set is* SG_k^**, provided that* $k \geq 4d$*.*

Remark. We note that a weaker variant of the theorem, with SG_k^* replaced with IP^*, is true for much simpler (or at least better studied) reasons. Recall that any nilrotation is distal. On the other hand, for any topological dynamical system (T, X) distality is equivalent to the condition that for any $x \in X$ and any open neighbourhood $U \ni x$, the set $\{n \mid T^n x \in U\}$ is IP^*. Since any Nil_d–Bohr_0 is (a superset of) a set of precisely this form, the claim follows. (A similar argument can be found in [16].)

IV.2 Polynomial maps

IV.2.1 Main results reformulated

In order to prove Theorem IV.1.2, we need a good understanding of "linear" maps of the form $\alpha \mapsto g^{n_\alpha}\Gamma$, where $\alpha \in \mathcal{F}$. As explained in Section I.1.8, it is more convenient to work with the more general setup of *polynomial* sequences $\mathrm{poly}(\mathcal{F} \to G_\bullet/\Gamma)$, where $G_\bullet/\Gamma$ is a nilmanifold. We mention a crucial fact, whose proof we defer to Section IV.2.4.

Theorem IV.2.1. *If $G_\bullet/\Gamma$ is a d-step nilmanifold, then the space $\mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet/\Gamma)$ is compact in the topology of pointwise convergence.*

We are now ready to formulate versions of Theorems IV.1.1 and IV.1.2 in the proper generality.

Theorem IV.2.2 (IV.1.1, strong version)**.** *Let $f\colon \mathcal{F}_\emptyset \to \mathbb{T}^m$ be a polynomial of degree d with $f(\emptyset) = 0$. Then, for any $\varepsilon > 0$, there exists $\alpha \in \mathcal{S}_d$, such that $\|f(\alpha)\| < \varepsilon$.*

Theorem IV.2.3 (IV.1.2, strong version)**.** *Let $G_\bullet/\Gamma$ be a nilmanifold of length d, and let $\bar{f} \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet/\Gamma)$ with $\bar{f}(\emptyset) = e\Gamma$. Then, for any open neigbourhood $e\Gamma \in U \subset G/\Gamma$, there exists $\alpha \in \mathcal{S}_k$ with $k \leq 4d$, such that $\bar{f}(\alpha) \in U$.*

Remark. Theorem IV.2.2 is ostensibly stronger than Theorem IV.1.1, and there is a similar relation between Theorems IV.2.3 and IV.1.2. While it is beyond the scope of our investigation to ask if Theorem IV.1.1 formally implies IV.2.2, we pause to give some examples where polynomials of the special form we are interested in are already dense in the family of all polynomials.

Example IV.2.4. Fix $\varepsilon_i > 0$ and $\theta \in \mathbb{R} \setminus \mathbb{Q}$, and let $f\colon \mathcal{F}_\emptyset \to \mathbb{T} = \mathbb{R}/\mathbb{Z}$ be a degree 1 polynomial with $f(\emptyset) = 0$. Then, there exist $(n_i)_{i=1}^\infty$ such that for all $\alpha \in \mathcal{F}_\emptyset$ we have $\|f(\alpha) - n_\alpha\theta\| < \varepsilon_\alpha$, where as usual $n_\alpha = \sum_{i\in\alpha} n_i$ and $\varepsilon_\alpha = \sum_{i\in\alpha} \varepsilon_i$.

Proof. By Proposition IV.2.7, f takes the form $f(\alpha) = \sum_{i\in\alpha} a_i$. Choose n_i so that $\|a_i - n_i\theta\| < \varepsilon_i$. □

Example IV.2.5. Fix $\varepsilon > 0$ and $r \in \mathbb{N}$. Let $\theta \in \mathbb{R} \setminus \mathbb{Q}$, and let $f\colon \mathcal{F}_\emptyset \to \mathbb{T}$ be a degree 2 polynomial with $f(\emptyset) = 0$. Then, there exist $(n_i)_{i=1}^r$ such that for all $\alpha \in \mathcal{F}_\emptyset$ with $\max\alpha \leq r$ we have

$$\left\|f(\alpha) - n_\alpha^2\theta\right\| < \varepsilon, \tag{IV.3}$$

where as usual $n_\alpha = \sum_{i\in\alpha} n_i$.

Proof. Using Proposition IV.2.7, $f(\alpha) = \sum_{\gamma \subset \alpha} a_\gamma$ (where $a_\gamma = 0$ if $|\gamma| > 2$ or $\gamma = \emptyset$). Subject to the choice of n_i's, write $b_{\{i,j\}} = 2n_i n_j \theta$, $b_{\{i\}} = n_i^2 \theta$, and $b_\gamma = 0$ otherwise, so that $n_\alpha^2 \theta = \sum_{\gamma \subset \alpha} b_\gamma$.

To obtain (IV.3), it will suffice to ensure that $\|b_\gamma - a_\gamma\| < \frac{\varepsilon}{2^r}$ for all γ, for a suitable choice of $(n_i)_{i=1}^r$. We claim that more is true, namely that $(b_\gamma)_\gamma$ is equidistributed in $\mathbb{T}^{\binom{r+1}{2}}$ (here, γ runs over all non-empty subsets of $\{1, 2, \dots, r\}$ of size ≤ 2). If this were no so, then by the multidimensional version of Weyl's equidistibution theorem, there would exist $k_\gamma \in \mathbb{Z}$, not all 0, such that $\sum_\gamma k_\gamma b_\gamma$ is a polynomial with integer coefficients in $(n_i)_i$, but this is impossible since $\theta \notin \mathbb{Q}$, and the non-trivial b_γ are distinct monomials. □

Similar statements hold for polynomials of higher degrees, the only difference being that the verification of equidistibution in the final step becomes more mundane.

IV.2.2 Connectivity

If the nilpotent group G is simply connected (which we take to mean, in particular, connected) then the exponential map $\exp\colon \mathfrak{g} \to G$ is a homeomorphism between the Lie algebra $\mathfrak{g}$ of G and G itself. If $G_\bullet$ is a prefiltration where each of the groups G_i is connected, then there exists a Mal'cev basis for $\mathfrak{g}$, where the description of G takes a particularly simple form (see [88]). Because of this, the assumption of simple connectivity is often made in the literature (notably [53], [54], [55]).

On the other hand, the definition of Nil_d–Bohr sets in [67, Def. 2.2] and Section I.1.7 makes no assumptions about connectivity whatsoever. The purpose of this section is to bridge this gap. We record that the assumption of simple connectivity can be freely added, and that it does not matter whether we work with genuine polynomial sequences or "linear sequences".

For a nilmanifold G/Γ (with no topological assumptions on G), open $U \subset G/\Gamma$ and (polynomial) sequence $g\colon \mathbb{N}_0 \to G/\Gamma$, denote the corresponding (Nil_d–Bohr_0) set

$$B(g, U) := \{n \in \mathbb{N}_0 \mid g(n)\Gamma \in U\}. \tag{IV.4}$$

Lemma IV.2.6. *Fix $d \geq 1$. Then, for a set $A \subset \mathbb{N}_0$, the following properties are equivalent to A being a Nil_d–Bohr_0 set:*

(i) There exists a d-step nilmanifold G/Γ (with G not necessarily simply connected), an open neighbourhood $e\Gamma \in U \subset G/\Gamma$, and a "linear" sequence $g(n) = a^n$ ($a \in G$) such that $A \supset B(g, U)$.

(ii) There exists a d-step nilmanifold G/Γ with G simply connected (in particular connected), an open neighbourhood $e\Gamma \in U \subset G/\Gamma$, and a "linear" sequence $g(n) = a^n$ ($a \in G$) such that $A \supset B(g, U)$.

(iii) There exists a length d nilmanifold $G_\bullet/\Gamma$ (with each G_i simply connected), an open neighbourhood $e\Gamma \in U \subset G/\Gamma$, and $g \in \mathrm{poly}(\mathbb{N}_0 \to G_\bullet/\Gamma)$, $g(0) = e$, such that $A \supset B(g, U)$.

Proof. Condition (i) is the one used to define Nil_d–Bohr_0 sets in Section I.1.7.

The equivalence between (ii) and (iii) follows from Proposition C.2 [55], which asserts that general polynomial sequences can be lifted to linear ones. Similar argument (without assumption of simple connectivity) appears also in [84, Proposition 3.14] and in the proof of Theorem B* in [83]. The equivalence between (i) and (ii) is established in [84, Section 1.11] by means of embedding an arbitrary nilmanifold in a simply connected one; see also [68, Section 3.1.2] □

In light of the above considerations, we can (and will) always assume that the groups G_i constituting the prefiltration are connected and simply connected.

Remark. Yet another definition of a Nil_d–Bohr_0 set is possible. Recall that the typical Nil_d–Bohr_0 set takes the form $\{n \in \mathbb{N}_0 \mid g(n)\Gamma \in V\Gamma\}$, where $e \in V \subset G$ is open, and $g \in \mathrm{poly}(\mathbb{N}_0 \to G_\bullet)$ with $g(0) = e$. Instead of the assumptions $e \in V$ and $g(0) = e$, we could put a milder restriction $g(0) \in V\Gamma$. Again, this does not lead to a more general notion. As noted in [68, Section 3.1.4.], upon replacing Γ with $\Gamma' = g(0)\Gamma g(0)^{-1}$, V with $V' = Vg(0)^{-1}$, and g with $g'(n) = g(n)g(0)^{-1}$, we recover $\{n \in \mathbb{N}_0 \mid g(n) \in V\Gamma\} = \{n \in \mathbb{N}_0 \mid g'(n) \in V'\Gamma'\}$, and $g'(0) = e$. In fact, the original definition in [67] uses this seemingly more general form.

IV.2.3 VIP-systems

A simple but already interesting instance of the definition of polynomial sequences is the abelian one, where $G = \mathbb{R}^m$, $\Gamma = \mathbb{Z}^m$ and the filtration is given by

$$G_0 = G_1 = \cdots = G_d = \mathbb{R}^m,\ G_{d+1} = G_{d+2} = \cdots = \{0\}.$$

In this situation, we will simply speak of a polynomial of degree d and keep the filtration implicit. We denote the set of all polynomial maps from A to $\mathbb{T}^m = \mathbb{R}^m/\mathbb{Z}^m$ by $\mathrm{poly}(A \to \mathbb{T}^m)$.

It is a general fact that if $G_i \supset G'_i$ are two prefiltrations then $\mathrm{poly}(A \to G_\bullet) \supset \mathrm{poly}(A \to G'_\bullet)$. In particular, any polynomial $A \to \mathbb{T}^m$ with respect to some prefiltration of length d is a polynomial of degree d, as above.

Such maps are a special case of a more general notion of a VIP-*system*, introduced in [12], well before the theory of polynomial maps between nilpotent groups flourished. (In our notation, a VIP-system is essentially a polynomial map $\mathcal{F}_\emptyset \to \Omega$, where Ω is an abelian group.)

The following structural result is well known, see [93, Proposition 2.5]. We provide a proof for the convenience of the reader; similar ideas will appear in the proof of Proposition IV.3.20.

Proposition IV.2.7 (Structure of polynomials $\mathcal{F} \to \mathbb{T}^m$)**.** *Suppose that $f\colon \mathcal{F}_\emptyset \to \mathbb{T}^m$ is a polynomial of degree d. Then f admits a representation of the form:*

$$f(\alpha) = \sum_{\substack{\gamma \subset \alpha \\ |\gamma| \leq d}} a_\gamma, \tag{IV.5}$$

where $\alpha_\gamma \in \mathbb{T}^m$ are constants. Conversely, any map of the form (IV.5) *is a polynomial of degree $\leq d$.*

Proof. Suppose first that f takes the form (IV.5). We show by induction on d that f is a polynomial of degree d; the case $d = 0$ is clear. For $d \geq 1$, the discrete difference relation $D_\beta f = f(\alpha \cup \beta) - f(\alpha)$ of (I.5) is satisfied for

$$D_\beta f(\alpha) := \sum_{\substack{\gamma \subset \alpha \\ |\gamma| \leq d-1}} \sum_{\substack{\delta \subset \beta \\ |\delta| \leq d - |\gamma|}} a_{\gamma \cup \delta},$$

which is again of the form (IV.5) with degree $d - 1$, so the claim follows.

Conversely, let $f\colon \mathcal{F}_\emptyset \to \mathbb{T}^m$ be a polynomial of degree $\leq d$. By a inclusion-exclusion argument, it is easy to construct g of the form (IV.5) such that $f(\alpha) = g(\alpha)$ if $|\alpha| \leq d$. Hence, replacing f with $f - g$ if necessary, we may assume that $f(\alpha) = 0$ whenever $|\alpha| \leq d$. We show, by induction on d this condition implies $f(\alpha) = 0$ for all $\alpha \in \mathcal{F}_\emptyset$. Because degree 0 polynomials are constant, the case $d = 0$ is clear; assume $d \geq 1$.

Choose any $\beta \in \mathcal{F}$ with $|\beta| = 1$. By the Definition I.1.15, there is a polynomial $D_\beta f$ of degree $d - 1$ such that

$$D_\beta f(\alpha) = f(\alpha \cup \beta) - f(\alpha), \qquad (\alpha \cap \beta = \emptyset).$$

If $|\alpha| \leq d-1$ then $D_\beta f(\alpha) = 0$. Using the inductive assumption, we conclude that $D_\beta f(\alpha) = 0$ for all $\alpha \in \mathcal{F}_\emptyset$, $\alpha \cap \beta = \emptyset$. Because β was arbitrary, if $f(\alpha) = 0$ for all $\alpha \in \mathcal{F}_\emptyset$ of a given size $|\alpha| = n$, then the same holds for $|\alpha| = n+1$; hence by induction $f(\alpha) = 0$ for all α. □

IV.2.4 Host-Kra cube groups

A useful approach to polynomial maps is obtained by introducing the notion of a *Host-Kra cube*, or more generally the *cube group* $\mathrm{HK}^k(G_\bullet)$. This notion was first introduced (though with a different name) by Host and Kra in [65], and is extensively used in a number of papers, including [53, 54]. It is also a basis for the work of Szegedy and Camarena [25] later refined by Gutman, Manners and Varjú [56, 57, 58]. We use some basic facts, using [53, Appendix E] and [54] as our main reference; an accessible introduction can be found in [50]. Throughout, $k \geq 1$ is an integer, and $G_\bullet$ is a prefiltration consisting of simply connected groups.

For any k, we may consider the *cube* $\{0,1\}^k$. It is often convenient to identify $\{0,1\}^k$ with the powerset $\mathcal{P}([k])$. In particular, $\{0,1\}^k$ carries a natural partial order where $\omega \leq \omega'$ if $\omega_i \leq \omega'_i$ for all i.

For $\omega \in \{0,1\}^k$ and $g \in G$, we define $g^{[\omega]} \in G^{\{0,1\}^k}$ by

$$g^{[\omega]}_\sigma = \begin{cases} g, & \text{if } \sigma \geq \omega, \\ e_G, & \text{otherwise.} \end{cases}$$

Hence, $g^{[\omega]}$ can be viewed as a cube with entries g on the upper face $\big\{\sigma \in \{0,1\}^k \;\big|\; \sigma \geq \omega\big\}$, and e_G elsewhere.

If the prefiltration $G_\bullet$ is given, we further define the *face group* $G^{[\omega]}$ to be the subgroup of $G^{\{0,1\}^k}$ generated by elements of the form $g^{[\omega]}$ with $g \in G_{|\omega|}$, where $|\omega| := |\{1 \leq i \leq k \mid \omega_i = 1\}|$. Finally, we define the Host-Kra cube group $\mathrm{HK}^k(G_\bullet)$ to be the group generated by all face groups $G^{[\omega]}$.

The faces mentioned above can be thought of as "upper faces". One can, for a face $F = \big\{\omega \in \{0,1\}^k \;\big|\; \omega_i = \sigma_i \text{ for } i \in I\big\}$ of codimension $\operatorname{codim} F = |I|$, consider the face group $G^{[F]}$ generated by elements $g^{[F]}$ given by

$$g^{[F]}_\sigma = \begin{cases} g, & \text{if } \sigma \in F, \\ e_G, & \text{otherwise,} \end{cases}$$

where $g \in G_{|I|}$. These are, however, contained in the group $\mathrm{HK}^k(G_\bullet)$ defined above (see [53], discussion after E.5).

A basic fact lying at the heart of many inductive arguments is the commutator relation:

$$[G^{[\sigma]}, G^{[\rho]}] \subset G^{[\sigma\cup\rho]}, \qquad \rho, \sigma \in \{0,1\}^k, \tag{IV.6}$$

which follows directly from the identity $[g^{[\sigma]}, h^{[\rho]}] = [g,h]^{[\sigma\cup\rho]}$ (see [53, Lemma E.5]).

For a (total) order $\prec$ on $\{0,1\}^k$, compatible with inclusion in the sense that if $\sigma \leq \rho$ coordinatewise, then also $\sigma \preceq \rho$ ($\sigma, \rho \in \{0,1\}^k$), we define ordered products $\prod_{\omega\in\Omega}^{\prec} \mathbf{x}_\omega := \mathbf{x}_{\omega_1}\mathbf{x}_{\omega_2}\dots\mathbf{x}_{\omega_r}$, where $\Omega = \{\omega_1 \prec \omega_2 \prec \dots \prec \omega_r\} \subset \{0,1\}^k$ (we use this with $\mathbf{x}_\omega \in G^{\{0,1\}^k}$).

Fact IV.2.8. If $\mathbf{g} = (g_\omega) \in \mathrm{HK}^k(G_\bullet)$, then the representation $\mathbf{g} = \prod_\omega^{\prec} \tilde{g}_\omega^{[\omega]}$ with $\tilde{g}_\omega \in G_{|\omega|}$ exists and is unique. Moreover, $\tilde{g}_\omega$ is a word in g_σ, $\sigma \leq \omega$.

Proof. See [53, Lemma E.6], or [54, Lemma 6.4]. □

The key reason for interest in the Host-Kra cube group is the characterisation of polynomial maps which they provide. Let $\mathcal{F}_\emptyset^{[k]}$ denote the set of parallelepipeds, i.e. cubes of the form $\{\alpha_\omega\}_{\omega\in\{0,1\}^k}$ with $\alpha_\omega = \alpha_0 \cup \bigcup_{i\in\omega} \alpha_i$ for some disjoint $\alpha_0, \alpha_1, \dots, \alpha_k \in \mathcal{F}_\emptyset$.

Proposition IV.2.9. *Let $G_\bullet$ be a length d prefiltration. A sequence $f\colon \mathcal{F}_\emptyset \to G$ is polynomial in the sense of Definition I.1.15 if and only if f maps $\mathcal{F}_\emptyset^{[k]}$ to $\mathrm{HK}^k(G_\bullet)$ for each k.*

Note that because $\mathrm{HK}^k(G_\bullet)$ are groups, Theorem I.1.19 for polynomials whose domain is $\mathcal{F}_\emptyset$ follows from Proposition IV.2.9 immediately.

To prove this proposition, it is convenient to isolate another fact on Host-Kra cube groups (implicit in [54, Proposition 6.5]). For $\mathbf{g}, \mathbf{g}' \in G^{\{0,1\}^k}$, denote by $\mathbf{h} = (\mathbf{g}, \mathbf{g}') \in \Gamma^{\{0,1\}^{k+1}}$ the result of "glueing" $\mathbf{g}$ and $\mathbf{g}'$, i.e. the cube with $\mathbf{h}_{\omega 0} = \mathbf{g}_\omega$ and $\mathbf{h}_{\omega 1} = \mathbf{g}'_\omega$.

Fact IV.2.10. For a cube $\mathbf{g} \in G^{\{0,1\}^k}$ the following are equivalent:

(i) $\mathbf{g} \in \mathrm{HK}^k(G_\bullet)$;

(ii) $(\mathbf{g}, \mathbf{g}) \in \mathrm{HK}^{k+1}(G_\bullet)$.

Also, the following are equivalent:

(i') $\mathbf{g} \in \mathrm{HK}^k(G_{\bullet+1})$;

(ii') $(\mathbf{e}, \mathbf{g}) \in \mathrm{HK}^{k+1}(G_{\bullet+1})$, where $\mathbf{e} = (e_G)_{\omega\in\{0,1\}^k}$.

Proof. Fix a compatible order $\prec$ on $\{0,1\}^k$. For the implication (i) $\Rightarrow$ (ii), write (using Fact IV.2.8) $\mathbf{g} = \prod_{\omega\in\{0,1\}^k}^{\prec} g_\omega^{[\omega]}$. Then $(\mathbf{g},\mathbf{g}) = \prod_{\omega\in\{0,1\}^k}^{\prec} g_\omega^{[\omega 0]} \in \mathrm{HK}^{k+1}(G_\bullet)$ if $\mathbf{g} \in \mathrm{HK}^k(G_\bullet)$. By the same token, for (i') $\Rightarrow$ (ii'), we have $(\mathbf{e},\mathbf{g}) = \prod_{\omega\in\{0,1\}^k}^{\prec} g_\omega^{[\omega 1]} \in \mathrm{HK}^{k+1}(G_\bullet)$ if $\mathbf{g} \in \mathrm{HK}^k(G_{\bullet+1})$.

Conversely, for (ii) $\Rightarrow$ (i), if $(\mathbf{g},\mathbf{g}) = \prod_{\omega\in\{0,1\}^{k+1}}^{\prec} g_\omega^{[\omega]} \in \mathrm{HK}^{k+1}(G_\bullet)$, then $\mathbf{g} = \prod_{\omega\in\{0,1\}^k}^{\prec} g_{\omega 0}^{[\omega]} \in \mathrm{HK}^k(G_\bullet)$. Finally, to prove (ii') $\Rightarrow$ (i'), write $(\mathbf{e},\mathbf{g}) = \prod_{\omega\in\{0,1\}^{k+1}}^{\prec} g_\omega^{[\omega]} \in \mathrm{HK}^{k+1}(G_\bullet)$. Because of uniqueness in Fact IV.2.8, we have $g_{\omega 0} = e$ for $\omega \in \{0,1\}^k$, so $\mathbf{g} = \prod_{\omega\in\{0,1\}^k}^{\prec} g_{\omega 1}^{[\omega]} \in \mathrm{HK}^k(G_{\bullet+1})$. □

Proof of Proposition IV.2.9. This follows by a standard modification of the proof of [54, Proposition 6.5]. The key point is that for each $k \geq 1$ and $(\alpha_\omega)_\omega \in \mathcal{F}_\emptyset^{[k]}$, the cube $\mathbf{g} = (f(\alpha_\omega))_{\omega\in\{0,1\}^k}$ can be written as the product $\mathbf{g} = (\mathbf{g}',\mathbf{g}')(\mathbf{e},\mathbf{h})$, where $\mathbf{g}' = (f(\alpha_\omega))_{\omega\in\{0,1\}^{k-1}}$ and $\mathbf{h} = (D_{\alpha_k} f(\alpha_\omega))_{\omega\in\{0,1\}^{k-1}}$.

Suppose that $f \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet)$, and proceed by induction on k and d. We have $\mathbf{g}' \in \mathrm{HK}^{k-1}(G_\bullet)$ and $\mathbf{h} \in \mathrm{HK}^{k-1}(G_{\bullet+1})$, with notation as above. Hence, by IV.2.10, $(\mathbf{g}',\mathbf{g}'), (\mathbf{e},\mathbf{h}) \in \mathrm{HK}^k(G_\bullet)$, so $\mathbf{g} \in \mathrm{HK}^k(G_\bullet)$.

Suppose conversely that f maps $\mathcal{F}_\emptyset^{[k]} \to \mathrm{HK}^k(G_\bullet)$ for all k, and proceed by induction on d. Then, again with notation as above, $\mathbf{g}' \in \mathrm{HK}^{k-1}(G_\bullet)$, so $(\mathbf{g}',\mathbf{g}') \in \mathrm{HK}^k(G_\bullet)$. Hence, $(\mathbf{e},\mathbf{h}) \in \mathrm{HK}^k(G_\bullet)$ and by another application of IV.2.10, $\mathbf{h} \in \mathrm{HK}^k(G_{\bullet+1})$. Since $D_{\alpha_k} f$ maps $\mathcal{F}_\emptyset^{[k]}$ to $\mathrm{HK}^k(G_{\bullet+1})$, by inductive assumption $D_{\alpha_k} f \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_{\bullet+1})$. Since α_k was arbitrary, it follows that $f \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet)$. □

IV.2.5 Host-Kra cubes and nilmanifolds

To deal with polynomial maps $\mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet/\Gamma)$, we introduce the following nilmanifold analogue of the Host-Kra group. For each k, we define $\mathrm{HK}^k(G_\bullet/\Gamma) \subset (G/\Gamma)^{\{0,1\}^k}$ to be the natural projection of $\mathrm{HK}^k(G_\bullet)$. Note that here we depart from the definitions in [53] (compare Definition E.8).

Lemma IV.2.11. *If $G_\bullet/\Gamma$ is a nilmanifold, then for each k the space $\mathrm{HK}^k(G_\bullet/\Gamma)$ is compact.*

Proof. By [53, Lemma E.10], $\Gamma^{\{0,1\}^k} \cap \mathrm{HK}^k(G_\bullet)$ is cocompact and discrete in $\mathrm{HK}^k(G_\bullet)$. Because the projection $\mathrm{HK}^k(G_\bullet) \to \mathrm{HK}^k(G_\bullet/\Gamma)$ factors through the quotient $\mathrm{HK}^k(G_\bullet)/\left(\Gamma^{\{0,1\}^k} \cap \mathrm{H}\right.$ the space $\mathrm{HK}^k(G_\bullet/\Gamma)$ is compact. □

The Host-Kra cube group has a "corner completion property", stating that it is always possible to complete a partial cube with a single missing entry. Here, by $\{0,1\}_*^k$ we denote the cube with missing upper corner $\{0,1\}^k \setminus \{1^k\}$. Note that any face, i.e.

a set of the form $\{\omega \in \{0,1\}^k \mid \omega_i = \sigma_i \text{ for } i \in I\}$, can be naturally be identified with $\{0,1\}^{k-|I|}$. Hence, if F is a face then for $(g_\omega)_{\omega\in F} \subset G^F$ it makes sense to ask if $(g_\omega)_\omega \in \mathrm{HK}^{k-|I|}(G_\bullet)$ (and likewise for G/Γ in place of G).

Lemma IV.2.12. *Let $G_\bullet/\Gamma$ be a length d nilmanifold. Suppose that $g_\omega \in G$, $\omega \in \{0,1\}^k_*$ are such that for any codimension 1 face F with $1^k \notin F$, the restricted cube $(g_\omega)_{\omega\in F}$ lies in $\mathrm{HK}^{k-1}(G_\bullet)$. Then, there exists $g_{1^k} \in G$, which completes the cube: $(g_\omega)_{\omega\in\{0,1\}^k} \in \mathrm{HK}^k(G_\bullet)$. Moreover, if $g_\omega \in \Gamma$ for all $\omega \in \{0,1\}^k_*$, then it is possible to choose $g_{1^k} \in \Gamma$.*

Proof. This is essentially [53, Lemma E.7]. Fix a compatible order $\prec$ of $\{0,1\}^k$. For each lower codimension 1 face $F = \{\omega \in \{0,1\}^k \mid \omega_j = 0\}$ (for some $1 \le j \le k$) we have by IV.2.8 a representation $(g_\omega)_{\omega\in F} = \prod^{\prec}_{\omega\in F} \tilde{g}^{[\omega]}_\omega|_F$. Because of uniqueness in IV.2.8, the coefficients $\tilde{g}^{[\omega]}_\omega$ do not depend of F. Hence, we may define $\mathbf{g} = \prod^{\prec}_{\omega\in\{0,1\}^k} \tilde{g}^{[\omega]}_\omega$; this is a completion of $(g_\omega)_{\omega\in\{0,1\}^k_*}$.

For the additional part, note that if $g_\omega \in \Gamma$ for all $\omega \in \{0,1\}^k_*$, then also $\tilde{g}_\omega \in \Gamma$. Hence, with the construction above, $g_{1^k} = \prod^{\prec}_\omega \tilde{g}_\omega \in \Gamma$. □

We are now ready to prove the analogue of Proposition IV.2.9 for nilmanifolds.

Proposition IV.2.13. *Let $G_\bullet/\Gamma$ be length d nilmanifold. A sequence $\bar{f}\colon \mathcal{F}_\emptyset \to G/\Gamma$ is polynomial in the sense of Definition I.1.21 if and only if $\bar{f}$ maps $\mathcal{F}^{[k]}_\emptyset$ to $\mathrm{HK}^k(G_\bullet/\Gamma)$ for each k.*

Proof. If $\bar{f} = \pi \circ f$ is a polynomial in the sense of Definition I.1.21, then f maps $\mathcal{F}^{[k]}_\emptyset$ to $\mathrm{HK}^k(G_\bullet)$, and hence $\bar{f}$ maps $\mathcal{F}^{[k]}_\emptyset$ to the projected image, $\mathrm{HK}^k(G_\bullet/\Gamma)$.

Suppose conversely that $\bar{f}$ maps $\mathcal{F}^{[k]}_\emptyset$ to $\mathrm{HK}^k(G_\bullet/\Gamma)$ for each k. We aim to construct $f \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet)$ such that $\bar{f} = \pi \circ f$. Our construction will be inductive, where $f(\alpha)$ is constructed only after all $f(\beta)$ with $|\beta| < |\alpha|$. To initiate the construction, pick arbitrary $f(\emptyset)$ with $\pi(f(\emptyset)) = \bar{f}(\emptyset)$.

Suppose we want to assign a value to $f(\alpha)$. Let $\alpha = \{a_1, a_2, \dots, a_k\}$, and write $\alpha_\omega = \{a_i \mid \omega_i = 1\}$, so that $(\alpha_\omega)_\omega$ defines a parallelepiped. Hence, we have the cube $\bar{\mathbf{g}} = (\bar{f}(\alpha_\omega))_\omega \in \mathrm{HK}^k(G_\bullet/\Gamma)$. By definition, this cube is the projection of some cube $\mathbf{g} = (g_\omega)_\omega \in \mathrm{HK}^k(G_\bullet)$. On the other hand, because $f(\beta)$ have been constructed for $\beta \subsetneq \alpha$, we have the partial cube $(f(\alpha_\omega))_{\omega\neq 1^k}$.

Consider the (partial) cube $(g_\omega^{-1} f(\alpha_\omega))_{\omega\in\{0,1\}^k_*}$. By construction, $g_\omega^{-1} f(\alpha_\omega) \in \Gamma$ for all $\omega \in \{0,1\}^k_*$. Thus, by Lemma IV.2.12, there is some $\gamma \in \Gamma$ which completes this partial cube; that is if we define $f(\alpha) = f(\alpha_{1^k}) = g_{1^k}\gamma$, then $(g_\omega^{-1} f(\alpha_\omega))_{\omega\in\{0,1\}^k} \in \mathrm{HK}^k(G_\bullet)$. Multiplying by $\mathbf{g}$, we conclude that $(f(\alpha_\omega))_{\omega\in\{0,1\}^k} \in \mathrm{HK}^k(G_\bullet)$.

This construction guarantees that f maps parallelepipeds of the special form $\alpha_\omega = \{a_i \mid \omega_i = 1\}$ into the Host-Kra cube groups. It remains to see that any parallelepiped can be suitably embedded into one of this special form.

Indeed, let $(\alpha_\omega)_{\omega\in\{0,1\}^k}$ be any parallelepiped, where $\alpha_\omega = \alpha_0 \cup_{i\in\omega} \alpha_i$ with α_i disjoint. Let $\{a_1, \dots, a_l\}$ be all the elements of $\bigcup_{i=0}^k \alpha_i$, and consider the projection map $P\colon G^{\{0,1\}^l} \to G^{\{0,1\}^k}$, mapping $\mathbf{g} = (g_\omega)_\omega$ to $\mathbf{h} = (h_\sigma)_\sigma$ with $h_\sigma = g_{\alpha_\sigma}$. Once we show that P maps $\mathrm{HK}^l(G_\bullet)$ to $\mathrm{HK}^k(G_\bullet)$, it will follow that f maps $\mathcal{F}_\emptyset^{[k]}$ to $\mathrm{HK}^k(G_\bullet)$. Because P preserves multiplication, it suffices to verify that it maps generators of $\mathrm{HK}^l(G_\bullet)$ to $\mathrm{HK}^k(G_\bullet)$. This follows from the observation that $P(g^{[\omega]}) = g^{[\sigma]}$, where $\sigma = \{1 \le i \le k \mid \alpha_i \cap \omega \neq \emptyset\}$ has size $\le |\omega|$, and thus P maps $G^{[\omega]}$ to $G^{[\sigma]} \subset \mathrm{HK}^k(G_\bullet)$. □

Corollary IV.2.14. *Theorem IV.2.1 holds.*

Proof. The set $\mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet/\Gamma)$ lies in the compact space $(G/\Gamma)^{\mathcal{F}_\emptyset}$, so it will suffice to check that it is closed.

Suppose that $\bar{f}_n \to \bar{f}$ pointwise. Then for each k, and for each parallelepiped $(\alpha_\omega)_{\omega\in\{0,1\}^k} \in \mathcal{F}_\emptyset^{[k]}$, the cube $(\bar{f}(\alpha_\omega))_\omega$ lies in the closure of $\mathrm{HK}^k(G_\bullet/\Gamma)$. But $\mathrm{HK}^k(G_\bullet/\Gamma)$ is already closed by Lemma IV.2.11, so we $\bar{f}$ maps $\mathcal{F}_\emptyset^{[k]}$ to $\mathrm{HK}^k(G_\bullet/\Gamma)$, so we are done. □

IV.3 $\mathcal{S}_k$-sequences

In this section, we develop some language to speak about sequences indexed by $\mathcal{S}_k$, the k-syndetic sets. The key insight here is that $\mathcal{S}_k$-indexed sequences admit a well behaved notion of a subsequence, which allows us to restrict to particularly structured sequences later on, through application of Proposition IV.3.17.

IV.3.1 IP sets revisited

We briefly return to the discussion of IP sets, which serve as a motivation and an analogue for the SG_k sets. Recall that an IP set is (a superset of) a set of finite sums $\mathrm{FS}(n_i) = \{n_\alpha \mid \alpha \in \mathcal{F}\}$ for some $n_i \in \mathbb{N}$, where $n_\alpha = \sum_{i\in\alpha} n_i$. One of the reasons for interest in the IP sets is the celebrated theorem of Hindman [63], stating that the class of IP sets is *partition regular*.

Barring spurious coincidences between different terms n_α and n_β with $\alpha \neq \beta$, the set $\mathrm{FS}(n_i)$ can for most intents and purposes be identified with the set $\mathcal{F}$ of finite subsets of $\mathbb{N}$. To make this idea more precise, recall that we endow $\mathcal{F}$ with the structure of a partial semigroup, where the operation is the disjoint union. Now, the

map $\alpha \mapsto n_\alpha$ is a morphism of partial semigroups. Moreover, any morphism of partial semigroups $\mathcal{F} \to \mathbb{N}$ takes the form $\alpha \mapsto n_\alpha = \sum_{i\in\alpha} n_i$ for some sequence n_i, and the IP sets are precisely the images of $\mathcal{F}$ in $\mathbb{N}$ by such morphisms.

This point of view makes it more convenient to speak of IP subsets of a given IP set. If $(\alpha_i)_{i=1}^{\infty}$ is a sequence of disjoint sets, then the map $\beta \mapsto \alpha_\beta := \bigcup_{i\in\beta} \alpha_i$ is a morphism of partial semigroups, which is in fact an isomorphism onto the image. Recall from I.1.5 that we call the image of such a morphism an IP *ring* and denote it by $\mathrm{FU}(\alpha_i)$. If f is an $\mathcal{F}$-indexed sequence and $\mathrm{FU}(\alpha_i)$ is an IP ring, it is natural to consider the "restriction" of f to $\mathrm{FU}(\alpha_i)$, given by $\tilde{f}(\beta) = f(\alpha_\beta)$,where as usual $\alpha_\beta = \bigcup_{i\in\beta} \alpha_i$. (Note that the definition is arranged so that the domain of $\tilde{f}$ is again $\mathcal{F}$.)

We can reformulate Hindman's theorem as follows. Suppose that a sequence $f\colon \mathcal{F} \to X$ is given, taking values in some finite set X. Hindman's theorem then ensures that for a suitable choice of the IP ring, the corresponding subsequence $\tilde{f}$ is constant. Using the succinct terminology of Zorin-Kranich [110], any finitely valued $\mathcal{F}$-sequence is *wlog* constant. Slightly more generally, Hindman's theorem is equivalent to the statement that for any sequence $f\colon \mathcal{F} \to X$ taking values in a compact space X, there exists β_i such that the limit $\mathrm{IP}-\lim_\alpha f(\beta_\alpha)$ exists.

IV.3.2 Basic definitions

We wish to adapt some of the ideas relevant to IP sets to the context of SG_k sets.

As suggested by the formulation of Question IV.1.3, we will be interested in polynomial sequences $\mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet/\Gamma)$, but only values at $\mathcal{S}_k$, the k-syndetic sets, will play a role. To emphasise this state of affairs, we will use the term *$\mathcal{S}_k$-sequence* to refer to an $\mathcal{F}_\emptyset$-indexed sequence which we only intend to evaluate on $\mathcal{S}_k$. If we consider an $\mathcal{F}_\emptyset$-indexed sequence with no $\mathcal{S}_k$ in mind, we refer to it as an $\mathcal{F}_\emptyset$-sequence, or simply a sequence.

Remark. It is tempting to dispose of $\mathcal{F}_\emptyset$ altogether, and work with sequences indexed by $\mathcal{S}_k$. Indeed, $\mathcal{S}_k$ is certainly a partial semigroup, so it makes sense to consider polynomial groups such as $\mathrm{poly}(\mathcal{S}_k \to G_\bullet)$, and much of the discussion in this section would carry through with minor modifications.

However, we pursue a different route. One of the reason is that there does not seem to be a satisfactory analogue of Proposition IV.2.7 for $\mathrm{poly}(\mathcal{S}_k \to \mathbb{R}/\mathbb{Z})$. Indeed, the algebraic structure of $\mathcal{S}_k$ is not strong enough to admit a good description of

the polynomials from $\mathcal{S}_k$. For instance, $\mathcal{S}_k$ has no non-degenerate parallelepipeds of dimension $> k+1$.

We will now introduce a notion of a subsequence suitable for the study of $\mathcal{S}_k$-sequences. Consider a sequence $f\colon \mathcal{F}_\emptyset \to X$, taking values in some space X. As a source of motivation, recall that for any IP ring $\mathrm{FU}(\alpha_i)$, we may construct a subsequence $\tilde{f}$ given by $\tilde{f}(\beta) = f(\alpha_\beta)$, where $\alpha_\beta = \bigcup_{i\in\beta} \alpha_i$. Suppose now that α_j are chosen so that α_β is in $\mathcal{S}_k$ for any β in $\mathcal{S}_k$. Then, almost tautologically, we have $\tilde{f}(\mathcal{S}_k) \subset f(\mathcal{S}_k)$, so whenever we are interested in proving statements such as "there exists $\alpha \in \mathcal{S}_k$ such that $f(\alpha) \in U$", we may equally well replace f with $\tilde{f}$.

The condition that α_β should be in $\mathcal{S}_k$ whenever β is in $\mathcal{S}_k$ will be satisfied in the case when α_j takes the form $\alpha_j = \{i_j, i_j + k, \dots, i_{j+k} - k\}$ for an increasing sequence $(i_j)_{j=1}^\infty$ with $i_{j+k} \equiv i_j \pmod{k}$. Indeed, it suffices to check that if $j < j' \leq j+k$, then the gaps of $\alpha_j \cup \alpha_{j'}$ are bounded by k. If $j = j+k$, then $\alpha_j \cup \alpha_{j'}$ is just a progression with step k, and if $j' < j+k$ then $\alpha_j \cup \alpha_{j'}$ consists of two overlapping progressions with step k. In either case, the bound on the gaps is clear.

We will show that any possible choice of α_j such that $\beta \mapsto \alpha_\beta$ preserves $\mathcal{S}_k$ will essentially be of the above form, with the inconsequential caveat that there is some additional freedom in the choice of $\alpha_1, \alpha_2, \dots, \alpha_{k+1}$. Although formally not necessary, we record the proof of this fact, so that the definition which follows is not unmotivated.

Lemma IV.3.1. *Let $k \in \mathbb{N}$. Suppose that $\mathrm{FU}(\alpha_i)$ is an* IP *ring such that $\alpha_\beta \in \mathcal{S}_k$ for any $\beta \in \mathcal{S}_k$. Then, there exists an increasing sequence $(i_j)_{j=1}^\infty$ with $i_{j+k} \equiv i_j \pmod{k}$ for all j, such that $\alpha_j = \{i_j, i_j + k, \dots, i_{j+k} - k\}$ for any $j > k+1$.*

Proof. The basic idea is to iterate over $\mathbb{N}$ in increasing order and show that if α_j were not of the aforementioned form, then we would encounter a gap $> k$ in α_β for some $\beta \in \mathcal{S}_k$.

Let i_* be such that $\max \alpha_{i_*}$ is minimal. For $t \geq t_* := \max \alpha_{i_*} + k + 1$, consider the set $\iota(t)$ of indices of i such that $\alpha_i \cap [t-k, t) \neq \emptyset$. In particular, if $\min \alpha_i < t$ and $i \notin \iota(t)$, then $t \notin \alpha_i$ because $\alpha_i \in \mathcal{S}_k$.

We claim that for each $t \geq t_*$, $\iota(t)$ is a connected interval of length k. Indeed, let n be sufficiently large that $\min \alpha_n \geq t$, and consider the set $\beta = [n] \setminus \iota(t)$. Then $\min \alpha_\beta \leq \min \alpha_{i_*} < t-k$ and $\max \alpha_\beta \geq \max \alpha_n > t$ and $[t-k, t) \cap \alpha_\beta = \emptyset$, so $\alpha_\beta \notin \mathcal{S}_k$. Hence, $\beta \notin \mathcal{S}_k$. Because $|\iota(t)| \leq k$, the only way for $\beta = [n] \setminus \iota(t)$ to have gap $> k$ is if $\iota(t)$ is a connected interval of length k, as claimed.

For each t and $0 \leq l \leq k$, denote by $i_l(t)$ the index such that $t - l \in \alpha_{i_l(t)}$ (each $t-l$ belongs to some α_i by the above claim, and $i_1(t), \ldots, i_k(t)$ are all distinct). Since both $\iota(t)$ and $\iota(t+1)$ are intervals of length k, one of the following holds:

(i) $i_0(t) = i_k(t)$; or

(ii) $i_k(t) = \min \iota(t)$ and $i_0(t) = i_k(t) + k$; or

(iii) $i_k(t) = \max \iota(t)$ and $i_0(t) = i_k(t) - k$.

We next show that (i) never actually occurs. Clearly, (ii) happens infinitely often. Hence, if (iii) happens at any point, then there exists t such that (iii) happens at t and the next of (ii) and (iii) to happen is (ii). Thus, at some $t' > t$, (ii) happens, and for $t < s < t'$, (i) happens. Hence, $\iota(t') = \iota(t+1) = \iota(t) \setminus \{i_k(t)\} \cup \{i_k(t) - k\}$. Because $\max \iota(t') = i_k(t) + k$, we have $i_0(t') = i_k(t)$; but this is impossible, since then $\alpha_{i_k(t)}$ would have gap $> k$.

Note that if (i) happens then $\iota(t+1) = \iota(t)$, and if (ii) happens then $\iota(t+1) = \iota(t) + 1$. Each $i \neq i_*$ appears as the element of some $\iota(t)$, and the only way for this to happen is if $i_* = 1$ and $\iota(t_*) = \{2, 3, \ldots, k+1\}$. Inspecting what happens in each of (i) and (ii), we may further conclude that each of the sets $\alpha_i \cap [t_*, \infty)$ is a progression of step k, and for $i > k+1$, $\alpha_i \subset [t_*, \infty)$ and $\min \alpha_{i+1} > \min \alpha_i$. It follows that α_i are of the desired form. □

We are now ready to state the definition of an $\mathcal{S}_l$-subsequence.

Definition IV.3.2. Let $f\colon \mathcal{F}_\emptyset \to X$ be an $\mathcal{S}_k$-sequence, $k \geq 1$. Then, a $\mathcal{S}_k$-subsequence of f is any sequence $\tilde{f}\colon \mathcal{F}_\emptyset \to X$ taking the form $\tilde{f}(\beta) = f(\alpha_\beta)$ where $\alpha_j = \{i_j, i_j + k, \ldots, i_{j+k} - k\}$ for some increasing sequence i_j such that $i_{j+k} \equiv i_j \pmod{k}$, and $\alpha_\beta := \bigcup_{i \in \beta} \alpha_i$.

Similarly, for any $l < k$, $k \geq 1$ or $k = l = 0$, an $\mathcal{S}_l$-subsequence of an $\mathcal{S}_k$-sequence f is a sequence $\tilde{f}\colon \mathcal{F}_\emptyset \to X$ taking the form $\tilde{f}(\beta) = f(\alpha_\beta)$ where $\alpha_i \in \mathcal{F}$ are pairwise disjoint and the map $\beta \mapsto \alpha_\beta := \bigcup_{i \in \beta} \alpha_i$ takes $\mathcal{S}_l$ to $\mathcal{S}_k$. Note that this definition depends on both l and k.

Remark. There is a slight mismatch in the above definition. Namely, the notion of $\mathcal{S}_k$-subsequence of an $\mathcal{S}_k$-sequence is not exactly the same as would be obtained by extrapolating the definition of an $\mathcal{S}_l$-subsequence ($l < k$) to $l = k$.

This could be amended by requiring, in the definition of $\mathcal{S}_l$-subsequence ($l < k$), not only that $\beta \mapsto \alpha_\beta$ maps $\mathcal{S}_l$ to $\mathcal{S}_k$, but also that α_i can be extended to $(\alpha_i)_{i=-\infty}^{\infty}$, $\alpha_i \subset \mathbb{Z}$, in such a way that α_β has gaps $\leq k$ if β has gaps $\leq l$ for $\beta \subset \mathbb{Z}$, finite.

Although this would make the definitions more consistent, they would also become more complicated, so we instead accept the mismatch for the sake of simplicity.

Example IV.3.3. As already remarked, $\mathcal{S}_0$-sequences can be identified with the usual $\mathbb{N}$-indexed sequences. Under this identification, if $f\colon \mathcal{F}_\emptyset \to X$ is an $\mathcal{S}_0$-sequence, then any subsequence $(f(\{i_j\}))_{j=1}^\infty$ (in the usual sense of the word) can be seen as corresponding to an $\mathcal{S}_0$-subsequence $\tilde{f}$ with $\tilde{f}(\{j\}) = f(\{i_j\})$.

Example IV.3.4. Let $f\colon \mathcal{F}_\emptyset \to \mathbb{T}$ be a linear sequence of the special form $f(\alpha) = |\alpha|/M \bmod 1$, viewed as an $\mathcal{S}_k$-sequence, $k \geq 1$. Pick the sequence $i_{l+ak} = l + Mak$ $(1 \leq l \leq k,\ 0 \leq a)$, so that $\alpha_j = \{i_j, i_j + k, \dots, i_{j+k} - k\}$ are arithmetic progressions of length M and step k. Then, $\tilde{f}(\beta) = f(\alpha_\beta) = 0 \in \mathbb{T}$ is an $\mathcal{S}_k$-subsequence of f. Note that for $k = 1$, this amounts to splitting $\mathbb{N}$ into intervals of length M.

Example IV.3.5. Let $f\colon \mathcal{F}_\emptyset \to \mathbb{T}$ be a general linear sequence $f(\alpha) = \sum_{i\in\alpha} a_i$ for some $a_i \in \mathbb{T}$. To begin with, view f as an $\mathcal{S}_1$-sequence. There exists a convergent subsequence of the partial sums: $\sum_{i=1}^{i_j-1} a_i \to c$ for some $i_j \to \infty$, increasing, and $c \in \mathbb{T}$.

Pick $\alpha_j = \{i_j, \dots, i_{j+1} - 1\}$. Then

$$f(\alpha_j) = \sum_{i\in\alpha_j} a_i = \sum_{i=1}^{i_{j+1}-1} a_j - \sum_{i=1}^{i_j-1} a_j =: \varepsilon_j \to 0 \quad \text{as} \quad j \to \infty.$$

We may now construct the $\mathcal{S}_1$-subsequence $\tilde{f}$ of f given by $\tilde{f}(\beta) = \sum_{j\in\beta} \varepsilon_j$. Note that, by suitable choice of $(i_j)_j$, we may ensure that the ε_j are as small as we please. Requiring that $\|\varepsilon_j\| < \delta 2^{-j}$ for some small $\delta > 0$, say, we may thus ensure that $\left\|\tilde{f}(\beta)\right\| < \delta$ for all $\beta \in \mathcal{F}_\emptyset$.

By a similar construction, it can be checked that if f is considered as an $\mathcal{S}_k$-sequence, we may again extract an $\mathcal{S}_k$-subsequence $\tilde{f}$ taking the form $\tilde{f}(\beta) = \sum_{j\in\beta} \varepsilon_j$, where $\varepsilon_j \to 0$ rapidly as $j \to \infty$. Because the details do not include any new ideas, we skip them.

Example IV.3.6. Let $f\colon \mathcal{F}_\emptyset \to \mathbb{T}$ be any $\mathcal{S}_1$-sequence. Take a large integer M, and consider the complete graph on $\mathbb{N}$, coloured so that the edge $\{i, i'\}$ (where $i < i'$) is assigned colour $0 \leq m < M$ precisely when $f(\{i, i+1, \dots, i'-1\}) \in \left[\frac{m}{M}, \frac{m+1}{M}\right) \bmod 1$. By Ramsey's theorem, this graph has an infinite monochromatic clique $\{i_1, i_2, \dots\}$ with some colour m. Putting, as usual, $\alpha_j = \{i_j, \dots, i_{j+1} - 1\}$, we obtain an $\mathcal{S}_1$-subsequence $\tilde{f}(\beta) = f(\alpha_\beta)$ such that for each $\beta \in \mathcal{F}$, $\tilde{f}(\beta) \in \left[\frac{m}{M}, \frac{m+1}{M}\right) \bmod 1$. In other words, the sequence $\tilde{f}$ is approximately constant. Note, however, that this phenomenon does not immediately generalise to $\mathcal{S}_k$-sequences with $k \geq 2$.

Example IV.3.7. Consider the degree 2 polynomial $\mathcal{S}_2$-sequence $f\colon \mathcal{F}_\emptyset \to \mathbb{T}$ given by

$$f(\alpha) = \frac{1}{2}\left|\{\{i,j\} \subset \alpha \mid 0 \leq |i-j| \leq 2\}\right| \bmod 1.$$

It is not difficult to check that if α is a (non-empty) arithmetic progressions with step 2, then $f(\alpha) = \frac{1}{2}$. Moreover, if α, α' are two disjoint progressions with step 2, then $f(\alpha \cup \alpha') = f(\alpha) + f(\alpha') + \frac{1}{2} = \frac{1}{2}$ of α, α' either overlap ($\min\alpha < \min\alpha' < \max\alpha < \max\alpha'$) or touch ($\max\alpha + 2 = \min\alpha'$); and $f(\alpha \cup \alpha') = f(\alpha) + f(\alpha') = 0$ if α and α' are distance > 2 apart. It follows that the only $\mathcal{S}_2$-subsequence of f is f itself.

We now show a few $\mathcal{S}_1$-subsequences which can be extracted from f. If $\alpha = \{i, i+1, i+2\}$ is an interval of length 3, then by direct computation $f(\alpha) = 0$. More generally, if α, α' are two disjoint intervals of length 3, then $f(\alpha \cup \alpha') = \frac{1}{2}$ if α, α' are adjacent, and $= 0$ otherwise. Putting $\alpha_j = \{3j, 3j+1, 3j+2\}$, we thus obtain the $\mathcal{S}_1$-subsequence $\tilde{f}(\beta) = \frac{1}{2}\left|\{\{i,j\} \subset \alpha \mid |i-j| = 1\}\right| \bmod 1$; for $\beta \in \mathcal{S}_1$, this simplifies to $\tilde{f}(\beta) = \frac{1}{2}(|\beta| - 1) \bmod 1$.

A slightly more involved construction takes $\alpha_j = \{5j+1, 5j+3, 5j+4, 5j+5\}$. One can check by hand that $f(\alpha_j) = f(\alpha_j \cup \alpha_{j+1}) = 0$. As a consequence of these identities, the $\mathcal{S}_1$-subsequence $\tilde{f}(\beta) = f(\alpha_\beta)$ is just the constant 0 sequence: $\tilde{f}(\beta) = 0$, $\beta \in \mathcal{F}_\emptyset$.

Having discussed a number of examples of sequences and their subsequences, we turn to properties that are preserved under the operation of taking a subsequence.

It is clear that if g is an $\mathcal{S}_l$-subsequence of an $\mathcal{S}_k$-sequence f then $g(\mathcal{S}_l) \subset f(\mathcal{S}_k)$. Moreover, if h is an $\mathcal{S}_m$-subsequence of the $\mathcal{S}_l$-sequence g, then h is also an $\mathcal{S}_m$-subsequence of f, hence the relation of being a subsequence is transitive.

Also, an $\mathcal{S}_l$-subsequence of a polynomial sequence to a filtered nilpotent group is again polynomial, with respect to the same filtration, and there is an analogous statement for polynomial maps to nilmanifolds. Instances of this appear in Examples IV.3.4, IV.3.5, IV.3.7; we record the proof below. Note that the "linear" sequences $\alpha \mapsto g^{n_\alpha}$ are preserved under taking subsequences (direct substitution).

Proposition IV.3.8. *Let $G_\bullet$ be a filtration on nilmanifold G, and suppose that an $\mathcal{S}_k$-sequence $f\colon \mathcal{F}_\emptyset \to G$ is a polynomial sequence as in Definition I.1.15. For any $l \leq k$, if $\tilde{f}$ is an $\mathcal{S}_l$-subsequence of f, then $\tilde{f}$ is a polynomial sequence.*

Likewise, if $G_\bullet/\Gamma$ is a nilmanifold and $\bar{f}\colon \mathcal{F}_\emptyset \to G/\Gamma$ is a polynomial sequence as in Definition I.1.21 then for any $l \leq k$, if $\tilde{\bar{f}}$ is an $\mathcal{S}_l$-subsequence of $\bar{f}$, then $\tilde{\bar{f}}$ is a polynomial sequence.

Proof. Recall that $\tilde{f}(\beta) = f(\alpha_\beta)$ for some IP ring $\mathrm{FU}(\alpha_i)$. Because of Proposition IV.2.13, it suffices to check that $\tilde{f}$ maps parallelepipeds $\mathcal{F}_\emptyset^{[m]}$ to the corresponding Host-Kra space $\mathrm{HK}^m(G_\bullet)$. But this is clear because $\beta \mapsto \alpha_\beta$ maps $\mathcal{F}_\emptyset^{[m]}$ to $\mathcal{F}_\emptyset^{[m]}$. In fact, the last statement holds without any assumptions on α_i apart from them being disjoint.

The second part of the statement can be seen either as an immediate consequence of the previous one, or proved using an analogous argument, with $\mathrm{HK}^m(G_\bullet/\Gamma)$ in place of $\mathrm{HK}^m(G_\bullet)$. □

IV.3.3 Asymptotic subsequences

We are also interested in an asymptotic notion of a subsequence. From this point, we restrict attention to sequences taking values in compact metric spaces (we could work in the larger generality of compact topological spaces, but we do not need to).

Definition IV.3.9. Let $f\colon \mathcal{F}_\emptyset \to X$ be an $\mathcal{S}_k$-sequence taking values in a compact metric space X. Then, a sequence $\tilde{f}\colon \mathcal{F}_\emptyset \to X$ is said to be an *asymptotic* $\mathcal{S}_l$-*subsequence* of f if it is a pointwise limit of $\mathcal{S}_l$-subsequences of f. That is, $\tilde{f}$ is an asymptotic $\mathcal{S}_l$-subsequence of f precisely when there exist $\mathcal{S}_l$-subsequences $\tilde{f}_n$ of f such that for each $\beta \in \mathcal{F}_\emptyset$, $\tilde{f}(\beta) = \lim_{n\to\infty} \tilde{f}_n(\beta)$.

Remark. Note that in the final line of Definition IV.3.9 above, we require convergence for all $\beta \in \mathcal{F}_\emptyset$, as opposed to only $\beta \in \mathcal{S}_l$. However, if we have convergence for $\beta \in \mathcal{S}_l$, then by a diagonal argument we may always assume that $\tilde{f}_n(\beta)$ converge also for $\beta \in \mathcal{S}_l$.

Example IV.3.10. We revisit previously mentioned examples.

The rational $\mathcal{S}_1$-sequence in Example IV.3.4 has the constant 0 sequence as an ordinary $\mathcal{S}_1$-subsequence.

In contrast, the general linear $\mathcal{S}_1$-sequence from Example IV.3.5 has the constant 0 sequence as an asymptotic $\mathcal{S}_1$-subsequence, but not an ordinary subsequence (unless the coefficients a_γ are chosen in a specific way).

Example IV.3.6 shows that any $\mathcal{S}_1$-sequence has an asymptotic subsequence which is constant on $\mathcal{S}_1$.

Remarks similar to the ones we made about subsequences apply also to asymptotic subsequences. If $\tilde{f}$ is an asymptotic $\mathcal{S}_l$-subsequence of an $\mathcal{S}_k$-sequence f then $\tilde{f}(\mathcal{S}_l) \subset \mathrm{cl}(f(\mathcal{S}_k))$. The property of being a polynomial map is also preserved under taking asymptotic subsequences.

Proposition IV.3.11. *Let $G_\bullet/\Gamma$ be a nilmanifold, and suppose that an $\mathcal{S}_k$-sequence $\bar{f}\colon \mathcal{F}_\emptyset \to G/\Gamma$ is a polynomial sequence as in Definition I.1.21. For any $l \leq k$, if $\bar{\tilde{f}}$ is an $\mathcal{S}_l$-subsequence of $\bar{f}$, then $\bar{\tilde{f}}$ is a polynomial sequence.*

Proof. This follows immediately from Proposition IV.3.8 combined with compactness of $\mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet/\Gamma)$ (Theorem IV.2.1). □

However, we stress that "linear" sequences $\alpha \mapsto g^{n_\alpha}\Gamma$ are not preserved by the operation of taking asymptotic subsequences, as already seen from Example IV.2.5.

Example IV.3.12. Let $G = \begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}$ be the 2-step nilpotent group consisting of upper-triangular matrices, and let $\Gamma = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}$ consist of the matrices in G with integer entries. Then G/Γ is a nilmanifold, known as the Heisenberg nilmanifold. Denote by L the set of all sequences $\bar{f}\colon \mathcal{F}_\emptyset \to G/\Gamma$ of the form $\bar{f}(\alpha) = h^{m_\alpha}\Gamma$ ($h \in G,\ m_i \in \mathbb{N}$).

Pick $g = \begin{pmatrix} 1 & 2\theta & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$, so that we may compute

$$g^n\Gamma = \begin{pmatrix} 1 & 2n\theta & n(n+1)\theta \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix}\Gamma = \begin{pmatrix} 1 & \{2n\theta\} & \{-n(n-1)\theta\} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\Gamma.$$

Assuming that θ is irrational, for any $\varepsilon > 0$ we may choose n_1, n_2 so that $\|n_1\theta\| < \varepsilon,\ \|n_1^2\theta\| < \varepsilon,\ \|n_2\theta\| < \varepsilon,\ \|n_2^2\theta\| < \varepsilon$ and $\|n_1n_2\theta - \psi\| < \varepsilon$ for some freely chosen ψ (this is an instance of Weyl's equidistribution theorem). Passing to a subsequence, we conclude that $\mathrm{cl}\, L$ (in the topology of pointwise convergence) contains a sequence $\bar{f}$ with $\bar{f}(\{1\}) = \bar{f}(\{2\}) = e\Gamma$ and $\bar{f}(\{1,2\}) = \begin{pmatrix} 1 & 0 & -2\psi \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\Gamma$. Suppose for the sake of contradiction that $\bar{f} \in L$, so $\bar{f}(\alpha) = h^{m_\alpha}\Gamma$. Then, $h^{m_1}, h^{m_2} \in \Gamma$, so $h^{m_1+m_2} \in \Gamma$ and $\bar{f}(\{1,2\}) = e\Gamma$, which is a contradiction.

Observation IV.3.13. Among the $\mathcal{S}_k$-sequences taking values in a given compact metric space X, the relation of being an asymptotic $\mathcal{S}_k$-subsequence is transitive and reflexive. Hence, it is a partial weak order.

More generally, if f is an $\mathcal{S}_k$-sequence taking values in a compact metric space X, g is an asymptotic $\mathcal{S}_l$-subsequence of f, and h is an asymptotic $\mathcal{S}_m$-subsequence of g, then h is an asymptotic $\mathcal{S}_m$-subsequence of f.

Proof. Reflexivity is direct from the definition. It suffices to take $\alpha_i = \{i\}$.

For transitivity, we begin with noting that it will suffice to prove that for f, g, h as above, if h is a (non-asymptotic) $\mathcal{S}_m$-subsequence of g, then h is also an asymptotic $\mathcal{S}_m$-subsequence of f (this is because the set of asymptotic $\mathcal{S}_m$-subsequences is the same as the closure of the set of $\mathcal{S}_m$-subsequences in $X^{\mathcal{F}_\emptyset}$).

Suppose that $g(\beta) = \lim_{n\to\infty} f_n(\beta)$ where $f_n(\beta) = f(\alpha_\beta^{(n)})$, and $h(\gamma) = g(\beta_\gamma)$. Then clearly $h(\gamma) = \lim_{n\to\infty} f'_n(\gamma)$, where $f'_n(\gamma) = f(\alpha_{\beta_\gamma}^{(n)})$ is an $\mathcal{S}_m$-subsequence of f. Hence, h is an asymptotic $\mathcal{S}_m$-subsequence of f. □

Remark. The relation of being $\mathcal{S}_k$-subsequence (let alone asymptotic $\mathcal{S}_k$-subsequence) is not, however, anti-symmetric. This is shown by a simple example of a pair of $\mathbb{Z}/k\mathbb{Z}$-valued $\mathcal{S}_k$-sequences $f(\alpha) = \min\alpha \pmod k$, $g(\alpha) = \min\alpha + 1 \pmod k$, which are easily seen to be $\mathcal{S}_k$-subsequences of one another (put $\alpha_i := \{i+1\}$ and $\alpha'_i = \{i+k-1\}$). More generally, it follows directly from Definition IV.3.9 that all $\mathcal{S}_k$-subsequences of f take the form $\alpha \mapsto \sigma(\min\alpha \pmod k)$, where σ is a permutation of $\mathbb{Z}/k\mathbb{Z}$ (identified with $\{1, 2, \ldots, k\}$), and any sequence of this form is an $\mathcal{S}_k$-subsequence of f (take $\alpha_{ak+l} = \{2ak + \sigma(l), (2a+1)k + \sigma(l)\}$). Hence, it can be checked that f is an $\mathcal{S}_k$-subsequence of any of its $\mathcal{S}_k$-subsequences.

IV.3.4 Stable sequences

We will often find ourselves in the position of working with an $\mathcal{S}_k$-sequence taking values in a compact metric space, where we may freely restrict to asymptotic $\mathcal{S}_k$-subsequences. Hence, it is of interest to enquire into the possible simplest objects that can be obtained through such restrictions. This motivates the following definition.

Definition IV.3.14. Let f be an $\mathcal{S}_k$-sequence taking values in a compact metric space X. Then f is said to be *stable* if for any asymptotic $\mathcal{S}_k$-subsequence g of f, f is again an asymptotic subsequence of g.

Example IV.3.15. Any constant sequence is automatically stable.

Example IV.3.5 shows that any linear $\mathcal{S}_1$-sequence $f\colon \mathcal{F}_\emptyset \to \mathbb{T}$ (with $f(\emptyset) = 0$) has the constant 0 sequence as an $\mathcal{S}_1$-subsequence. Hence, the only stable linear $\mathcal{S}_1$-sequence is the constant 0 sequence.

On the other hand, Example IV.3.7 shows that there are non-trivial quadratic $\mathcal{S}_2$-sequences $f\colon \mathcal{F}_\emptyset \to \mathbb{T}$ which are stable. To extract a simpler subsequence from f, one has to pass to an $\mathcal{S}_1$-subsequence.

Lemma IV.3.16. *Let $f\colon \mathcal{F}_\emptyset \to X$ be an $\mathcal{S}_k$-sequence taking values in a compact metric space X. Then, there exists an asymptotic $\mathcal{S}_k$-subsequence g of f which is stable in the sense of Definition IV.3.14.*

Proof. This is standard application of the Kuratowski-Zorn Lemma. Consider the set of all asymptotic $\mathcal{S}_k$-subsequences of f, weakly partially ordered by the relation of being an $\mathcal{S}_k$-subsequence. Clearly, any minimal element for this ordering will be a stable $\mathcal{S}_k$-sequence, so it remains to check that any chain has a lower bound. If $\mathcal{C}$ is a chain, then by a standard diagonalising argument, we may assume that for each $\alpha \in \mathcal{F}_\emptyset$, the sequence $\{h(\alpha)\}_{h\in\mathcal{C}}$ converges to some limit $g(\alpha)$. It is now routine to check that g is an asymptotic $\mathcal{S}_k$-subsequence of f. □

To formulate (and prove) a structural result for stable $\mathcal{S}_k$-sequences, it will be useful to introduce a way of "shuffling" $\mathcal{S}_k$-sequences. Let π be a permutation of $\{1, 2, \dots, k\}$, and let us extend π to a permutation of $\mathbb{N}$ by $\pi(ak + l) = \pi(l) + ak$ $(0 \le a,\ 1 \le l \le k)$. Now, put $\alpha_i := \{\pi(i)\}$ and define the $\mathcal{S}_k$-sequence $\tilde{f}_\pi$ by $\tilde{f}_\pi(\beta) = f(\alpha_\beta)$. In plainer terms, $\tilde{f}_\pi(\beta)$ is obtained from f by permuting $\mathbb{N}$ according to π. We will call such sequence a *shuffle* of f. Note that a shuffle of a shuffle is again a shuffle, and the original sequence is a shuffle of any of its shuffles.

Remark. Note that in general, the shuffle $\tilde{f}_\pi$ need not be an $\mathcal{S}_k$-subsequence of f, since $\beta \mapsto \alpha_\beta$ (with α_i as above) usually does not map $\mathcal{S}_k$ to $\mathcal{S}_k$. However, it will later be of importance that if $g(\gamma)$ is an $\mathcal{S}_k$-subsequence of $\tilde{f}_\pi$, taking the form $g(\gamma) = \tilde{f}_\pi(\beta_\gamma)$ where $|\beta_j| \ge 2$ for all j, then g *is* an $\mathcal{S}_k$-subsequence of f. Indeed, it will suffice to check that for j, j' with $j < j' \le j + k$, the set $\alpha_{\beta_{\{j,j'\}}} = \{\pi(i) \mid i \in \beta_j \cup \beta_{j'}\}$ has gaps $\le k$. By definition, $\beta_j = \{i_j, \dots, i_{j+k} - k\}$ and $\beta_{j'} = \{i_{j'}, \dots, i_{j'+k} - k\}$. Since π maps pairs of points at distance exactly k apart into pairs of points exactly k apart, in the case $j' = j + k$ we are done. If $j' < j + k$, then there is some $i \in \beta_j$ such that $i_{j'} < i < i_{j'} + k$, so $\pi(i')$ is at distance $\le k$ from one of $\pi(i_{j'}), \pi(i_{j'} + k)$.

In the sequel, we only use a special case of the following proposition, which deals with a more specific type of a subsequence.

Proposition IV.3.17 (Structure theorem for stable sequences)**.** *Let f be a stable $\mathcal{S}_k$-sequence taking values in a compact metric space X. Suppose that g is an asymptotic $\mathcal{S}_k$-subsequence of f. Then g is a shuffle of f.*

In the special case when g takes the form $g(\beta) = f(\alpha_\beta)$, where $\alpha_j = \{i_j, i_j + k, \dots, i_{j+k} - k\}$ for an increasing sequence $(i_j)_{j=1}^\infty$ with $i_j \equiv j \pmod k$, we have a stronger conclusion $g = f$.

We will obtain the above result as a consequence of the following more precise statement. It will be convenient to use the notion of ultrafilters and the corresponding limits, which are discussed in Appendix B. Using ultrafilters, we can extract particularly simple asymptotic $\mathcal{S}_k$-subsequences.

Lemma IV.3.18 (Ultrafilter restriction lemma)**.** *Let k be an integer, and let $p = (p_1, p_2, \dots, p_k)$ be a k-tuple of non-principal ultrafilters with $n \equiv j \pmod k$ for p_j-almost all n. For a sequence $i = (i_j)_{j=1}^{\infty}$ with $i_j \equiv j \pmod k$, consider the intervals $\alpha_j(i)$ given by*

$$\alpha_j(i) = \{i_j, i_j + k, \dots, i_{j+k} - k\},$$

and put as usual $\alpha_\beta(i) := \bigcup_{j\in\beta} \alpha_j(i)$.

For an $\mathcal{S}_k$-sequence f taking values in a compact metric space X, define the $\mathcal{S}_k$-sequence $\tilde{f}_p$ as

$$\tilde{f}_p(\beta) := p_1\text{-}\lim_{i_1}\ p_2\text{-}\lim_{i_2}\ p_3\text{-}\lim_{i_3}\dots p_1\text{-}\lim_{i_{k+1}} p_2\text{-}\lim_{i_{k+2}}\dots f(\alpha_\beta(i)). \tag{IV.7}$$

(Above, the limit of i_m is taken along $p_{m \bmod k}$. The limits are taken over all relevant indices i_j, i.e. those with $j = j' + l$, $j \in \beta$, $0 \le l \le k$; there are finitely many of those.)

Then $\tilde{f}_p$ is an asymptotic $\mathcal{S}_k$-subsequence of f.

Proof. It will suffice to construct, for any sequence of neighbourhoods U_β of $\tilde{f}_p(\beta)$ (with $U_\beta = X$ for all but finitely many β), an increasing sequence $i = (i_j)_{j=1}^{\infty}$ such that $f(\alpha_\beta(i)) \in U_\beta$ for each β. For convenience, let $p_m := p_{m \bmod k}$ for $m > k$.

We construct i_j inductively. Unwinding the definitions of limits, for each β, there exists a sequence of (families of) sets $A^{(0)}_\beta$, $A^{(1)}_\beta(i_1)$, $A^{(2)}_\beta(i_1, i_2), \dots, A^{(r)}_\beta(i_1, i_2, \dots, i_r)$, (with r dependent on β) such that, firstly, for each l we have $A^{(l)}_\beta(i_1, i_2, \dots, i_l) \in p_{l+1}$ and, secondly, if $i_{l+1} \in A_\beta(i_1, i_2, \dots, i_l)$ for each $0 \le l \le r$, then $f(\alpha_\beta(i)) \in U_\beta$.

By assumption, $\{i \in \mathbb{N} \mid i \equiv j \pmod k\} \in p_j$, so without loss of generality we may suppose that for any l, and any $i \in A^{(l)}_\beta(i_1, i_2, \dots, i_l)$ we have $i \equiv l + 1 \pmod k$ and $i > i_l$. For simplicity of notation, let us put $A^{(l)}_\beta(i_1, i_2, \dots, i_l) = \{i \in \mathbb{N} \mid i \equiv l + 1 \pmod k,\ i > i_l\}$ for all $l > r$, so that $A^{(l)}_\beta$ are defined for all r.

It now becomes clear how the sequence i_j needs to be constructed. Let

$$A^{(l)}(i_1, \dots, i_l) := \bigcap_{\beta : U_\beta \neq X} A^{(l)}_\beta(i_1, \dots, i_l),$$

where the intersection is taken over $\beta \in \mathcal{F}$ with $U_\beta \neq X$. Begin with taking any $i_1 \in A^{(0)}$. This can be done because $\emptyset \neq A^{(0)} \in p_1$. In general, for each l, take arbitrary

$i_{l+1} \in A^{(l)}(i_1, i_2, \ldots, i_l)$, which can always be done because the sets $A^{(l)}(i_1, i_2, \ldots, i_l)$ are guaranteed to be in p_l, and in particular be non-empty. □

Lemma IV.3.18 implies Proposition IV.3.17. Pick a k-tuple of ultrafilters p as in Lemma IV.3.18. By Lemma IV.3.18, $\tilde{f}_p$ is an asymptotic $\mathcal{S}_k$-subsequence of f, and by stability f is an asymptotic $\mathcal{S}_k$-subsequence of $\tilde{f}_p$.

Let g be a (non-asymptotic) $\mathcal{S}_k$-subsequence of $\tilde{f}_p$. We claim that g is a shuffle of $\tilde{f}_p$. Once this is shown, it will follow that also all asymptotic $\mathcal{S}_k$-subsequences of $\tilde{f}_p$ are shuffles of $\tilde{f}_p$ (this is because there are finitely many shuffles of $\tilde{f}_p$). In particular, f is a shuffle of $\tilde{f}_p$, and consequently all (asymptotic) $\mathcal{S}_k$-subsequences of f are shuffles of f, as needed. Hence, it remains to prove the claim.

By definition, g takes the form $g(\gamma) = \tilde{f}_p(\beta_\gamma)$. Here, the sequence β_m takes the form $\beta_m = \{j_m, j_m + k, \ldots, j_{m+k} - k\}$ where $(j_m)_{m=1}^{\infty}$ is a sequence of integers and $j_m \bmod k$ depends only on $m \bmod k$. Replacing g with a shuffle, we may assume that $j_m \equiv m \pmod{k}$.

For a sequence $i = (i_k)_{k=1}^{\infty}$, let $\alpha_\beta(i)$ be as in the Lemma IV.3.18. We may now observe that

$$\alpha_{\beta_m}(i) = \bigcup_{j \in \beta_m} \{i_j, \ldots, i_{j+k} - k\} = \{i_{j_m}, \ldots, i_{j_{m+k}} - k\}.$$

Let $i'_m = i_{j_m}$ and $\alpha'_m = \{i'_m, \ldots, i'_{m+k} - k\}$, and let α'_γ be defined accordingly. The key point is that if we disregard inconsequential indices i_j with $j \neq j_m$, then the limit defining $g(\gamma) = \tilde{f}_p(\beta_\gamma)$ becomes identical with the limit defining $\tilde{f}_p(\gamma)$. More precisely, we may write:

$$\begin{aligned} g(\gamma) = \tilde{f}_p(\beta_\gamma) &= p_1 - \lim_{i_1} \; p_2 - \lim_{i_2} p_3 - \lim_{i_3} \ldots f(\alpha_{\beta_\gamma}(i)) \\ &= p_1 - \lim_{i_1} \; p_2 - \lim_{i_2} p_3 - \lim_{i_3} \ldots f(\alpha'_\gamma(i)) \\ &= p_1 - \lim_{i'_1} \; p_2 - \lim_{i'_2} p_3 - \lim_{i'_3} \ldots f(\alpha'_\gamma(i)) \\ &= p_1 - \lim_{i_1} \; p_2 - \lim_{i_2} p_3 - \lim_{i_3} \ldots f(\alpha_\gamma(i)) = \tilde{f}_p(\gamma). \end{aligned}$$

Note that if $j_m \equiv m \mod k$ for all m to begin with, then the application of the shuffle inside the proof is unnecessary; we immediately find $g = \tilde{f}_p$. Applying the shuffle which takes $\tilde{f}_p$ to f, we conclude that the analogous statement holds f, i.e. all $\mathcal{S}_k$-subsequences of f of the special form as above are again f. In particular, by inspection of formula (IV.7), we conclude that $\tilde{f}_p = f$. □

Corollary IV.3.19. *Let A be a SG_k set, and pick any $m \in \mathbb{N}$. Then $A \cap m\mathbb{N}$ is a SG_k set. Likewise, if B is a SG_k^* set, then $B \cap m\mathbb{N}$ is a SG_k^* set.*

Note that similar facts for IP sets are well known.

Proof. By the definition of A being a SG_k set, there is an $\mathcal{S}_k$-sequence n_α such that $n_\alpha \in A$ for $\alpha \in \mathcal{S}_k$. Passing to an $\mathcal{S}_k$-subsequence, we may assume without loss of generality that the sequence $(n_\alpha \bmod m)$ is stable (note that because the target space is finite, there is no need to use asymptotic subsequences). Thus, for any i we have

$$n_i \equiv n_{\{i,i+k\}} \equiv n_i + n_{i+k} \equiv 2n_i \pmod{m},$$

whence $n_i \equiv 0 \pmod{m}$, and more generally $n_\alpha \equiv 0 \pmod{m}$ for all $\alpha \in \mathcal{S}_k$.

For the second statement, it suffices to check that $B \cap m\mathbb{N} \cap A \neq \emptyset$ whenever A is SG_k. But this is clear, since $m\mathbb{N} \cap A$ is SG_k and B is SG_k^*. □

IV.3.5 Stable polynomials

Among all $\mathcal{F}_\emptyset$-sequences, we will be particularly interested in polynomial maps into the torus. Recall that an (asymptotic) subsequence of a polynomial map is again polynomial because of Proposition IV.2.13. Hence, we may in most cases assume that the polynomial we are working with is stable by passing to a subsequence. In this situation, we can obtain the following refinement of Proposition IV.2.7. For a set $\gamma \in \mathcal{F}_\emptyset$, we define the diameter $\operatorname{diam}(\gamma) = \max \gamma - \min \gamma$. By a slight abuse of notation, we write $\gamma + k$ for the sumset $\{i + k \mid i \in \gamma\}$.

Proposition IV.3.20 (Structure of stable polynomials $\mathcal{F}_\emptyset \to \mathbb{T}^m$)**.** *Suppose that $f\colon \mathcal{F}_\emptyset \to \mathbb{T}^m = \mathbb{R}^m/\mathbb{Z}^m$ is a polynomial of degree $d \leq k$, which is a stable $\mathcal{S}_k$-sequence (in the sense of Definition IV.3.14). Then f admits a representation of the form:*

$$f(\alpha) = \sum_{\gamma \subset \alpha} a_\gamma, \tag{IV.8}$$

where $\alpha_\gamma \in \mathbb{T}^m$ are constants, which further satisfy $a_\gamma = 0$ if $\operatorname{diam}(\gamma) > k$ or $|\gamma| > d$, and are periodic in the sense that $a_{\gamma+k} = a_\gamma$.

Proof. It is already shown in Proposition IV.2.7 that f admits a representation as in (IV.8) with $a_\gamma = 0$ if $|\gamma| > d$.

The periodicity condition $f(\alpha) = f(\alpha + k)$ follows immediately from applying Proposition IV.3.17 to the $\mathcal{S}_k$-subsequence of f given by the shift: $g(\alpha) = f(\alpha + k)$. (Alternatively, it is also a consequence of the form of the limit in (IV.7) and the fact that $f = \tilde{f}_p$). Since the coefficients a_γ are uniquely determined by f, we also have $a_\gamma = a_{\gamma+k}$.

For the vanishing of coefficients a_γ with $\mathrm{diam}(\gamma) > k$, we proceed by induction on d, the case $d = 1$ being trivial. Note that if the claim holds for some d, then for any degree d stable polynomial f and for any α, β with $\max \alpha < \min \beta - k$ we have $f(\alpha \cup \beta) = f(\alpha) + f(\beta) - f(\emptyset)$.

To prove the claim for arbitrary $d \leq k$, fix some $j_1 < j_2 - k$. By an inclusion-exclusion type argument it will suffice to prove that for any $\alpha \in \mathcal{F}_\emptyset$ with $\alpha \subset (j_1, j_2)$ we have

$$\sum_{\substack{\gamma \subset \alpha \cup \{j_1, j_2\} \\ j_1, j_2 \in \gamma}} a_\gamma = 0. \tag{IV.9}$$

Indeed, for $\delta \subset [j_1, j_2]$ with $j_1, j_2 \in \delta$ we have the identity

$$a_\delta = \sum_{\alpha \subset \delta \setminus \{j_1, j_2\}} (-1)^{|\delta \setminus \alpha|} \sum_{\substack{\gamma \subset \alpha \cup \{j_1, j_2\} \\ j_1, j_2 \in \gamma}} a_\gamma,$$

so (IV.9) implies that $a_\delta = 0$.

We first consider the case $\alpha = \emptyset$. Let $j_3 = j_2 + k$. Since $\Delta_{j_3} f$ is a polynomial of degree $d - 1$ we may write:

$$\Delta_{j_3} f(\{j_1, j_2\}) = \Delta_{j_3} f(\{j_1\}) + \Delta_{j_3} f(\{j_2\}) - \Delta_{j_3} f(\emptyset)$$

which can be rewritten in simpler terms as:

$$\begin{aligned} f(\{j_1, j_2, j_3\}) - f(\{j_1, j_2\}) = f(\{j_1, j_3\}) - f(\{j_1\}) \\ + f(\{j_2, j_3\}) - f(\{j_2\}) - f(\{j_3\}) + f(\emptyset). \end{aligned}$$

Using stability of f combined with Proposition IV.3.17, we may replace each occurrence of $\{j_2, j_3\}$ or $\{j_3\}$ with $\{j_2\}$. The above equation now simplifies to

$$f(\{j_1, j_2\}) = f(\{j_1\}) + f(\{j_2\}) - f(\emptyset).$$

Writing out f in coordinates from (IV.8) and cancelling repeating terms, this gives the sought after formula (IV.9).

We now consider arbitrary $\alpha \subset (j_1, j_2)$. The map $\Delta_\alpha f$ is a polynomial of degree $d - 1$ so by the inductive hypothesis

$$\Delta_\alpha f(\{j_1, j_2\}) = \Delta_\alpha f(\{j_1\}) + \Delta_\alpha f(\{j_2\}) - \Delta_\alpha f(\emptyset).$$

In particular, we have

$$\begin{aligned} f(\alpha \cup \{j_1, j_2\}) = f(\alpha \cup \{j_1\}) + f(\alpha \cup \{j_2\}) \\ + f(\{j_1, j_2\}) - f(\{j_1\}) - f(\{j_2\}) - f(\alpha) + f(\emptyset). \end{aligned}$$

Using the previous step, we may simplify this to

$$f(\alpha \cup \{j_1, j_2\}) = f(\alpha \cup \{j_1\}) + f(\alpha \cup \{j_2\}) - f(\alpha),$$

which is equivalent to (IV.9). □

IV.4 Basic results

IV.4.1 Abelian case

We observe that with the theory developed so far, the case $d = 1$ of our main result becomes trivial. This case is already well known (compare Example IV.3.5), but we discuss it here as a source of motivation.

Recall that any 1-step nilmanifold is in fact a compact abelian Lie group, and hence a product of a torus $\mathbb{T}^m$ and a finite abelian group. In all cases, it can be viewed as a submanifold of a (possibly higher-dimensional) torus, so we may restrict our attention to tori (see also Section IV.2.2).

Proof of Theorem IV.1.2, case $d = 1$. We need to check that for any linear (i.e. polynomial degree 1) $\mathcal{S}_1$-sequence, $f\colon \mathcal{F}_\emptyset \to \mathbb{T}^m$ with $f(\emptyset) = 0$, there are points $f(\alpha)$ with $\alpha \in \mathcal{S}_1$ arbitrarily close to 0. In fact, we prove somewhat more, namely that the constant sequence 0 is an asymptotic $\mathcal{S}_1$-subsequence of f, and hence such sets α are in rich supply.

Let g be a stable asymptotic $\mathcal{S}_1$-subsequence of f, as introduced in Section IV.3.4. By Proposition IV.3.20, we may write g in the form:

$$g(\alpha) = \sum_{i \in \alpha} a_i = |\alpha|\, t,$$

where $t = a_1$. Stability of g further implies that

$$t = g(\{1\}) = g(\{1, 2\}) = 2t,$$

and hence $t = 0$. If follows that $g(\alpha) = 0$ for each α, as needed. □

IV.4.2 Case $d = 2$

We now move on to quadratic polynomials, which already shows some of our main ideas.

Proof of Theorem IV.1.2, case $d = 2$. It will suffice to show that for any nilmanifold $G_\bullet/\Gamma$ of length 2 and any $\bar{f} \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet/\Gamma)$ with $\bar{f}(\emptyset) = e\Gamma$, there exist $\alpha \in \mathcal{S}_2$ such that $\bar{f}(\alpha)$ is arbitrarily close to $e\Gamma$. Hence, we prove that any Nil_2–Bohr_0 set is SG_2^*.

Step 1 (Model problem)**.** Let $f\colon \mathcal{F}_\emptyset \to \mathbb{T}^m$ be a polynomial of degree 2 with $f(\emptyset) = 0$. Then, for any $\varepsilon > 0$, there exists $\alpha \in \mathcal{S}_2$ such that $\|f(\alpha)\| < \varepsilon$.

Proof. The form of the claim is such that we can freely restrict f to an asymptotic $\mathcal{S}_2$-subsequence. Hence, we may without loss of generality assume that f is stable in the sense of Definition IV.3.14. Now, by Proposition IV.3.20, f takes the form:

$$f(\alpha) = \sum_{\gamma \subset \alpha} a_\gamma,$$

where $\alpha_\gamma \in \mathbb{T}$ are constants, which further satisfy $a_\gamma = 0$ if $\mathrm{diam}(\gamma) > 2$ or $|\gamma| > 2$, and a_γ are periodic in the sense that $a_{\gamma+2} = a_\gamma$.

Taking into account above properties of a_γ, there are only 6 meaningful coefficients, namely $a_1, a_2, a_{12}, a_{23}, a_{13}, a_{24}$. (We omit the curly brackets to avoid obfuscating notation; hence e.g. $a_{13} \equiv a_{\{1,3\}}$.)

We record some of the relations among the α_γ, which easily follow from stability:

$$a_{13} = -a_1, \qquad a_{24} = -a_2, \qquad a_{12} + a_{23} = 0. \tag{IV.10}$$

More exactly, these follow from the identities $f(\{1,3\}) = f(\{1\})$, $f(\{2,4\}) = f(\{2\})$ and $f(\{1,2,3,4\}) = f(\{1,2\})$ respectively.

Our general strategy at this point is to start with a sufficiently *generic* set $\alpha \in \mathcal{S}_2$, and check that it can be perturbed to a set α' such that $f(\alpha') \simeq 0$. A set α will be highly generic if it contains any possible *pattern* a large number of times.

To make these ideas precise, let us say that a pattern of length M is a set $\pi \in \mathcal{S}_2$ such that $\pi \subset \{1, 2, \dots, M\}$. We will say that π *appears in* α *at position* n if $n \equiv 0 \pmod 2$ and $(\alpha - n) \cap \{1, 2, \dots, M\} = \pi$. (Note that this condition depends not only on α and π but also on M.) Every pattern will appear in *some* $\alpha \in \mathcal{S}_2$ (e.g. $\alpha = \pi$), but there are some patterns which appear only boundedly many times. For instance, the length 3 pattern $\pi = \{1\}$ can appear only once in any $\alpha \in \mathcal{S}_2$, since if π appears at position n, then $n + 1 = \max \alpha$. Thus, we say that α is (M, N)-generic if for any pattern π of length M one of the following is true: either π appears at most $C = C(\pi)$ times in *any* $\beta \in \mathcal{S}_2$, or π appears at least N times in α.

It is a simple but useful fact that for any M, N there exist $\alpha \in \mathcal{S}_2$ which are (M, N)-generic. We will discuss a related issue in more detail in Section IV.5. Note

also that if $\alpha, \alpha' \in \mathcal{S}_2$ are close in the sense that $|\alpha \triangle \alpha'| \leq L$ for some integer L, then if α is (M, N)-generic, then also α' is (M, N')-generic, where $N' = N - C(L, M)$ for some constant C independent of N. We are interested in the regime where $N \to \infty$, but M takes a fairly small value. In fact, it will suffice to take $M = 5$.

The class of generic sets is slightly too large for our purposes, for somewhat mundane reasons. We will call a set α *well-formed* if additionally $\max \alpha \equiv \min \alpha \pmod 2$, and work mostly with well-formed α. For instance, $\alpha = \{1, 2, 3, 5\}$ is well-formed and $f(\alpha) = 3a_1 + a_2 + 2a_{13} + a_{12} + a_{23} = a_1 + a_2$, while $\beta = \{1, 2, 3, 4\}$ is not well-formed and $f(\beta) = 2a_1 + 2a_2 + a_{13} + a_{24} + 2a_{12} + a_{23} = a_1 + a_2 + a_{12}$; the appearance of the coefficient a_{12} would cause problems.

Informally, we would like to consider the set of all possible values of $f(\alpha)$ for all highly generic and well-formed α. Hence, we define Σ to be the set of those $s \in \mathbb{T}$ such that for any $\varepsilon > 0$ and any N there exists $(5, N)$-generic, well-formed α such that $\|f(\alpha) - s\| < \varepsilon$. Note that Σ is closed. Secondly, we want to take Δ to be the set of the perturbations of $f(\alpha)$ which are always feasible if α is sufficiently generic. Precisely, we declare $t \in \Delta_0$ if and only if there exists constants L and N_0 such that for all $N \geq N_0$ and all N-generic, well-formed α, we can find well-formed α' with $|\alpha \triangle \alpha'| \leq L$ and $f(\alpha') = f(\alpha) + t$. Finally, put $\Delta := \mathrm{cl}(\Delta_0)$. We note some simple properties of Σ and Δ.

Firstly, we have $\Sigma + \Delta = \Sigma$ and Δ is a group. Because $0 \in \Delta$ and Σ is closed, to prove the first equality it will suffice to prove the inclusion $\Sigma + \Delta_0 \subset \Sigma$. Take $s \in \Sigma$ and $t \in \Delta_0$. For any N, ε we may select $(5, N)$-generic, well-formed α such that $\|f(\alpha) - s\| < \varepsilon$. We may (assuming that N is large enough) find α' with $|\alpha \triangle \alpha'| < L$ for a constant L dependent only on t such that $f(\alpha') = f(\alpha) + t$. Hence, $\|f(\alpha') - (s + t)\| < \varepsilon$, and α' is N' generic, where $N' = N - C(L) \to \infty$ as $N \to \infty$. This finishes the proof of the first equality. A proof that $\Delta + \Delta = \Delta$ follows along similar lines. If $t, t' \in \Delta_0$ then for sufficiently generic α, there are α', α'' such that $|\alpha \triangle \alpha''| \leq |\alpha \triangle \alpha'| + |\alpha' \triangle \alpha''| \leq L(t) + L(t')$ and $f(\alpha'') = f(\alpha') + t' = f(\alpha) + t + t'$. To prove that Δ is a group, it remains to check that $\Delta = -\Delta$. Take $t \in \Delta$. There exists a sequence n_j such that $n_j t \to 0$ as $j \to \infty$. Hence $(n_j - 1)t \to -t \in \Delta$, as needed.

Secondly, we exhibit generating sets for Σ and Δ. For any well-formed α, we claim that $f(\alpha) \in \mathbb{Z}a_1 + \mathbb{Z}a_2$. Indeed, the analogous statement involving a_γ for all $\gamma \in \{1, 2, 12, 23, 13, 24\}$ would obviously be true. We may eliminate a_{13} and a_{24} using (IV.10). Next, we may notice that the expression for $f(\alpha)$ contains the same number of occurrences of a_{12} and a_{23}, so these can also be eliminated using (IV.10) ($f(\alpha) = a_1 + a_2$ when α is an interval, and removing an element from α preserves the

difference between the number of times a_{12} and a_{23} appear in $f(\alpha)$; this is the only point where we use the fact that α is well-formed). Hence, $\Sigma \subset \mathrm{cl}(\mathbb{Z}a_1 + \mathbb{Z}a_2)$.

We next consider Δ. If α is $(5, N)$-generic and N is sufficiently large, then we may find an occurrence of the length 5 pattern $\{2, 4\}$. We may form α' by replacing this pattern with $\{2, 3, 4\}$ (i.e. put $\alpha' := \alpha \cup \{n + 3\}$, where n is the position where the original pattern appears in α). An easy calculation shows that $f(\alpha') - f(\alpha) = a_{12} + a_{23} + a_1 = a_1$. Hence $a_1 \in \Delta$. By a symmetric argument, $a_2 \in \Delta$. Thus, $\Delta \supset \mathrm{cl}(\mathbb{Z}a_1 + \mathbb{Z}a_2)$.

Combining the two inclusions, we conclude that $\Sigma = \Delta = \mathrm{cl}(\mathbb{Z}a_1 + \mathbb{Z}a_2)$. This finishes the argument: Since $0 \in \Sigma$, we may find some $\alpha \in \mathcal{S}_2$ (which incidentally is highly-generic and well-formed) such that $\|f(\alpha)\| < \varepsilon$, where ε is as small as we please. □

Remark. Our proof can be used to find $\mathcal{S}_0$-subsequence which is identically 0, rather than a single set where the sequence takes the value 0. Note, however, that we are unable to get an $\mathcal{S}_1$-subsequence.

Remark. A more condensed proof of Step 1 is possible. Let f be a stable degree 2 polynomial with coefficients α_γ, as above. Consider the set

$$\alpha = \{1, 2, \dots, 6r + 2\} \setminus (\{3, 9, \dots, 6r_1 - 3\} \cup \{6, 12, \dots, 6r_2\})$$

for some $r \geq r_1, r_2 \geq 0$. Then, a short calculation shows $f(\alpha) = (r_1+1)a_1+(r_2+1)a_2$. Clearly, for any $\varepsilon > 0$, there are r_1, r_2 such that $\|f(\alpha)\| \leq \varepsilon$. Unfortunately, this succinct argument does not generalise well to $d \geq 3$.

Step 2 (Reduction to model setting)**.** Let $\bar{f} \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet/\Gamma)$ be a polynomial $\mathcal{S}_2$-sequence with $\bar{f}(\emptyset) = e\Gamma$, where $G_\bullet/\Gamma$ is a length 2 nilmanifold. Define the nilmanifold $\tilde{G}_\bullet/\tilde{\Gamma}$ by $\tilde{G}_0 = \tilde{G}_1 = \tilde{G}_2 = G_2$, $\tilde{G}_3 = \tilde{G}_4 = \cdots = \{e_G\}$, $\tilde{\Gamma} = \Gamma \cap G_2$.

Then, $\tilde{G}_\bullet/\tilde{\Gamma}$ is a length 2 nilmanifold, and there exists a polynomial sequence $\bar{g} \in \mathrm{poly}(\mathcal{F}_\emptyset \to \tilde{G}_\bullet/\tilde{\Gamma})$ such that $\iota \circ \bar{g}$ is an $\mathcal{S}_2$-asymptotic subsequence of $\bar{f}$, where $\iota \colon \tilde{G}/\tilde{\Gamma} \to G/\Gamma$ is the natural inclusion $g\tilde{\Gamma} \mapsto g\Gamma$.

Proof. It is clear that $\tilde{G}_\bullet$ is a filtration, and that $\tilde{\Gamma}$ is discrete. Each quotient $\tilde{G}/\tilde{\Gamma} \cap \tilde{G}_j$ is either $G_2/\Gamma \cap G_2$ or trivial, so $\tilde{G}_\bullet$ is evidently $\tilde{\Gamma}$-rational. The content of this step is that $\bar{f}$ has an asymptotic subsequence which takes values in $G_2\Gamma$. (See Lemma IV.5.2 for details.)

Let $\tilde{\bar{f}}$ be a stable asymptotic $\mathcal{S}_2$-subsequence $\bar{f}$. We claim that indeed $\tilde{\bar{f}}(\alpha) \in G_2\Gamma$ for all $\alpha \in \mathcal{F}_\emptyset$. By definition, $\tilde{\bar{f}}$ is the projection of some $\tilde{f} \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet)$. For any disjoint $\alpha, \beta \in \mathcal{F}_\emptyset$, we have $\tilde{f}(\alpha \cup \beta) = \tilde{f}(\alpha)\tilde{f}(\beta)\Delta_\beta \tilde{f}(\alpha)$ with $\Delta_\beta \tilde{f}(\alpha) \in G_2$.

Because G_2 is normal, there is a well-defined projection map $\pi\colon G/\Gamma \to G/G_2\Gamma$ given by $g\Gamma \mapsto gG_2\Gamma$. The map $\bar{h} = \pi \circ \tilde{\bar{f}}$ is easily verified to be a degree 1 polynomial:

$$\bar{h}(\alpha \cup \beta) = \tilde{f}(\alpha \cup \beta)G_2\Gamma = \tilde{f}(\alpha)\tilde{f}(\beta)\Delta_\beta \tilde{f}(\alpha)G_2\Gamma = \tilde{f}(\alpha)\tilde{f}(\beta)G_2\Gamma = \bar{h}(\alpha) + \bar{h}(\beta).$$

We also have $\bar{h}(\emptyset) = eG_2\Gamma$. Moreover, $\bar{h}$ is a stable $\mathcal{S}_2$-sequence, because $\tilde{\bar{f}}$ is stable. Repeating the argument from the case $d = 1$, we find that $\bar{h}(\alpha) = eG_2\Gamma$ for all $\alpha \in \mathcal{F}_\emptyset$. Thus, $\tilde{\bar{f}}$ takes values in $G_2\Gamma$, as needed. □

Combining the two steps easily finishes the proof. Start with an $\mathcal{S}_2$-sequence $\bar{f} \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet/\Gamma)$ with $\bar{f}(\emptyset) = e\Gamma$, where $G_\bullet/\Gamma$ is a length 2 nilmanifold. Let $\bar{g}$ be the asymptotic $\mathcal{S}_2$-subsequence constructed Step 2. Now, $\bar{g} \in \mathrm{poly}(\mathcal{F}_\emptyset \to \tilde{G}_\bullet/\tilde{\Gamma})$, where $\tilde{G}_0 = G_2$ is abelian. We may without loss of generality suppose that $\tilde{G}_\bullet/\tilde{\Gamma}$ is a torus, equipped with the standard filtration of length 2.

Applying Step 1 to $\bar{g}$, we find that $e\tilde{\Gamma} \in \mathrm{cl}\,\{\bar{g}(\alpha) \mid \alpha \in \mathcal{S}_2\}$, whence we conclude that $e\Gamma \in \mathrm{cl}\left\{\bar{f}(\alpha) \;\middle|\; \alpha \in \mathcal{S}_2\right\}$, as needed. □

IV.5 Main results

IV.5.1 Robust version and induction

We now approach the proof of Theorem IV.1.2 for general $d \geq 1$. To begin with, we state a more robust version, which is better suited for an inductive proof.

Theorem IV.5.1 (IV.1.2, robust version)**.** *Let $G_\bullet/\Gamma$ be a nilmanifold of length d, and let $\bar{f} \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet/\Gamma)$ with $\bar{f}(\emptyset) = e\Gamma$. Let $0 \leq k \leq 4d$ be an integer, and put $l = k - 4d$. Then, there exists a asymptotic $\mathcal{S}_l$-subsequence of $\bar{f}$ which is constantly equal to $e\Gamma$.*

Note that the strong version of Theorem IV.1.2, IV.2.3, follows directly from the above statement by putting $k = 4d$.

We will prove the above robust version of our main theorem by induction on the complexity of $G_\bullet/\Gamma$. The following lemma gives a useful description of polynomial sequences taking values on subnilmanifolds. We will often use this identification implicitly.

Lemma IV.5.2. *Let $G_\bullet/\Gamma$ be a nilmanifold of length d, and let $r \leq d$. Let the filtration $\tilde{G}_\bullet$ on $\tilde{G} = G_r$ be given by $\tilde{G}_i = G_r$ for $i \leq r$ and $\tilde{G}_i = G_i$ for $i \geq r$, and let $\tilde{\Gamma} := G_r \cap \Gamma$. Then, there is a natural bijective correspondence between maps*

$\bar{f} \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet/\Gamma)$ *taking values in* $G_r\Gamma = \{g\Gamma \mid g \in G_r\} \subset G/\Gamma$ *on one side and polynomial maps* $\bar{h} \in \mathrm{poly}(\mathcal{F}_\emptyset \to \tilde{G}_\bullet/\tilde{\Gamma})$ *on the other side. More precisely,* $\bar{h}$ *corresponds to* $\bar{f} = \iota \circ \bar{h}$*, where* $\iota\colon G_r/(G_r \cap \Gamma) \to G/\Gamma$ *is the natural inclusion given by* $g(G_r \cap \Gamma) \mapsto g\Gamma$*.*

Proof. Given $\bar{h} \in \mathrm{poly}(\mathcal{F}_\emptyset \to \tilde{G}_\bullet/\tilde{\Gamma})$, we may lift it to $h \in \mathrm{poly}(\mathcal{F}_\emptyset \to \tilde{G}_\bullet)$ so that $\bar{h}(\alpha) = h(\alpha)\tilde{\Gamma}$. Since $\tilde{G}_0 = G_r \subset G_0$, there is a natural way to consider h as a map to G_0, and under this identification $h \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet)$. Thus, we may take $\bar{f} \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet/\Gamma)$ given by $\bar{f}(\alpha) := h(\alpha)\Gamma$. It is clear that $\bar{f}$ takes values in $G_r\Gamma$, and $\bar{f} = \iota \circ \bar{h}$.

For the other direction, assume that $\bar{f} \in \mathrm{poly}(\mathcal{F}_\emptyset \to G_\bullet/\Gamma)$ taking values in $G_r\Gamma$ is given. For each $\alpha \in \mathcal{F}_\emptyset$, there is unique $\bar{h}(\alpha) \in G_r/(G_r \cap \Gamma)$ such that $\iota(\bar{h}(\alpha)) = \bar{f}(\alpha)$: one can take arbitrary $g_r \in G_r$ such that $\bar{f}(\alpha) = g_r\Gamma$ and put $\bar{h}(\alpha) = g_r(G_r \cap \Gamma)$ (it is easy to see that g_r is unique up to multiplication by and element of $G_r \cap \Gamma$). Hence, we have a map $\bar{h}\colon \mathcal{F}_\emptyset \to G_r/(G_r \cap \Gamma)$, and it remains to check that $\bar{h}$ is a polynomial.

Take any parallelepiped $(\alpha_\omega)_{\omega\in\{0,1\}^k} \in \mathcal{F}_\emptyset^{[k]}$. Let us consider the cubes $\bar{\mathbf{h}} = (\bar{h}(\alpha_\omega))_\omega$ and $\bar{\mathbf{g}} = (\bar{f}(\alpha_\omega))_\omega \in \mathrm{HK}^k(G_\bullet/\Gamma)$. We claim that $\bar{\mathbf{g}}$ can be lifted to a cube $\mathbf{g} = (g_\omega)_\omega \in \mathrm{HK}^k(G_\bullet)$ so that $g_\omega \in G_r$ for all ω. Once this is accomplished, we can construe $\mathbf{g}$ as an element of $\mathrm{HK}^k(\tilde{G}_\bullet)$, so that $\bar{\mathbf{h}}$ is the projection of $\mathbf{g}$, and in particular $\bar{\mathbf{h}} \in \mathrm{HK}^k(\tilde{G}_\bullet/\tilde{\Gamma})$.

Begin with any lift $\mathbf{g} \in \mathrm{HK}^k(G_\bullet)$ of $\bar{\mathbf{g}}$, and recall that $\mathbf{g}$ can be written as

$$\mathbf{g} = \prod_{\omega\in\{0,1\}^k}^{\prec} \tilde{g}_\omega^{[\omega]} = \tilde{g}_{\omega_1}^{[\omega_1]}\tilde{g}_{\omega_2}^{[\omega_2]}\dots\tilde{g}_{\omega_{2^k}}^{[\omega_{2^k}]},$$

where the product is taken in some fixed order compatible with the order induced by inclusion. Note that we are still free replace $\mathbf{g}$ by any element of $\mathrm{HK}^k(G_\bullet) \cap \Gamma^{\{0,1\}^k}$ (by a slight abuse of notation, we retain the same symbol $\mathbf{g}$). We will inductively alter $\mathbf{g}$ in this way, so that $\tilde{g}_\omega \in G_r$, where we consider ω in the order of increasing size $|\omega|$. Supposing that $g_\sigma \in G_r$ for $\sigma \prec \omega$, we may expand

$$g_\omega = \prod_{\sigma\preceq\omega}^{\prec} \tilde{g}_\sigma = \left(\prod_{\sigma\prec\omega}^{\prec} \tilde{g}_\sigma\right)\tilde{g}_\omega \in G_r\tilde{g}_\omega.$$

If $|\omega| \geq r$, then we automatically have $\tilde{g}_\omega \in G_{|\omega|} \subset G_r$, whence also $g_\omega \in G_r$, so we are done. Otherwise, $|\omega| < r$, and $g_\omega \in G_r\Gamma \cap G_{|\omega|}$. Writing $g_\omega = g_r\gamma$ with $g_r \in G_r$, $\gamma \in \Gamma$, we have $\gamma \in \Gamma \cap G_{|\omega|}$. After multiplying $\mathbf{g}$ by the inverse of $\gamma^{[\omega]} \in \mathrm{HK}^k(G_\bullet)$, we may assume that $\gamma = e$. (Note that this does not alter $\tilde{g}_\sigma$ unless $\sigma \succeq \omega$). Hence, we obtain $g_\omega \in G_r$ as desired. □

We will deduce Theorem IV.5.1 from the following inductive step.

Proposition IV.5.3. *Let $\bar{f} \in \mathrm{poly}(\mathcal{F} \to G_\bullet/\Gamma)$, where $G_\bullet/\Gamma$ is a nilmanifold of length d, and let $k \geq 0$ be an integer. Let r be the first index such that $G_{r+1} \subsetneq G_0$. Put $l = k - 2r$, and suppose that $l \geq 0$.*

Then, there exists an asymptotic $\mathcal{S}_l$-subsequence $\bar{g}$ of $\bar{f}$ such that g takes values in $G_{2r}\Gamma \simeq G_{2r}/(G_{2r} \cap \Gamma)$.

Proof of Theorem IV.5.1, assuming Proposition IV.5.3. Take $\bar{f} \in \mathrm{poly}(\mathcal{F} \to G_\bullet/\Gamma)$ as in Theorem IV.5.1. By repeated application of Proposition IV.5.3, for $n \geq 0$ we construct asymptotic $\mathcal{S}_{l_n}$-subsequences $\bar{f}^{(n)}$, where (up to the identification discussed in Lemma IV.5.2) $\bar{f}^{(n)} \in \mathrm{poly}(\mathcal{F} \to G_\bullet^{(n)}/\Gamma^{(n)})$. We begin with $G_i^{(0)} = G_i$ and $\Gamma^{(0)} = \Gamma$, $l^{(0)} = k$ and $r^{(0)} \geq 1$.

For each $n \geq 0$, the nilmanifold $G_\bullet^{(n+1)}/\Gamma^{(n+1)}$ takes the form $G_0^{(n+1)} = G_1^{(n+1)} = \cdots = G_{2r_n}^{(n+1)} = G_{2r_n}$ and $G_i^{(n)} = G_i$ for $i > 2r^{(n)}$, and $\Gamma^{(n+1)} = \Gamma \cap G_0^{(n+1)}$. We may finally take $l^{(n+1)} = l^{(n)} - 2r^{(n)} = k - \sum_{j=0}^{n} 2r^{(j)}$. The construction ensures that $r^{(n+1)} \geq 2r^{(n)}$. The subsequence at step n can be constructed, provided that $l^{(n)} \geq 0$.

We are specifically interested in the first time n such that $r^{(n)} \geq d + 1$, whence $G^{(n)}$ is trivial and $f^{(n)}$ is constantly equal e_G. For such n, we have $2r^{(n-1)} \leq 2d$, so $l^{(n)} \geq k - 4d \geq 0$ by assumption. □

IV.5.2 Reduction to an abelian problem

We deduce Proposition IV.5.3 from a model problem in the abelian setting. The proof of the following proposition is the most technical element of this chapter, and occupies most of Section IV.6.

Proposition IV.5.4. *Let k, d be integers with $k \leq d + 1$, and put $l = k - d - 1 \geq 0$. Let $f \colon \mathcal{F}_\emptyset \to \mathbb{T}^m$ be a polynomial map of degree d into a torus.*

Then, the constant sequence $0 \in \mathbb{T}^m$ is an asymptotic $\mathcal{S}_l$-subsequence of f.

Remark. Any finite abelian group A can be embedded in a torus, so the above proposition applies equally to polynomial maps $f \colon \mathcal{F}_\emptyset \to \mathbb{T}^m \times A$.

Proof of Proposition IV.5.3 assuming Proposition IV.5.4. Let $\bar{f} \in \mathrm{poly}(\mathcal{F} \to G_\bullet/\Gamma)$ as in Proposition IV.5.3. It will suffice to find an $\mathcal{S}_l$-asymptotic subsequence $\bar{g}$ of $\bar{f}$ which takes values in $G_{2r}\Gamma$.

We have a natural projection map $\pi \colon G/\Gamma \to G/G_{2r}\Gamma$, given by $g\Gamma \mapsto gG_{2r}\Gamma$. Since $[G_0, G_0] = [G_r, G_r] \subset G_{2r}$, the quotient $G/G_{2r}\Gamma = (G/G_{2r})/(G_{2r}\Gamma/G_{2r})$ is a compact connected abelian Lie group, hence a torus.

Note that $\pi \circ \bar{f} \colon \mathcal{F}_\emptyset \to G/\Gamma G_{2r}$ is a polynomial with respect to the length $2r$ filtration $\tilde{G}_j = G_j/G_{2r}$ for $j \leq 2r$ (and $\tilde{G}_j = \{e\}$ for $j > 2r$). We may identify $G/\Gamma G_{2r}$ as embedded in a torus, equipped with the standard length $2r-1$ filtration, and thus construe $\pi \circ \bar{f}$ as a degree $2r-1$ polynomial into a torus.

Hence, we may apply Proposition IV.5.4 to extract an asymptotic $\mathcal{S}_{k-2r}$-subsequence of $\pi \circ \bar{f}$ which is identically 0. By a standard diagonal argument, we may assume that this asymptotic $\mathcal{S}_{k-2r}$-subsequence takes the form $\pi \circ \bar{g}$ where $\bar{g}$ is an asymptotic $\mathcal{S}_{k-2r}$-subsequence of $\bar{f}$. This means precisely that $\bar{g}$ takes values in $G_{2r}\Gamma$, so we are done. □

IV.5.3 A counterexample

It is not clear if Theorem IV.5.1 is sharp, in the sense that it is no longer true for a larger value of l. Indeed, if the answer to Question IV.1.3 is positive, one would expect the theorem to hold with $l = k - d$.

Here, we give an example showing that the bounds in Proposition IV.5.4 and in Theorem IV.1.1 are close to being sharp.

Example IV.5.5. Let k be an integer, and put $d = k+1$. Consider the sequence

$$f(\alpha) = t \sum_{\substack{\emptyset \neq \gamma \subset \alpha \\ \mathrm{diam}(\gamma) \leq k}} (-1)^{|\gamma|} \pmod 1,$$

where $t \in \mathbb{T} \setminus \{0\}$ is a constant. Since any set γ with $\mathrm{diam}(\gamma) \leq k$ automatically has $|\gamma| \leq k+1$, thus defined f is evidently a polynomial of degree d with $f(\emptyset) = 0$.

We claim that for any $\alpha \in \mathcal{S}_k$, we have $f(\alpha) = -t$. In particular, the constant 0 sequence is not an asymptotic $\mathcal{S}_0$-subsequence of f.

We proceed by induction on $|\alpha|$. If $|\alpha| = 1$ then we have $f(\alpha) = -t$ directly by definition. If $|\alpha| \geq 2$, writing $j = \max \alpha$ and $\alpha = \alpha' \cup \{j\}$, we may compute that

$$f(\alpha) = f(\alpha') - t \sum_{\delta \subset \alpha \cap [j-k, j)} (-1)^{|\delta|} = f(\alpha') = -t,$$

by the inclusion-exclusion and the inductive assumption.

Example IV.5.6. Fix integers $d \leq k$. We construct a degree d polynomial $f \in \mathrm{poly}(\mathcal{F}_\emptyset \to \mathbb{T})$ such that if f is viewed as an $\mathcal{S}_k$-sequence, then for $l \geq k - d + 2$ the constant sequence 0 is *not* an asymptotic $\mathcal{S}_l$-subsequence f.

Generalising the construction from Example IV.5.6, let f be given by

$$f(\alpha) = t \sum_{\substack{\emptyset \neq \gamma \subset \alpha \\ \operatorname{diam}(\gamma) \leq k \\ |\gamma| \leq d}} (-1)^{|\gamma|} \pmod 1,$$

where $t \in \mathbb{T} \setminus \{0\}$. Ostensibly, f is a polynomial of degree d and $f(\emptyset) = 0$. Suppose now that $\mathrm{FU}(\alpha_i)$ is an IP ring such $\beta \mapsto \alpha_\beta$ maps $\mathcal{S}_l$ sets to $\mathcal{S}_k$ sets, where l is such as above. We will show that $f(\alpha_\beta)$ is far from 0 for some $\beta \in \mathcal{S}_l$. Indeed, we will show that $f(\alpha_i) = -t$ for sufficiently large i, namely i sufficiently large that $\min \alpha_i > \max \alpha_1, \max \alpha_2, \ldots, \max \alpha_l$.

Consider any $\gamma \subset \alpha_i$ with $\operatorname{diam}(\gamma) \leq k$. We will show these properties already imply that $|\gamma| \leq d$. Let $n = \min \gamma$ so that $\gamma \subset [n, n+k]$. For any residue class $m \in [l]$, $m \not\equiv i \pmod l$, there needs to be some $j \equiv m \pmod l$ such that $\alpha_j \cap [n, n+k] \neq \emptyset$; else for $\beta = \{m, m+l, m+2l, \ldots, m+bl\}$ with large b, the set α_β would fail to have gaps $\leq k$ although β has gaps $\leq l$. Thus, by a counting argument we have $|\gamma| \leq (k+1) - (l-1) \leq d$. If follows that

$$f(\alpha_i) = -t + t \sum_{\substack{\gamma \subset \alpha_i \\ \operatorname{diam}(\gamma) \leq k}} (-1)^{|\gamma|} = -t \pmod 1,$$

where the last equality is a general fact about $\mathcal{S}_k$ sets which is proved by simple induction on $|\alpha_i|$.

Example IV.5.7. Fix $k \geq 3$, and consider a "generic" degree 2 sequence

$$f(\alpha) = \sum_{\substack{\emptyset \neq \gamma \subset \alpha \\ |\gamma| \leq 2}} a_\gamma,$$

where $a_\gamma \in \mathbb{T}^{\binom{k+1}{2}}$ obey the following "stability" conditions: $a_{\gamma+k} = a_\gamma$, $a_\gamma = 0$ if $\operatorname{gap}(\gamma) > k$, $a_i = -a_{i,i+k}$, $a_{i,j} + a_{j,i+k} = 0$ and $\{a_{i,j} \mid 1 \leq i \leq j \leq k\}$ are linearly independent. (To shorten the notation, we write $a_{i,j}$ for $a_{\{i,j\}}$.)

Put $l = k-1$. We will show that the constant 0 sequence is not an asymptotic $\mathcal{S}_l$-subsequence of f. More precisely, we claim that if $\tilde{f}(\beta) = f(\alpha_\beta)$ is an $\mathcal{S}_k$-subsequence of f and we write out the expansion

$$\tilde{f}(\beta) = \sum_{\substack{\emptyset \neq \delta \subset \beta \\ |\delta| \leq 2}} b_\delta,$$

then for infinitely many j, the coefficient $b_{j,j+l}$ is of a rather special form $a_{i_1,i_2} + \epsilon_1 a_{i_1',i_2} + \epsilon_2 a_{i_1,i_2'} \neq 0$, where $\epsilon_1, \epsilon_2 \in \{0,1\}$ and $i_1' \not\equiv i_1 \pmod k$, $i_2' \not\equiv i_2 \pmod k$. Note that we have $b_{j,j+l} = \sum_{i \in \alpha_j, i' \in \alpha_{j+l}} a_{i,i'}$.

Fix a sufficiently large j. We may assume that $\min \alpha_j < \min \alpha_{j+l}$ and $\max \alpha_j < \max \alpha_{j+l}$. Since the set $\alpha_j \cup \alpha_{j+l}$ has gaps bounded by k, there is the least $i_1 \in \alpha_j$ such that $[i_1, i_1 + k] \cap \alpha_{j+l} \neq \emptyset$. By the same token, there is the largerst $i_2 \in \alpha_{j+l}$ such that $[i_2 - k, i_2] \cap \alpha_j \neq \emptyset$.

Each length k interval $[a, a+k) \subset [i_1, i_2]$ contains an element of α_j, α_{j+l}, as well as an element of one of α_m for m in each of the residue classes modulo l different than j. By an inductive argument reminiscent of Observation IV.3.1, we see that $[i_1, i_2] \cap \alpha_j$ and $[i_1, i_2] \cap \alpha_j'$ are arithmetic progressions of step k (same applies to $[i_1, i_2] \cap \bigcup_{n \equiv m \bmod k} \alpha_n$ for each $m \not\equiv j \pmod k$). Because of the stability conditions we imposed, the total contribution to $b_{j,j+l}$ from the part of α_j, α_{j+l} contained in $[i_1, i_2]$ is

$$\sum_{\substack{i \in [i_1,i_2] \cap \alpha_j, \\ i' \in [i_1,i_2] \cap \alpha_{j+l}}} a_{i,i'} = \sum_{\substack{i < i' < i+k \\ i_1 \leq i,\ i' < i_2}} (a_{i,i'} + a_{i',i+k}) + a_{i_1^*,i_2} = a_{i_1^*,i_2} = a_{i_1,i_2^*},$$

where $i_1^* = \max [i_1, i_2] \cap \alpha_j$ and $i_2^* = \min [i_1, i_2] \cap \alpha_{j+l}$.

By another counting argument, we see that the interval $[i_2^* - k, i_2^*)$ may contain at most one additional element $i_1' < i_1$ of α_j, and likewise interval $(i_1^*, i_1^* + k]$ may contain one additional element i_2' of α_{j+l}. These give the contributions a_{i_1',i_2^*} and $a_{i_1^*,i_2'}$ accordingly. We may finally bring the result to the required form, using periodicity of a_γ.

Remark. We expect that for any $d \leq k+1$, there exists a degree d polynomial $\mathcal{S}_k$-sequence which does not have the constant 0 sequence as asymptotic $\mathcal{S}_l$-subsequence for $l \geq k - d + 1$. However, exhibiting concrete examples of such sequences proves problematic.

IV.6 Model problem

In this section we prove Proposition IV.5.4. Together with previous considerations, this will finish the proof of our main result, Theorem IV.1.2. Towards the end, we also explain how to adapt the argument to prove Theorem IV.1.1.

Let $f \colon \mathcal{F}_\emptyset \to \mathbb{T}$ be a degree d polynomial viewed as an $\mathcal{S}_k$-subsequence, as in Proposition IV.5.4. Suppose further that $k \geq d+1$ and let $l = k - d - 1$. Passing to

an asymptotic $\mathcal{S}_k$-subsequence if necessary, we may assume that f is stable. Recall that by Corollary IV.3.20, f takes the form

$$f(\beta) = \sum_{\gamma \subset \beta} a_\gamma, \tag{IV.11}$$

$$a_\gamma = 0 \text{ if } |\gamma| \geq d \text{ or } \operatorname{diam} \gamma > k. \tag{IV.12}$$

Our main idea, much as in the case $d = 2$ discussed in Section IV.4, is to begin with a suitably generic IP ring $\mathrm{FU}(\alpha_i)$ such that $\beta \mapsto a_\beta$ maps $\mathcal{S}_l$ to $\mathcal{S}_k$, and to perturb α_i slightly to ensure that $f(\alpha_\beta)$ is small for all $\beta \in \mathcal{F}_\emptyset$. Throughout, (α_i) denotes a sequence of disjoint $\mathcal{S}_k$ sets.

IV.6.1 Patterns

We define a *pattern of length* M to be a collection of disjoint sets $\pi = (\pi_i)_{i=1}^{\infty}$ which are either in $\mathcal{S}_k$ or empty, such that $\beta \mapsto \pi_\beta$ maps $\mathcal{S}_l$ sets to $\mathcal{S}_k$, provided that π_i are all non-empty for $i \in \beta$. We say that (π_i) *occurs* at position $n \equiv 0 \pmod{k}$ in (α_i) if for each i, we have $(\alpha_i - n) \cap \{1, 2, \ldots, M\} = \pi_i$. (Note that this definition depends on M as well as on α_i and π_i).

We say that (α_i) is *well-formed* if the following conditions are satisfied:

(i) for each i, $\min \alpha_i \equiv \max \alpha_i \pmod{k}$,

(ii) for each i, $\min \alpha_{i+l+1} > \max \alpha_i + k$,

(iii) $\beta \mapsto \alpha_\beta$ maps $\mathcal{S}_l$ to $\mathcal{S}_k$.

(Note that we are working with fixed k, l and this definition is specific to those values.)

We say that (α_i) is (M, N)-generic if for any pattern π of length at most M, one of the following holds:

(i) There is a constant $C = C(\pi)$ such that (π_i) occurs at most C times in any well-formed sequence (β_i),

(ii) The pattern π occurs at least N times in (α_i).

We apply this definition in the regime where M is fixed and $N \to \infty$. In fact, it is enough to take $M = 3k$. The following observation shows that the above definition is not vacuous.

Claim IV.6.1. For any (M, N), there exists a well-formed sequence of sets (α_i) which is (M, N)-generic.

Proof. We need to construct (α_i) such that each pattern (π_i) of bounded length, which may potentially occur numerous times in some well-formed (β_i), occurs many times in (α_i). Our strategy is to begin with a class of patterns whose numerous occurrences may be easily guaranteed, and then gradually extending this class. The construction is (implicitly) inductive, but we reuse the same symbol (α_i) at each step.

Taking α_i to be long arithmetic progressions with step k, is easy to ensure that (α_i) has many occurrences of any pattern of the form

$$\pi_i = \begin{cases} \{(i \bmod k), (i \bmod k) + k, \ldots, (i \bmod k) + mk\}, & \text{if } j \leq i \leq i + l' \\ \emptyset, & \text{otherwise,} \end{cases} \tag{IV.13}$$

where $l' = l$ or $l' = l - 1$, and m is bounded. It is not difficult to alter (α_i) so as to change any occurrence of the pattern (IV.13) into a pattern of the similar form

$$\pi_i = \begin{cases} \{a_i, a_i + k, a_i + 2k, \ldots, a_i + m'k\}, & \text{if } j \leq i \leq i + l' \\ \emptyset, & \text{otherwise,} \end{cases} \tag{IV.14}$$

where $1 \leq a_i \leq k$ are arbitrary and $m' = m - O(1)$. Hence, we may assume that (α_i) contains many occurrences of patterns (IV.14). Applying similar reasoning, we may also convert occurrences of (IV.14) into patterns

$$\pi_i = \begin{cases} \{a_i, a_i + k, \ldots, a_i + t_i k, b_i + t_i k, \ldots, b_i + m''k\}, & \text{if } j \leq i \leq i + l' \\ \emptyset, & \text{otherwise,} \end{cases} \tag{IV.15}$$

where $1 \leq b_i \leq k$ are arbitrary and $m'' = m' - O(1)$ (but we claim no control over the t_i).

Finally, take any pattern (π_i) which occurs numerous times in some well-formed (β_i). There is some index j such that if $\pi_i \neq \emptyset$ then $j \leq i \leq j + l$. We may further assume that $\pi_i \neq \emptyset$ precisely for $j \leq i \leq j + l'$, where $l' = l$ or $l' = l - 1$. Letting $a_i = \min \pi_i \bmod k$ and $b_i = \max \pi_i \bmod k$, we see that any occurrence of the pattern (IV.15) may be converted into an occurrence of the sought pattern π, supposing (as we may) that m'' is sufficiently large with respect to π.

□

IV.6.2 Perturbations

Fix an arbitrary metric on $\mathbb{T}^{\mathcal{F}_\emptyset}$ which is compatible with the product structure. We define Σ to be the set of those $s \in \mathbb{T}^{\mathcal{F}_\emptyset}$ such that for any $\varepsilon > 0$ and any N there exists

$(3k, N)$-generic, well-formed (α_i) such that $\|f(\alpha_\xi) - s_\xi\| < \varepsilon$ (where ξ stands for a dummy variable[1]).

We further define the set of possible "perturbations" $\Delta \subset \mathbb{T}^{\mathcal{F}_\emptyset}$. For $t = (t_\xi)_{\xi \in \mathcal{F}_\emptyset}$, we declare $t \in \Delta_0$ if and only if there exists constants L and N_0 such that for all $N \geq N_0$, all $(3k, N)$-generic, well-formed (α_i), we can find well-formed (α_i') with $\sum_{i=1}^{\infty} |\alpha_i \triangle \alpha_i'| \leq L$ and $f(\alpha_\xi') = f(\alpha_\xi) + t_\xi$. Finally, put $\Delta := \mathrm{cl}(\Delta_0)$.

Because of (IV.11), any $(s_\xi) \in \Sigma$ admits a representation $s_\xi = \sum_{\gamma \subset \xi} c_\gamma$ for some constants c_γ. More precisely, if $s_\xi = f(\alpha_\xi)$, then c_γ are given by $c_\gamma = \sum_\delta a_\delta$, where the sum runs over δ with $\delta \subset \alpha_\gamma$ but $\delta \not\subset \alpha_{\gamma'}$ for $\gamma' \subsetneq \gamma$. Stability of f implies that $c_\gamma = 0$ if $|\gamma| > d$ or $\mathrm{diam}(\gamma) > k$. Likewise, any $t \in \Delta$ admits a representation of the same form $t_\xi = \sum_{\gamma \subset \xi} b_\gamma$, where $b_\gamma = 0$ if $|\gamma| > d$ or $\mathrm{diam}(\gamma) > k$. Let $A \subset \mathbb{T}^{\mathcal{F}_\emptyset}$ be the set of all $(c_\gamma)_\gamma$ as above, and let B accordingly be the set of all $(b_\gamma)_\gamma$.

The sets Σ and A are rather closely connected. Indeed, the map $\mathbb{T}^{\mathcal{F}_\emptyset} \to \mathbb{T}^{\mathcal{F}_\emptyset}$ given by $\{\xi \mapsto c_\xi\} \mapsto \{\xi \mapsto \sum_{\gamma \subset \xi} c_\xi\}$ is a bijection (as verified by an inclusion-exclusion argument) and an isomorphism of groups. Same applies to Δ and B.

We make some simple observations concerning the sets just defined. Just as before, we have $\Delta + \Sigma = \Sigma$ and $\Delta + \Delta = \Delta$, and moreover Δ is a closed group. The argument is essentially the same as in Section IV.4.2. For the same reasons, we have that $A + B = A$ and B is a closed group. We now study Δ and Σ in more detail.

Claim IV.6.2. For any $\kappa, \gamma \in \mathcal{F}_\emptyset$ with $d \geq |\gamma| \geq |\kappa|$ and $\mathrm{diam}(\kappa) \leq l$, $\mathrm{diam}(\gamma) \leq k$, there exists $t = (t_\xi)_\xi \in \Delta$ such that

(i) $t_\beta = 0$ unless $\beta \supseteq \kappa$,

(ii) $t_\kappa \in a_\gamma + \sum_{\delta \supsetneq \gamma} \mathbb{Z} a_\delta$.

Proof. Without loss of generality, we may assume that $\gamma \subset [k, 2k]$. We begin by choosing a pattern $\pi = (\pi_j)$ of length $3k$ such that

(i) $\pi_j \neq \emptyset$ for $j \in \kappa$,

(ii) $\bigcup_{j=1}^{\infty} \pi_j \cap \gamma = \emptyset$,

(iii) (π_j) occurs in all $(3k, N)$-generic (α_i) for sufficiently large N.

[1]Hence, strictly speaking we mean $\left\|(f(\alpha_\xi) - s_\xi)_{\xi \in \mathcal{F}_\emptyset}\right\| < \varepsilon$.

Such π can be constructed greedily, assigning each $m \in [1, 3k] \setminus \gamma$ to consecutive sets π_j, $j = \min\kappa, \ldots, \min\kappa+l$. (It is at this point that we are using that $l+d+1 \leq k$.)

Since $|\gamma| \geq |\kappa|$, we may partition $\gamma = \bigcup_{j\in\kappa} \gamma_j$ into $|\kappa|$ non-empty sets. Let $\gamma_j = \emptyset$ for $j \notin \kappa$. Fix one such partition once and for all.

Suppose that (α_i) is a well-formed, $(3k, N)$-generic sequence of sets, for some large N. Pick, arbitrarily, some n such that (π_j) appears at position n. For any $\sigma \subset \gamma$, we may consider the distortion of (α_i) given by $\alpha_i^\sigma = \alpha_i \cup (\gamma_i \cap \sigma + n)$. Note that the union is disjoint, and we have

$$(\alpha_i^\sigma - n) \cap [3k] = \pi_i \cup (\gamma_i \cap \sigma).$$

Clearly, thus obtained α_i^σ differs from α_i only in boundedly many places. Hence, the difference $f(\alpha_\xi^\sigma) - f(\alpha_\xi)$ belongs to Δ. Note that this difference depends only on π and σ, but not on α (this is the reason why we work with patterns of length $3k$ rather that k). More generally, summing over all σ (with appropriate choice of signs) we conclude that

$$t = (t_\xi)_\xi \in \Delta, \qquad t_\xi = \sum_{\sigma\subset\gamma} (-1)^{|\sigma|} f(\alpha_\xi^\sigma).$$

We now study the coefficients t_β for different sets β. (Only $\beta \subseteq \kappa$ will play an important role.) We have

$$t_\beta = \sum_{\sigma\subset\gamma} (-1)^{|\sigma|} f(\alpha_\beta^\sigma) = \sum_{\sigma\subset\gamma} (-1)^{|\sigma|} \sum_{\delta\subset\alpha_\beta^\sigma} a_\delta = \sum_{\delta\subset\alpha_\beta^\gamma} a_\delta \sum_{\substack{\sigma\subset\gamma \\ \delta\subset\alpha_\beta^\sigma}} (-1)^{|\sigma|}.$$

For fixed $\delta \subset \alpha_\beta^\kappa$, denote by $\sigma(\delta)$ the unique minimal set $\sigma \subset \gamma$ such that $\delta \subset \alpha_\beta^\sigma$ ($\sigma(\delta)$ does not depend on β). We now have

$$t_\beta = \sum_{\delta\subset\alpha_\beta^\gamma} a_\delta \sum_{\sigma(\delta)\subset\sigma\subset\gamma} (-1)^{|\sigma|}.$$

The inner sum is identically 0, unless $\sigma(\delta) = \gamma$, when it is equal to $(-1)^{|\gamma|}$. Note that we always have $\sigma(\delta) \subset \gamma_\beta$; indeed, $\alpha_\beta^\sigma = \alpha_\beta^{\sigma\cap\gamma_\beta}$. Hence, $t_\beta = 0$ unless $\kappa \subset \beta$. When $\kappa \subset \beta$, we may rewrite the requirement $\sigma(\delta) = \gamma$ as $\gamma + n \subset \delta$ (here, n is the index where (π_i) appears in (α_i)), and hence

$$t_\beta = (-1)^{|\gamma|} \sum_{(\gamma+n)\subset\delta\subset\alpha_\beta\cup(\gamma+n)} a_\delta.$$

We are specifically interested in $\beta = \kappa$. It is clear that the above sum includes $a_\gamma = a_{\gamma+n}$, and all the remaining summands take the form a_δ with $\delta \supsetneq \gamma$. Thus, after multiplying by $(-1)^{|\gamma|}$, the constructed (t_ξ) satisfies the required conditions. $\square$

Let $\vec{e}_\kappa$ denote the "base vector" $(e_{\kappa,\xi})$ with $e_{\kappa,\kappa} = 1$ and $e_{\kappa,\xi} = 0$ for $\xi \neq \kappa$.

Claim IV.6.3. The set A (and hence also B) is contained in

$$S = \overline{\text{span}}\left\{a_\gamma \vec{e}_\kappa \mid d \geq |\gamma| \geq |\kappa|\,,\ \text{diam}(\gamma) \leq k,\ \text{diam}(\kappa) \leq l\right\},$$

the smallest closed subgroup containing all elements $a_\gamma \vec{e}_\kappa$.

Proof. Fix some well-formed $(\alpha_i)_i$, and write $f(\alpha_\xi)$ as $f(\alpha_\xi) = \sum_{\kappa \subset \xi} c_\kappa$. One can check that this formula holds with c_κ given by

$$c_\kappa = \sum_{\substack{\gamma \subset \alpha_\kappa \\ \gamma \not\subseteq \alpha_\lambda \text{ for } \lambda \subsetneq \kappa}} a_\gamma.$$

The above sum only contains elements a_γ with $|\gamma| \geq |\kappa|$, and we may eliminate the terms with $|\gamma| > d$ because these vanish. Finally, if $\text{diam}(\kappa) > l$ then (because (α_i) is well-formed), we have $\text{diam}(\gamma) > k$ for all γ in the above sum, and hence the corresponding a_γ vanish.

It follows that $(c_\xi)_\xi \in S$. Since S is closed, also $A \subset S$. Finally, S is a group, so $A + B = A$, $A \subset S$ implies $B \subset S$. □

We next bootstrap Claim IV.6.2 to a more precise statement.

Claim IV.6.4. In the situation of Claim IV.6.2, there exists $b = (b_\xi)_\xi \in B$ such that

(i) $b_\beta = 0$ unless $\beta = \kappa$,

(ii) $b_\kappa = a_\gamma$.

Proof. It follows from Claim IV.6.2 that we may find for any γ, κ with $|\gamma| \geq |\kappa|$ and $\text{diam}(\kappa) \leq l$ a vector $(t_\xi)_\xi \in \Delta$ such that

$$t_\kappa = a_\gamma + \sum_{\delta \supsetneq \gamma} k_\delta a_\delta$$

for some integers k_δ. Applying this inductively with δ in place of γ and using that $a_\delta = 0$ for sufficiently large δ, we may eliminate all the summands a_δ above, so that $t_\kappa = a_\gamma$. Writing out t_κ in the coordinates, we thus find $b = (b_\xi)_\xi \in B$ such that $b_\beta = 0$ unless $\beta \supseteq \kappa$, and $b_\kappa = a_\gamma$.

Using a form of Gaussian elimination together with Claim IV.6.3, we may now produce for any $\varepsilon > 0$ an element $b \in B$ such that $b_\kappa = a_\gamma$ and $\|b_\beta\| < \varepsilon$ if $\beta \not\supseteq \kappa$ or $\beta \in \{\kappa_1, \kappa_2, \ldots, \kappa_r\}$, for any finite list $\{\kappa_j\}_{j=1}^r$. Finally, using the compactness of B, we may use the above procedure to produce $b \in B$ such that $b_\kappa = a_\gamma$ and $b_\beta = 0$ if $\beta \neq \kappa$. □

IV.6.3 Final step

We are now in position to finish the proof of Proposition IV.5.4. Combining Claims IV.6.4 and IV.6.3, we see that $A = B$, and hence $\Sigma = \Delta$. But this means that Σ contains the constant sequence $\beta \mapsto 0$. Hence, there is some (α_i) such that $\beta \mapsto \alpha_\beta$ maps $\mathcal{S}_l$ to $\mathcal{S}_k$ (which additionally happens to be well-formed and generic) so that $\beta \mapsto f(\alpha_\beta)$ is as close to the constant sequence $\beta \mapsto 0$ (in the product topology of $\mathbb{T}^{\mathcal{F}_\emptyset}$) as we wish. Thus, we can extract an asymptotic $\mathcal{S}_l$-subsequence of f which is identically 0, which was our goal.

IV.6.4 Proof of Theorem IV.1.1

Having proved Proposition IV.5.4 (and hence Theorem IV.1.2) we discuss the proof of Theorem IV.1.1. We are in the same situation as in Proposition IV.5.4, with the exception that we have $k = d$, and we need to take $l = 0$.

We apply the same argument as above, with some simplifications. A pattern is now (again) just a single set $\pi \in \mathcal{S}_k$. The appropriate version of Claim IV.6.1 is easily proved. The key difference in Claim IV.6.2 is that we can now prove it with $k = d$ and $l = 0$ (we may simply put $\pi = [1, 3k] \setminus \gamma$, with notation therein). Neither Claim IV.6.3, not the remainder of the argument ever use the relation between k, l and d, so the reasoning carries through.

Chapter V

Automatic sequences and generalised polynomials

V.1 Introduction

In this chapter we study combinatorial properties of generalised polynomial sequences. As explained in Section I.1.9, such sequences are closely related to nil–Bohr sets studied the previous Chapter IV, however our focus will be rather different. We recall that generalised polynomials are polynomial-like expressions, involving the floor function, such as $\sqrt{5}n \left\lfloor \sqrt{2}n^2 + \pi \left\lfloor \sqrt{3}n \right\rfloor^2 \right\rfloor$. For a precise definition, see Section I.1.9.

Despite some superficial similarities, the class of generalised polynomials differs in fundamental ways from the class of polynomials. For instance, the set of zeroes of a generalised polynomial may be infinite without the polynomial being 0, which is the case for example for $f(n) = n/2 - \lfloor n/2 \rfloor$. As a more striking example, the characteristic function of the Fibonacci numbers ($F_0 = 0,\ F_1 = 1,\ F_{i+2} = F_{i+1} + F_i$) can also be expressed as a generalised polynomial (see Example V.7.1 and Lemma V.4.4 for details, including the construction of the generalised polynomial formula for the said function). We refer to generalised polynomials that take non-zero values on a set of density 0 as *sparse generalised polynomials*.

In a series of papers [59], [60], [61], Håland and Knuth studied the equidistribution modulo 1 of certain generalised polynomials of a specific form. In a more general situation, Bergelson and Leibman [16] studied the distribution of values of arbitrary generalised polynomials. They proved, among other things, that any generalised polynomial $g\colon \mathbb{Z} \to \mathbb{R}^d$ is well-distributed inside a (parametrised) piecewise polynomial surface (for precise definitions, see [16]). A very general result on equidistibution modulo 1 was obtained by Leibman [85].

Much of the existing theory does not distinguish between two generalised polynomials which differ on a set of density 0. In particular, most of the results mentioned above are vacuous for sparse generalised polynomials, which from this point of view are indistinguishable from the constant zero function.

Presently, we propose to study some combinatorial structures which are related to sparse generalised polynomials. Our focus is restricted mainly to $\{0,1\}$-valued sequences, which can be identified with subsets of $\mathbb{Z}$ in a natural way. We will say that a set $E \subset \mathbb{Z}$ (resp. $\mathbb{N}$) is generalised polynomial if its characteristic function is a generalised polynomial; the set E is sparse if its characteristic function is sparse, i.e. $d(E) = 0$. We ask which combinatorial structures can be found in sparse generalised polynomial sets.[1]

Recall that an IP set is a set containing the set of finite sums of a sequence of positive integers, i.e. a set of the form $\left\{\sum_{i\in\alpha} n_i \;\middle|\; \emptyset \neq \alpha \subset \mathbb{N},\ |\alpha| < \infty\right\}$, where $(n_i)_{i\in\mathbb{N}} \subset \mathbb{N}$. We are now ready to state the first of our main results, which is also the most technically demanding one.

Theorem V.1.1. *Suppose that E is a sparse generalised polynomial set. Then E does not contain an* IP *set.*

Our methods rely on the link between the generalised polynomials and orbits on nilmanifolds that was established by Bergelson and Leibman (Theorem I.1.22). Specifically, we show that the set of times at which a linear orbit on a nilmanifold visits a semialgebraic set cannot be an IPS set, unless the set has nonempty interior (Theorem V.5.1). (The notion of an IPS set generalises the notion of a shifted IP set, see Definition V.4.6.)

The main application for Theorem V.1.1 is a classification of automatic sequences given by generalised polynomials. Roughly speaking, a sequence is k-automatic, if its n-th term can be computed by a finite state machine from base k digits of n. (A precise definition is given below.) A wider and closely related class of k-regular sequences [3, 5] was introduced by Allouche and Shallit. At this point suffice it to say that finitely valued regular sequences are precisely the automatic sequences, but regular sequences may not be finitely valued. For the definition (and the motivation), see Definition V.2.5 and the remainder of Section .

[1] Note that this definition depends on whether we are considering the set E as a subset of $\mathbb{Z}$ or $\mathbb{N}$, and a generalised polynomial set $E \subset \mathbb{N}$ may a priori not be a generalised polynomial set when considered as a subset of $\mathbb{Z}$. It will always be clear from context which meaning we have in mind.

The problem of demonstrating that a certain sequence is or is not automatic or regular has been widely studied, particularly for sequences of arithmetic origin. In [5, Theorem 6.2] it is proved that the sequence $(f(n))_{n\geq 0}$ given by $f(n) = \lfloor \alpha n + \beta \rfloor$ for real numbers α, β is regular if and only if α is rational. The method used there does not immediately generalise to higher degree polynomials in n, but the proof implicitly uses rotation on a circle by an angle of $2\pi\alpha$. Replacing the rotation on a circle by a skew product transformation on a torus (as in Furstenberg's proof of Weyl's equidistribution theorem), we easily obtain the following result. (For more on regular sequences, see Section V.2).

Theorem V.1.2. *Let $p \in \mathbb{R}[x]$ be a polynomial. Then the sequence $f(n) = \lfloor p(n) \rfloor$, $n \geq 0$ is regular if and only if all the coefficients of p except possibly for the constant term are rational.*

In fact, we show the stronger statement that for an integer $m \geq 2$ the sequence $\lfloor f(n) \rfloor \bmod m$ is not automatic unless all the coefficients of p except for the constant term are rational, in which case it is periodic. (A sequence $f\colon \mathbb{N}_0 \to \Omega$ is periodic if there exists $d > 0$ such that $f(n) = f(n+d)$ for all $n \in \mathbb{N}_0$.)

Another closely related motivating example comes from the classical Fibonacci word[2] $w_{\mathrm{Fib}} \in \{0,1\}^{\mathbb{N}_0}$. There are several ways to define it, each shedding light from a different direction.

(i) *Morphic word.* Define the sequence of words $w_0 := 0$, $w_1 := 01$ and $w_{i+2} := w_{i+1}w_i$. Then w_{Fib} is the (coordinate-wise) limit of w_i as $i \to \infty$.

(ii) *Sturmian word.* Explicitly, $w_{\mathrm{Fib}}(n) = 2 + \lfloor (n-1)\varphi \rfloor - \lfloor n\varphi \rfloor$.

(iii) *Fib-automatic sequence.* If $n = \sum_{i\geq 2} v_i F_i$ where $v_i \in \{0,1\}$ there is no i with $v_i = v_{i+1} = 1$, then $w_{\mathrm{Fib}}(n) = v_1$.

The equivalence of (i) and (ii) is well known, see e.g. [87]. The representation $\ldots v_3 v_2 v_1$ of n in (iii), is known as the Zeckendorf representation; it exist for each n and is unique. The notion of automaticity using the Zeckendorf representation (or for that matter a representation from a much wider class) in place of the usual base k representation of the input n was introduced and studied by Shallit in [101], where among other things the equivalence of (i) and (iii) is shown. We return to this subject in Section V.9.3.

Hence, w_{Fib} gives a non-trivial example of a sequence which is given by a generalised polynomial and satisfies a variant of automaticity related to the Zeckendorf

[2]We will freely identify words in $\Omega^{\mathbb{N}_0}$ with functions $\mathbb{N}_0 \to \Omega$.

representation. It is natural to ask if similar examples exist for the usual notion of k-automaticity. Motivated by Theorem V.1.1, we believe the answer to this should be negative. The following conjecture was the initial motivation for the line of research discussed in this chapter.

Conjecture V.1.3. *Suppose that a sequence $f\colon \mathbb{N}_0 \to \mathbb{R}$ is simultaneously automatic and generalised polynomial. Then f is ultimately periodic.*

(We say that a sequence f is *ultimately periodic* if it coincides with a periodic sequence except at a finite set.)

We are able to partially resolve this conjecture. First of all, we prove that the conjecture holds except on a set of density zero. In fact, in order to obtain such a result, we only need a specific property of automatic sequences. For the purpose of stating the next theorem, let us say that a sequence $f\colon \mathbb{N}_0 \to X$ is *weakly periodic* if for any restriction of f to an arithmetic sequence, $f'(n) = f(an+b)$, $a \in \mathbb{N},\ b \in \mathbb{N}_0$, there exist $q \in \mathbb{N}$, $r, s \in \mathbb{N}_0$ with $r \neq s$, such that $f'(qn+r) = f'(qn+s)$. Of course, any periodic sequence is weakly periodic, but not conversely. All automatic sequences are weakly periodic, which follows from the finiteness of kernels (see Lemma V.3.1). Another non-trivial example is the characteristic function of the square-free numbers.

Theorem V.1.4. *Suppose that a sequence $f\colon \mathbb{N}_0 \to \mathbb{R}$ is weakly periodic and generalised polynomial. Then there exists a periodic function b and a set $Z \subset \mathbb{N}_0$ of upper Banach density zero such that $f(n) = b(n)$ for $n \in \mathbb{N}_0 \setminus Z$.*

This is already sufficient to rule out automaticity of many natural examples of generalised polynomials. In particular, sequences such as $\lfloor \sqrt{2}n \lfloor \sqrt{3n} \rfloor \rfloor \bmod 10$ or $\left\lfloor \sqrt{2}n \left\lfloor \sqrt{3n} \right\rfloor^2 + \sqrt{5}n + \sqrt{7} \right\rfloor \bmod 10$ are not automatic. For details and more examples, see Corollary V.3.6.

To obtain stronger bounds on the size of the exceptional set Z, we need to make use of Theorem V.1.1, and exploit some of the finer properties of automatic sequences. Towards this end we develop a structure theory for sparse automatic sequences, i.e. those which take non-zero values on a set of integers of zero Banach density. In particular, we show that the set where a sparse automatic sequence takes non-zero values is either extremely small or combinatorially rich (see Theorem V.4.7). Conversely, we can show that sparse generalised polynomials must be free of similar combinatorial structures. As a consequence, we prove the following result.

Theorem V.1.5. *Suppose that a sequence $f\colon \mathbb{N}_0 \to \mathbb{R}$ is automatic and generalised polynomial. Then there exists a periodic function b and a set $Z \subset \mathbb{N}$ such that $f(n) = b(n)$ for $n \in \mathbb{N}_0 \setminus Z$ and*

$$\sup_M |Z \cap [M, M+N)| = O\left(\log^C N\right)$$

as $N \to \infty$ for a certain constant C.

In fact, we obtain a much more precise structural description of the exceptional set Z. See Theorem V.4.5 for details. Similar techniques allow us to show non-automaticity of some sparse generalised polynomials, for which Theorem V.1.2 is vacuously satisfied. For instance, the sequence given by

$$n \mapsto \begin{cases} 1, & \text{if } \left\| \sqrt{2} n \left\lfloor \sqrt{3} n \right\rfloor \right\| < n^{-c}, \\ 0, & \text{otherwise} \end{cases}$$

is not automatic, provided that c is small enough. (Recall that $\|x\|$ denotes the distance from $\mathbb{Z}$.) For details, see Proposition V.4.11.

While Theorem V.1.5 does not resolve Conjecture V.1.3, our proof thereof greatly restricts the number of possible counterexamples. In fact, in order to prove Conjecture V.1.3, it would suffice to prove that the characteristic sequence of powers of an integer $k \geq 2$ given by

$$g_k(n) = \begin{cases} 1, & \text{if } n = k^t \text{ for some } t \geq 0, \\ 0, & \text{otherwise} \end{cases}$$

is not generalized polynomial.

Theorem V.1.6. *Let $k \geq 2$ be an integer. Then one of the following statements holds:*

(i) All sequences which are simultaneously k-automatic and generalised polynomial are ultimately periodic.

(ii) The characteristic sequence g_k of the powers of k is generalised polynomial.

Unfortunately, we are now unable to decide which of the two possibilities in the above theorem holds.

Prompted by Theorem V.1.6, we are also interested in "natural" examples of sparse generalised polynomials. One of the simplest examples is the set of integers n such that $\{n\varphi\} < 1/n$, where $\varphi = (1+\sqrt{5})/2$ is the golden ratio. This set is closely connected to the Fibonacci numbers. It is not difficult to show that sets of the form

$\{n \in \mathbb{N} \mid 0 < \{q(n)\} < n^{-c}\}$ are sparse generalised polynomials if $q(n)$ is a generalised polynomial and c is rational and positive. For sequences of this form, a much simpler proof of Theorem V.1.1 is possible (Proposition V.7.10).

Because of the intimate connection between continued fractions and Diophantine approximation, it is not particularly surprising that for a quadratic irrational $\beta > 1$ of norm ± 1, the set of best integer approximations to the powers of β, $\{\langle\langle \beta^i \rangle\rangle \mid i \in \mathbb{N}_0\}$, is generalised polynomial (see Example V.7.1 and Proposition V.8.4). It is considerably more surprising that a similar result holds also for certain algebraic integers of degree three (see Theorem V.8.6).

Theorem V.1.7. *Assume that $\beta > 1$ is either*

(i) a root of an equation $\beta^2 = a\beta \pm 1$, where a is an integer; or

(ii) the unique real root of an equation $\beta^3 = a\beta^2 + b\beta + 1$, where a, b are integers and either ($a \geq 0$ and $0 \leq b \leq a + 1$) or ($a \geq 2$ and $b = -1$).

Then the set $\{\langle\langle \beta^i \rangle\rangle \mid i \in \mathbb{N}_0\}$ is generalised polynomial.

In light of these examples, we find it pertinent to ask for which $\beta > 1$ is the set $\{\langle\langle \beta^i \rangle\rangle \mid i \in \mathbb{N}_0\}$ generalised polynomial (Question V.9.3). The question is of particular interest if β is an algebraic integer, or, more specifically, a Pisot unit.

The final main result of this chapter says that any sufficiently sparse set is generalised polynomial. For this reason, it seems unlikely that a comprehensive understanding of sparse generalised polynomials is possible. This result is perhaps expected. In [16, Example 3.4.2], the set $S_a = \{n \in \mathbb{N} \mid 0 < \langle\langle an \rangle\rangle < 1/n\}$ is considered for $a \in \mathbb{R}$. The authors note that for the choice of $a = \sum_{n=1}^{\infty} 2^{-(2^n-1)}$, we have $2^{2^n-1} \in S_a$ for $n \in \mathbb{N}$. However, in this construction it is not clear if S_a contains any other elements. We show that the construction can be modified in order to produce the following general result.

Theorem V.1.8. *For any sequence of integers $(n_i)_{i \geq 0}$ such that $n_i \geq 2$ and*

$$\liminf_{i \to \infty} \frac{\log n_{i+1}}{\log n_i} > 1,$$

the set $E = \{n_i \mid i \in \mathbb{N}_0\}$ is generalised polynomial.

This chapter is organised as follows. The relevant definitions and background on automatic sequences are given in Section V.2 (other topics are covered in Chapter I). In Section V.3 we cover the results in the dense setting, and prove Theorems V.1.2 and V.1.4. We then proceed to attack Theorem V.1.5. In Section V.4, this theorem is reduced to a slightly stronger version of Theorem V.1.1 and a structure theorem for sparse automatic sequences. We then proceed to prove these theorems in Sections V.5 and V.6 respectively. Theorem V.1.8 is also obtained obtained with V.6. In Section V.7 we discuss some relevant examples of generalised polynomial sequences and automatic sequences. In Section V.8, we discuss sequences with exponential growth, obtaining among other results Theorem V.1.7. Finally, in Section V.9 we give some concluding remarks.

V.2 Automatic sequences

In this section, we will introduce the basic properties of automatic sequences. For more information, we refer the reader to the canonical book of Allouche and Shallit [4]. This material will also be relevant to Chapter VI.

Whenever A is a (finite) set, we denote[3] by A^* the free monoid with basis A. It consists of finite words in A, including the empty word ϵ, with the operation of concatenation. We denote the concatenation of two words $v, w \in A^*$ by vw and we denote the length of a word $w \in A^*$ by $|w|$. In particular, $|\epsilon| = 0$. We say that a word $v \in A^*$ is a factor of a word $w \in A^*$ if there exist words $u, u' \in A^*$ such that $w = uvu'$. We denote by $w^{\mathrm{R}} \in A^*$ the reversal of the word $w \in A^*$ (the word in which the elements of A are written in the opposite order).

Let $k \geq 2$ be an integer and denote by $\Sigma_k = \{0, 1, \ldots, k-1\}$ the set of digits in base k. For $w \in \Sigma_k^*$, we denote by $[w]_k$ the integer whose expansion in base k is w, i.e. if $w = v_l v_{l-1} \cdots v_1 v_0$, $v_i \in \Sigma_k$, then $[w]_k = \sum_{i=0}^{l} v_i k^i$. Similarly, for an integer $n \geq 0$, we write $(n)_k \in \Sigma_k^*$ for the base k representation of n (without an initial zero). (In particular, $(0)_k = \epsilon$.)

The class of automatic sequences consists, informally speaking, of finite-valued sequences $(a_n)_{n\geq 0}$ whose values a_n are obtained via a finite procedure from the digits of base k expansion of an integer n.

The most famous example of an automatic sequence is arguably the Thue-Morse sequence, first discovered by Prouhet in 1851. Let $s_2(n)$ denote the sum of digits of

[3]This piece of notation is occasionally referred to as the Kleene star.

the base 2 expansion of an integer n. Then the Thue-Morse sequence $(t_n)_{n\geq 0}$ is given by $t_n = 1$ if $s_2(n)$ is odd and $t_n = 0$ if $s_2(n)$ is even.

To formally introduce the notion of automatic sequences, we begin by finite automata.

Definition V.2.1. A *deterministic finite k-automaton with output* (which we will just call a k-automaton) $\mathcal{A} = (S, s_\bullet, \delta, \Omega, \tau)$ consists of the following data:

(i) a finite set of states S;

(ii) an initial state $s_\bullet \in S$;

(iii) a transition map $\delta\colon S \times \Sigma_k \to S$;

(iv) an output set Ω;

(v) an output map $\tau\colon S \to \Omega$.

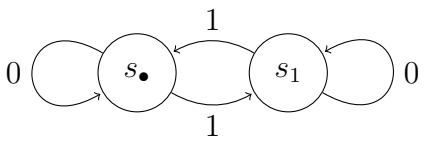
We extend the map δ to a map $\delta\colon S \times \Sigma_k^* \to S$ (denoted by the same letter) by the recurrence formula

$$\delta(s, \epsilon) = s, \quad \delta(s, wv) = \delta(\delta(s, w), v), \quad s \in S, w \in \Sigma_k^*, v \in \Sigma_k.$$

We call a sequence *k-automatic* if it can be produced by a k-automaton in the following manner: one starts at the initial state of the automaton, follows the digits of the base k expansion of an integer n, and then uses the output function to print the n-th term of the sequence. This is stated more precisely in the following definition.

Definition V.2.2. A sequence $(a_n)_{n\geq 0}$ with values in a finite set Ω is *k-automatic* if there exists a k-automaton $\mathcal{A} = (S, s_\bullet, \delta, \Omega, \tau)$ such that $a_n = \tau\left(\delta(s_\bullet, (n)_k)\right)$. We call a set E of nonnegative integers *automatic* if the characteristic sequence (a_n) of E given by $a_n = [\![n \in E]\!]$ is automatic.

The values of the Thue-Morse sequence are given by the 2-automaton



with nodes depicting the states of the automaton, edges describing the transition map and with $\tau(s_\bullet) = 0$ and $\tau(s_1) = 1$. Thus, the Thue-Morse sequence is 2-automatic.

In the definition above, the automaton reads the digits starting with the most significant one. In fact, we might equally well demand that the digits be read starting with the least significant digit or that the automaton produces the correct answer even if the input contains some leading zeroes. Neither of these modifications changes the notion of automatic sequences [4, Theorem 5.2.3] (though of course for most sequences we would need to use a different automaton to produce a given automatic sequence).

There is a number of equivalent definitions of automatic sequences connecting them to different branches of mathematics (stated for example in terms of algebraic power series over finite fields or letter-to-letter projections of fixed points of uniform morphisms of free monoids). We will need one such definition that has a combinatorial flavour and is expressed in terms of the k-kernel.

Definition V.2.3. The *k-kernel* $\mathcal{N}_k((a_n))$ of a sequence $(a_n)_{n\geq 0}$ is the set of its subsequences of the form

$$\mathcal{N}_k((a_n)) = \{(a_{k^l n+r})_{n\geq 0} \mid l \geq 0, 0 \leq r < k^l\}.$$

Automaticity of a sequence is equivalent to the finiteness of its kernel.

Proposition V.2.4. *[4, Theorem 6.6.2] Let $(a_n)_{n\geq 0}$ be a sequence. Then the following conditions are equivalent:*

(i) The sequence (a_n) is k-automatic.

(ii) The k-kernel $\mathcal{N}_k((a_n))$ is finite.

For the Thue-Morse sequence we have relations $t_{2n} = t_n$, $t_{2n+1} = 1 - t_n$, and hence one easily sees that the 2-kernel $N_2((t_n))$ consists of only two sequences $\mathcal{N}_2((t_n)) = \{t_n, 1 - t_n\}$. This gives another argument for the 2-automaticity of the Thue–Morse sequence.

An automatic sequence by definition takes only finitely many values. In 1992 Allouche and Shalit [3], [5] generalised the notion of automatic sequences to a wider class of k-regular sequences that are allowed to take values in a possibly infinite set. The definition of regular sequences is stated in terms of the k-kernel. For simplicity, we state the definition over the ring of integers, though it could also be introduced over a (noetherian) ring.

Definition V.2.5. Let $(a_n)_{n\geq 0}$ be a sequence of integers. We say that the sequence (a_n) is *k-regular* if its k-kernel $\mathcal{N}_k((a_n))$ spans a finitely generated abelian subgroup of $\mathbb{Z}^{\mathbb{N}_0}$.

For example, the following sequences are easily seen to be 2-regular: $(t_n)_{n\geq 0}$, $(n^3+5)_{n\geq 0}$, $(s_2(n))_{n\geq 0}$. (The corresponding subgroups spanned by the 2-kernel have rank 2, 4, and 2. In the case of $t = (t_n)_{n\geq 0}$, the subgroup spanned by the 2-kernel is free abelian with basis consisting of t and the constant sequence $(1)_{n\geq 0}$.) In fact, every k-automatic (integer-valued) sequence is obviously k-regular, and the following converse result holds.

Theorem V.2.6. *[4, Theorem 16.1.5] Let $(a_n)_{n\geq 0}$ be a sequence of integers. Then the following conditions are equivalent:*

(i) The sequence (a_n) is k-automatic.

(ii) The sequence (a_n) is k-regular and takes only finitely many values.

Corollary V.2.7. *[4, Corollary 16.1.6] Let $(a_n)_{n\geq 0}$ be a sequence of integers that is k-regular and let $m \geq 1$ be an integer. Then the sequence $(a_n \bmod m)$ is k-automatic.*

A convenient tool for ruling out that a given sequence is automatic is provided by the Pumping Lemma.

Lemma V.2.8. *[4, Lemma 4.2.1] Let $(a_n)_{n\geq 0}$ be a k-automatic sequence. Then, there exists a constant N such that for any $w \in \Sigma_k^*$ with $|w| \geq N$ and any $L \leq |w| - N$ there exist $u_0, u_1, v \in \Sigma_k^*$, $v \neq \epsilon$, with $w = u_0 v u_1$ such that $L \leq |u_0| \leq L + N - |v|$, and a_n takes the same value for all $n \in \{[u_0 v^t u_1]_k \mid t \in \mathbb{N}_0\}$.*

The final issue that we need to discuss is the dependence of the notion of k-automaticity on the base k. While the Thue-Morse sequence is 2-regular, and is also easily seen to be 4-regular, it is not 3-regular. This follows from the celebrated result of Cobham [28]. We say that two integers $k, l \geq 2$ are *multiplicatively independent* if they are not both powers of the same integer (equivalently, $\log k / \log l \notin \mathbb{Q}$).

Theorem V.2.9. *[4, Theorem 11.2.2] Let $(a_n)_{n\geq 0}$ be a sequence with values in a finite set Ω. Assume that the sequence (a_n) is simultaneously k-automatic and l-automatic with respect to two multiplicatively independent integers $k, l \geq 2$. Then (a_n) is eventually periodic.*

We will have no use for Cobham's Theorem. We will however use the following much easier related result.

Theorem V.2.10. *[4, Theorem 6.6.4] Let $(a_n)_{n\geq 0}$ be a sequences with values in a finite set Ω. Let $k, l \geq 2$ be two multiplicatively* dependent *integers. Then the sequence (a_n) is k-automatic if and only if it is l-automatic.*

V.3 Density 1 results

V.3.1 Polynomial sequences

Our first purpose in this section is to prove Theorem V.1.2. Recall that we aim to show that the sequence $n \mapsto \lfloor p(n) \rfloor$ is not regular, where $p \in \mathbb{R}[x]$ has at least one irrational coefficient other than the constant term. We will show more, namely that the sequence $m \mapsto \lfloor p(n) \rfloor \bmod m$ is not automatic for any $m \geq 2$. In fact, we will only need to work with the weaker property of weak periodicity, defined in the introduction.

Lemma V.3.1. *Any automatic sequence is weakly periodic.*

Proof. Let f be a k-automatic sequence. Since restriction of a k-automatic sequence to an arithmetic progression is again k-automatic ([4, Thm. 6.8.1]), it will suffice to find $q \in \mathbb{N}$, $r, r' \in \mathbb{N}_0$ with $r \neq r'$ such that $f(qn + r) = f(qn + r')$.

The k-kernel $\mathcal{N}_k(f)$ of f, consisting of the functions $f(k^t n + r)$ for $0 \leq r < k^t$, is finite. Pick t sufficiently large that $k^t > |\mathcal{N}_k(f)|$. By the pigeonhole principle, there exist $r \neq r'$ such that $f(k^t n + r) = f(k^t n + r')$. □

Proposition V.3.2 (Polynomial sequences are not weakly periodic)**.** *Let $p(x) \in \mathbb{R}[x]$ be a polynomial, and let an $m \geq 2$ be an integer. Then, the sequence $n \mapsto \lfloor p(n) \rfloor \bmod m$ is weakly periodic if and only if it is periodic. This happens precisely when all non-constant coefficients of $p(x)$ are rational.*

Proof. If all coefficients of $p(x)$ are rational, except possibly the constant term, then the sequence $n \mapsto \lfloor p(n) \rfloor \bmod m$ is easily checked to be periodic, hence weakly periodic.

Suppose now that at least one non-constant coefficient of $p(x)$ is irrational. Replacing $p(x)$ with $p(hx + r)$ for multiplicatively large h if need be, we may assume that the leading coefficient of $p(x)$ is irrational. We will prove marginally more than claimed, namely that for any $0 \leq l < m$, the sequence f given by

$$f(n) = \begin{cases} 1, & \text{if } \lfloor p(n) \rfloor \equiv l \pmod{m}, \\ 0, & \text{otherwise,} \end{cases} \tag{V.1}$$

fails to be weakly periodic. For a proof by contradiction, suppose this claim is false for some choice of l.

It will be convenient to expand $p(x)/m = \sum_{i=0}^{d} a_i \binom{x}{i}$, where $d = \deg p$ and $\binom{x}{i} = x(x-1)(x-2)\dots(x-i+1)/i!$. Note that $a_d \in \mathbb{R} \setminus \mathbb{Q}$ and

$$f(n) = \begin{cases} 1, & \text{if } \left\lfloor \frac{1}{m} p(n) \right\rfloor \bmod 1 \in \left[\frac{l}{m}, \frac{l+1}{m} \right), \\ 0, & \text{otherwise.} \end{cases} \tag{V.2}$$

We will represent the sequence $p(n)$ dynamically. Let $X = \mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, and define the self-map $T\colon X \to X$ by

$$(x_1, x_2, x_3, \dots, x_d) \mapsto (x_1 + a_d, x_2 + x_1 + a_{d-1}, \dots, x_d + x_{d-1} + a_1). \tag{V.3}$$

A standard computation shows that for $z = (0, 0, \dots, 0, a_0)$ we have

$$(T^n z)_j = z_j + \sum_{i \geq 1} a_{d-j+i} \binom{n}{i}, \tag{V.4}$$

and in particular $(T^n z)_d = p(n)/m$. Putting $A = \mathbb{T}^{d-1} \times \left[\frac{l}{m}, \frac{l+1}{m} \right)$ we thus find

$$f(n) = \begin{cases} 1, & \text{if } T^n z \in A, \\ 0, & \text{otherwise.} \end{cases} \tag{V.5}$$

Since f is weakly periodic, we may find q and $r \neq r'$ such that $f(qn + r) = f(qn + r')$. The dynamical system (X, T) is known to be totally minimal (this follows easily from the results e.g. in [35, Section 4.4.3]). In particular, for any point $y \in \operatorname{cl} A$ we may find a sequence $(n_i)_{\geq 1}$ such that $T^{qn_i + r} z \to y$ and $T^{qn_i + r} z \in A$. It follows that the points $T^{qn_i + r'} z$ converge to $T^{r'-r} y$ and lie in A. Thus, $T^{r'-r}(\operatorname{cl} A) \subset \operatorname{cl} A$. In light of total minimality, this is only possible if $\operatorname{cl} A = X$ or $\operatorname{cl} A = \emptyset$; but this is absurd. □

Corollary V.3.3. *With the notation as in Proposition V.3.2, the sequence $n \mapsto \lfloor p(n) \rfloor \bmod m$ is automatic if and only if it is periodic.*

Proof. Immediate from Proposition V.3.2 and Lemma V.3.1. □

Proof of Theorem V.1.2. Suppose first that all non-constant coefficients of $p(n)$ are rational, and fix k. Let $h \in \mathbb{N}$ be such that $hp(n)$ has integer coefficients, possibly except the constant term. Then $f_1(n) = \lfloor hp(n) \rfloor$ is an integer valued polynomial, hence k-regular ($\mathcal{N}_k(f_1)$ is contained in the $\deg p + 1$ dimensional $\mathbb{Z}$-module consisting of $\mathbb{Z}$-valued polynomials of degree $\leq \deg p$). Also, $f_2(n) = \lfloor hp(n) \rfloor - hf(n) = \lfloor h \{p(n)\} \rfloor$ is

periodic, hence k-automatic, hence k-regular. It follows that $f(n) = \frac{1}{h}(f_1(n) - f_2(n))$ is regular.

Conversely, suppose that $f(n)$ is regular. Then, by Theorem V.2.6, for any choice of $m \geq 2$, $f(n) \bmod m$ is automatic. Now, it follows from Corollary V.3.3 that all non-constant coefficients of $p(n)$ are rational. □

V.3.2 Generalised polynomials

Having dealt with the case of polynomial maps, we move on to a more general context. Our next goal is the proof of Theorem V.1.4. We begin with abstracting and generalising some of the key steps from the proof of Theorem V.1.2.

Recall that a set of integers is *thick* if it contains arbitrarily long segments, and *syndetic* if it has bounded gaps; any thick set intersects any syndetic set.

We will need the classical fact that a minimal system (X, T) with X connected is totally minimal. (Otherwise, there exists $d \in \mathbb{N}$ and a closed subset $\emptyset \neq Y \subsetneq X$ such that $T^d(Y) \subset Y$. By the Kuratowski-Zorn Lemma there exists a minimal Y with this property. The sets $Y, T(Y), T^2(Y), \ldots, T^{d-1}(Y)$ are disjoint, closed, and their union is X. This contradicts X being connected.)

Lemma V.3.4 (Totally minimal sequences are not weakly periodic)**.** *Let (X, T) be a totally minimal dynamical system. Let $A \subset X$, be a set with $\operatorname{cl} A = \operatorname{cl}\operatorname{int} A \neq \emptyset, X$ and let $z \in X$. Suppose that $f\colon \mathbb{N}_0 \to \{0, 1\}$ is a sequence such the set of n with $f(n) = [\![T^n z \in A]\!]$ is thick. Then f is not weakly periodic.*

Proof. Suppose for the sake of contradiction that f were weakly periodic. In particular there are some $q \in \mathbb{N}$, $r, r' \in \mathbb{N}_0$ with $r \neq r'$ such that $f(qn + r) = f(qn + r')$. Put $d = r' - r$.

We will show that that $T^d(\operatorname{cl} A) \subset \operatorname{cl} A$. Since T is continuous and $\operatorname{cl}\operatorname{int} A = \operatorname{cl} A$, it will suffice to prove that $T^d(\operatorname{int} A) \subset \operatorname{cl} A$. Once this is accomplished, the contradiction follows immediately, because (X, T^d) is minimal, while $\operatorname{cl} A \neq \emptyset, X$.

Pick any $y \in \operatorname{int} A$ and an open neighbourhood V of $T^d y$; we aim to show that $V \cap A \neq \emptyset$. Put $U = T^{-d}V \cap \operatorname{int} A$, and consider the set S of those n for which $T^{qn+r} z \in U$. Since (X, T^q) is minimal and $U \neq \emptyset$, S is syndetic. Let R_0 be the set of those n for which $f(n) = [\![T^n z \in A]\!]$ and put $R = \{n \in \mathbb{N}_0 \mid qn + r \in R_0\}$ and $R' = \{n \in \mathbb{N}_0 \mid qn + r' \in R_0\}$.

Since R_0 is thick, $R \cap R'$ is also thick, and because S is syndetic, $S \cap R \cap R'$ is non-empty. Pick any $n \in S \cap R \cap R'$, and put $x = T^{qn+r} z$. Since $n \in S$, we have $x \in U \subset A$, and so $T^d x \in V$. Since $n \in R$, we have $f(qn + r) = [\![x \in A]\!] = 1$, and

hence also $f(qn+r')=1$. Finally, since $n \in R'$, we have $1 = f(qn+r') = [\![T^d x \in A]\!]$, meaning that $T^d x \in V \cap A$. In particular, $V \cap A \neq \emptyset$, which was our goal. □

Remark. Some mild topological restrictions on the target set A are, of course, necessary in the above lemma. Note that any open, non-dense and non-empty subset of X will satisfy the stated assumptions.

The assumption that the map T is totally minimal is essential. Indeed, take X to be the Thue-Morse shift, i.e. the closed orbit under the shift map of the Thue-Morse sequence. Let

$$A = \{(a_n)_{n\in\mathbb{N}_0} \mid a_{2k} = a_{2k+1} \text{ for some } k \in \mathbb{N}_0\}$$
$$B = \{(a_n)_{n\in\mathbb{N}_0} \mid a_{2k-1} = a_{2k} \text{ for some } k \in \mathbb{N}_0\}.$$

Since the Thue-Morse sequence (t_n) has the property $t_{2n} \neq t_{2n+1}$ for all n and since the Thue-Morse word contains no cubes (i.e. no occurences of factors of the form www with $w \in \Sigma_k^*$, $w \neq \epsilon$), we see that $A \cap B \neq \emptyset$, $X = A \cup B$ and A and B are clopen. Let $z = (t_n) \in X$ be the Thue-Morse sequence. Then the function $f(n) = [\![T^n z \in A]\!]$ is periodic with period 2.

The analogue of the representation of a polynomial sequence using a skew rotation on the torus in (V.5) is provided by the Bergelson-Leibman Theorem I.1.22. Note that for nilsystems, ergodicity and minimality are equivalent, and the boundary of any semialgebraic subset of a nilmanifold has zero Haar measure. We are now ready to state and prove the main result of this section, from which Theorem V.1.4 easily follows.

Theorem V.3.5. *Let $g\colon \mathbb{Z} \to \mathbb{R}$ be a generalised polynomial taking finitely many values, and let $f\colon \mathbb{N}_0 \to \mathbb{R}$ be a weakly periodic sequence which agrees with g on a thick set $R \subset \mathbb{N}_0$. Then, there exists a set $Z \subset R$ with $d^*(Z) = 0$, such that the common restriction of f and g to $R \setminus Z$ is periodic.*

Proof. It follows from Theorem I.1.22 that we may find a minimal nilsystem (X, T), $z \in X$, a partition $X = \bigcup_{j=1}^r S_j$ with $\operatorname{int} \partial S_j = \emptyset$ (in fact, S_j semialgebraic) for all j, and constants c_j, such that

$$g(n) = \sum_{j=1}^{r} [\![T^n z \in S_j]\!]\, c_j. \tag{V.6}$$

If X is not totally minimal, then (using the fact that nilmanifolds have finitely many connected components) we may find $a \in \mathbb{N}$ such that for any $b \in \mathbb{Z}$, $g'_{a,b}(n) =$

$g(an+b)$ has representation as in (V.6) on a totally minimal nilsystem. Clearly, $f'_{a,b}(n) = f(an+b)$ is weakly periodic and agrees with $g'_{a,b}(n)$ on the thick set $R'_{a,b} = \{n \in \mathbb{N}_0 \mid an+b \in R\}$. Thus, it will suffice to prove the theorem under the additional assumption that (X,T) is totally minimal.

We may write

$$g(n) = \sum_{j=1}^{r} [\![T^n z \in \operatorname{int} S_j]\!]\, c_j + h(n), \tag{V.7}$$

where $h(n) = 0$ unless $T^n z \in \bigcup_{j=1}^r \partial S_j$. In particular (by Corollary I.1.5) the set $Z \subset \mathbb{N}_0$ of n with $h(n) \neq 0$ has upper Banach density 0. Note that $R \setminus Z$ is then thick.

For $1 \leq j \leq r$, put $g'_j(n) = [\![T^n z \in \operatorname{int} S_j]\!]$ and $f'_j(n) = [\![f(n) = c_j]\!]$. Then $g'_j(n) = f'_j(n)$ for $n \in R \setminus Z$. By Lemma V.3.4, this is only possible if for each j, the set $\operatorname{int} S_j$ is either empty or dense. Since $\mu_X(X \setminus \bigcup_{j=1}^r \operatorname{int} S_j) = 0$, there is some i such that $\operatorname{int} S_i$ is dense, and $\operatorname{int} S_j = \emptyset$ for $j \neq i$. Denoting by $Z' \supset Z$ the set of $n \in R$ with $T^n z \in X \setminus \operatorname{int} S_i$ we have $d^*(Z') = 0$ and $f(n) = g(n) = c_i$ for $n \in R \setminus Z'$, as needed. □

Proof of Theorem V.1.4. Direct application of Theorem V.3.5 with $f = g$ and $R = \mathbb{N}_0$ □

It is not a trivial matter to determine whether a given generalised polynomial is periodic away from a set of density 0, although it can be accomplished by the techniques in [16], [85]. In order to give explicit examples, we restrict to generalised polynomials of a specific form, which is somewhat more general than the one considered in Proposition V.3.2.

Corollary V.3.6. *Suppose that $q\colon \mathbb{Z} \to \mathbb{R}$ is a generalised polynomial with the property that $\lambda g(an) \bmod 1$ is equidistributed in $[0,1)$ for any $\lambda \in \mathbb{Q} \setminus \{0\}$ and $a \in \mathbb{N}$, and let $m \geq 2$. Then, the sequence $f(n) = \lfloor q(n) \rfloor \bmod m$ is not automatic.*

Proof. Suppose $f(n)$ were automatic. By Theorem V.1.4, there exist $a \in \mathbb{N}$ and $Z \subset \mathbb{N}_0$ with $d^*(Z) = 0$, such that $f(an)$ is constant for $n \in \mathbb{N}_0 \setminus Z$. Hence, there is some $0 \leq l < m$ such that $\frac{1}{m} q(an) \in \left[\frac{l}{m}, \frac{l+1}{m}\right)$ for $n \in \mathbb{N}_0 \setminus Z$, contradicting the equidistribution assumption. □

The uniform distribution of generalised polynomials has been extensively studied by Håland-Knutson [59],[60],[61], and later a very general theory was developed by Bergelson and Leibman [16], [85]. In view of the results in [59], it is fair to say that

for a "generic" generalised polynomial $q(n)$ is equidistributed modulo 1. Hence, the assumptions on $q(n)$ in Corollary V.3.6 are relatively mild.

To make the last remark precise, let us define the (multi-)set of coefficients of a generalised polynomial q as follows. If $q(n) = \sum_j \alpha_j n^j$ is a polynomial, then the coefficients of $q(n)$ are the non-zero terms among the α_j. If $q(n) = r_1(n) + r_2(n)$ or $q(n) = r_1(n) \cdot r_2(n)$ where r_1, r_2 are generalised polynomials, then the coefficients of $q(n)$ are the union of the coefficients of $r_1(n)$ and $r_2(n)$. Finally, if $q(n) = p(n) \lfloor r(n) \rfloor^d$, then the coefficients of $q(n)$ are the union of the coefficients of $r(n)$ and the coefficients of $p(n)$. The set of the coefficients will depend on the choice of representation of the generalised polynomial at hand; we fix one such choice. We cite a slightly simplified version of the main theorem of [59].

Theorem V.3.7. *Suppose that $q(n)$ is a generalised polynomial, and all of the products of the coefficients of $q(n)$ are $\mathbb{Q}$-linearly independent. Then $q(n)$ is equidistributed modulo 1.*

As an example of an application, we conclude that $\lfloor \sqrt{2} n \lfloor \sqrt{3} n \rfloor \rfloor \bmod 10$ is not an automatic sequence (because $1, \sqrt{2}, \sqrt{3}, \sqrt{6}$ are linearly independent over $\mathbb{Q}$).

V.4 Sparse sets

In this section, we begin the investigation of sparse sequences. Recall that we call a sequence $f \colon \mathbb{N}_0 \to \{0, 1\}$ *sparse* if it is the characteristic function of a set with upper density 0. Note that for such sparse sequences, Theorem V.1.4 conveys no useful information. Conversely, in light of Theorem V.1.4, to prove Conjecture V.1.3, it would suffice to verify it for sparse sequences; this observation will be made precise in the Proof of Theorem V.1.5 below.

V.4.1 Arid sets

To formulate the main result of this section, from which Theorem V.1.5 easily follows, it is convenient to introduce a piece of terminology.

Definition V.4.1 (Arid sets)**.** Let $k \geq 2$, $r \geq 0$. A set $A \subset \Sigma_k^*$ is *arid* (of rank $\leq r$) if it is a finite union of *basic arid sets*

$$A = \left\{ v_0 w_1^{l_1} v_1 w_2^{l_2} \dots w_r^{l_r} v_r \;\middle|\; l_1, \dots, l_r \in \mathbb{N}_0 \right\}, \tag{V.8}$$

where $v_0, \dots, v_r \in \Sigma_k^*$, and $w_1, \dots, w_r \in \Sigma_k^*$. Accordingly, an arid set $E \subset \mathbb{N}_0$ (of rank $\leq r$) base k is a set of the form $\{[u]_k \mid u \in A\}$ where $A \subset \Sigma_k^*$ is arid (of rank $\leq r$). A sequence $f\colon \mathbb{N}_0 \to \{0,1\}$ is arid if the set $\{n \in \mathbb{N}_0 \mid f(n) = 1\}$ is arid.

Alternatively, one can write the basic k-arid set A given by (V.8) as

$$A = v_0\{w_1\}^* v_1 \{w_2\}^* \dots \{w_r\}^* v_r,$$

where, as before, $\{w\}^* = \left\{w^l \;\middle|\; l \in \mathbb{N}_0\right\}$.

Lemma V.4.2. *Any k-arid set is k-automatic.*

Proof. It is clear that any k-arid set is given by a regular expression (i.e. one built up using concatenation, set theoretic operation, and the Kleene star). Hence, it is k-automatic by Kleene's theorem [4, Thm. 4.1.5]. Alternatively, in this simple case one can construct the required automata by hand. □

Remark. Let $a \geq 1$ be an integer. Then the notions of k-arid (of rank $\leq r$) and k^a-arid (of rank $\leq r$) sets coincide. Indeed, basic k^a-arid sets are obviously basic k-arid sets, and k-arid can be written as finite unions of basic k^a-arid sets. We will use this observation several times.

Lemma V.4.3. *Let $E \subset \mathbb{N}_0$ be arid of rank (exactly) r. Then*

$$|E \cap [N]| = \Theta(\log^r N).$$

Moreover, we also have the estimate

$$\max_{M \in \mathbb{N}_0} |E \cap [M, M+N)| = \Theta(\log^r N).$$

Proof. It will suffice to deal with the basic arid sets, given by

$$E = \left\{[v_0 w_1^{l_1} v_1 w_2^{l_2} \dots w_r^{l_r} v_r]_k \;\middle|\; l_1, \dots, l_r \in \mathbb{N}_0\right\}. \tag{V.9}$$

We begin with some standard reductions. Replacing w_i with suitably chosen powers (and altering v_i accordingly) we may assume that all w_i have the same length a. Replacing k with k^a, by the preceding remark, we may assume that $|w_i| = 1$ for each i. Because r is minimal, if $w_i = w_{i+1}$ for some i then v_i is not a power of w_i. Finally, we may assume (in both parts of the statement) that $N = k^L$ is a large power of k, and (in the latter part) that $M = k^L M'$ is divisible by N.

The upper bound $|E \cap [N]| \ll L^r$ follows immediately from counting the r-tuples $(l_1, \dots, l_r)$ with $\sum_{i=1}^r l_i + \sum_{i=0}^r |v_i| \leq L$. For the lower bound $|E \cap [N]| \gg L^r$, it

suffices to observe that all $v_0 w_1^{l_1} v_1 w_2^{l_2} \dots w_r^{l_r} v_r$ for $l_1, \dots, l_r \in \mathbb{N}$ are distinct (the corresponding terms where some l_i is 0 constitute a arid set of rank $r-1$). Indeed, it is easy to check that if $v_0 w_1^{l_1} v_1 w_2^{l_2} \dots w_r^{l_r} v_r = v_0 w_1^{l'_1} v_1 w_2^{l'_2} \dots w_r^{l'_r} v_r$ and i is the least index with $l_i \neq l'_i$ then v_i is both a power of w_i and w_{i+1}, contrary to previously made assumption.

Finally, to obtain the bound $|E \cap [M, M+N)| \ll L^r$, note that an integer $n = [v_0 w_1^{l_1} v_1 w_2^{l_2} \dots w_r^{l_r} v_r]_k \in E \cap [M, M+N)$ is now uniquely determined by its terminal L digits, which may be chosen in $\ll L^r$ ways. □

For future reference, we also record several facts concerning generalised polynomial sets. The class of generalised polynomial sets is clearly closed under Boolean operations. To see this, note that if we denote the characteristic functions of sets E_1 and E_2 by f_1 and f_2, then the characteristic function of $E_1 \cup E_2$ (resp. $E_1 \cap E_2$, $\mathbb{Z} \setminus E_1$) is $f_1 f_2$ (resp. $f_1 + f_2 - f_1 f_2$, $1 - f_1$) and hence is generalised polynomial if so are f_1 and f_2. Furthermore, the set of values at which a generalised polynomial vanishes is a generalised polynomial set. More precisely, we can state the following lemma.

Lemma V.4.4. *If h is a generalised polynomial, then so is the function g given by the formula*

$$g(n) = \begin{cases} 1, & \text{if } h(n) = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Likewise, for any $a < b$,

$$g'(n) = \begin{cases} 1, & \text{if } a \leq h(n) < b, \\ 0, & \text{otherwise,} \end{cases}$$

is a generalised polynomial.

Proof. Because $h(n)$ takes countably many values for $n \in \mathbb{N}_0$, there exists $\theta \in \mathbb{R}$ such that $\theta h(n) \in \mathbb{R} \setminus \mathbb{Q}$ unless $h(n) = 0$. Now, a short computation verifies that

$$g(n) = \lfloor 1 - \{\theta h(n)\} \rfloor .$$

For the latter claim, after rescaling we may assume that $a = 0$ and $b = 1$. Then $g'(n) = [\![\lfloor h(n) \rfloor = 0]\!]$, and we may apply the construction above to $h'(n) = \lfloor h(n) \rfloor$. □

An immediate corollary of the lemma is that finite sets, being the zero sets of polynomials, are generalised polynomial. Finally, if E is a generalised polynomial set, then so is the set of its multiples $kE = \{kn \mid n \in E\}$ for $k \in \mathbb{Z}$.

We are now ready to formulate the main result of this section.

Theorem V.4.5. *Suppose that a sparse sequence $f\colon \mathbb{N}_0 \to \{0,1\}$ is simultaneously k-automatic and generalised polynomial. Then, f is arid.*

Proof of Theorem V.1.5, assuming V.4.5. Suppose that $f\colon \mathbb{N}_0 \to \mathbb{R}$ is automatic and generalised polynomial. Let $p(n)$ be the periodic function such that the set $Z = \{n \in \mathbb{N}_0 \mid f(n) = p(n)\}$ has $d^*(Z) = 0$, whose existence is guaranteed by Theorem V.1.4. Denote

$$f'(n) = \begin{cases} 1, & \text{if } f(n) = p(n), \\ 0, & \text{otherwise.} \end{cases}$$

Then, by Lemma V.4.3, $f'(n)$ is generalised polynomial and automatic. By Theorem V.4.5, the sequence $f'(n)$ is arid, so we have $|Z \cap [M, M+N)| = \sum_{n=M}^{M+N} f'(n) = O\left(\log^r N\right)$ for some $r \in \mathbb{N}_0$ as $N \to \infty$. □

V.4.2 Proof strategy

We now approach the proof of Theorem V.4.5. The first ingredient is a structure theorem for sparse automatic sequences. It is convenient to introduce some additional terminology, namely a slight refinement of the notion of an IP set introduced in Section I.1.5, allowing for a variable sequence of shifts.

Definition V.4.6 (IPS set)**.** For a sequence $(n_i)_{i\in\mathbb{N}} \subset \mathbb{N}$ and shifts $(N_t)_{t\geq 1} \subset \mathbb{N}_0$, the corresponding set of *shifted finite sums* is

$$\mathrm{FS}(n_i; N_t) = \{n_\alpha + N_t \mid t \in \mathbb{N},\ \alpha \in \mathcal{F},\ \max\alpha \leq t\}, \tag{V.10}$$

where $n_\alpha = \sum_{i\in\alpha} n_i$. Any set containing $\mathrm{FS}(n_i; N_t)$ for some (n_i), (N_t) is called an IPS *set*.

Remark. If $N_t = 0$ (resp. $N_t = \text{const.}$) then the set $\mathrm{FS}(n_i; N_t)$ in (V.10) is an IP set (resp. IP_+ set). In general, an IPS set need not be an IP_+ set.

Remark. An equivalent (though apparently weaker) definition of an IPS set can be phrased as follows: A set $E \subset \mathbb{N}$ is IPS if there exist sequences $(n_i)_{i\in\mathbb{N}} \subset \mathbb{N}$, $(N_t)_{t\geq 1} \subset \mathbb{N}_0$ and $(t_0(\alpha))_{\alpha\in\mathcal{F}} \subset \mathbb{N}$ such that $n_\alpha + N_t \in E$ for any $\alpha \in \mathcal{F}$ and $t \geq t_0(\alpha)$.

Remark. IPS sets are very natural to consider from the point of view of ultrafilters.

It is well known that a set $E \subset \mathbb{N}$ is an IP set if and only if it is belongs to an ultrafilter $p \in \beta\mathbb{N}$ that is idempotent (i.e. $p + p = p$). One can immediately conclude that a set is an IP_+ set if and only if it belongs to an ultrafilter of the form $n+p = p+n$ for some $n \in \mathbb{N}$ and some idempotent $p \in \beta\mathbb{N}$. One can show that a set is an IPS set

if and only if it belongs to an ultrafilter of the form $r+p$ for some $r \in \beta\mathbb{N}$. The proofs are quite immediate. (For the latter statement, note that $E \subset \mathbb{N}$ is an IPS set if and only if there is a sequence $(n_i)_{i\in\mathbb{N}} \subset \mathbb{N}$ such that the family $\{(E-n_\alpha)\cap\mathbb{N}\}_{\alpha\in\mathcal{F}}$ has the finite intersection property. We choose $p, r \in \beta\mathbb{N}$ such that $p+p=p$, $\mathrm{FS}(n_i) \in p$, and $(E-n_\alpha)\cap\mathbb{N} \in r$.)

A set of the form $p+\beta\mathbb{N}$ (resp. $\beta\mathbb{N}+p$ or $\beta\mathbb{N}+p+\beta\mathbb{N}$) is called a right (resp. left or two-sided) ideal generated by p. Thus a set is an IP_+ set if and only if it belongs to an ultrafilter lying in the right ideal generated by an idempotent ultrafilter and a set is an IPS set if and only if it belongs to an ultrafilter lying in the left ideal (or equivalently the two-sided ideal) generated by an idempotent ultrafilter.

For more details about $\beta\mathbb{N}$, see Appendix B. We will not use the ultrafilter interpretation of IPS sets in what follows.

We are now ready to phrase the structure theorem alluded to before, which is interesting in its own right. The proof is given in Section V.6.

Theorem V.4.7. *Let $f\colon \mathbb{N}_0 \to \{0,1\}$ be a sparse automatic sequence. Then, either f is arid, or there exists an* IPS *set E such that $f(n)=1$ for all $n \in E$.*

In light of Theorem V.4.7, combined with Theorem I.1.22, the relevance of the behaviour of nilsequences along IPS sets becomes clear. The following theorem is the main result of Section V.5.

Theorem V.4.8. *Let (X,T) a minimal nilsystem, and let $x \in X$. Suppose that $S \subset X$ is semialgebraic, and that the set $\{n \in \mathbb{N}_0 \mid T^n x \in S\}$ is* IPS*. Then* $\operatorname{int} S \neq \emptyset$.

Proof of Theorem V.4.5 assuming V.4.7 and V.4.8. Let $f\colon \mathbb{N}_0 \to \{0,1\}$ be sparse, automatic and generalised polynomial. By Theorem I.1.22, there exists a minimal nilsystem (X,T) and a semialgebraic set $S \subset X$, such that $f(n) = [\![T^n x \in S]\!]$.

Suppose for the sake of contradiction that f is not arid. Then, by Theorem V.4.7, the set $\{n \in \mathbb{N}_0 \mid T^n x \in S\}$ is IPS. Hence, by Theorem V.4.8, $\operatorname{int} S \neq \emptyset$, and in particular $\mu_X(S) > 0$. On the other hand, by the Ergodic Theorem I.1.4, we have $\mu_X(S) = d(\{n \in \mathbb{N}_0 \mid f(n)=1\}) = 0$, which is a contradiction. □

V.4.3 Comments and applications

Any IPS set contains a translate of an IP_r set (i.e. the set of all finite sums of a finite sequence $(n_i)_{i=1}^r$) for any r. It is therefore natural to ask if Theorem V.4.8 can be strengthened by replacing IPS sets with IP_r sets. The answer is negative, as shown by the following example.

Example V.4.9. Let $E = \left\{n \in \mathbb{N}_0 \mid \left\|n\sqrt{2}\right\| < 1/\sqrt{n}\right\}$. Then E is a sparse generalised polynomial set and E contains sets of the form $\{km \mid 1 \leq k \leq r\}$ for arbitrary r. In particular, E contains an IP_r set.

Proof. By standard Diophantine approximation, there exist arbitrarily large m such that $\left\|m\sqrt{2}\right\| < 1/m$. Then $km \in E$ for $k \leq m^{1/3}$. The fact that E is sparse and given by a generalised polynomial follows from Proposition V.7.2. □

Remark. A weaker version of Theorem V.1.1 follows immediately by an adaptation of the methods in [16]. Indeed, adapting the proof of Theorem C therein, one concludes that if E is a sparse generalised polynomials set, then for all n except for a set of (upper Banach) density 0 the set $\mathbb{N} \setminus (E - n)$ is IP^*, whence $E - n$ contains no IP set. Unfortunately, this weaker statement is not as useful in applications; in particular is not sufficient for our purposes

In Theorem V.4.7, it is also not possible to replace IPS sets with IP sets, as shown by Example V.7.8. However, under slightly stronger assumptions we obtain the following variant.

Proposition V.4.10. *Let $E \subset \mathbb{N}_0$ be a k-automatic set. Assume that for every $w \in \Sigma_k^*$ there is an integer $n \in E$ such that w is a factor of $(n)_k$. Then E is an IP_+ set.*

This has the following amusing application, whose proof we include in Section V.7. Similar results can be shown in greater generality.

Proposition V.4.11. *There exists a constant $c > 0$ such that the following holds. Let $1, \alpha, \beta$ be algebraic numbers linearly independent over $\mathbb{Q}$ and let $\varepsilon(n) \geq 0$ be a sequence with $n^{-c} \ll \varepsilon(n)$ and $\varepsilon(n) \to 0$ as $n \to \infty$. Then the set*

$$E = \{n \in \mathbb{N}_0 \mid \|\alpha n \lfloor \beta n \rfloor\| < \varepsilon(n)\}$$

is not automatic.

V.5 Sparse generalised polynomials

In this section, we prove Theorem V.4.8 . In the process, we obtain results concerning the distribution of fractional parts of polynomial sequences, such as Propositions V.5.7 and V.5.8, which are possibly of independent interest. Throughout, G is a nilpotent Lie group, $G_\bullet$ denotes the lower central series, G° denotes the connected component

of indetity, and $G_i^\circ := G_i \cap G^\circ$ (which is a filtration on G°, not necessarily equal to the lower central series in general).

It will be convenient to reformulate Theorem V.4.8 . By Theorem I.1.22, a sparse generalised polynomial set E arises as $u(n) = p(g^n a)$ for a nilmanifold $X = G/\Gamma$, $g \in G$ acting on X ergodically, a piecewise polynomial map $p\colon X \to \mathbb{R}$, and $a \in X$. Let $S = \{x \in X \mid p(x) = 1\}$. We do not assume that G is connected, however we may and will assume that the connected component of identity G° of G is *simply connected*. Since E is sparse, Corollary I.1.5 shows that $\operatorname{int} S = \emptyset$. (As a side remark, we note that it is also easy to see that a sparse genereralised polynomial set has upper Banach density zero.) Thus Theorem V.4.8 is equivalent to the following statement.

Theorem V.5.1. *Let $X = G/\Gamma$ be a nilmanifold, let $g \in G$ be such that the left translation by g is ergodic, $a \in G$, and let $S \subset X$ be a semialgebraic subset with* $\operatorname{int} S = \emptyset$. *Then there does not exist an* IPS *set E such that*

$$g^n a\Gamma \in S, \qquad n \in E. \tag{V.11}$$

The proof of Theorem V.5.1 will occupy most of this section.

V.5.1 Preliminaries

In dealing with IPS sets, the following lemma will be useful.

Lemma V.5.2 (Partition regularity for IPS sets)**.** *Suppose that $E \subset \mathbb{N}$ is an* IPS *set. Then for any finite partition $E = E_1 \cup E_2 \cup \ldots \cup E_r$ one of the sets E_j is an* IPS *set.*

Proof. This is clear from the ultrafilter interpretation of IPS sets. We give a direct proof below. Suppose that $(n_i)_{i\in\mathbb{N}}$, $(N_t)_{t\geq 1}$ and $(t_0(\alpha))_{\alpha\in\mathcal{F}}$ are such that $n_\alpha + N_t \in E$ for $\alpha \in \mathcal{F}$ and $t \geq t_0(\alpha)$. Restricting N_t to subsequence by means of a familiar diagonal argument, we may ensure that for any α there is some $j = j(\alpha)$ such that $n_\alpha + N_t \in E_j$ for all $t \geq t_0(\alpha)$. The claim now follows directly from Hindman's theorem. □

The following lemmas are classical.

Lemma V.5.3. *Let G be a nilpotent group (not necessarily Lie), and let $H < G$ be a subgroup of G such that $HG_2 = G$. Then $H = G$.*

For the proof, see e.g. [84, Lemma 2.5], [82, Lemma 3.4], or [110, Lemma 1.11]. The following lemma allows, under reasonably weak assumptions, us to restrict our attention to polynomials with coefficients in the connected component of identity G°. We point out that there are several other ways to navigate this technical point; while this is possibly not the most elegant solution, it is a relatively simple one.

Lemma V.5.4. *Let G/Γ be a connected nilmanifold, and assume further that all the nilmanifolds G_i/Γ_i are connected, where $\Gamma_i = \Gamma \cap G_i$. Let $g \in \mathrm{poly}(\mathbb{Z} \to G_\bullet)$ be a polynomial sequence. Then there exists a polynomial sequence $h \in \mathrm{poly}(\mathbb{Z} \to G_i^\circ)$, such that $g(n)\Gamma = h(n)\Gamma$.*

This follows from [110, Lemma 4.24].

Remark. We do not claim that h is a polynomial sequence with respect to the lower central series on G°, since possibly $(G^\circ)_i \subsetneq (G_i)^\circ$.

We will also need the following basic fact about invariant algebraic sets.

Lemma V.5.5. *Suppose that $A \subset \mathbb{R}^d$ is an algebraic set and $P\colon \mathbb{R}^d \to \mathbb{R}^d$ is a surjective polynomial map. If $P^{-1}(A) \subset A$, then also $P(A) = A$.*

Proof. Consider the descending sequence of nonempty algebraic sets

$$A \supset P^{-1}(A) \supset P^{-2}(A) \supset P^{-3}(A) \supset \dots \tag{V.12}$$

By Hilbert's Basis Theorem, there is n such that $P^{-n}(A) = P^{-n-1}(A)$. Since P is surjective, we get $P(A) = A$. □

V.5.2 Initial reductions

Our proof of Theorem V.5.1 is indirect: we exploit the existence of an orbit satisfying (V.11) to construct algebraically invariant objects whose existence is inreasingly difficult to sustain, up to the point when we obtain a blatant contradiction. We begin with a relatively straightforward reduction.

Step 3. Assume Theorem V.5.1 is false. Then there exists a nilmanifold G/Γ such that every G_i/Γ_i is connected, where $\Gamma_i = \Gamma \cap G_i$, an element $g \in G$ acting ergodically by left translations, a semialgebraic subset $S \subset G/\Gamma$ with empty interior, and an IP set $E \subset \mathbb{N}$ such that

$$g^n\Gamma \in S, \qquad n \in E. \tag{V.13}$$

Proof. Suppose that Theorem V.5.1 fails, so that (V.11) holds. By [110, Lemma 4.5], there exists a discrete cocompact subgroup $\Gamma < \tilde{\Gamma} < G$ such $[\tilde{\Gamma} : \Gamma] < \infty$, and such that $G_i/\tilde{\Gamma}_i$ is connected for every i, where $\tilde{\Gamma}_i := \tilde{\Gamma} \cap G_i$. We replace Γ by $\tilde{\Gamma}$ and S by its image under the map $G/\Gamma \to G/\tilde{\Gamma}$ (this is still a semialgebraic set with empty interior).

Let $(n_i)_{i\in\mathbb{N}}$, $(N_t)_{t\geq 1}$ and $(t_0(\alpha))_{\alpha\in\mathcal{F}}$ be such that $n_\alpha + N_t \in E$ for $t \geq t_0(\alpha)$, so that

$$g^{n_\alpha+N_t} a\Gamma \in S.$$

Replacing S by its closure, which is again a semialgebraic set with empty interior, we may assume that S is closed. Restricting N_t to a subsequence if necessary, we may assume that $g^{N_t}a\Gamma \to b\Gamma$ as $t \to \infty$ for some $b \in G$. It follows that for any IP set $E' \subset \mathrm{FS}(n_i)$ we have

$$g^n b\Gamma = \lim_{t\to\infty} g^{n+N_t} a\Gamma \in S, \qquad n \in E'.$$

Finally, note that the map $g \mapsto gb^{-1}$ induces an isomorphism $G/\Gamma \to G/\Gamma'$, where $\Gamma' = b\Gamma b^{-1}$. Let $S' \subset G/\Gamma'$ be the image of S under this map. We find that $g^n\Gamma' \in S'$ for $n \in E'$, as needed. □

V.5.3 Fractional parts and limits

Let G/Γ be a nilmanifold with a specified Mal'cev basis. A technical difficulty in dealing with fractional parts in G stems from the fact that, for a sequence $g_n \in G$, the limit of $g_n\Gamma$ does not determine the limit of $\{g_n\}$ if the limit lies on the boundary of the fundamental domain. The following lemma partially overcomes this difficulty.

Lemma V.5.6. *Let $X = G/\Gamma$ be a connected nilmanifold. Then for any sequence $(g_n)_{n\geq 0}$ in G such that $\lim_{n\to\infty} g_n\Gamma = e\Gamma$, there exists a choice of a Mal'cev basis $\mathcal{X}$ of X and a subsequence (g'_n) of (g_n) such that $\lim_{n\to\infty}\{g'_n\} = e$.*

Likewise, for any $(g_\alpha)_{\alpha\in\mathcal{F}}$ in G such that $\mathrm{IP}{-}\lim_{a\in\mathcal{F}} g_\alpha\Gamma = e\Gamma$, there exists a choice of a Mal'cev basis $\mathcal{X}$ and an IP ring $\mathcal{G}$ such that $\mathrm{IP}{-}\lim_{\alpha\in\mathcal{G}}\{g_\alpha\} = e$.

Proof. Since G/Γ is connected, we have an isomorphism $G/\Gamma \simeq G^\circ/(\Gamma \cap G^\circ)$, and hence we may assume that G is connected. Recall that a choice of a Mal'cev basis $\mathcal{X} = (e_1, \dots, e_k)$ determines the map $\tilde{\tau} = (\tilde{\tau}_1, \dots, \tilde{\tau}_k)\colon G \to \mathbb{R}^k$. By induction, we will prove that for each $1 \leq r \leq k+1$ there exists a Mal'cev basis $\mathcal{X}$ such that, after passing to a subsequence (g'_n) of (g_n), we have

$$\lim_{n\to\infty}\{g'_n\} = h$$

for $h \in \operatorname{cl} D = \tilde{\tau}^{-1}([0,1]^k)$ such that $\tilde{\tau}_i(h) = 0$ for $1 \leq i < r$. For $r = 1$, the claim is trivially satisfied, and the claim for $r = k+1$ implies the claim in the statement of the lemma.

If the claim holds for r, then

$$g'_n = e_r^{t_r(n)} e_{r+1}^{t_{r+1}(n)} \dots e_k^{t_k(n)} \gamma(n),$$

where $t_i(n) \in [0,1)$, $\gamma(n) \in \Gamma$, and $t_i(n) \to \tilde{\tau}_i(h)$ as $i \to \infty$. Since $h \in \Gamma$, we have $\tilde{\tau}_i(h) \in \{0,1\}$. If $\tilde{\tau}_r(h) = 0$, we are done. Otherwise,

$$\begin{aligned} g'_n &= (e_r^{-1})^{1-t_r(n)} e_{r+1}^{t_{r+1}(n)} \dots e_k^{t_k(n)} \gamma(n) \\ &= (e_r^{-1})^{1-t_r(n)} e_{r+1}^{t'_{r+1}(n)} \dots e_k^{t'_k(n)} \gamma'(n), \end{aligned}$$

where $t'_i(n) \in [0,1)$ and $\gamma'(n) \in \Gamma$. Replacing $\mathcal{X}$ with $\mathcal{X}' = (e'_1, \dots, e'_k)$ with $e'_r = e_r^{-1}$ and $e'_i = e_i$ for $i \neq r$, we find the sought Mal'cev basis for $r+1$.

The proof of the claim concerning IP limits is completely analogous. □

We are now ready to perform the first substantial step in the proof of Theorem V.5.1. We construct an algebraic set inside G with surprising invariance properties. This is helpful largely because working in G is easier than in G/Γ.

Step 4. Suppose that Theorem V.5.1 is false. Then there exists a nilmanifold G/Γ with a Mal'cev basis $\mathcal{X}$ such that every G_i/Γ_i is connected, where $\Gamma_i = \Gamma \cap G_i$, an element $g \in G$ acting ergodically by left translations, a nonempty algebraic subset $R \subset G^\circ$ with empty interior, and an IP set $E = \operatorname{FS}(n_i)$ such that R is preserved under the operations

$$P_\alpha \colon G^\circ \to G^\circ, \qquad x \mapsto g^{n_\alpha} x \, [g^{n_\alpha}]^{-1} \tag{V.14}$$

for all $\alpha \in \mathcal{F}$.

Proof. Let the notation be as in Step 3. We will subsequently pass to IP subsets of $E = \operatorname{FS}(n_i)$ to ensure better properties.

After replacing E with a smaller IP set, we may assume that $\operatorname{IP-lim}_{\alpha \in \mathcal{F}} g^{n_\alpha} \Gamma = e\Gamma$. (Indeed, we pass to an IP subring on which the sequence is convergent and apply the IP Recurrence Theorem I.1.13 to see that it converges to $e\Gamma$.) Hence, by Lemma V.5.6 there exists a choice of a Mal'cev basis such that, possibly after shrinking E again, we have

$$\operatorname{IP-lim}_{\alpha \in \mathcal{F}} \{g^{n_\alpha}\} = e. \tag{V.15}$$

For a fixed β and sufficiently large $\alpha > \beta$, we aim to compute $\{g^{n_{\alpha\cup\beta}}\}$ in terms of g^{n_β} and $\{g^{n_\alpha}\}$. We begin by noting that

$$\{g^{n_{\alpha\cup\beta}}\} = \{g^{n_\beta}\,\{g^{n_\alpha}\}\} = g^{n_\beta}\,\{g^{n_\alpha}\}\,[g^{n_\beta}]^{-1}\,\gamma(\alpha,\beta), \tag{V.16}$$

where $\gamma(\alpha,\beta) = [g^{n_\beta}]\,[g^{n_\beta}\,\{g^{n_\alpha}\}]^{-1} \in \Gamma$ takes, for a fixed β, only finitely many possible values. For a fixed β, we may use Hindman's theorem to pass to an IP subring so that $\gamma(\alpha,\beta)$ does not depend on the choice of $\alpha > \beta$. Using a standard diagonal argument, we replace E by an IP subset so that $\gamma(\alpha,\beta) = \gamma(\beta)$ does not depend on $\alpha > \beta$ for all $\beta \in \mathcal{F}$. Letting $\alpha \to \infty$ in (V.16), we get

$$\mathrm{IP}\text{-}\lim_{\alpha\in\mathcal{F}} \{g^{n_{\alpha\cup\beta}}\} = \{g^{n_\beta}\}\,\gamma(\beta).$$

If β is sufficiently large, this implies that $\gamma(\beta) = e$. Passing to an IP subset of E, we may assume that $\gamma(\beta) = e$ for all $\beta \in \mathcal{F}$. Hence, applying (V.16) again, we have

$$\{g^{n_{\alpha\cup\beta}}\} = g^{n_\beta}\,\{g^{n_\alpha}\}\,[g^{n_\beta}]^{-1} = P_\beta(\{g^{n_\alpha}\}) \tag{V.17}$$

for any $\alpha, \beta \in \mathcal{F}$ with $\beta < \alpha$.

We claim that $P_\beta \circ P_\alpha = P_{\alpha\cup\beta}$ for all $\alpha > \beta$. Indeed, both maps take the form $x \mapsto g^{n_{\alpha\cup\beta}} x \gamma$ for some $\gamma \in \Gamma$ and by (V.17) agree on points $\{g^{n_\lambda}\}$ for $\lambda > \alpha$. Therefore, the two maps are equal. We will use this fact below.

Denote by $\widetilde{S} \subset G^\circ$ the Zariski closure of the copy of S contained in the fundamental domain for $\mathcal{X}$. Then $\tilde{S}$ is a algebraic subset of G° with empty interior. For any $\beta \in \mathcal{F}$, let R_β denote the set of $x \in \tilde{S}$ such that $P_\beta(x) \in \tilde{S}$. Clearly, these sets are algebraic, and by (V.17) we have $\{g^{n_\alpha}\} \in R_\beta$ for $\alpha > \beta$.

Let $R = \bigcap_{\alpha\in\mathcal{F}} R_\alpha$. By Hilbert's Basis Theorem, there is a finite collection $\mathcal{B} \subset \mathcal{F}$ such that $R = \bigcap_{\beta\in B} R_\beta$. Note that R is nonempty as $\{g^{n_\alpha}\} \in R$ for α such that $\alpha > \beta$ for $\beta \in B$. It remains to show that R is preserved under the maps P_α. Suppose that $x \in R$ and $\alpha > \beta$ for all $\beta \in B$. We will show that $P_\alpha(x) \in R$. To this end, it suffices to check that $P_\alpha(x) \in R_\beta$ for $\beta \in B$, which amounts to $P_\alpha(x) \in \tilde{S}$ and $P_\beta(P_\alpha(x)) \in \tilde{S}$. The first condition follows immediately from $x \in R_\alpha$. Since $P_\beta \circ P_\alpha = P_{\alpha\cup\beta}$, the latter follows from $x \in R_{\alpha\cup\beta}$. Thus $P_\alpha(R) \subset R$ for α large enough. Passing yet again to an IP subset, we ensure that the claim holds for all α. □

V.5.4 Fractional parts of polynomials

Our next aim is to show that the set R constructed in Step 4 needs to be invariant under a wider range of operations. For this purpose, we study the possible algebraic relations between polynomials and their fractional parts.

Proposition V.5.7. *Let G/Γ be a nilmanifold equipped with a Mal'cev basis $\mathcal{X}$. Let $g, h \in \mathrm{poly}(\mathbb{Z} \to G^\circ_\bullet)$. Suppose that E is an* IP *set and $P\colon G^\circ \times G^\circ \to \mathbb{R}$ is a polynomial map such that*

$$P\left(g(n), \{h(n)\}\right) = 0, \qquad n \in E. \tag{V.18}$$

Then there exists an IP *set $E' \subset E$ such that*

$$P\left(g(m), \{h(n)\}\right) = 0, \qquad m \in \mathbb{Z},\ n \in E'. \tag{V.19}$$

Remark. If g is extended to a polynomial mapping $g\colon \mathbb{R} \to G^\circ$, then the same argument gives $P\left(g(m), \{h(n)\}\right) = 0$ for $m \in \mathbb{R}$, $n \in E'$.

Proof. We may assume that g is defined on $\mathbb{R}$. Consider the map $Q\colon \mathbb{R} \times G^\circ \to \mathbb{R}$ given by $Q(m, y) = P(g(m), y)$. This is a polynomial map. Expand Q as $Q(x, y) = \sum_{k=0}^{d} x^k Q_k(y)$ for polynomials $Q_k\colon G^\circ \to \mathbb{R}$. By the assumption, $Q(n, \{h(n)\}) = 0$ for $n \in E$.

Looking at the leading term in the equation

$$\sum_{k=0}^{d} n^k Q_k\left(\{h(n)\}\right) = 0, \quad n \in E, \tag{V.20}$$

and using the fact that Q_k are bounded functions on the fundamental domain D of G/Γ, we conclude that

$$\lim_{E \ni n \to \infty} Q_d\left(\{h(n)\}\right) = 0. \tag{V.21}$$

Take $(n_i)_{i \in \mathbb{N}}$ such that $n_\alpha \in E$ for $\alpha \in \mathcal{F}$. In particular, as a special case of (V.21), for any fixed $\beta \in \mathcal{F}$ we have

$$\mathrm{IP}\text{-}\lim_{\alpha \in \mathcal{F}} Q_d\left(\{h\left(n_\alpha + n_\beta\right)\}\right) = 0. \tag{V.22}$$

By the IP polynomial recurrence theorem for nilrotations (see [84, Theorem D]), there exists an IP ring $\mathcal{F}_\beta$ (which may be chosen to be contained in any previously specified IP ring) such that

$$\mathrm{IP}\text{-}\lim_{\alpha \in \mathcal{F}_\beta} h(n_\alpha + n_\beta)\Gamma = h(n_\beta)\Gamma. \tag{V.23}$$

We would like to take now the fractional parts in the preceding limit. A slight technical difficulty stems from the fact that the map $g \mapsto \{g\}$ is discontinuous. To overcome this problem, let $\varepsilon > 0$ be sufficiently small so that if $x \in G$ lies in the fundamental domain, then $d(x, x\gamma) \geq \varepsilon$ for all $\gamma \in \Gamma \setminus \{e\}$. After replacing E with

an IP subset E' in the original statement, we may assume that all the points $\{h(n_\alpha)\}$ for $\alpha \in \mathcal{F}$ lie within a ball of radius $\varepsilon/10$. This will be useful shortly.

Applying (V.23), possibly after refining $\mathcal{F}_\beta$ further, we find that

$$\mathrm{IP}-\lim_{\alpha\in\mathcal{F}_\beta} \{h(n_\alpha + n_\beta)\} = \{h(n_\beta)\}\,\gamma(\beta), \tag{V.24}$$

where $\gamma(\beta) \in \Gamma$. Now, since two limit points of $\{h(n_\alpha)\}$ cannot differ by a factor of $\gamma \in \Gamma \setminus \{e\}$, we conclude that

$$\mathrm{IP}-\lim_{\alpha\in\mathcal{F}_\beta} \{h(n_\alpha + n_\beta)\} = \{h(n_\beta)\}\,. \tag{V.25}$$

Because Q_d is continuous, (V.25) yields $Q_d(\{h(n_\beta)\}) = 0$, where we recall that β was arbitrary. Reasoning inductively, we obtain similarly $Q_k(\{h(n_\beta)\}) = 0$ for all k and $\beta \in \mathcal{F}$. Consequently, $Q(m, \{h(n)\}) = 0$ for all $m \in \mathbb{Z}$ and $n \in E'$, as desired. □

Step 5. Suppose that Theorem V.5.1 is false. Then there exists a nilmanifold G/Γ with a Mal'cev basis $\mathcal{X}$ such that every G_i/Γ_i is connected, where $\Gamma_i = \Gamma \cap G_i$, an element $g \in G$ acting ergodically by left translations, a nonempty algebraic subset $R \subset G^\circ$ with empty interior, and an IP set E such that R is preserved under the operations

$$x \mapsto x\,\{g^n\}\,, \qquad n \in E.$$

Proof. Assume the notation is as in Step 4. Let us consider for $x \in R$ the set

$$M_x = \left\{(m,n) \in \mathbb{Z}\times\mathbb{Z} \;\middle|\; g^m x g^{-m}\,\{g^n\} \in R\right\}.$$

Since multiplication in G° is given by polynomial formulae and R is an algebraic set, the condition $(m,n) \in M_x$ is equivalent to a system of polynomial equations in $g^m x g^{-m}$ and $\{g^n\}$. By Step 4, we have $(n,n) \in M_x$ for all $n \in E$. We first apply Lemma V.5.4 to replace the polynomial sequence g^n by a polynomial sequence $h \in \mathrm{poly}(\mathbb{Z} \to G^\circ_\bullet)$ with $\{g^n\} = \{h(n)\}$. Then Lemma V.5.7 shows that M_x contains $\mathbb{Z} \times E'$ for an IP set E' (which may be chosen inside a given IP subset of E).

For $x \in R$ let us also consider the set

$$P_x = \left\{h \in G^\circ \;\middle|\; g^m x g^{-m} h \in R \text{ for all } m \in \mathbb{Z}\right\}.$$

and let $P = \bigcap_{x\in R} P_x$. These sets are algebraic, and $\{g^n\} \in P_x$ precisely when $\mathbb{Z}\times\{n\} \subset M_x$. By Hilbert's Basis Theorem, we have $P = \bigcap_{x\in R_0} P_x$ for a finite subset R_0 of R.

Iterating the argument above, we find an IP set E' such that $\mathbb{Z}\times E' \subset \bigcap_{x\in R_0} M_x$. Hence $\{g^n\} \in P$ for $n \in E'$. In particular, $R\{g^n\} \subset R$ for $n \in E'$, and so R is preserved by the map $x \mapsto x\{g^n\}$. □

V.5.5 Group generated by fractional parts

We are now ready to take the last step in the proof of Theorem V.5.1. Our strategy is to show that the set R appearing in Step 5 in invariant under multiplication by all of G°, which is clearly absurd. In order to facilitate this strategy, we need the following fact.

Proposition V.5.8. *Let $X = G/\Gamma$ be a connected nilmanifold with a fixed Mal'cev basis and such that every G_i/Γ_i is connected, where $\Gamma_i = \Gamma \cap G_i$. Let $g\colon \mathbb{Z} \to G$ be a polynomial sequence such that the sequence $(g(n)\Gamma)_{n\geq 0}$ is dense in X. Suppose that $H \subset G^\circ$ is a Lie subgroup such that $\{g(n)\} \in H$ for infinitely many n. Then $H = G^\circ$.*

Step 6. Theorem V.5.1 holds true.

Proof (assuming Propositon V.5.8). Suppose Theorem V.5.1 were false, and assume notation as in conclusion of Step 5. Denote

$$H = \{h \in G^\circ \mid Rh \subset R\} = \bigcap_{x\in R} \{h \in G^\circ \mid xh \in R\}.$$

It is immediate from the definition that H is algebraic and closed under multiplication and taking inverses (by Lemma V.5.5). Hence, H is a Lie subgroup of G°. Furthermore, $\{g^n\} \in H$ for $n \in E$. Applying Proposition V.5.8, we conclude that $H = G^\circ$. In particular, $R = G^\circ$, which is in contradiction with $\operatorname{int} R = \emptyset$. □

Proof of Proposition V.5.8. Let E be an infinite set with $\{g(n)\} \in H$ for $n \in E$.

By Lemma V.5.4, there exists a polynomial sequence $h \in \operatorname{poly}(\mathbb{Z} \to G^\circ_\bullet)$ such that $\{g(n)\} = \{h(n)\}$. Write, using Proposition I.1.20,

$$h(n) = g_0^{\binom{n}{0}} g_1^{\binom{n}{1}} g_2^{\binom{n}{2}} \dots g_d^{\binom{n}{d}},$$

where $g_i \in G_i^\circ$.

In order to prove the claim, it is sufficient by Lemma V.5.3 to show that $H(G^\circ)_2 = G^\circ$. The quotient $G^\circ/(G^\circ)_2$ can be identified with $\mathbb{R}^d$ in such a way that $\Gamma \cap G^\circ$ maps to $\mathbb{Z}^d$. Let $\pi\colon G^\circ \to \mathbb{R}^d$ be the corresponding projection. Put $\theta(n) = \pi(h(n))$, $\xi(n) = \pi(\{h(n)\}) = \{\theta(n)\}$ and $V = \pi(H) < \mathbb{R}^d$. From the form of $h(n)$, we see that $\theta(n)$ is a polynomial in n, $\xi(n)$ is bounded, and $\xi(n) \in V$ for $n \in E$. We need to show that $V = \mathbb{R}^d$.

Suppose for the sake of contradiction that $V \neq \mathbb{R}^d$. Write $\theta(n) = (\theta_1(n), \dots, \theta_d(n))$. Then there exists a non-trivial linear relation

$$\sum_{i=1}^{d} \lambda_i \{\theta_i(n)\} = 0, \qquad n \in E, \tag{V.26}$$

where $\lambda_i \in \mathbb{R}$ are not all zero. This can be rewritten as

$$\sum_{i=1}^{d} \lambda_i \left[\theta_i(n)\right] = \sum_{i=1}^{d} \lambda_i \theta_i(n), \qquad n \in E. \tag{V.27}$$

The maps $\theta_i \colon \mathbb{Z} \to \mathbb{R}$ are polynomials, so writing

$$-\sum_{i=1}^{d} \lambda_i \theta_i(n) = \sum_{j=0}^{D} \kappa_j n^j,$$

we obtain

$$\sum_{i=1}^{d} \lambda_i \left[\theta_i(n)\right] + \sum_{j=0}^{D} \kappa_j n^j = 0, \qquad n \in E. \tag{V.28}$$

This amounts to saying that there is a non-trivial linear relation between the vectors $([\theta_i(n)])_{n\in E}$ for $i = 1, \dots, d$ and $(n^j)_{n\in E}$ for $j = 0, 1, \dots, D$. Because these vectors are integer-valued, if such a relation exists, there is also such a relation with integer coefficients. Take one such relation,

$$\sum_{i=1}^{d} l_i \left[\theta_i(n)\right] + \sum_{j=0}^{D} k_j n^j = 0, \qquad n \in E, \tag{V.29}$$

where $l_i \in \mathbb{Z}$ and $k_j \in \mathbb{Z}$ for all i, j, and not all of l_i, k_j are zero. Dropping the integer parts in (V.29) at the cost of introducing a bounded error, we get

$$\sum_{i=1}^{d} l_i \theta_i(n) + \sum_{j=0}^{D} k_j n^j = O(1), \qquad n \in E. \tag{V.30}$$

This is only possible if the left hand side of (V.30) is constant (as a polynomial, hence for all $n \in \mathbb{Z}$). It follows that $\theta(n) \bmod \mathbb{Z}^d$ for $n \in \mathbb{Z}$ takes values in the proper subtorus $\left\{x \in \mathbb{R}^d/\mathbb{Z}^d \;\middle|\; \sum l_i x_i = c\right\}$, which contradicts the fact that the sequence $(g(n)\Gamma)_{n\geq 0}$ is dense in G/Γ. □

V.6 Sparse automatic sets

The aim of this section is to provide proofs of the several combinatorial results on (sparse) automatic sequences that have been used in the previous sections, specifically Theorem V.4.7. We begin with some preparatory results.

V.6.1 Density of symbols

First of all, we note that density of symbols for an automatic sequence is often uniform. A set $E \subset \mathbb{N}_0$ has *uniform density* $d = d(E) = d^*(E)$ if $|E \cap [M, M+N)|/N \to d$ as $N \to \infty$, uniformly in M. For an automaton $A = (S, s_\bullet, \delta, \Omega, \tau)$, a *terminal component* is an automaton $A' = (S', s'_\bullet, \delta', \Omega, \tau')$, where $S' \subset S$ is nonempty, preserved under $\delta(\cdot, j)$ for all $j \in \Sigma_k$ and minimal with respect to these properties, $s'_\bullet \in S'$, and δ', τ' are restrictions of δ and τ respectively.

Lemma V.6.1. *Let $a\colon \mathbb{N}_0 \to \Omega$ be a k-automatic sequence generated by an automaton $\mathcal{A} = (S, s_0, \delta, \Omega, \tau)$ reading input from the most significant digit, ignoring the initial zeros. For $y \in \Omega$, let $\rho_y \geq 0$. Then, the following conditions are equivalent:*

(i) For any $y \in \Omega$, the set $\{n \in \mathbb{N}_0 \mid a(n) = y\}$ has density ρ_y.

(ii) For any $y \in \Omega$, the set $\{n \in \mathbb{N}_0 \mid a(n) = y\}$ has uniform density ρ_y.

(iii) For any sequence $\tilde{a}'\colon \Sigma_k^* \to \Omega$ produced by an automaton $\mathcal{A}'$ induced by a terminal component of $\mathcal{A}$, for any $y \in \Omega$ we have

$$\left|\left\{u \in \Sigma_k^L \;\middle|\; \tilde{a}'(u) = y\right\}\right|/k^L \to \rho_y \text{ as } L \to \infty.$$

Proof. It is clear that (ii) implies (i). We will further show that (i) implies (iii) and (iii) implies (ii). Throughout, it will be convenient to assume that $\Omega = \{0, 1\}$, which we may do without loss of generality.

Suppose that (i) holds, and take some $\tilde{a}'$ like in (iii). There is some $v \in \Sigma_k^*$ such that $\tilde{a}'(u) = a([vu]_k)$, whence

$$\frac{1}{k^L} \sum_{u \in \Sigma_k^L} \tilde{a}'(u) = \frac{1}{k^L} \sum_{n=[v]_k k^L}^{([v]_k+1)k^L-1} a(n) \to \rho \tag{V.31}$$

as $L \to \infty$.

Suppose now that (iii) holds. For any N, M and L we have

$$\frac{1}{N} \sum_{n=M}^{M+N-1} a(n) = \frac{1}{N} \sum_{m=\lfloor M/k^L \rfloor}^{\lfloor (M+N)/k^L \rfloor} \sum_{n=0}^{k^L-1} a(mk^L + n) + O(k^L/N). \tag{V.32}$$

For any $m \in \mathbb{N}_0$, consider the sequence $\tilde{a}'_m\colon \Sigma_k^* \to \Omega$ given by $\tilde{a}'_m(u) = a([(m)_k u]_k)$, so that if $u \in \Sigma_k^L$ then $a(mk^L + [u]_k) = \tilde{a}'_m(u)$. Note that $\tilde{a}'_m$ is produced by the automaton $\mathcal{A}'$ induced by the set of states in $\mathcal{A}$ reachable from the initial state $s'_0 = \delta(s_0, (m)_k)$.

If $\delta(s_0, (m)_k)$ lies in a terminal component of $\mathcal{A}$, then we may use (iii) to estimate the inner sums in (VI.20):

$$\sum_{n=0}^{k^L-1} a(mk^L + n) = k^L\rho + o(k^L)$$

as $N \to \infty$ (where the error term is uniform with respect to m, since there are only finitely many possible sequences a'_m). It is an easy exercise to check that the set of $m \in \mathbb{N}_0$ such that $\delta(s_0, (m)_k)$ does not lie in a terminal component of $\mathcal{A}$ has upper Banach density 0. Estimating the inner sums in (VI.20) corresponding to such m trivially by $O(k^L)$, and letting $L \to \infty$ slowly, we conclude that

$$\frac{1}{N}\sum_{n=M}^{M+N-1} a(n) = \rho + o(1)$$

as $N \to \infty$, uniformly in M. Hence, (ii) holds. □

As a consequence, for automatic sets, in the definition of sparsity it does not matter if whether we work with density or Banach density.

Corollary V.6.2. *Let $E \subset \mathbb{N}_0$ be an automatic set. Then, $d(E) = 0$ if and only if $d^*(E) = 0$.*

V.6.2 Dichotomy for sparse automatic sets

For a k-automaton $\mathcal{A} = (S, s_\bullet, \delta, \{0,1\}, \tau)$ with output $\{0,1\}$, we call a state $s \in S$ *promising* if there exists a word $w \in \Sigma_k^*$ such that $\tau(\delta(s, w)) = 1$.

Proposition V.6.3. *Let $\mathcal{A} = (S, s_\bullet, \delta, \{0,1\}, \tau)$ be a k-automaton with output $\{0,1\}$. Then either*

(i) there exists a promising state $s \in S$ and $v_1, v_2 \in \Sigma_k^*$ such that $v_1 \neq v_2$, $|v_1| = |v_2|$, and $\delta(s, v_1) = \delta(s, v_2) = s$; or

(ii) the set of $w \in \Sigma_k^*$ such that $\tau(\delta(s_\bullet, w)) = 1$ is arid.

Remark. Similar results are implicit in [1, Lemma 6.7] and [30, Proposition 7.9]. For the convenience of the reader, we provide a self-contained proof.

Proof. Without loss of generality, we may assume that for every state s there is a word $w \in \Sigma_k^*$ such that $\delta(s_\bullet, w) = s$ (so $\mathcal{A}$ is connected as a non-directed graph). Let $S^+ \subset S$ denote the set of promising states. We construct a directed multigraph

G with the vertex set S^+ and with directed edges from $s \in S^+$ to $s' \in S^+$ for each $v \in \Sigma_k$ such that $s' = \delta(s, v)$. Given a state $s \in S$, there is a natural way to identify paths of length l in the automaton $\mathcal{A}$ beginning in s and words in Σ_k^l.

For $s_1, s_2 \in S^+$, write $s_1 \sim s_2$ if s_1 and s_2 are in the same strongly connected component of G, i.e. if there exist $v_1, v_2 \in \Sigma_k^*$ such that $\delta(s_1, v_1) = s_2$ and $\delta(s_2, v_2) = s_1$. We see that $\sim$ is an equivalence relation. For s_1, s_2 in S^+ we further write that $s_1 \prec s_2$ if there exist $v_1, v_2 \in \Sigma_k^*$ such that $v_1 \neq v_2$, $|v_1| = |v_2|$ and $\delta(s_1, v_1) = \delta(s_1, v_2) = s_2$. The relation $\prec$ is transitive and is invariant with respect to $\sim$ (i.e. if $s_1 \sim s_1'$, $s_2 \sim s_2'$, and $s_1 \prec s_2$, then $s_1' \prec s_2'$), and therefore induces a transitive relation on the set of connected components $S^+/\!\sim$, which we continue to denote by $\prec$. Assume now that the condition (i) does not hold. This means precisely that $\prec$ is a partial strict order (on S^+ or $S^+/\!\sim$).

Recall that a *cycle graph* of length $n \geq 1$ is a graph C on n vertices $\{v_1, \ldots, v_n\}$ with exactly n edges, going from v_i to v_{i+1} for $1 \leq i \leq n$, where $v_{n+1} = v_1$.

Claim. Any strongly connected component C of G is either a cycle or consists of a single vertex with no edges.

Proof of the claim. If C has a single vertex, then the claim follows immediately, so assume this is not the case. Strong connectivity now implies that any vertex lies on a cycle. For any $s \in C$, any two cycles γ_1, γ_2 from s to s begin with the same edge (otherwise some multiples of these two cycles would produce two different paths from s to s of the same length). Because of strong connectivity any edge in C can be prolonged to a cycle, so no vertex in C has two outgoing edges. Thus, each vertex has outdegree exactly 1. It follows that C is a disjoint union of cycles. Since C is connected, it is a cycle graph. □

Any path from the initial vertex $s_\bullet$ to a vertex s with $\tau(s) = 1$ passes only through promising states and has the form $\gamma = u_1 v_1 u_2 \cdots v_{r-1} u_r$, where $r \geq 1$, $v_i \in \Sigma_k$, $u_i \in \Sigma_k^*$ (possibly empty), and the segment of the path corresponding to u_i is contained entirely in a strongly connected component C_i, while v_i correspond to edges between strongly connected components. Since $S^+/\!\sim$ is strictly ordered by the relation $\prec$, r does not exceed the number of strongly connected components $|S^+/\!\sim|$. Furthermore, since C_i are cycle graphs, any u_i has the form $u_i = u_i' w_i^{l_i}$, $l_i \geq 0$, where w_i describes a cycle in C_i, $|w_i| = |C_i|$ and $|u_i'| \leq |C_i|$. Hence, any such path γ is uniquely determined by the following data:

(i) the integer $1 \leq r \leq |S^+/\!\sim|$;

(ii) the digits v_i, $1 \leq i \leq r-1$;

(iii) the words u'_i and w_i, $1 \leq i \leq r$;

(iv) the multiplicities l_i, $1 \leq i \leq r$.

There are only finitely many choices of the data (i)-(iii). Hence every path from $s_\bullet$ to a vertex s with $\tau(s) = 1$ takes one of finitely many forms

$$\gamma = v'_1 w_1^{l_1} v'_2 w_2^{l_2} \cdots v'_{r-1} w_r^{l_r}, \quad l_i \geq 0,$$

where $v'_1 = u'_1$, $v'_i = v_{i-1}u'_i$ for $2 \leq i \leq r$. This ends the proof of the proposition. □

With the above proposition, we can easily derive our main result.

Proof of Theorem V.4.7. Assume that $E \subset \mathbb{N}_0$ is sparse but not arid; we need to show that E is IPS. Let $\mathcal{A} = (S, s_\bullet, \delta, \{0,1\}, \tau)$ be a k-automaton with output $\{0,1\}$ which produces the characteristic sequence of E when reading digits starting from the most significant one, allowing for initial zeros. Of course, we may additionally assume that every state is reachable from the initial state, i.e. for every $s \in S$ there exists $v \in \Sigma_k^*$ such that $\delta(s_\bullet, v) = s$.

Since E is not arid, neither is the set $A = \{w \in \Sigma_k^* \mid \tau(\delta(s_\bullet), w) = 1\}$. Hence, by Proposition V.6.3, there exists a state $s_0 \in S$ and $v_1, v_2 \in \Sigma_k^*$ of equal length $l = |v_1| = |v_2|$ such that $[v_1]_k < [v_2]_k$ and $\delta(s_0, v_1) = \delta(s_0, v_2) = s_0$. Pick $u_0, u_1 \in \Sigma_k^*$ so that $s_0 = \delta(s_\bullet, u_0)$, and $\tau(\delta(s_0, u_1)) = 1$.

The set A contains all words of the form $w = u_0 v_{j_1} v_{j_2} \ldots v_{j_t} u_1$, where $j_i \in \{1,2\}$ and $t \in \mathbb{N}$. It remains to notice that the set of $[w]_k$ for w as above is precisely the IPS set $\mathrm{FS}(n_i; N_t)$ where $N_t = [u_0 v_1^t u_1]_k$ and $n_i = ([v_2]_k - [v_1]_k)k^{(i-1)l+|u_1|}$. □

We may use Theorem V.4.7 to easily derive a result on growth of automatic sequences.

Corollary V.6.4. *Let $(a_n)_{n\geq 0}$ be an automatic sequence with values in $\{0,1\}$ and let $\nu(N) = |\{0 \leq n < N \mid a_n = 1\}|$. Then either*

(i) there exist $c, \alpha > 0$ such that $\nu(N) \geq cN^\alpha$ for large enough N; or

(ii) there exist $c, l > 0$ such that $\nu(N) \leq c(\log N)^l$ for large enough N.

Proof. Let $E = \{n \in \mathbb{N}_0 \mid a_n = 1\}$. By Theorem V.4.7, E is either arid or an IPS set. In the former case, by Lemma V.4.3, Condition (ii) holds. In the latter case, Condition (i) follows easily from the proof of Theorem V.4.7. □

Remark. Corollary V.6.4 can also be obtained directly by observing that the sequence $\nu(k^L)$ is obeys a linear recurrence with non-negative coefficients, and using the Perron-Frobenius theorem. This observation is standard and exploited e.g. in [4, Chpt. 3].

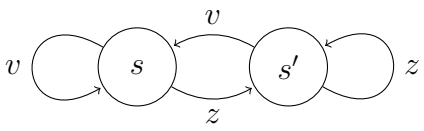
Remark. The sequence $\nu(N)$ is regular. The possible rates of growth of regular sequences are actively studied, see e.g. [8], [9].

V.6.3 IP rich automatic sets

Finally, we prove the variant of Theorem V.4.7 mentioned in Section V.4.

Proof of Proposition V.4.10. Let $\mathcal{A} = (S, s_\bullet, \delta, \{0,1\}, \tau)$ be a k-automaton that produces the characteristic sequence of E by reading the digits of n *starting with the least significant one*, allowing for leading zeroes. We will denote the word $0 \cdots 0 \in \Sigma_k^*$ with n zeroes by 0^n. We begin by proving the following claim.

Claim. There exist states $s, s' \in S$ with $\tau(s) = 1$, $l \in \mathbb{N}$, and a word $v \in \Sigma_k^l$ that is not a power of 0 such that for $z = 0^l$ we have $\delta(s, z) = s'$, $\delta(s, v) = s$, $\delta(s', z) = s'$, $\delta(s', v) = s$. This is portrayed below:



Proof of the claim. Let $n = |S|$ be the number of states in $\mathcal{A}$. We first show a weaker statement, namely that there is a state s with $\tau(s) = 1$ such that if $s' = \delta(s, 0^n)$ denotes the state reached from s after reading n zeros, then we can return to s along a path not consisting only of zeroes, that is $\delta(s', v) = s$ for some $v \in \Sigma_k^*$ that is not a power of 0.

To prove this, we construct a word $w = w_1 w_2 \cdots w_{n^2}$ as follows. Enumerate all pairs in $S \times S$ as (s_i, s_i') for $1 \leq i \leq n^2$. In the first step, if s_1' is reachable from s_1, let w_1 describe any path between the two, so that $\delta(s_1, w_1) = s_1'$; otherwise let $w_1 = \epsilon$. In general, if $w_1, \ldots, w_{i-1}$ have been defined, choose w_i such that $\delta(s_i, w_1 w_2 \cdots w_{i-1} w_i) = s_i'$ if possible (i.e. if s_i' is reachable from $\delta(s_i, w_1 w_2 \cdots w_{i-1})$), and $w_i = \epsilon$ otherwise.

By assumption, there exists some $x, y \in \Sigma_k^*$ such that for $s = \delta(s_\bullet, xwy)$ we have $\tau(s) = 1$. Applying the same assumption with $w1$ in place of w, we may ensure that y is not a power of 0. It remains to show that we can return from $s' = \delta(s, 0^n)$ to s. For $0 \leq i \leq n^2$, denote by $r_i = \delta(s_\bullet, x w_1 w_2 \cdots w_i)$ the intermediate states on the

path from $s_\bullet$ to s labelled xwy, in particular $r_0 = \delta(s_\bullet, x)$. The construction of w is arranged so that for any i with $s_i = r_0$, we have $r_i = \delta(r_{i-1}, w_i) = s'_i$, provided that s'_i is reachable from r_{i-1}.

Choose $1 \leq j \leq n^2$ such that $s_j = r_0$ and $s'_j = s'$. Since s is reachable from r_{j-1} and s' is reachable from s, s' is reachable from r_{j-1}. Hence, the construction of w guarantees that $r_j = \delta(r_{j-1}, w_j) = s'$. In particular, $\delta(s', v) = s$, where $v = w_{j+1} \dots w_{n^2} y$. Note that v is not a power of 0. This proves the weaker version of the claim.

To prove the stronger statement, note first that since S has only n states, there exist $0 \leq i < j \leq n$ such that $\delta(s, 0^i) = \delta(s, 0^j)$. Let $m > i$ be any integer divisible by $(j - i)$ and put $s'' = \delta(s, 0^m)$. Since m is divisible by $(j - i)$, we have $s'' = \delta(s'', 0^m)$. Because s' is reachable from s'' (actually, $\delta(s'', 0^n) = s'$), there is a word u (equal to $0^n v$, hence not a power of 0), such that $\delta(s'', u) = s$. Take $v' = (0^m u)^m$ and $l = m(|u| + m)$. The states s, s'' and the word v' (of length l) satisfy all the required conditions, namely $\delta(s, 0^l) = s''$, $\delta(s, v') = s$, $\delta(s'', 0^l) = s''$, $\delta(s'', v') = s$, and $\tau(s) = 1$. □

To finish the proof of Proposition V.4.10, we may assume that all states are reachable from $s_\bullet$. Choose states s and s' and words v and $z = 0^l$ as in the statement of the claim. Let $u \in \Sigma_k^*$ be such that $\delta(s_\bullet, u) = s$. For any word $w = uv_1v_2 \cdots v_r$, where $v_i \in \{v, z\}$ for $1 \leq i < r$ and $v_r = v$, we have $\delta(s_\bullet, w) = s$, whence $[w]_k^{\mathrm{R}} \in E$. It follows that E contains $\mathrm{FS}(n_i; N)$, where $N = [u]_k^{\mathrm{R}}$ and $n_i = k^{(i-1)l+|u|}[v]_k^{\mathrm{R}}$. □

V.6.4 Proof of Theorem V.1.6

In this section, we derive Theorem V.1.6 from Theorem V.1.5. Our argument is purely combinatorial and can be entirely phrased in terms of finite automata with no further recourse to dynamics.

Proposition V.6.5. *Let $A \subset \Sigma_k^*$ be an infinite arid set. Then there exists $v \in \Sigma_k^*$ such that $A \cap v\Sigma_k^*$ takes the form*

$$A \cap v\Sigma_k^* = \bigcup_{i=1}^{p} \{vw^l u_i \mid l \in \mathbb{N}_0\},$$

where $p \geq 1$, $v, w, u_i \in \Sigma_k^*$ and $w \neq \epsilon$. (In particular, $A \cap v\Sigma_k^*$ is arid of rank 1.)

Likewise, $A \cap \Sigma_k^* v$ takes the form

$$A \cap \Sigma_k^* v = \bigcup_{i=1}^{p'} \{u'_i (w')^l v' \mid l \in \mathbb{N}_0\},$$

where $p' \geq 1$, $v', w', u'_i \in \Sigma_k^*$ *and* $w' \neq \epsilon$.

Proof. Since the notion of an arid set is preserved under the reversal operation, it is sufficient to prove the former statement. For $B \subset \Sigma_k^*$ and $v \in \Sigma_k^*$, put $v^{-1}B = \{u \in \Sigma_k^* \mid vu \in B\}$. If B is arid of rank $\leq r$, then so is $v^{-1}B$.

Claim. Let $B \subset \Sigma_k^*$ be arid of rank r, and let $x_1, x_2, y \in \Sigma_k^*$ with $a = |y| = |x_2|$ and $y \neq x_2$. Then for sufficiently large m (depending on B, x_1, x_2, y), $(x_1 y^m x_2)^{-1}B$ is arid of rank $\leq (r-1)$.

Proof. Replacing B with $x_1^{-1}B$, we may assume that $x_1 = \epsilon$.

In analogy with Remark V.4.1, note that there is a natural way to identify $\Sigma_{k^a}^*$ with a subset of Σ_k^*, and any arid set $B \subset \Sigma_k^*$ is a finite union of translates $B_i v_i$ ($v_i \in \Sigma_k^*$) of arid sets $B_i \subset \Sigma_{k^a}^*$. Hence, it will suffice to show that if $B \subset \Sigma_{k^a}^*$ is arid of rank r then for sufficiently large m, $B \cap y^m x_2 \Sigma_k^*$ is arid of rank $\leq (r-1)$. We may now replace k with k^a and assume that $|y| = |x_2| = 1$.

It will suffice to prove the claim for B of the form

$$B = \left\{ v_0 w_1^{l_1} v_1 w_2^{l_2} \dots w_r^{l_r} v_r \;\middle|\; l_1, \dots, l_r \in \mathbb{N} \right\}$$

where $w_i \neq \epsilon$ for all i (note that l_i here are required to be strictly positive; any arid set of rank r is a union of such sets and an arid set of rank $\leq (r-1)$). Now, if $m > |v_0 w_1|$ then either $B \cap y^m x_2 \Sigma_k^* = \emptyset$ (in which case we are trivially done) or $B \cap y^m x_2 \Sigma_k^* \neq \emptyset$ and w_1 is a power of y. In the latter case, we further conclude that x_2 appears in $v_1 w_2$ (else B would have rank $\leq (r-1)$), which is necessarily of the form $y^b x_2 v'_1$ with $b \in \mathbb{N}_0$. Hence

$$(y^m x_2)^{-1} B = \left\{ v'_1 w_2^{l_2 - 1} \dots w_r^{l_r} v_r \;\middle|\; l_2, \dots, l_r \in \mathbb{N} \right\}$$

is arid of rank $\leq (r-1)$. □

The proof of the proposition is now a simple induction on the rank r of A. Since A is infinite, $r \geq 1$.

If $r = 1$, then A takes the form $\bigcup_{i=1}^{r} \left\{ v_i w_i^l u_i \;\middle|\; l \in \mathbb{N}_0 \right\}$, where $w_i \neq \epsilon$ for at least one i, say $i = 1$. Then $A \cap v\Sigma_k^*$ takes the required form for $v = v_1 w_1^m$ for m large enough.

If $r > 1$, then we may find a rank 2 basic arid set

$$B = \left\{ v_0 w_1^{l_1} v_1 w_2^{l_2} v_2 \;\middle|\; l_1, l_2 \in \mathbb{N}_0 \right\}$$

contained in A. Without loss of generality, we may assume that $|w_1| = |w_2| > |v_1|$. Apply the above Claim with $x_1 = v_0$, $y = w_1^{l_1}$ and x_2 equal to the first $|y|$ symbols of $v_1 w_2^{l_2}$, where $l_2 \geq l_1 \geq 2$. Note that $y \neq z_2$, because otherwise by an elementary computation one could show that the rank of B is 1. Then for m large enough $A' = (x_1 y^m x_2)^{-1} A$ is arid of rank $\leq (r-1)$ and infinite. By the inductive assumption, there exists $v' \in \Sigma_k^*$ such that $A' \cap v'\Sigma_k^*$ takes the required form. It remains to take $v = x_1 y^m x_2 v'$. □

Corollary V.6.6. *Let E be an infinite k-arid set. Then there exist integers $n \geq 1$, $r \geq 0$, $p \geq 1$, and words $v_1, \ldots, v_p, w, u \in \Sigma_k^*$, $w \neq \epsilon$ such that*

$$E \cap (n\mathbb{Z} + r) = \bigcup_{i=1}^{p} \{[v_i w^l u]_k \mid l \in \mathbb{N}_0\}.$$

Proof. Follows immediately from Proposition V.6.5. □

Proposition V.6.7. *If the set $\{k^l \mid l \geq 0\}$ is not generalised polynomial, then neither is any infinite k-arid set.*

Proof. Assume we know that $P = \{k^l \mid l \geq 0\}$ is not generalised polynomial. Then neither is any set of the form $P_t = \{k^{tl} \mid l \geq 0\}$ for $t \geq 1$ since $P = \bigcup_{j=0}^{t-1} k^j P_t$.

Suppose that there exists an infinite k-arid set which is generalised polynomial. Since the class of generalised polynomial sets contains all arithmetic progressions and is closed under finite intersections, Corollary V.6.6 allows us to assume that

$$E = \bigcup_{i=1}^{p} \{[v_i w^l u]_k \mid l \geq 0\}$$

for some $p \geq 1, v_1, \ldots, v_p, w, u \in \Sigma_k^*$, $w \neq \epsilon$. Let $s = |u|, t = |w|$ and note that

$$[v_i w^l u]_k = [u]_k + k^s [w]_k \frac{k^{tl} - 1}{k^t - 1} + [v_i]_k k^{tl+s}.$$

Let g be a generalised polynomial such that $E = \{n \in \mathbb{N}_0 \mid g(n) = 0\}$ and assume further that g is a restriction of a generalised polynomial of a real variable that has no further zeroes in $\mathbb{R}_{>0} \setminus \mathbb{N}$. (To this end, replace $g(n)$ by $g(n)^2 + \|n\|^2$.) Then an easy computation shows that the polynomial

$$h(n) = g\left(k^s \frac{n - [w]_k}{k^t - 1} + [u]_k\right)$$

has as its zero set

$$B = \{n \in \mathbb{N} \mid h(n) = 0\} = \bigcup_{i=1}^{p} \{b_i k^{tl} \mid l \geq 0\}$$

where $b_i = [w]_k + (k^t - 1)[v_i]_k$, $i = 1, \ldots, p$.

The set $C = \{n \in \mathbb{N}_0 \mid b_1 n \in B\}$ is also generalised polynomial and it has the form

$$C = \bigcup_{i=1}^{p} \{c_i k^{tl} \mid l \geq 0\}$$

with $c_1 = 1$ and $c_i = b_i k^{tl_i}/b_1$ where $l_i \geq 0$ is the smallest integer such that b_1 divides $b_i k^{tl_i}$. (If there is no such integer, the corresponding term is not present.)

Let $m \geq 1$ be such that $c_i < k^{tm}$ for $i = 1, \ldots, p$. Replacing the set $\{c_i k^{tl} \mid l \geq 0\}$ by the union

$$\{c_i k^{tl} \mid l \geq 0\} = \bigcup_{j=0}^{m-1} \{c_i k^{tj} k^{mtl} \mid l \geq 0\}$$

and replacing k by k^{mt}, we may assume that

$$C = \bigcup_{i=1}^{p} \{c_i k^{l} \mid l \geq 0\}$$

with $c_1 = 1$ and $1 \leq c_i < k^2$.

Consider the set $D = \{n \in C \mid n \equiv 1 \pmod{k^2 - 1}\}$. The set D is generalised polynomial and an integer $c_i k^l \in C$ can be an element of D only if $c_i \equiv 1 \pmod{k^2-1}$ or $c_i \equiv k \pmod{k^2 - 1}$. Since $1 \leq c_i \leq k^2 - 1$, this gives $c_i = 1$ or $c_i = k$ and whether the latter possibility is realized or not, we have $D = \{k^{2l} \mid l \geq 0\}$. This is a contradiction with our remark that no set of the form $P_t = \{k^{tl} \mid l \geq 0\}$, $t \geq 1$, is generalised polynomial (note that during the proof we have replaced k by its power). □

We are now ready to finish the proof of Theorem V.1.6.

Proof of Theorem V.1.6. The two statements in Theorem V.1.6 are of course mutually exclusive. Assume now that there exists a sequence (a_n) which is k-automatic, generalised polynomial, and not ultimately periodic. By Theorem V.1.5, it nevertheless coincides with a periodic sequence (b_n) except at a set of density zero. Consider the set $C = \{n \in \mathbb{N}_0 \mid a_n \neq b_n\}$. This set is k-automatic, generalised polynomial, sparse, and infinite. By Theorem V.4.5, C is then arid and hence by Proposition V.6.7 the set $\{k^l \mid l \geq 0\}$ is generalised polynomial as well. □

V.7 Examples

We now consider some specific examples of sequences for which we can prove or disprove that they are automatic or given by generalised polynomials. Results discussed here do not rely on Theorem V.4.8.

V.7.1 Small fractional parts

In this section, we study the sequences of the form

$$g(n) = \begin{cases} 1, & \text{if } 0 < q(n) < \varepsilon(n), \\ 0, & \text{otherwise,} \end{cases} \tag{V.33}$$

where $q(n) \geq 0$ is a generalised polynomial and $\varepsilon(n) \to 0$, which include the most natural examples of sparse generalised polynomials. The most frequent application is when $q(n) = \|r(n)\|$ or $q(n) = \{r(n)\}$ for an (unbounded) generalised polynomial $r(n)$. In this section, it will be more convenient to regard generalised polynomials as functions on $\mathbb{N}_0$. This causes no problems since we can always extend them to $\mathbb{Z}$.

Example V.7.1. Take $a \in \mathbb{N}$ and let $(n_i)_{i\geq 0}$ be the sequence given by $n_0 = 0,\ n_1 = 1$ and $n_{i+2} = an_{i+1} + n_i$. Then $E = \{n_i \mid i \in \mathbb{N}_0\}$ is generalised polynomial. For $a = 1$, this is the set of Fibonacci numbers.

Proof. Let $\alpha \in \mathbb{R}$ be the real number with continued fraction expansion $\alpha = [a; a, a, \dots]$, i.e.

$$\alpha = a + \cfrac{1}{a + \cfrac{1}{a + \dots}} = \frac{a + \sqrt{a^2+4}}{2}. \tag{V.34}$$

Let

$$E' = \left\{ n \in \mathbb{N} \;\middle|\; \|n\alpha\| < \frac{\sqrt{5}}{2\sqrt{a^2+4}n} \right\}.$$

Note that E' is generalised polynomial by Lemma V.4.4, so it will suffice to show that the symmetric difference $E \triangle E'$ is finite.

The finite continued fractions $[a; a, a, \dots, a]$ (with $i-1$ occurences of a after the semicolon) have the form m_i/n_i with this rational number written in lowest terms. Using a well-known formula for the error term in the continued fraction approximations (see e.g. [70, Thm. 3.1]), we find that

$$n_i \|n_i\alpha\| = n_i^2 \sum_{l=0}^{\infty} \frac{(-1)^l}{n_{i+l}n_{i+l+1}} \xrightarrow[i\to\infty]{} \sum_{l=0}^{\infty} \frac{(-1)^l}{\alpha^{2l+1}} = \frac{1}{\alpha + 1/\alpha} = \frac{1}{\sqrt{a^2+4}}, \tag{V.35}$$

because $n_{i+1}/n_i \to \alpha$ as $i \to \infty$. Hence, for sufficiently large i we have $n_i \in E'$.

Conversely, suppose that a positive integer n is in E'. Then (since $\sqrt{a^2+4} \geq \sqrt{5}$) there exists an integer $m \in \mathbb{N}$ such that

$$\left|\alpha - \frac{m}{n}\right| < \frac{1}{2n^2}.$$

By a classical theorem of Legendre (see e.g. [70, Thm. 5.1]), we have $m/n = m_i/n_i$, so that $m = km_i$, $n = kn_i$ for an integer $k \in \mathbb{N}$. It follows that $n \|n\alpha\| = k^2 n_i \|n_i\alpha\|$, which by (V.35) exceeds $2/\sqrt{a^2+4}$ if $k \geq 2$ and i is sufficiently large. Hence, for all but finitely many choices of $n \in E'$ we have $k = 1$, whence $n = n_i \in E$ as desired. □

Remark. A slightly more careful reasoning shows that in the case of Fibonacci numbers a simpler result is true. Let $\varphi = (1+\sqrt{5})/2$. Then

$$\{F_i \mid i \in \mathbb{N}\} = \left\{ n \in \mathbb{N} \;\middle|\; \|n\varphi\| < \frac{1}{2n} \right\}.$$

Remark. In fact, this result holds for any choice of $n_0, n_1 \in \mathbb{Z}$. This follows from Proposition V.8.2 below. A similar argument works also for sequences $(n_i)_{i\geq 0}$ given by the recurrence $n_{i+2} = a_{i+2}n_{i+1} + n_i$, where $a_i \geq 2$ is a bounded sequence of integers. Then n_i is the sequence of denominators in the best rational approximations of a badly approximable real number $\alpha = [0; a_2, a_3, \ldots]$.

Proposition V.7.2. *Suppose $q(n) \geq 0$ and $p(n) > 0$ are generalised polynomials with $p(n) \to \infty$ as $n \to \infty$, $b < 0$ is a rational number and $\varepsilon(n) = p(n)^b$. Then $g(n)$ given by* (V.33) *is a sparse generalised polynomial.*

Proof. That $g(n)$ is a generalised polynomial follows from Lemma V.4.4. To see that $g(n)$ is sparse, use Theorem I.1.22 to represent $q(n)$ as $F(g^n x)$ where G/Γ is a nilmanifold on which $g \in G$ acts ergodically, $x \in G/\Gamma$, and $F\colon G/\Gamma \to \mathbb{R}$ is piecewise polynomial. By Corollary I.1.5, the sets $\{n \in \mathbb{N} \mid 0 < q(n) < \varepsilon\}$ have density $\mu_{G/\Gamma}(F^{-1}((0,\varepsilon)))$, which tends to 0 as $\varepsilon \to 0$. Since $\varepsilon(n) \to 0$, it follows that $g(n)$ is sparse. □

We move on to an example of a fairly explicit class of sparse generalised polynomial sequences. We later discuss how the argument adapts to other sequences of the form (V.33).

Proposition V.7.3. *For any sufficiently small $c > 0$ the following holds. Let $1, \alpha, \beta \in \mathbb{R}$ be algebraic numbers linearly independent over $\mathbb{Q}$, and let $\varepsilon(n) \to 0$ be a sequence with $\varepsilon(n) \gg n^{-c}$. Then the sequence given by*

$$f(n) = \begin{cases} 1, & \text{if } \|n\alpha \lfloor n\beta \rfloor\| < \varepsilon(n), \\ 0, & \text{otherwise} \end{cases}$$

is a sparse generalised polynomial with $\sum_{n=0}^{N-1} f(n) \gg N^{1-c}$. Moreover, for any $a \in \mathbb{N}, b \in \mathbb{Z}$, we also have $\sum_{n=0}^{N-1} f(an+b) \gg N^{1-c}$ (where the implicit constant depends on a and b).

Remark. In particular, in the situation above, any arithmetic progression $a\mathbb{N} + b$ contains infinitely many n with $f(n) = 1$.

If $\varepsilon(n)$ is a rational power of a generalised polynomial, then it follows from V.7.2 that $f(n)$ is a generalised polynomial.

A key tool which we will use is the quantitative equidistribution theorem of Green and Tao for nilrotations I.1.24, whose immediate consequence cited below we use as a black box.

Theorem V.7.4 (Green-Tao, [54])**.** *Let $g(n)$ be a polynomial sequence on a nilmanifold G/Γ where G is connected. Fix a Mal'cev basis $\mathcal{X}$ of G. Then there exists a constant $A > 0$ such that for any $0 < \delta < 1/2$ and any N, one of the following conditions holds:*

(i) for each $x \in G/\Gamma$, there exist δN values of $0 \le n < N$ with $d_{\mathcal{X}}(g(n)\Gamma, x) \le \delta$;

(ii) there exists a horizontal character η with $\|\eta\| \ll \delta^{-A}$ such that

$$\max_{0 \le n < N} \|\eta(g(n+1)) - \eta(g(n))\| \ll \delta^{-A} N^{-1}.$$

For linear orbits $g(n) = a^n$, $a \in G$, the condition (V.7.4.ii) in Theorem V.7.4 asserts that the projection of a to the torus $G/G_2\Gamma$ satisfies an approximate linear relation over $\mathbb{Z}$. A convenient criterion which can be used to rule that out is provided by the following classical result of Schmidt. As before, we only cite the special case which we shall use.

Theorem V.7.5 (Schmidt [100])**.** *Let $\alpha_1, \ldots, \alpha_n$ be n algebraic numbers linearly independent over $\mathbb{Q}$, and let $\varepsilon > 0$. Then for $k \in \mathbb{Z}^n$ we have*

$$\left\| \sum_{i=1}^{n} k_i \alpha_i \right\| \gg \|k\|^{-n-\varepsilon}$$

where the implicit constant depends on $\alpha_1, \ldots, \alpha_n$ and ε.

Remark. In fact, we only use the above with $\varepsilon = 1$, and any bound of the form $\|\sum_{i=1}^{n} k_i\alpha_i\| \gg \|k\|^{-C}$ would work equally well for our purposes, where C is a constant dependent at most on n. We note that the same result with C dependent on the degrees of α_i is elementary. If we allowed the constant c in Proposition V.7.3 to depend on α, β, then the weaker bound would suffice for our purposes.

Proof of Proposition V.7.3. It is clear by the same argument as in the proof of Proposition V.7.2 that $f(n)$ is sparse; what remains to be proved is the bound on its growth. We only deal with $\sum_{n=0}^{N-1} f(n)$, the estimation of $\sum_{n=0}^{N-1} f(an+b)$ is analogous. We may assume assume that $\varepsilon(N) \sim N^{-c}$.

We can explicitly describe a nilmanifold on which $f(n)$ can be realised in a way analogous to Theorem I.1.22. Denote

$$[x, y, z] \equiv \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix},$$

and define

$$G = \{[x, y, z] \mid x, y, z \in \mathbb{R}\}, \qquad \Gamma\{[x, y, z] \mid x, y, z \in \mathbb{Z}\}.$$

Then, G/Γ is the Heisenberg nilmanifold with a natural choice of a Mal'cev basis $[1, 0, 0]$, $[0, 1, 0]$, $[0, 0, 1]$.

The horizontal characters on G/Γ take the form $\eta_k([x, y, z]) = k_1 x + k_2 y$ with $k = (k_1, k_2) \in \mathbb{Z}^2$ and $\|\eta_k\| = \|k\|_2$. Let g be the polynomial sequence

$$g(n) = [-n\alpha, n\beta, 0] = [-\alpha, \beta, 0]^n [0, 0, \alpha\beta]^{\binom{n}{2}},$$

so that $\{g(n)\} = [\{-n\alpha\}, \{n\beta\}, \{n\alpha \lfloor n\beta \rfloor\}]$. Let $F\colon G/\Gamma \to [0, 1)$ be the map $F([x, y, z]) = \|z\|$ for $x, y, z \in [0, 1)^3$. It follows from the way the metric on G/Γ is defined that F is Lipschitz. With this setup, $f(n) = [\![F(g(n)) < \varepsilon(n)]\!]$.

Let N be a sufficiently large integer, and put $z = [0, 0, 0]\Gamma$. There exists $\rho_N \gg \varepsilon(N)$ such that $F(B(z, \rho_N)) \subset [0, \varepsilon(N))$, where $B(z, \rho)$ denotes the ball with radius ρ centred at z. Hence, if $d(g(n)\Gamma, z) < \rho_N$ for some $0 \leq n < N$, then $f(n) = 1$. If $f(n) = 1$ for at least $N\rho_N \gg N^{1-c}$ values of $0 \leq n < N$, we are done. Suppose this is not the case. It now follows from Theorem V.7.4 that there exists a horizontal character η with $\|\eta\| \ll \rho_N^{-A} \ll N^{Ac}$ such that

$$\max_{0 \leq n < N} \|\eta(g(n+1)) - \eta(g(n))\| \ll N^{-1+Ac}, \tag{V.36}$$

where A is an absolute constant. Picking $l \in \mathbb{Z}^2$ so that $\eta = \eta_l$, we conclude from (V.36) and Theorem V.7.5 that

$$N^{-3Ac} \ll \|l\|^{-3} \ll \|-l_1\alpha + l_2\beta\| \ll N^{-1+Ac}. \tag{V.37}$$

As long as $c < 1/(4A)$, we reach a contradiction. □

Remark. The above argument is not specific to the sequence $[\![n\alpha \lfloor n\beta \rfloor < \varepsilon(n)]\!]$; in fact, the term $n\alpha \lfloor n\beta \rfloor$ could be replaced by a fairly arbitrary generalised polynomial $q(n)$. The difficulty lies in the need to impose suitable "quantative irrationality" conditions of $q(n)$.

In the situation of Example V.7.3, this was easily accomplished since we had a direct access to the representation of $q(n)$ in terms of an orbit on a nilmanifold. For more general sequences, one needs to resort to the construction in [16].

V.7.2 IP rich sequences

In light of Theorem V.4.7, any sparse but not very sparse automatic set contains an IPS set. It turns out that a slightly stronger condition of containing (a translate of) an IP set leads to interesting consequences.

In contrast to Theorem V.4.7, there are sparse automatic sets which are neither arid nor IP as will be clear from the following example. However, many natural examples of sparse but not arid automatic sets are IP.

Example V.7.6. Let $k \geq 2$, and let $\mathcal{B} \subset \Sigma_k^*$ be a finite set of "prohibited" words of length $\leq t$. A word $u \in \Sigma_k^*$ is $\mathcal{B}$-free if u contains no $b \in \mathcal{B}$ as a substring, and accordingly $n \in \mathbb{N}_0$ is $\mathcal{B}$-free if its base k expansion $(n)_k$ is $\mathcal{B}$-free. Denote by $F_{\mathcal{B}} \subset \mathbb{N}_0$ the set of $\mathcal{B}$-free integers.

(i) The set $F_{\mathcal{B}}$ is k-automatic.

(ii) If $\mathcal{B} \neq \emptyset$, then $F_{\mathcal{B}}$ is sparse.

(iii) If $\sum_{b\in\mathcal{B}} k^{-|b|} < 1/8t$, then $F_{\mathcal{B}}$ is not arid.

(iv) If each $b \in \mathcal{B}$ contains at least two non-zero digits, then $F_{\mathcal{B}}$ is IP.

(v) If some $b \in \mathcal{B}$ consists only of 0's, then $F_{\mathcal{B}}$ is not IP_+.

Proof.

(i) It is not difficult to explicitly describe the k-automaton which computes the characteristic function of $F_{\mathcal{B}}$; alternatively one may use Kleene's theorem.

(ii) A an elementary computation shows that if $t = \max_{b\in\mathcal{B}} |b|$, then

$$\frac{1}{N}\sum_{n<N} f(n) \leq 2\left(1 - \frac{|\mathcal{B}|}{k^t}\right)^{\left\lfloor \frac{\log N}{t\log k}\right\rfloor} = o(1), \text{ as } N \to \infty.$$

(iii) Construct an undirected graph $G = (V, E)$ (allowing loops), where $V = \Sigma_k^t$, and $\{u, v\} \in E$ if uv and vu are both $\mathcal{B}$-free. If $u_1, u_2, \ldots, u_r$ is a path in G, then $u_1u_2\ldots u_r$ is $\mathcal{B}$-free. If G contains a path u_1, w, u_2 of length 3 with $u_1 \neq u_2$, then for any $i_1, \ldots, i_r \in \{1, 2\}$, the word $u_{i_1}wu_{i_2}w\ldots u_{i_r}w$ is $\mathcal{B}$-free, and hence $F_{\mathcal{B}}$ is not arid (note that if $u \in \Sigma_k^*$ is $\mathcal{B}$-free then $[u]_k$ is $\mathcal{B}$-free). Thus, it remains to check that G contains a length 3 path with distinct endpoints; for the sake of contradiction suppose this is not the case.

On one hand, any vertex $u \in V$ can have at most 1 neighbour (including itself if $\{u, u\}$ is an edge), and it is an easy exercise in extremal combinatorics to show that $|E| \leq |V| = k^t$. On the other hand, given $b \in \mathcal{B}$, the number of pairs $(u, v) \in V^2$ such that b appears in uv or vu is $< 4tk^{2t-|b|}$, so

$$|E| \geq \frac{1}{2}k^{2t} - 4tk^{2t}\sum_{b\in\mathcal{B}} k^{-|b|} > \frac{1}{4}k^{2t} > k^t,$$

(note that assumption implies $k^t \geq 8$), which is the sought contradiction.

(iv) Let $n_i = k^{it}$. Then $\mathrm{FS}(n_i) \subset F_{\mathcal{B}}$.

(v) Suppose there were an IP set E and $m \in \mathbb{Z}$ with $E + m \subset F_{\mathcal{B}}$. We may assume without loss of generality that $m \geq 0$ (else, find $n \in E$ such that $E' := (E - n) \cap E$ is and IP set and $m' := m + n \geq 0$). Now, for any t, there is some $n \in E$ divisible by k^t. Hence, for sufficiently large t, the expansion of $n + m$ contains a block of $t - \lceil \log_k m \rceil$ zeros, contradicting the assumption that $E + m \subset F_{\mathcal{B}}$.

□

Example V.7.7. The sequence

$$f_{\mathrm{Fib}}(n) = \begin{cases} 0, & \text{if binary expansion of } n \text{ contains } 11, \\ 1, & \text{otherwise} \end{cases}$$

is 2-automatic, and is the characteristic function of a set which is sparse, not arid, and IP.

Example V.7.8 (Baum-Sweet)**.** The Baum-Sweet sequence ([6]), given by

$$f_{\mathrm{BS}}(n) = \begin{cases} 0, & \text{if binary expansion of } n \text{ contains } 10^l1 \text{ with } l \in 2\mathbb{N}_0 + 1, \\ 1, & \text{otherwise} \end{cases}$$

is 2-automatic, and is the characteristic sequence of a set which is sparse, not arid, and IP.

In dealing with polynomials, it is often useful to exploit the fact that sufficiently high discrete derivatives vanish. More precisely, if q is a polynomial of degree at most $d-1$, then for any $n_1, \ldots, n_d$ we have

$$\sum_{\alpha \subseteq \{1,\ldots,d\}} (-1)^{|\alpha|} q\left(n_\alpha\right) = 0,$$

with the usual convention $n_\alpha = \sum_{i\in\alpha} n_i$.

A similar relation holds also for generalised polynomials. Such a relation is stated in [19, Theorem 2.42] in terms of limits along ultrafilters. Below, we give a statement purely in terms of IP limits, which is easily deduced from [19] or proven directly.

Theorem V.7.9. *For any generalised polynomial q and an* IP *ring $\mathcal{G}_0$, there exist $\lambda \in \mathbb{R}$, $d \in \mathbb{N}_0$ (equal to the degree of q), and families of* IP *subrings $\mathcal{G}_1$, $\mathcal{G}_2(\beta_1), \ldots, \mathcal{G}_d(\beta_1, \ldots, \beta_{d-1})$ of $\mathcal{G}_0$, $\beta_i \in \mathcal{F}$, such that*

$$\sum_{\emptyset \neq \alpha \subseteq \{1,\ldots,d\}} (-1)^{|\alpha|} q\left(\sum_{i\in\alpha} n_{\beta_i}\right) = -\lambda \tag{V.38}$$

whenever $\beta_1 \in \mathcal{G}_1,\ \beta_2 \in \mathcal{G}_2(\beta_1),\ \ldots,\ \beta_d \in \mathcal{G}_d(\beta_1, \ldots, \beta_{d-1})$.

Using the above relation, we are able to prove a special case of Theorem V.1.1 with negligible effort.

Proposition V.7.10. *Let $q(n) \geq 0$ be a generalised polynomial, $\varepsilon(n) \geq 0$ with $\varepsilon(n) \to 0$ as $n \to \infty$, and let g be given by* (V.33)*. Then there exists no* IP *set E such that $g(n) = 1$ for $n \in E$.*

Proof of Proposition V.7.10. Suppose that $(n_\alpha)_{\alpha\in\mathcal{F}}$ were an IP set such that $g(n_\alpha) = 1$ for $\alpha \in \mathcal{F}$. Apply Theorem V.7.9 to find $\lambda \in \mathbb{R}$ and families of IP rings $\mathcal{G}_j(\beta_1, \ldots, \beta_{j-1})$ such that

$$\sum_{\emptyset \neq \alpha \subseteq \{1,\ldots,d\}} (-1)^{|\alpha|} q\left(\sum_{i\in\alpha} n_{\beta_i}\right) = -\lambda \tag{V.39}$$

for $\beta_1 \in \mathcal{G}_1,\ \beta_2 \in \mathcal{G}_2(\beta_1),\ \ldots,\ \beta_d \in \mathcal{G}_d(\beta_1, \ldots, \beta_{d-1})$. Passing to the IP limit with $\beta_d, \ldots, \beta_2$ (in this order), we obtain $-q(n_{\beta_1}) = -\lambda$. This proves that

$$0 < \lambda < \varepsilon(n_{\beta_1}),$$

which leads to a contradiction. □

Corollary V.7.11. *If $p(x) \in \mathbb{R}[x]$ has at least one irrational coefficient and $\varepsilon(n) \to 0$, then the set $\{n \in \mathbb{N} \mid \|p(n)\| < \varepsilon(n)\}$ does not contain an* IP *set.*

While the property of being IP is arguably somewhat rare among sparse automatic sets, recall that Proposition V.4.10 asserts that an automatic set is IP_+ under some relatively mild assumptions. We are now ready to prove Proposition V.4.11.

Proof of Proposition V.4.11. Suppose that the set E given by were k-automatic. Then, by Proposition V.7.3 the set E would satisfy the assumptions of Proposition V.4.10, provided that c is taken small enough. Thus, E would contain a translate of an IP set, contradicting Proposition V.7.10. □

V.7.3 Very sparse sequences

It is difficult to say anything substantial about generalised polynomial sequences which are extremely sparse. Indeed, as we presently show, any sufficiently sparse set is generalised polynomial.

We will need the following elementary lemma. We thank Aled Walker for simplifying our argument.

Lemma V.7.12. *For any $\delta > 0$, there exists $q_0 = q_0(\delta)$ such that for all $q \geq q_0$ and all $x \geq 0$, the interval $[x, x + q^\delta)$ contains an integer n coprime to q.*

Proof. Using an inclusion-exclusion argument, we find that

$$\sum_{\substack{x \leq n < x + q^\delta \\ (n,q)=1}} 1 = \sum_{d \mid q} \left(\mu(d) \frac{q^\delta}{d} + O(1) \right) = q^\delta \sum_{d \mid q} \frac{\mu(d)}{d} + O(\tau(q)) = q^\delta \frac{\varphi(q)}{q} + O(\tau(q)), \tag{V.40}$$

where μ is the Möbius function, φ is the Euler totient function, and τ is the number of divisors function. Using the standard estimates $\varphi(q) \gg q^{1-\delta/3}$ and $\tau(q) \ll q^{\delta/3}$, we conclude that the expression in (V.40) is positive for q large enough. □

Proposition V.7.13. *Let C, D be integers satisfying $5 \leq C < D \leq \frac{1}{2}(C-1)^2$. Suppose that the sequence $(n_i)_{i\geq 0}$ of integers satisfies $n_0 \geq 2$ and $n_i^D < n_{i+1} < n_i^{2D}$ for $i \geq 0$. Then there exists $\alpha \in (0,1)$ such that the symmetric difference of the sets $E = \{n_i \mid i \geq 0\}$ and*

$$E' = \left\{ n \in \mathbb{N} \;\middle|\; \frac{1}{4}n^{-C+1} \leq \|n\alpha\| \leq \frac{1}{2}n^{-C+1} \right\}$$

is finite.

Proof. We inductively construct a sequence of integers $(m_i)_{i\geq 0}$ such that the intervals $I_i = [\frac{m_i}{n_i} + \frac{1}{4}n_i^{-C}, \frac{m_i}{n_i} + \frac{1}{2}n_i^{-C}]$ are nested (i.e. $I_{i+1} \subset I_i$) and moreover $(m_i, n_i) = 1$ for sufficiently large i. Choose $m_0 = 0$. If m_i has been constructed, then the possible choices of m_{i+1} such that the nesting condition is satisfied form an interval of length $\gg n_{i+1}/n_i^C \gg n_{i+1}^{1-C/D}$. Because $C < D$, by Lemma V.7.12, for sufficiently large i there is a choice of m_{i+1} with $(m_{i+1}, n_{i+1}) = 1$.

Once the sequence $(m_i)_{i\geq 0}$ has been constructed, pick α so that $\{\alpha\} = \bigcap_{i\geq 0} I_i$. By construction, for this choice of α we have $E \subseteq E'$. Let $\alpha = [0; a_1, a_2, \dots]$ be the continued fraction expansion of α, and let $\frac{p_j}{q_j} = [0; a_1, \dots, a_j]$ be the corresponding convergents. We recall an estimate for the error term of the approximation: $\frac{1}{2}\frac{1}{q_j q_{j+1}} \leq |\alpha - \frac{p_j}{q_j}| \leq \frac{1}{q_j q_{j+1}}$. Since $C \geq 2$, the classical theorem of Legendre implies that any approximation $\frac{m}{n}$ with $(m,n) = 1$ and $|\alpha - \frac{m}{n}| \leq \frac{1}{2}n^{-C}$ is equal to $\frac{p_j}{q_j}$ for some j. Furthermore, in this case $q_{j+1} \geq q_j^{C-1}$. In particular, for each i, there is $j(i)$ such that $\frac{m_i}{n_i} = \frac{p_{j(i)}}{q_{j(i)}}$ and $q_{j(i)+1} \geq q_{j(i)}^{C-1}$.

Suppose now that $n \in E' \setminus E$. Then there exists m (not necessarily coprime to n) with $\frac{1}{4}n^{-C} \leq |\alpha - \frac{m}{n}| \leq \frac{1}{2}n^{-C}$, whence $\frac{m}{n} = \frac{p_j}{q_j}$ for some j. If $\frac{m}{n}$ were equal to $\frac{m_i}{n_i}$ for some i sufficiently large that $(n_i, m_i) = 1$, then contradiction would follow immediately: $n^{-C} \leq \frac{1}{4}n_i^{-C} \leq |\alpha - \frac{m_i}{n_i}| = |\alpha - \frac{m}{n}| \leq \frac{1}{2}n^{-C}$. The cases when q_j is bounded account for finitely many possible values of n. Hence, we may assume that j lies strictly between $j(i)$ and $j(i+1)$ for some i which is sufficiently large so that $(m_k, n_k) = 1$ for $k \geq i$. We then have the inequalities

$$q_j \geq q_{j(i)+1} \geq q_{j(i)}^{C-1} = n_i^{C-1}$$

and

$$n_{i+1} = q_{j(i+1)} \geq q_{j+1} \geq q_j^{C-1} \geq n_i^{(C-1)^2},$$

which imply that $n_i^{2D} > n_{i+1} \geq n_i^{(C-1)^2}$, contradicting $2D \leq (C-1)^2$. □

We are now ready to prove another one of our main results, whose formulation we reiterate.

Theorem (V.1.8)**.** *For any sequence of positive integers* $(n_i)_{i\geq 0}$*,* $n_i \geq 2$ *such that*

$$\liminf_{i\to\infty} \frac{\log n_{i+1}}{\log n_i} > 1,$$

the set $E = \{n_i \mid i \in \mathbb{N}_0\}$ *is generalised polynomial.*

Proof. Observe first that for a positive integer k we have

$$\liminf_{i\to\infty} \frac{\log n_{i+k}}{\log n_i} \geq k \liminf_{i\to\infty} \frac{\log n_{i+1}}{\log n_i},$$

and therefore by slicing the sequence $(n_i)_{i\geq 0}$ into the subsequences $(n_{ki+j})_{i\geq 0}$ for $j = 0, 1, \ldots, k-1$, we may assume that the value of $\liminf_{i\to\infty} \frac{\log n_{i+1}}{\log n_i}$ is arbitrarily large.

We now construct a sequence $(n'_j)_{j\geq 0}$ of integers with $n_j'^6 < n'_{j+1} < n_j'^{12}$ for sufficiently large j and such that $(n_i)_{i\geq 0}$ is its subsequence. We construct n'_j inductively, keeping track of $j(i)$ such that $n'_{j(i)} = n_i$. For $i = 0$, pick $j(0) = 0$ and $n'_0 = n_0$. If $j(i)$ and $(n'_j)_{0\leq j\leq j(i)}$ have been constructed, we define $j(i+1) = j(i) + l$ and $n'_{j(i)+k} = \left\lfloor \exp((\log n_{i+1})^{k/l} (\log n_i)^{1-k/l}) \right\rfloor$ for $0 \leq k \leq l$, where l remains to be chosen. This ensures that $\frac{\log n'_{j(i)+k+1}}{\log n'_{j(i)+k}} = \left(\frac{\log n_{i+1}}{\log n_i}\right)^{1/l} + o(1)$ as $i \to \infty$. Putting $A = \frac{\log n_{i+1}}{\log n_i}$, and letting l be any integer in the interval $(\frac{\log A}{\log 11}, \frac{\log A}{\log 7})$, we obtain (for sufficiently large i) the inequality $6 < \frac{\log n'_{j(i)+k+1}}{\log n'_{j(i)+k}} < 12$. This interval indeed contains at least one integer (for sufficiently large i), provided that the value of $\liminf_{i\to\infty} \frac{\log n_{i+1}}{\log n_i}$ is large enough.

Applying Proposition V.7.13 with $C = 5, D = 6$, we conclude that the set $E' = \left\{n'_j \;\middle|\; j \geq 0\right\}$ is generalised polynomial. By the same argument, the set $E'' = \left\{n''_j \;\middle|\; j \geq 0\right\}$ is also generalised polynomial, where $n''_{j(i)} = n_i$ and $n''_j = n'_j + 1$ if $j \neq j(i)$ for all i. It follows that the set $E' \cap E''$, which coincides with E except possibly for a finite set, is generalised polynomial, as required. □

V.8 Exponential sequences

In this section, we deal with sequences with exponential growth, such as the Fibonacci or Tribonacci numbers. Recall that in the case of the Fibonacci numbers, the corresponding set is generalised polynomial and not automatic. It is natural to ask to what extent this behaviour extends to other sequences of similar type. Specifically, we show that there is no sequence obeying a linear recurrence such that the set of its values is automatic, except the trivial examples of bounded sequences. Secondly, we show that a class of sequences obeying a third order recurrence relation gives rise to generalised polynomial sets.

V.8.1 Automaticity of recursive sequences

Proposition V.8.1. *Let $(a_m)_{m\geq 0}$ be an $\mathbb{N}$-valued sequence satisfying a linear recurrence of the form*

$$a_{m+D} = \sum_{i=1}^{D} c_i a_{m+D-i}, \quad m \geq 0 \tag{V.41}$$

with integer coefficients c_i. Suppose that $E = \{a_m \mid m \in \mathbb{N}_0\}$ is k-automatic for some k. Then E is a finite union of the following standard sets: linear progressions $\{am + b \mid m \in \mathbb{N}_0\}$, exponential progressions $\left\{ak^{dm} + b \;\middle|\; m \in \mathbb{N}_0\right\}$ and finite sets.

Proof. We first claim there exists a representation of E as a finite union

$$E = \bigcup_{i=1}^{K_{\text{lin}}} L_i \cup \bigcup_{i=1}^{K_{\text{poly}}} P_i \cup \bigcup_{i=1}^{K_{\text{exp}}} E_i \cup F, \tag{V.42}$$

where F is finite, $L_i = \{a_i m + b_i \mid m \in \mathbb{N}_0\}$ are arithmetic progressions, $P_i = \{p_i(m) \mid m \in \mathbb{N}_0\}$ are value sets of polynomials $p_i(x) \in \mathbb{Z}[x]$ with $\deg p_i \geq 2$, and E_i have exponential growth in the sense that $|E_i \cap [N]| \ll \log N$.

In order to prove this claim, begin by noting that any restriction of (a_m) to an arithmetic progression $a_m^{(h,r)} = a_{hm+r}$ obeys some (minimal length) linear recurrence

$$a_{m+D'}^{(h,r)} = \sum_{i=1}^{D'} c_i^{(h,r)} a_{m+D'-i}^{(h,r)}, \quad m \in \mathbb{N}_0$$

with $D' = D'(h,r) \leq D$. Moreover, there exists a choice of h such that each of that each $c_m^{(h,r)}$ is either identically zero or non-degenerate, in the sense that the associated characteristic polynomial $q^{(h,r)}(x) = x^{D'} - \sum_{i=1}^{D'} c_i^{(h,r)} x^{D'-i}$ has no pair of roots $\lambda, \mu \in \mathbb{C}$ such that λ/μ is a root of unity (see e.g. [36, Thm 1.2] for a much stronger statement). Hence, for the purpose of showing the existence of representation (V.42), we may assume that (a_m) is non-degenerate. Suppose also that D is minimal, and let $\lambda_1, \ldots, \lambda_r$ be the roots of $q(x) = x^D - \sum_{i=1}^{D} c_i x^{D-i}$, with $|\lambda_1| \geq |\lambda_2| \geq \ldots$. Note that $|\lambda_1| \geq 1$.

If $|\lambda_1| > 1$, then by the result of Evertse [37] and van der Poorten and Schlickewei [106] (see [36, Thm 2.3]), we have $a_m = |\lambda_1|^{m+o(m)}$ as $m \to \infty$. Hence, E has exponential growth, and we are done.

Otherwise, if $|\lambda_1| = 1$, then for all j we have $|\lambda_j| = 1$ or $\lambda_j = 0$. Kronecker's theorem [77] (or a standard Galois theory argument) shows that if λ is an algebraic

integer all of whose conjugates have absolute value 1, then λ is a root of unity. Using the general formula for the solution of a linear recurrence, we may write:

$$a_m = \sum_{j=1}^{r} \lambda_j^m p_j(m) = \sum_{j=1}^{r} b_j(m) p_j(m),$$

where $p_j(x)$ are polynomials and $b_j(m)$ are periodic. Splitting $\mathbb{N}_0$ into arithmetic progressions where $b_j(m)$ are constant, we conclude that E is a finite union of value sets of polynomials. This again produces a representation of the form (V.42).

Such a representation is not unique. Splitting P_i into a finite number of subprogressions and discarding those which are redundant, we may assume that $P_i \cap L_j = \emptyset$ for any i, j. Likewise, we may assume that $E_i \cap L_j = F \cap F_j = \emptyset$ for any i, j. Fix one representation subject to these restrictions. The set

$$E' = \bigcup_{i=1}^{K_{\mathrm{poly}}} P_i \cup \bigcup_{i=1}^{K_{\mathrm{exp}}} E_i \cup F = E \setminus \bigcup_{i=1}^{K_{\mathrm{lin}}} L_i$$

is again k-automatic; it will suffice to show that E' is a union of the standard sets mentioned above.

We claim that $K_{\mathrm{poly}} = 0$, i.e. the representation of E uses no polynomial progressions of degree ≥ 2. Suppose for the sake of contradiction that $P = \{p(m) \mid m \in \mathbb{N}_0\}$ appears is one of the sets P_i, and write $p(m) = \sum_{i=0}^{d} c_i m^i$, where $c_i \in \mathbb{Z}$. Replacing $p(m)$ with $p(m+r)$ for a suitably chosen $r \in \mathbb{N}_0$, we may assume that $c_i > 0$ for $0 \leq i \leq d$. For sufficiently large t, we have $p(k^t) = [u_d 0^{t-t_0} u_{d-1} 0^{t-t_0} u_{d-2} \dots u_1 0^{t-t_0} u_0]_k$, where t_0 is a constant and u_i is the base k expansion of c_i, padded by 0's so as to have $|u_i| = t_0$. Since $p(k^t) \in E'$, from the Pumping Lemma V.2.8 it follows that there is $l \in \mathbb{N}$ such that for any $s_1, \dots, s_d \in \mathbb{N}$ it holds that

$$n(s_1, \dots, s_d) := [u_d 0^{ls_d} u_{d-1} 0^{ls_{d-1}} \dots u_1 0^{ls_1} u_0]_k \in E'.$$

For sufficiently large S and an absolute constant δ to be determined later, consider the set

$$Q(S) = \{n(s_1, \dots, s_d) \mid s_i \in \mathbb{N}, s_1 + \dots + s_d = S,\ s_2 + \dots + s_d \leq \delta S\},$$

and put $N(S) := n(S - d + 1, 1, \dots, 1) = \min Q(S)$. Note that $N(S) = k^{lS+O(1)}$ and that $\max Q(S) = N(S) + O(N(S)^\delta)$. For fixed T_0 and $T \to \infty$, we shall consider the cardinality of the set $Q(T_0, T) = \bigcup_{T_0 \leq S \leq T} Q(S)$. By an elementary counting argument, we find

$$|Q(T_0, T)| \gg T^d \gg T^2. \tag{V.43}$$

To obtain an upper bound, we separately estimate $|Q(S) \cap P_i|$ and $|Q(T_0, T) \cap E_j|$ for each i, j.

Suppose that $n, n' \in Q(S) \cap P_i$ with $n' > n$, so in particular $n = p_i(m)$ and $n' = p_i(m')$ for some $m, m' \gg N(S)^{1/\deg p_i}$. We then have the chain of inequalities:

$$N(S)^{\delta} \gg n' - n = p_i(m') - p_i(m) \geq \min_{x \in [m,m']} |p_i'(x)| \gg N(S)^{\frac{\deg p_i - 1}{\deg p_i}},$$

which is a contradiction for sufficiently large S, provided that $\delta < \frac{\deg p_i - 1}{\deg p_i}$ (which will hold if we put $\delta = \frac{1}{3}$). Thus, $|Q(S) \cap P_i| \leq 1$.

As for $Q(T_0, T) \cap E_j$, from the bounds on growth of E_j we immediately have

$$\left| \bigcup_{T_0 \leq S \leq T} Q(T_0, T) \cap E_i \right| \ll |E_i \cap [2N(T)]| \ll T. \tag{V.44}$$

In total, using (V.43) and (V.44) we find that

$$|Q(T_0, T)| \leq \sum_{S=T_0}^{T} \sum_{i=1}^{K_{\text{poly}}} |Q(S) \cap P_i| + \sum_{i=1}^{K_{\text{exp}}} |Q(T_0, T) \cap E_i| + O(1) \ll T, \tag{V.45}$$

contradicting the previously obtained bound. It follows that indeed $K_{\text{poly}} = 0$.

Since E' contains no polynomial or linear progressions, we have $|E' \cap [N]| \ll \log N$. It follows from Theorem V.4.7 (cf. also Corollary V.6.4 and Lemma V.4.3) that E' must be arid of rank 1 and base k. Since all basic arid sets of rank 1 are of the form described in the formulation of the theorem, we are done. □

V.8.2 Exponentially sparse generalised polynomial sets

In order to construct a generalised polynomial that takes value zero exactly on the set of Fibonacci numbers, we used the theory of continued fractions. In this section, we will do this more generally for certain linear recurrence sequences of degree 2 and 3. We need to begin, however, by the following basic result.

Proposition V.8.2. *Let β be a Pisot number (or a Pisot–Vijayaraghavan number), i.e. a real algebraic integer $\beta > 1$ such that all its Galois conjugates have absolute value smaller than one. Let $P(x) = x^d - c_1 x^{d-1} - \ldots - c_d$ be the minimal polynomial of β. Let $(R_i)_{i \geq 0}$ be a sequence of integers satisfying the linear recurrence relation*

$$R_{i+d} = c_1 R_{i+d-1} + \ldots + c_d R_i, \quad i \geq 0.$$

Assume that $(R_i)_{i \geq 0}$ is not identically zero. Then the following conditions are equivalent:

(i) The set $\{R_i \mid i \in \mathbb{N}_0\}$ *is generalised polynomial.*

(ii) The set $\{\langle\!\langle \beta^i \rangle\!\rangle \mid i \in \mathbb{N}_0\}$ *is generalised polynomial.*

In particular, the question whether the set $\{R_i \mid i \in \mathbb{N}_0\}$ *is generalised polynomial depends only on the recurrence relation and not on the initial values of* $(R_i)_{i\geq 0}$.

Proof. From the form of the linear recurrence relation for $(R_i)_{i\geq 0}$, we see that $R_i = u\beta^i + o(1)$ for some $u \in \mathbb{R}$. Note that $u \neq 0$. In fact, otherwise $R_i = o(1)$ and since R_i takes integer values we see that $R_i = 0$ for large i. It follows that R_i is identically zero (note that $c_d \neq 0$).

Since β is an algebraic integer, we know that $\sum_{\beta'}(\beta')^i$ is an integer, the sum being taken over all the Galois conjugates β' of β. Since $|\beta'| < 1$ for $\beta' \neq \beta$, we see that $\beta^i = \langle\!\langle \beta^i \rangle\!\rangle + o(1)$. It follows that

$$R_i = u\langle\!\langle \beta^i \rangle\!\rangle + o(1)$$

and the claim follows immediately from the following lemma. □

Lemma V.8.3. *Let* $u \in \mathbb{R}$, $u \neq 0$, *and let* $(R_i)_{i\geq 0}$ *and* $(S_i)_{i\geq 0}$ *be integer-valued sequences such that*

$$R_i = uS_i + o(1).$$

Then the set $\{R_i \mid i \in \mathbb{N}_0\}$ *is generalised polynomial if and only if the set* $\{S_i \mid i \in \mathbb{N}_0\}$ *is generalised polynomial.*

Proof. It follows from the relation $R_i = uS_i + o(1)$ that for sufficiently large i, an integer m is of the form $m = S_i$ if and only if $\langle\!\langle um \rangle\!\rangle = R_i$ and $\|um\| < |u|/2$. We immediately see from this that if the set $\{R_i \mid i \in \mathbb{N}_0\}$ is generalised polynomial, then so is $\{S_i \mid i \in \mathbb{N}_0\}$. Since the argument is symmetric with respect to (R_i) and (S_i), we conclude the claim. □

V.8.3 Quadratic Pisot numbers

In this subsection we will generalise Example V.7.1, where we have shown that the set of Fibonacci numbers is generalised polynomial.

Proposition V.8.4. *Let* β *be a quadratic Pisot unit, i.e. a quadratic Pisot number of (field) norm* ± 1. *Then the set* $\{\langle\!\langle \beta^i \rangle\!\rangle \mid i \in \mathbb{N}_0\}$ *is generalised polynomial.*

Proof. Let β' be the Galois conjugate of β. Since β is a quadratic Pisot unit, it satisfies the equation $\beta^2 - a\beta - 1 = 0$ for $a \geq 1$ (if the norm of β is -1) or the equation $\beta^2 - a\beta + 1 = 0$ for $a \geq 3$ (if the norm of β is 1). In Example V.7.1, we have constructed a (non-zero) sequence (n_i) satisfying the recurrence $n_{i+2} = an_{i+1} + n_i$ for $i \geq 0$ and such that its set of values is generalised polynomial. Hence in the case when the norm of β is -1, the claim follows from Proposition V.8.2.

If the norm of β is 1, we need to argue a bit more carefully. The continued fractions expansion of β is $\beta = [a-1; \overline{1, a-2}]$. Let $(q_i)_{i\geq 0}$ be the sequence of denominators of convergents of this continued fraction. We get the recurrence relations $q_0 = q_1 = 1$ and

$$q_{2i+2} = (a-2)q_{2i+1} + q_{2i},$$
$$q_{2i+3} = q_{2i+2} + q_{2i+1}.$$

By induction, we get from this the relations

$$q_{2i+4} = aq_{2i+2} + q_{2i},$$
$$q_{2i+5} = aq_{2i+3} + q_{2i+1}.$$

Therefore, the sequences q_{2i} and q_{2i+1} satisfy the same linear recurrence relations as the sequence β^i.

Consider the set $E' = \{n \in \mathbb{N} \mid \|n\beta\| < 1/(2n)\}$. As in Example V.7.1, we argue that E' is generalised polynomial. By the well-known error formula for the approximation by convergents (cf. e.g. [35, (3.12)]), we get that

$$\frac{q_i(q_{i+2} - q_i)}{q_{i+1}q_{i+2}} < q_i \|q_i\beta\| < \frac{q_i}{q_{i+1}}.$$

Applying the recurrence relations above together with the bounds $q_{i+2} \geq aq_i$, $q_{2i+2} \geq (a-2)q_{2i+1}$ (that also follow from the recurrence relations), we get the inequalities

$$\frac{a-2}{a} < q_{2i} \|q_{2i}\beta\|$$

and

$$\frac{1}{a} < q_{2i+1} \|q_{2i+1}\beta\| < \frac{1}{a-2}.$$

Assume now that $a \geq 4$. Applying Legendre's Theorem and arguing as in Example V.7.1, we obtain the inclusions

$$\{q_{2i+1} \mid i \in \mathbb{N}_0\} \subset E' \subset \{kq_{2i+1} \mid i \in \mathbb{N}_0, 1 \leq k \leq \lfloor\sqrt{a/2}\rfloor\}.$$

Since $\lfloor\sqrt{a/2}\rfloor^2 \le a/2 < a-1$ and $q_{2i+3} \ge (a-1)q_{2i+1}$, we see that the set

$$E'' = E' \setminus \bigcup_{k=2}^{\lfloor\sqrt{a/2}\rfloor} kE'$$

coincides with the set $\{q_{2i+1} \mid i \in \mathbb{N}_0\}$. Furthermore, it is clear that the set E'' is generalised polynomial. Thus the set $\{q_{2i+1} \mid i \in \mathbb{N}_0\}$ is generalised polynomial and the claim follows from Proposition V.8.2.

It remains to treat the case $a = 3$. In this case $\beta^2 = 3\beta - 1$, and hence $(\beta^2)^2 = 7\beta^2 - 1$. Applying the previous case to β^2, we see that the set $\{\langle\langle \beta^{2i} \rangle\rangle \mid i \in \mathbb{N}_0\}$ is generalised polynomial. By Proposition V.8.2, so is the set $\{\langle\langle \beta^{2i+1} \rangle\rangle \mid i \in \mathbb{N}_0\}$ (since $\langle\langle \beta^{2i+1} \rangle\rangle$ and $\langle\langle \beta^{2i} \rangle\rangle$ satisfy for large i the same linear recurrence). Thus, our claim holds in this case as well. □

V.8.4 Cubic Pisot numbers

In order to replace the Fibonacci numbers by a linear recurrence sequence of higher degree, we need to use higher dimensional Diophantine approximation. We briefly review the topic below. Let θ be a point in $\mathbb{R}^d$, $d \ge 1$, and let N denote a norm on $\mathbb{R}^d$. Let

$$N_0(\theta) = \inf_{p \in \mathbb{Z}^d} N(\theta - p)$$

denote the distance from θ to a nearest lattice point. We say that an integer $q \ge 1$ is a *best approximation* of θ with respect to the norm N if $N_0(q\theta) < N_0(k\theta)$ for $1 \le k \le q-1$. If $\theta \notin \mathbb{Q}^d$, there are infinitely many best approximations of θ. We order them in an increasing sequence $(q_i)_{i \ge 0}$, $1 = q_0 < q_1 < \dots$. We call this sequence the *sequence of best approximations*. For $d = 1$, the sequence is well known as the sequence of denominators of convergents of the continued fraction expansion of θ (in the latter sequence the first term possibly appears twice). For $d \ge 2$ and general $\theta \in \mathbb{R}^d$ there is no satisfactory theory of best approximations. However, the problem has been studied for specific choices of θ, notably by Lagarias [78], Chekhova-Hubert-Messaoudi [26], Chevallier [27], and Hubert-Messaoudi [69] in the case when the coordinates of θ lie in some cubic number fields. We follow the presentation of [69].

Consider the polynomial $P(x) = x^3 - ax^2 - bx - 1$ with integer coefficients a, b satysfying ($a \ge 0$ and $0 \le b \le a+1$) or ($a \ge 2$ and $b = -1$). Assume further that the polynomial P has a unique real root $\beta > 1$ and a pair of complex conjugate roots $\alpha, \bar{\alpha}$. (The condition on a and b arises from the work of Akiyama [2] who classified

Pisot units of degree 3 satisfying the so-called Property (F) concerning finiteness of certain β-expansions.) Then β generates an imaginary cubic Pisot field. We will show that the set $\{\langle\langle \beta^i \rangle\rangle \mid i \in \mathbb{N}_0\}$ is generalised polynomial. To this end, we use the results of Hubert and Messaoudi who studied the best approximations of the point $\theta = (\beta^{-1}, \beta^{-2})$ [69, Theorem 1].

Theorem V.8.5 ([69])**.** *Let the polynomial $P(x) = x^3 - ax^2 - bx - 1$ with roots $\beta, \alpha, \bar{\alpha}$ be as above. Define the sequence $(R_i)_{i\geq 0}$ by the recurrence formula $R_0 = 1$, $R_1 = a$, $R_2 = a^2 + b$,*

$$R_i = aR_{i-1} + bR_{i-2} + R_{i-3}, \quad i \geq 3.$$

Then there exists a norm N on $\mathbb{R}^2$ (called a Rauzy norm*) such that the sequence of best approximations of the point $\theta = (\beta^{-1}, \beta^{-2})$ with respect to the norm N coincides except for finitely many elements with the sequence (R_i). Furthermore, there exists a sequence of real numbers $(m_q)_{q\geq 1}$ satisfying the following properties:*

(i) There exists a constant $c > 0$ such that for all $q \geq 1$ and $p \in \mathbb{Z}^2$ if $N(q\theta - p) < c$, then

$$N(q\theta - p) = N_0(q\theta) = m_q.$$

(ii) If $q < R_i$, then $m_{R_i} < m_q$.

(iii) For $i \geq 0$ we have $m_{R_i} = m_1 |\alpha|^n$.

(iv) We have $\liminf_{q\to\infty} m_q = 0$.

(v) We have

$$\beta^i = R_i + \frac{b\beta + 1}{\beta^2} R_{i-1} + \frac{1}{\beta} R_{i-2}, \quad i \geq 2.$$

(vi) The norm N is given by the formula

$$N(x) = |(\alpha + b/\beta)x_1 + x_2/\beta|, \quad x = (x_1, x_2) \in \mathbb{R}^2.$$

Proof. This is proven in [69, Theorem 1, Theorem 3, Corollary 2, Proposition 7, Lemma 4, Lemma 2, and formula (3)]. The sequence m_q is denoted $N(\delta(q))$ there. □

Remark. The choice of the initial values of (R_i) might seem unnatural. Note however that it corresponds to $R_0 = 1$, $R_{-1} = R_{-2} = 0$.

We are now ready to prove the following theorem.

Theorem V.8.6. *Let P be a polynomial $P(x) = x^3 - ax^2 - bx - 1$ with integer coefficients a, b satisfying ($a \geq 0$ and $0 \leq b \leq a+1$) or ($a \geq 2$ and $b = -1$). Assume that P has a unique real root β and that $\beta > 1$. Then the set $\{\langle\langle \beta^i \rangle\rangle \mid i \in \mathbb{N}_0\}$ is generalised polynomial.*

Proof. By Proposition V.8.2, instead of $\langle\langle \beta^i \rangle\rangle$, we may study the sequence R_i of the previous theorem. It is easy to verify that the sequence (R_i) is strictly increasing for $i \geq 6$. We will start by constructing a generalised polynomial $g\colon \mathbb{Z} \to \mathbb{R}$ that is increasing and which takes the value $g(q) = m_q^{-2}$ at points $q = R_i$ with i sufficiently large. The sequence R_i satisfies a third order linear recurrence with characteristic polynomial P and hence

$$R_i = u\beta^i + v\alpha^i + w\bar{\alpha}^i$$

for some $u, v, w \in \mathbb{C}$. Hence, $\lim_{i\to\infty}(R_i - u\beta^i) = 0$, and thus for i large enough we have

$$R_{i-1} = \langle\langle R_i/\beta \rangle\rangle, \quad R_{i-2} = \langle\langle R_i/\beta^2 \rangle\rangle.$$

Define a generalised polynomial g by the formula

$$g(q) = m_1^{-2}\left(q + \frac{b\beta+1}{\beta^2}\langle\langle q/\beta \rangle\rangle + \frac{1}{\beta}\langle\langle q/\beta^2 \rangle\rangle\right), \quad q \in \mathbb{Z}.$$

Then by Theorem V.8.5.(V.8.5.iii) and V.8.5.(V.8.5.v), we have $g(R_i) = m_1^{-2}|\beta|^i = m_{R_i}^{-2}$ for large i. (Note that $|\alpha| = |\beta|^{-1/2}$.) Furthermore, g is increasing.

Consider the set

$$S = \{q \geq 1 \mid N_0(q\theta) < g(q)^{-1/2}\}.$$

Since the sequence $\{R_i \mid i \in \mathbb{N}_0\}$ is up to finitely many terms the sequence of best approximations of θ, applying Theorem V.8.5.(V.8.5.i) and V.8.5.(V.8.5.ii) and the fact that g is increasing and $g(R_i) = m_{R_i}^{-2}$ tends to 0 as $i \to \infty$, we see that S coincides with the set $\{R_i \mid i \in \mathbb{N}_0\}$ up to a finite set. In order to show that S is generalised polynomial, it is enough to write the expression $N_0(q\theta)$ in terms of a generalised polynomial in q. We will do that under the additional assumption that $N_0(q\theta)$ is small. Assume that $N_0(q\theta) < \mathrm{Im}(\alpha)/2$ and let $p = (p_1, p_2) \in \mathbb{Z}^2$ be such that $N_0(q\theta) = N(q\theta - p)$. By Theorem V.8.5.(V.8.5.vi), we have

$$N(q\theta - p) = |(\alpha + b/\beta)(q\beta^{-1} - p_1) + (q\beta^{-2} - p_2)/\beta| < \mathrm{Im}(\alpha)/2.$$

By looking at the imaginary part of the expression, it follows that $|q\beta^{-1} - p_1| < 1/2$, and hence p_1 is uniquely determined as $p_1 = \langle\langle q\beta^{-1} \rangle\rangle$. Since p minimises $N(q\theta - p)$, we also have

$$p_2 = \langle\langle \beta(\alpha + b/\beta)(q\beta^{-1} - p_1) + q\beta^{-2} \rangle\rangle.$$

Therefore we have

$$\begin{aligned} N_0(q\theta) \leq & |(\alpha + b/\beta)(q\beta^{-1} - \langle\!\langle q\beta^{-1} \rangle\!\rangle) + \\ & + (q\beta^{-2} - \langle\!\langle \beta(\alpha + b/\beta)(q\beta^{-1} - \langle\!\langle q\beta^{-1} \rangle\!\rangle) + q\beta^{-2} \rangle\!\rangle)/\beta| \end{aligned}$$

and equality holds if furthermore $N_0(q\theta) < \mathrm{Im}(\alpha)/2$. Call the expression on the right hand side of this inequality $h(q)$. Rewriting the norm in $\mathbb{C}$ in terms of the coordinates, we see that $h(q)^2$ is given by a generalised polynomial. Therefore the set

$$T = \{q \geq 1 \mid h(q)^2 < g(q)^{-1}\}$$

coincides with S up to a finite set. Hence, the set $\{R_i \mid i \in \mathbb{N}_0\}$ is generalised polynomial. □

V.9 Concluding remarks

We conclude with some general remarks and questions which naturally appeared during the work on this project.

To begin with, we note that the difficulty in understanding the behaviour of arbitrary (possibly sparse) generalised polynomials should not come as a surprise. Indeed, many deep questions in number theory can be reduced to the task of deciding whether a certain generalised polynomial is identically 0 or not. We cite one example which we find particularly striking.

Conjecture V.9.1 (Littlewood)**.** *Let $\alpha, \beta \in \mathbb{R}$ and $\varepsilon > 0$. Then the generalised polynomial $f\colon \mathbb{N} \to \{0, 1\}$ given by*

$$f(n) = [\![n \, \|n\alpha\| \, \|n\beta\| < \varepsilon]\!]$$

is not identically 0*.*

It is known by the work of Einsiedler, Katok and Lindenstrauss [34] that the set of (α, β) for which the above fails is either empty or of Hausdorff dimension 0, but the conjecture remains unresolved.

V.9.1 Small fractional parts

Question V.9.2. *Does there exist a generalised polynomial $q(n)$ which does not take any value* mod1 *infinitely many times and a non-increasing sequence $\varepsilon(n) \geq 0$, such that*

$$[\![\|q(n)\| \leq \varepsilon(n)]\!]$$

is an automatic sequence, not eventually 0*?*

Remark. The answer is negative when $q(n)$ is a genuine polynomial and $\varepsilon(n)$ is constant (Proposition V.3.3) , or when $q(n) = n\alpha \lfloor n\beta \rfloor$ and $\varepsilon(n) = n^{-o(1)}$ with α, β algebraic and linearly independent (Proposition V.7.3; see also Remark V.7.1). The condition that $q(n)$ mod 1 does not take the same value infinitely often is added to remove trivial examples such as $q(n) \in \mathbb{Q}[n] + \mathbb{R}$, or $q(n) = g(n)h(n)$ where $g(n)$ is any sparse ($\{0,1\}$-valued) generalised polynomial and $h(n) \to 0$ as $n \to \infty$ with $g(n) = 1$.

V.9.2 Exponential sequences

By Theorem V.1.6, the question of whether or not there exist non-trivial sequences which are simultaneously generalised polynomial and automatic reduces to deciding if the set $\{k^t \mid t \in \mathbb{N}\}$ is generalised polynomial; we suspect it is not. Conversely, by Example V.7.1, the set of Fibonacci numbers, which can be described (up to finitely many exceptions) as $\{\langle\langle \varphi^t \rangle\rangle \mid t \in \mathbb{N}\}$ where $\varphi = (1+\sqrt{5})/2$, is generalised polynomial. Similarly, in Theorem V.8.6 we show the set $\{\langle\langle \beta^t \rangle\rangle \mid t \in \mathbb{N}\}$ is generalised polynomial, where β is a certain non-totally real cubic Pisot unit. This suggests the following questions.

Question V.9.3. *For which $\lambda > 1$ is the set*

$$E_\lambda := \left\{ \langle\langle \lambda^t \rangle\rangle \;\middle|\; t \in \mathbb{N} \right\}$$

generalised polynomial?

We note that the set of such λ includes at least all the Pisot units of degree 2, which is essentially shown in Example V.7.1, and some Pisot units of degree 3.

An intimately related (cf. Proposition V.8.2) and perhaps more natural question is the following one.

Question V.9.4. *For which linear recurrence sequences $n = (n_i)_{i \geq 0}$ with values in $\mathbb{N}_0$ is the set*

$$E := \{n_i \mid i \geq 0\}$$

generalised polynomial?

Looking at the aforementioned examples from the perspective of recursive relations with non-constant coefficients (as in Remark V.7.1), we may also ask the following question.

Question V.9.5. *Let n_i be a sequence given by $n_0 = 0,\ n_1 = 1$ and the recursive relation*

$$n_i = a_i n_{i-1} + b_i n_{i-2},$$

with $a_i, b_i \in \mathbb{Z}$ bounded in absolute value. For which $a = (a_i)_i,\ b = (b_i)_i$ is the set

$$E_{a,b} = \{n_i \mid i \in \mathbb{N}\}$$

generalised polynomial?

V.9.3 Morphic words

The class of morphic words is a natural extension of the class of automatic sequences. Let Ω be a finite set. Any morphism φ of the monoid Ω^* extends naturally to $\Omega^{\mathbb{N}_0}$. A word $w \in \Omega^{\mathbb{N}_0}$ (which we identify with a function $\mathbb{N}_0 \to \Omega$) is a *pure morphic word* if it is a fixed point of a non-trivial morphism of Ω^*. Similarly, a morphic word is the image $\pi \circ w \colon \mathbb{N}_0 \to \Omega'$ of a morphic word w under a coding $\pi \colon \Omega \to \Omega'$ (i.e. any set-theoretic map, not necessarily injective). Morphic words are connected with automatic sequences via the fact that k-automatic sequences are precisely the morphic words coming from k-uniform morphisms. Here, a morphism $\varphi \colon \Omega^* \to \Omega^*$ is k-uniform if $|\varphi(u)| = k$ for all $u \in \Omega$.

We have already encountered possibly the most classical example of a non-uniform morphic word, the Fibonacci word. Recall from the Introduction that the Fibonacci word w_{Fib} was defined as the limit of the words $w_0 := 0$, $w_1 := 01$ and $w_{i+2} := w_{i+1}w_i$. Directly from this definition, it is easy to see that w_{Fib} is fixed by the morphism $\varphi \colon \Omega^{\mathbb{N}_0} \to \Omega^{\mathbb{N}_0}$ given by $\varphi(0) = 01$ and $\varphi(1) = 0$.

Recall also that w_{Fib} is a Sturmian word. Here, a *Sturmian word* is one of the form $f(n) = \lfloor \alpha(n+1) + \rho \rfloor - \lfloor \alpha n + \rho \rfloor - \lfloor \alpha \rfloor$ where $\alpha, \rho \in \mathbb{R}$ with $\alpha \notin \mathbb{Q}$ (for w_{Fib} we may take $\alpha = -\varphi$ and $\rho = 0$). Some, but not all of these sequences give rise to morphic words; see [20] for details (cf. also [108], [38], [21]).

In analogy with Conjecture V.1.3, one could ask about a classification of all morphic words which are given by generalised polynomials. We believe that examples such as the Fibonacci word are essentially the only possible ones.

Question V.9.6. *Assume that a sequence $f \colon \mathbb{N}_0 \to \Omega \subset \mathbb{R}$ is both a morphic word and a generalised polynomial. Is it true that f is a linear combination of a number of Sturmian morphic words and an eventually periodic sequence?*

V.9.4 Regular sequences

We finish by presenting a generalisation of Conjecture V.1.3 to regular sequences. We call a function $f\colon \mathbb{N}_0 \to \mathbb{Z}$ a quasi-polynomial if there exists an integer $m \geq 1$ such that the sequences f_j given by $f_j(n) = f(mn+j)$, $0 \leq j \leq m-1$, are polynomials in n. We say that a function $f\colon \mathbb{N}_0 \to \mathbb{Z}$ is ultimately a quasi-polynomial if it coincides with a quasi-polynomial except at a finite set.

Question V.9.7. *Assume that the sequence $f\colon \mathbb{N}_0 \to \mathbb{Z}$ is both regular and generalised polynomial. Is it then true that f is ultimately a quasi-polynomial?*

If f takes only finitely many values, then all the polynomials inducing f_j are necessarily constant, and so in this case the question coincides with Conjecture V.1.3.

Chapter VI

Uniformity of automatic sequences

VI.1 Introduction

In this chapter, we estimate the Gowers norms of some classical automatic sequences, which were introduced in the previous Chapter V. For the sake of clarity, we focus on two specific examples, bearing the names of Thue-Morse and Rudin-Shapiro. However, similar methods can be applied in greater generality.

We begin with the Thue-Morse sequence, which we have already encountered[1] in Section V.2. Among several ways to define it, the most basic one uses the recursive relations:

$$t(0) = 1, \qquad t(2n) = t(n), \qquad t(2n+1) = -t(n),$$

Alternatively, one can use the explicit formula $t(n) = (-1)^{s_2(n)}$, where $s_2(n)$ denotes the sum of digits of n base 2. Thus, the signs of the initial 32 terms of $t(n)$ are

$$+--+-++--++-+--+-++-+--++--+-++--+-++-+--+ .$$

Arguably, the Thue-Morse sequence is very structured. In particular, its subword complexity (i.e. number of distinct subsequences of a given length) has linear rate of growth, which is a general feature of automatic sequences. On the other hand, there are also ways in which it can be construed as pseudorandom.

Mauduit and Sarközy [92] studied several measures of pseudorandomness for the Thue-Morse sequence, and showed that $t(n)$ is highly uniform according to some but not all of those measures. In particular, it is shown that for any positive integers

[1] Strictly speaking, the sequence we consider here is different than the one considered in V.2 and in most literature, as it takes values ± 1 rather than $0, 1$. This minor change makes the results slightly easier to state, and the relation to the Rudin-Shapiro sequence more apparent.

a, b, M, N with $a(M-1)+b < N$, we have

$$\sum_{n=0}^{M-1} t(an+b) = O(N^{\log 3/\log 4}), \tag{VI.1}$$

where the implied constant is absolute. In fact, this easily follows from the bound obtained by Gelfond [47]:

$$\sup_{\alpha\in\mathbb{R}} \left| \sum_{n=0}^{N-1} t(n)e(\alpha n) \right| = O(N^{\log 3/\log 4}), \tag{VI.2}$$

where as usual $e(x) := e^{2\pi i x}$. However, for all sufficiently large integers N, there exist positive integers M and h with $M + h \leq N$ such that

$$\left| \sum_{n=0}^{M-1} t(n)t(n+h) \right| \geq cN, \tag{VI.3}$$

where $c > 0$ is an absolute constant. (We may take $c = 1/12$ and $h = 1$.) Thus, the Thue-Morse sequence does not correlate with arithmetic progressions but can have some large self-correlations.

In a different direction, $t(n)$ is believed to look highly random when restricted to certain subsequences. Several conjectures to this effect, in a slightly more general situation, are known as the Gelfond Problems [47].

Let $\mathcal{P}$ denote the set of prime numbers and $\pi(N) = |\mathcal{P} \cap [N]|$. Gelfond conjectured that it should hold that

$$|\{p \in [N] \cap \mathcal{P} \mid t(p) = +1\}| = \frac{1}{2}\pi(N) + O(N^{1-c}), \tag{VI.4}$$

where $c > 0$ is an absolute constant. This was proved only relatively recently by Mauduit and Rivat [90].

Let $p(x) \in \mathbb{Q}[x]$ be a polynomial with $p(\mathbb{Z}) \subset \mathbb{Z}$, and extend $t(n)$ to $\mathbb{Z}$ by putting $t(-n) = t(n)$. Another of Gelfond's conjectures asserts that we should have

$$|\{n \in [N] \mid t(p(n)) = +1\}| = \frac{1}{2}N + O(N^{1-c}), \tag{VI.5}$$

for an absolute constant $c > 0$. This is only known for polynomials of degree 2 by work of Mauduit and Rivat [89]. In fact, for $p(n) = n^2$, a much stronger result is shown in [33], implying in particular that $t(n^2)$ is normal (i.e. that each block of ± 1's of length l appears in $t(n^2)$ with frequency $1/2^l$). Namely, for each l and each $\epsilon_0, \ldots, \epsilon_{l-1} \in \{-1, +1\}$ we have

$$\left|\left\{n \in [N] \;\middle|\; t((n+i)^2) = \epsilon_i \text{ for } 0 \leq i < l\right\}\right| = \frac{1}{2^l}N + O(N^{1-c}), \tag{VI.6}$$

where the constant c depends only on l.

Finally, let $\alpha \in \mathbb{R}_{>0} \setminus \mathbb{Z}$. It is conjectured (see e.g. [32]) that the sequence $t(\lfloor n^\alpha \rfloor)$ should be normal for any such α. This is confirmed in the quantitative sense by Müllner and Spiegelhofer [94], who showed that for any $1 < \alpha < 3/2$, for any l and $\epsilon_i \in \{-1, +1\}$ for $0 \leq i < l$ it holds that

$$|\{n \in [N] \mid t(\lfloor (n+i)^\alpha \rfloor) = \epsilon_i \text{ for } 0 \leq i < l\}| = \frac{1}{2^l} N + O(N^{1-c}), \tag{VI.7}$$

where the constant c depends on l and α.

In general, it is believed that $t(n)$ restricted to any of the aforementioned sequences should be a normal sequence (we refer e.g. to [32], which is also an excellent reference for related results).

Our main result pertains to a different notion of uniformity, namely the one corresponding to Gowers norms.

Theorem VI.1.1. *Let $t\colon \mathbb{N}_0 \to \{\pm 1\}$ denote the Thue-Morse sequence. For any $s \in \mathbb{N}$, there exists $c = c(s) > 0$ such that $\|t\|_{U^s[N]} = O(N^{-c})$ as $N \to \infty$.*

As discussed in Section I.1.11, a key reason for interest in the Gowers uniformity norms is their usefulness in counting linear patterns. In particular, as an immediate corollary of Theorem VI.1.1 we conclude via the Generalised von Neumann Theorem discussed in Section I.1.11 that the number of k-term arithmetic progressions in $\{n \in [N] \mid t(n) = +1\}$ is $N^2/2k + O(N^{2-c})$.

We also remark that sequences with small Gowers norms do not correlate with polynomials phases: if $p \in \mathbb{R}[x]$ with $\deg p = s - 1$ then $\mathbb{E}_{n<N} f(n)e(p(n)) \ll \|f\|_{U^s[N]}$. Hence, Theorem VI.1.1 implies that $t(n)$ is a fully oscillating sequence in the terminology of [39]. As an application, for any dynamical system with quasi-discrete spectrum (X, T) and any $f_1, \dots, f_l \in \mathcal{C}(X)$, $q_1, \dots, q_l \in \mathbb{Q}[x]$ with $q_i(\mathbb{N}_0) \subset \mathbb{N}_0$ for $1 \leq i \leq l$, we have

$$\lim_{N\to\infty} \mathbb{E}_{n<N} t(n) \prod_{i=1}^{l} f_i(T^{q_i(n)}x) = 0 \tag{VI.8}$$

for any point $x \in X$.

A subtly different type of uniformity norms $\|\cdot\|_{U(s)}$ on $l^\infty(\mathbb{Z})$ is introduced and studied in [66]. In fact, it follows from results obtained there that $\|t\|_{U(s)} = 0$ for all $s \in \mathbb{N}$ (cf. Proposition 2.21), and Theorem VI.1.1 can be construed as an analogue of this result. As an example application (cf. remarks after Theorem 2.25), this implies

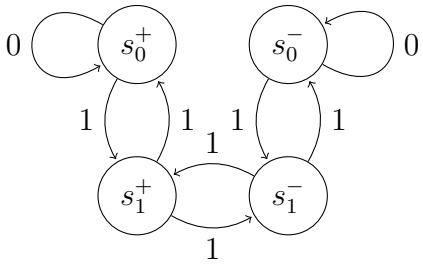
that for any measure preserving system $(X, T, \mathcal{B}, \mu)$, any $f_1, \ldots, f_l \in L^\infty(X)$ we have

$$\mathop{\mathbb{E}}_{n<N} t(n) \prod_{i=1}^{l} f_i(T^{in}x) \to 0 \quad \text{in } L^2(\mu) \text{ as } N \to \infty. \tag{VI.9}$$

A slightly more complicated sequence we deal with carries the name of Rudi-Shapiro. It is recursively given by

$$r(0) = 1,\ r(2n) = r(n),\ r(4n+1) = r(n),\ r(4n+3) = -r(2n+1),$$

or explicitly by $r(n) = (-1)^{f_{11}(n)}$, where $f_{11}(n)$ denotes the number of times the pattern 11 appears in the binary expansion of n. The automaton producing $r(n)$ is given by



where the initial state is $s_\bullet = s_0^+$ and $\tau(s_i^\pm) = \pm 1$. The first signs of $r(n)$ are

$$+++-++-+++++---+-+++-++-+---+++-+\,.$$

Much like in the case of the Thue-Morse sequence, various pseudorandomness properties of the Rudin-Shapiro sequence have long been studied. In [92] it is shown that $r(n)$ does not correlate with arithmetic progressions, but has large self-correlations. More precisely, we have

$$\sum_{n=0}^{M-1} r(an+b) = O(N^{1/2}), \tag{VI.10}$$

for any a, b, M with $a(M-1)+b < N$, while there exist M, h with $M + h \leq N$ such that

$$\left| \sum_{n=0}^{M-1} r(n)r(n+h) \right| \geq cN, \tag{VI.11}$$

where one can take $c = 1/6$ if N is sufficiently large.

There are also exist results asserting pseudorandomness of the restrictions of $r(n)$ to certain subsequences. In the case of the primes, in analogy with (VI.4), we have

$$|\{p \in [N] \cap \mathcal{P} \mid r(p) = +1\}| = \frac{1}{2}\pi(N) + O(N^{1-c}), \tag{VI.12}$$

where $c > 0$ is an absolute constant [91]. Likewise, in analogy to (VI.7), it follows as a special case from results in [31] that

$$|\{n \in [N] \mid r(\lfloor n^\alpha \rfloor) = +1\}| = \frac{1}{2}N + o(N). \tag{VI.13}$$

for any $1 < \alpha < 7/5$. (It is not known if $r(\lfloor n^\alpha \rfloor)$ is normal.)

We show that $r(n)$ is highly Gowers uniform. As an application we conclude that the number of k-term arithmetic progressions in $\{n \in [N] \mid r(n) = +1\}$ is $N^2/2k + O(N^{2-c})$.

Theorem VI.1.2. *Let $r\colon \mathbb{N}_0 \to \{\pm 1\}$ denote the Rudin-Shapiro sequence. For any $s \in \mathbb{N}$, there exists $c = c(s) > 0$ such that $\|r\|_{U^s[N]} = O(N^{-c})$ as $N \to \infty$.*

VI.2 Thue-Morse sequence

The purpose of this section is to prove Theorem VI.1.1, asserting that the uniformity norms of the Thue-Morse sequence $t(n) = (-1)^{s_2(n)}$ are small. Throughout this section, let $s \in \mathbb{N}_{\geq 2}$ be fixed. (We may assume that $s \geq 2$ since $\|t\|_{U^2[N]} \gg \|t\|_{U^1[N]}$.) It will be convenient to study somewhat more general averages

$$A(L, \mathbf{r}) := \mathop{\mathbb{E}}_{n,\mathbf{h}} \prod_{\omega \in \{0,1\}^s} t(n + \omega \cdot \mathbf{h} + r_\omega), \tag{VI.14}$$

where $n, \mathbf{h}$ parametrize the cubes $\{n + \omega \cdot \mathbf{h} \mid \omega \in \{0,1\}^s\} \subset [2^L]$, $L \in \mathbb{N}_0$ and $\mathbf{r} = (r_\omega)_{\omega \in \{0,1\}^s}$ with $r_\omega \in \mathbb{Z}$. Note that if $\mathbf{r} = \mathbf{0}$, then (VI.14) defines $\|t\|^{2^s}_{U^s[2^L]}$.

Lemma VI.2.1. *The averages $A(L, \mathbf{r})$ satisfy the recursive relation*

$$A(L, \mathbf{r}) = (-1)^{|\mathbf{r}|} \mathop{\mathbb{E}}_{\mathbf{e}} A(L-1, \delta(\mathbf{r}; \mathbf{e})) + O(2^{-L}), \tag{VI.15}$$

where the average is taken over $\mathbf{e} = (e_i)_{i=0}^s \in \{0,1\}^{s+1}$ and $\delta(\mathbf{r}; \mathbf{e})$ is given by

$$\delta(\mathbf{r}; \mathbf{e})_\omega = \left\lfloor \frac{r_\omega + (1, \omega) \cdot \mathbf{e}}{2} \right\rfloor = \left\lfloor \frac{r_\omega + e_0 + \sum_{i=1}^s \omega_i e_i}{2} \right\rfloor. \tag{VI.16}$$

Proof. For any cube $Q = \{n + \omega \cdot \mathbf{h} \mid \omega \in \{0,1\}^s\}$, there exists a unique choice of a cube $Q' = \{n' + \omega \cdot \mathbf{h}' \mid \omega \in \{0,1\}^s\}$ and a vector $\mathbf{e} \in \{0,1\}^s$ such that $Q = \{2(n' + \omega \cdot \mathbf{h}') + (1, \omega) \cdot \mathbf{e} \mid \omega \in \{0,1\}^s\}$; we simply take $n = 2n' + e_0$ and $h_i = 2h_i' + e_i$. Moreover, if $Q \subset [2^L]$ is chosen uniformly at random, then $Q' \subset [2^{L-1}]$ with

probability $1 - O(2^{-L})$, and conversely if $Q' \subset [2^{L-1}]$ and $\mathbf{e} \in \{0,1\}^{s+1}$ are chosen uniformly at random then $Q \subset [2^L]$ with probability $1 - O(2^{-L})$. It follows that

$$\begin{aligned} A(L, \mathbf{r}) &= \mathop{\mathbb{E}}_{\mathbf{e}} \mathop{\mathbb{E}}_{n', \mathbf{h}'} \prod_{\omega \in \{0,1\}^s} t(2n' + 2\omega \cdot \mathbf{h}' + (1, \omega) \cdot \mathbf{e} + r_\omega) + O(2^{-L}) \\ &= \mathop{\mathbb{E}}_{\mathbf{e}} \mathop{\mathbb{E}}_{n', \mathbf{h}'} \prod_{\omega \in \{0,1\}^s} (-1)^{(1,\omega)\cdot \mathbf{e} + r_\omega} t(n' + \omega \cdot \mathbf{h}' + \delta(\mathbf{r}; \mathbf{e})_\omega) + O(2^{-L}) \\ &= \mathop{\mathbb{E}}_{\mathbf{e}} (-1)^{S(\mathbf{e})} A(L-1, \delta(\mathbf{r}; \mathbf{e})) + O(2^{-L}), \end{aligned}$$

where $\delta(\mathbf{r}; \mathbf{e})$ is defined by (VI.16) and $S(\mathbf{e}) = \sum_\omega r_\omega + 2^s e_0 + 2^{s-1} \sum_{i=1}^s e_i \equiv |\mathbf{r}| \bmod 2$ (expectation is over $\mathbf{e} \in \{0,1\}^s$ and $\{n' + \omega \cdot \mathbf{h}' \mid \omega \in \{0,1\}^s\} \subset [2^{L-1}]$). □

Lemma VI.2.1 motivates us to introduce a random walk $\mathcal{W}_{\mathrm{TM}}$ on a directed graph $G = (V, E)$ defined as follows. The set of vertices is $V = V_+ \cup V_-$ where $V_\pm = \left\{(\mathbf{r}, \pm 1) \;\middle|\; \mathbf{r} \in \mathbb{Z}^{\{0,1\}^s}\right\}$. The transition probabilities are given by

$$P\left((\mathbf{r}, \pm 1); (\mathbf{r}', \pm(-1)^{|\mathbf{r}|})\right) = \mathbb{P}_{\mathbf{e}}\left(\delta(\mathbf{r}; \mathbf{e}) = \mathbf{r}'\right), \tag{VI.17}$$

where $\mathbf{e} = (e_i)_{i=0}^s$ is uniformly distributed in $\{0,1\}^{s+1}$ and $\delta(\mathbf{r}; \mathbf{e})$ is given by (VI.16). (By convention, the two occurrences of the symbol $\pm$ both denote the same sign.) The remaining transition probabilities (i.e. those where the signs do not agree) are declared to be identically 0.

The set E of (directed) edges of G consists of the pairs $(v, v') \in V^2$ with $P(v, v') > 0$; hence the edge $(\mathbf{r}, \pm 1) \to (\mathbf{r}', \pm(-1)^{|\mathbf{r}|})$ is present if and only if there exists $\mathbf{e} \in \{0,1\}^{s+1}$ such that $\delta(\mathbf{r}; \mathbf{e}) = \mathbf{r}'$ (with $\delta(\mathbf{r}; \mathbf{e})$ given by (VI.16)). We will be particularly interested in the graph G_0 supported on the vertices V_0 reachable from the initial vertex $v_0 = (\mathbf{0}, +1)$.

We note that $\mathcal{W}_{\mathrm{TM}}$ comes with a natural symmetry $\mathcal{R} \colon V \to V$ given by $(\mathbf{r}, \pm 1) \mapsto (\mathbf{r}, \mp 1)$. We have $\mathcal{R}(\mathcal{R}(v)) = v$ and $P(\mathcal{R}(v), \mathcal{R}(v')) = P(v, v')$ for all $v, v' \in V$. In particular, $\mathcal{R}$ preserves the edges of G.

Denote further by $P^{(l)}(v, v')$ the probability of reaching vertex v' after l steps, starting from v. Iterating Lemma VI.2.1 we obtain the following formula.

Corollary VI.2.2. *The averages $A(L, \mathbf{r})$ satisfy for any $l < L$ the recursive relation*

$$A(L, \mathbf{r}) = \sum_{\mathbf{r}', \sigma} P^{(l)}\big((\mathbf{r}, +1), (\mathbf{r}', \sigma)\big) \sigma A(L - l, \mathbf{r}') + O(2^{-(L-l)}), \tag{VI.18}$$

where the sum runs over all pairs $(\mathbf{r}', \sigma) \in V$ *which are reachable from* $(\mathbf{r}, +1)$*. In particular,*

$$\|t\|_{U^s[2^L]}^{2^s} = \sum_{\mathbf{r}'} \left(P^{(l)}\left(v_0, (\mathbf{r}', +1)\right) - P^{(l)}\left(v_0, (\mathbf{r}', -1)\right)\right) A(L-l, \mathbf{r}') + O(2^{-(L-l)}), \tag{VI.19}$$

where the sum runs over all $\mathbf{r}'$ *such that at least one of* $(\mathbf{r}', \pm 1)$ *belongs to* V_0*.*

Proof. Apply Lemma VI.2.1 l times. Note that the recurrence relation (VI.15) can be equivalently written as

$$A(L, \mathbf{r}) = \sum_{\mathbf{r}', \sigma} P\big((\mathbf{r}, +1), (\mathbf{r}', \sigma)\big)\sigma A(L-1, \mathbf{r}') + O(2^{-L}),$$

which is the same as (VI.18) for $l = 1$. For general $l \in \mathbb{N}$, the main term follows directly from how $P^{(l)}(v, v')$ are defined. The total error term is $\ll 2^{-L} + 2^{(-L-1)} + \cdots + 2^{-(L-l)} \ll 2^{-(L-l)}$. □

Recall that a directed graph is strongly connected if there exists a directed path from any vertex to any other vertex. A graph is aperiodic if the greatest common divisor of all cycles present in the graph equals 1.

Proposition VI.2.3. *Let* G_0 *be the graph constructed above. Then* G_0 *is finite, strongly connected, aperiodic, and preserved by* $\mathcal{R}$*.*

Proof. Aperiodicity follows immediately from the observation that G_0 contains a loop at $(\mathbf{0}, +1)$ since $\delta((\mathbf{0}, +1); \mathbf{0}) = (\mathbf{0}, +1)$.

If G contains an edge from $v = (\mathbf{r}, \pm 1)$ to $v' = (\mathbf{r}', \pm 1)$ then $r'_\omega = \delta(\mathbf{r}; \mathbf{e}) = \left\lfloor \frac{r_\omega + (1,\omega)\cdot \mathbf{e}}{2} \right\rfloor$ for some $\mathbf{e} \in \{0,1\}^{s+1}$, and in particular $0 \le (1, \omega) \cdot \mathbf{e} \le |\omega| + 1$. An elementary inductive argument now shows that if $(\mathbf{r}, \pm 1) \in V_0$ then $0 \le r_\omega \le |\omega|$, which proves finiteness.

Similarly, taking $\mathbf{e} = \mathbf{0} \in \{0,1\}^{s+1}$, we see that any vertex $v = (\mathbf{r}, \pm 1)$ has an edge to some $v' = (\mathbf{r}', \pm 1)$ with $|\mathbf{r}'| \le |\mathbf{r}|/2$. Repeating this argument, we may find a path from any $v \in V_0$ to one of $(\mathbf{0}, +1)$, $(\mathbf{0}, -1)$. Thus, to prove that G_0 is strongly connected, it will suffice to show that there exists a path from $(\mathbf{0}, -1)$ to $(\mathbf{0}, +1)$, which (in light of symmetry) is equivalent to $(\mathbf{0}, -1) \in V_0$. Since G is symmetric under $\mathcal{R}$, this will also imply that $\mathcal{R}(V_0) = V_0$.

It remains to show that $(\mathbf{0}, -1) \in V_0$. We do this by explicitly constructing the path from $(\mathbf{0}, +1)$ to $(\mathbf{0}, -1)$. Let $\mathbf{r}^{(0)} = \mathbf{r}^{(s+1)} = \mathbf{0}$, and for $j = 1, 2, \ldots, s$ let

$\mathbf{r}^{(j)} = (r_\omega^{(j)})$ be given by

$$r_\omega^{(j)} = \begin{cases} 1, & \text{if } \omega_1 = \omega_2 = \cdots = \omega_i = 1, \\ 0, & \text{otherwise.} \end{cases}$$

We claim that for each $j = 0, 1, \ldots, s-1$, there is an edge from $(\mathbf{r}^{(j)}, +1)$ to $(\mathbf{r}^{(j+1)}, +1)$, and for $j = s$ there is an edge from $(\mathbf{r}^{(s)}, +1)$ to $(\mathbf{r}^{(s+1)}, -1)$.

For $j = 0$, define $\mathbf{e}^{(0)}$ by $e_0^{(0)} = e_1^{(0)} = 1$ and $e_i^{(0)} = 0$ for $i \neq 0, 1$. Direct computation shows that $\delta(\mathbf{r}^{(0)}; \mathbf{e}^{(0)})_\omega = \left\lfloor \frac{1+\omega_1}{2} \right\rfloor = 1$ if $\omega_1 = 1$ and 0 otherwise. We also have $\left|\mathbf{r}^{(0)}\right| = 0$. Hence, G contains an edge from $(\mathbf{r}^{(0)}, +1)$ to $(\mathbf{r}^{(1)}, +1)$.

For $1 \leq j \leq s-1$, let $e_{j+1}^{(j)} = 1$ and $e_i^{(j)} = 0$ for $i \neq j+1$. We compute that $\delta(\mathbf{r}^{(j)}; \mathbf{e}^{(j)})_\omega = \left\lfloor \frac{r_\omega^{(j)} + \omega_{j+1}}{2} \right\rfloor = 1$ if $\omega_{j+1} = 1$ and $r_\omega^{(j)} = 1$, and 0 otherwise. Also, $\left|\mathbf{r}^{(j)}\right| = 2^{s-j} \equiv 0 \pmod 2$. Hence, G contains an edge from $(\mathbf{r}^{(j)}, +1)$ to $(\mathbf{r}^{(j+1)}, +1)$.

Finally, for $j = s$, let $\mathbf{e}^{(s)} = \mathbf{0}$. Computation similar to the one above shows that G contains an edge from $(\mathbf{r}^{(s)}, +1)$ to $(\mathbf{r}^{(s+1)}, -1)$. □

Corollary VI.2.4. *There exists a constant $c > 0$ such that $A(L, \mathbf{0}) = O(2^{-cL})$.*

Proof. Because G is aperiodic and strongly connected, the Frobenius-Perron Theorem implies that there exists a stationary distribution $\pi \colon V_0 \to [0, 1]$ such that for each $v, v' \in V_0$ we have $P^{(l)}(v, v') \to \pi(v')$ with exponential convergence rate:

$$\max_{v,v' \in V_0} \left| P^{(l)}(v, v') - \pi(v') \right| = O(2^{-cl}) \tag{VI.20}$$

for some $c > 0$. Because $\mathcal{R}(V_0) = V_0$, by symmetry we have $\pi(\mathcal{R}(v)) = \pi(v)$ for all $v \in V_0$. Combining this with (VI.20), we obtain

$$\max_{v,v'} \left| P^{(l)}(v, v') - P^{(l)}(v, \mathcal{R}(v')) \right| = O(2^{-cl}). \tag{VI.21}$$

Using this estimate and the trivial bound $|A(L, \mathbf{r})| \leq 1$ in (VI.19), we arrive at

$$A(L, \mathbf{0}) = O(2^{-cl} + 2^{-(L-l)}). \tag{VI.22}$$

It remains to put $l = L/2$ (say) to conclude that $A(L, \mathbf{0}) = O(2^{-c'L})$ with $c' = \min(1/2, c/2) > 0$. □

Remark. Computation of the best possible value of the constant c in the argument above is essentially equivalent to computing the spectral gap for the matrix $(P(v, v'))_{v,v' \in V_0}$. In particular, for fixed s, this is a computationally tractable problem.

To finish the proof of Theorem VI.1.1, we will need a standard observation concerning Gowers norms of functions restricted to and interval. We will use it several times in the sequel.

Observation. Let $J = n_0 + [M] \subset [N]$ be an interval, let $f\colon [M] \to \mathbb{C}$ be any sequence, and put $f'(n) := f(n-n_0)1_J(n)$. Then $\|f'\|_{U^s[N]} = \|1_J\|_{U^s[N]} \cdot \|f\|_{U^s[M]}$.

Proof. This follows by splitting the expected value in (I.7) depending on whether or not the set $\{n + \omega \cdot \mathbf{h} \mid \omega \in \{0,1\}^s\}$ is entirely contained in J. □

Proof of Theorem VI.1.1. Split $[N]$ into intervals $I_j = [m_j 2^{L_j}, (m_j+1)2^{L_j})$ where $L_1 > L_2 > \dots$. We then have

$$\|t\|_{U^s[N]} = \left\| \sum_j 1_{I_j} t \right\|_{U^s[N]} \leq \sum_j \left\| 1_{I_j} t \right\|_{U^s[N]}. \tag{VI.23}$$

The cubes $\{n + \omega \cdot \mathbf{h} \mid \omega \in \{0,1\}^s\} \subset I_j$ are precisely the translations of the cubes $\{n + \omega \cdot \mathbf{h} \mid \omega \in \{0,1\}^s\} \subset [2^{L_j}]$ by $2^{L_j} m_j$. Since $t(2^{L_j} m_j + n) = t(2^{L_j} m_j) t(n)$ for $n \in [2^{L_j}]$, we conclude that

$$\left\| 1_{I_j} t \right\|_{U^s[N]} = \left\| 1_{I_j} \right\|_{U^s[N]} \cdot \|t\|_{U^s[2^{L_j}]} \ll \frac{2^{L_j}}{N} \|t\|_{U^s[2^{L_j}]}.$$

Inserting this into (VI.23) and applying Corollary VI.2.4 yields (with $c' = c/2^s$ using notation therein) we obtain

$$\|t\|_{U^s[N]} \ll \sum_j 2^{L_j(1-c')}/N \ll N^{-c'}. \quad \square$$

VI.3 Rudin-Shapiro sequence

We now move on to the Rudin-Shapiro sequence $r(n) = (-1)^{f_{11}(n)}$, and embark upon the proof of Theorem VI.1.2. Our argument is similar to the one in Section VI.2, although slightly more technical. Complications arise because of the fact that the 2-kernel

$$\mathcal{N}_2(r) = \left\{ n \mapsto r(2^l n + m) \;\middle|\; 0 \leq m < 2^l \right\} = \{\pm r(n), \pm r(2n+1)\}$$

contains other functions apart from $\pm r(n)$, which forces us to deal with averages more general than those in (VI.14).

A key feature of $r(n)$ which allows our argument to work is that $\mathcal{N}_2(r)$ is symmetric, i.e. $\mathcal{N}_2(r) = -\mathcal{N}_2(r)$. Denote $\mathcal{N}_2^+(r) = \{r(n), r(2n+1)\}$, and fix from now on the value of $s \in \mathbb{N}_{\geq 2}$. We will study the averages

$$A(L, \mathbf{a}, \mathbf{r}) := \mathop{\mathbb{E}}_{n,\mathbf{h}} \prod_{\omega \in \{0,1\}^s} a_\omega(n + \omega \cdot \mathbf{h} + r_\omega), \tag{VI.24}$$

where $\{n + \omega \cdot \mathbf{h} \mid \omega \in \{0,1\}^s\} \subset [2^L]$, $L \in \mathbb{N}_0$, $\mathbf{a} = (a_\omega)_{\omega \in \{0,1\}^s}$ with $a_\omega \in \mathcal{N}_2^+(r)$ and $\mathbf{r} = (r_\omega)_{\omega \in \{0,1\}^s}$ with $r_\omega \in \mathbb{Z}$.

We record a recurrence relation analogous to Lemma VI.2.1. Now, it will be more convenient to consider l consecutive steps.

Lemma VI.3.1. *For any l, the averages $A(L, \mathbf{a}, \mathbf{r})$ obey the recursive relation:*

$$A(L, \mathbf{a}, \mathbf{r}) = \mathop{\mathbb{E}}_{\mathbf{e}} A(L - l, \delta^{(l)}(\mathbf{a}; \mathbf{r}, \mathbf{e}), \delta^{(l)}(\mathbf{r}; \mathbf{e})) + O(2^{-(L-l)}), \tag{VI.25}$$

where the average is taken over $\mathbf{e} = (e_i)_{i=0}^s \in [2^l]^{s+1}$, $\delta^{(l)}(\mathbf{a}; \mathbf{r}, \mathbf{e})$ is given by

$$\delta^{(l)}(\mathbf{a}; \mathbf{r}, \mathbf{e})_\omega(n) = a_\omega\left(2^l n + \left(r_\omega + (1, \omega) \cdot \mathbf{e} \bmod 2^l\right)\right), \tag{VI.26}$$

and $\delta^{(l)}(\mathbf{r}; \mathbf{e})$ is given by

$$\delta^{(l)}(\mathbf{r}; \mathbf{e})_\omega = \left\lfloor \frac{r_\omega + (1, \omega) \cdot \mathbf{e}}{2^l} \right\rfloor. \tag{VI.27}$$

Proof. Fix the value of l. Any cube $Q = \{n + \omega \cdot \mathbf{h} \mid \omega \in \{0,1\}^s\}$ can be uniquely written as $\left\{2^l(n' + \omega \cdot \mathbf{h}') + (1, \omega) \cdot \mathbf{e} \;\middle|\; \omega \in \{0,1\}^s\right\}$, where $\mathbf{e} \in [2^l]^{s+1}$. Up to errors of the order of $O(2^{-(L-l)})$, choosing $Q \subset [2^L]$ uniformly at random is equivalent to choosing $\mathbf{e} \in [2^l]^{s+1}$ and $Q' = \{n' + \omega \cdot \mathbf{h}' \mid \omega \in \{0,1\}^s\} \subset [2^{L-l}]$ uniformly at random, hence

$$\begin{aligned} A(L, \mathbf{a}, \mathbf{r}) &= \mathop{\mathbb{E}}_{\mathbf{e}} \mathop{\mathbb{E}}_{n', \mathbf{h}'} \prod_{\omega \in \{0,1\}^s} a_\omega(2^k(n' + \omega \cdot \mathbf{h}') + (1, \omega) \cdot \mathbf{e} + r_\omega) + O(2^{-(L-l)}) \\ &= \mathop{\mathbb{E}}_{\mathbf{e}} \mathop{\mathbb{E}}_{n', \mathbf{h}'} \prod_{\omega \in \{0,1\}^s} \delta^{(l)}(\mathbf{a}; \mathbf{r}, \mathbf{e})_\omega(n' + \omega \cdot \mathbf{h}' + \delta^{(l)}(\mathbf{r}; \mathbf{e})_\omega) + O(2^{-(L-l)}), \end{aligned}$$

which is precisely the stated formula. □

For $l = 1$, we write δ for $\delta^{(1)}$. Note that the symbol δ is used to denote several different functions, but this will not lead to ambiguity because it can always be inferred from the arguments which function is meant.

Like in the previous section, we introduce a random walk $\mathcal{W}_{\mathrm{RS}}$ on a graph $G = (V, E)$, which is associated to the averages $A(L, \mathbf{a}, \mathbf{r})$. The set of vertices V consists

of triples $(\mathbf{a}, \mathbf{r}, \pm 1)$, where $a_\omega \in \mathcal{N}_2^+(r)$ and $r_\omega \in \mathbb{Z}$. For $v = (\mathbf{a}, \mathbf{r}, \sigma) \in V$, we write $A(L, v) = \sigma A(L, \mathbf{a}, \mathbf{r})$.

Using the fact that $\mathcal{N}_2(r) = \mathcal{N}_2^+(r) \cup (-\mathcal{N}_2^+(r))$, we see that for any $\mathbf{a}$ with $a_\omega \in \mathcal{N}_2(r)$, we can find $\bar{\mathbf{a}} = (\bar{a}_\omega)_{\omega \in \{0,1\}^s}$ with $\bar{a}_\omega \in \mathcal{N}_2^+(r)$ and $\sigma = \pm 1$, such that $\prod_\omega a_\omega(x_\omega) = \sigma \prod_\omega \bar{a}_\omega(x_\omega)$. In particular, for any $\mathbf{r}$ we have $A(L, \mathbf{a}, \mathbf{r}) = A(L, v)$ for any L, where $v = (\bar{\mathbf{a}}, \mathbf{r}, \sigma) \in V$.

Let $l \in \mathbb{N}$ be fixed. Then, for $\mathbf{e} \in [2^l]^{s+1}$ and $v = (\mathbf{a}, \mathbf{r}, \sigma) \in V$, we let $\delta^{(l)}(v; \mathbf{e}) \in V$ denote the vertex constructed above, corresponding to the averages $A(L, \delta^{(l)}(\mathbf{a}; \mathbf{r}, \mathbf{e}), \delta^{(l)}(\mathbf{r}; \mathbf{e}))$. (In other words, $\delta^{(l)}(v; \mathbf{e}) = (\mathbf{a}', \mathbf{r}', \sigma')$, where we put $a'_\omega = \pm \delta^{(l)}(\mathbf{a}; \mathbf{r}, \mathbf{e})_\omega$, $r'_\omega = \delta^{(l)}(\mathbf{r}; \mathbf{e})$ and choose $\sigma' = \pm 1$ so that $\sigma' \prod_\omega a_\omega(x_\omega) = \sigma \prod_\omega \delta^{(l)}(\mathbf{a}; \mathbf{r}, \mathbf{e})_\omega(x_\omega)$).

The transition probabilities are given by $P(v, v') = \mathbb{P}_{\mathbf{e} \in \{0,1\}^{s+1}}(\delta(v; \mathbf{e}) = v')$ for $v, v' \in V$, so that (VI.25) for $l = 1$ is equivalent to

$$A(L, v) = \sum_{v' \in V} P(v, v') A(L - 1, v') + O(2^{-L}). \tag{VI.28}$$

The edge from v to v' is present in the edge set E of G if $P(v, v') > 0$. More generally, for arbitrary $l \in \mathbb{N}$ we have

$$A(L, v) = \sum_{v' \in V} P^{(l)}(v, v') A(L - l, v') + O(2^{-(L-l)}), \tag{VI.29}$$

where $P^{(l)}(v, v') = \mathbb{P}_{\mathbf{e} \in [2^l]^{s+1}}(\delta^{(l)}(v; \mathbf{e}) = v')$ denotes the probability of transition from v to v' in l steps. Accordingly, a path of length l from $v = (\mathbf{a}, \mathbf{r}, +1)$ to $v' \in V$ exists if and only if the average $A(L - l, v')$ is present on the right hand side of (VI.25), meaning that there exists $\mathbf{e} = (e_i)_{i=0}^k$, $0 \le e_i < 2^l$, such that $\delta^{(l)}((\mathbf{a}, \mathbf{r}, +1); \mathbf{e}) = v'$.

Following the same reasoning as before, we note that G has a natural symmetry $\mathcal{R} \colon V \to V$ given by $(\mathbf{a}, \mathbf{r}, \sigma) \mapsto (\mathbf{a}, \mathbf{r}, -\sigma)$, which preserves the transition probabilities. We will denote by V_0 the set of vertices reachable from the initial vertex $v_0 = ((r)_{\omega \in \{0,1\}^s}, \mathbf{0}, +1)$ and by G_0 the induced graph.

Proposition VI.3.2. *Let G_0 be the graph constructed above. Then G_0 is finite, strongly connected, aperiodic, and preserved by $\mathcal{R}$.*

Proof. Finiteness, aperiodicity, and strong connectedness follow from essentially the same argument as in Propositions VI.2.3. It remains to prove that $\mathcal{R}(v_0)$ is reachable from v_0.

Pick any $l \ge s + 2$, and $e_i = 2^{i-1}$ for $i = 1, 2, \dots, s$; we leave $0 \le e_0 < 2^s$ undefined for the time being. It follows from Lemma VI.3.1 and subsequent discussion that G_0 contains a path of length l from v_0 to $v_1 = (\mathbf{a}, \mathbf{r}, \sigma)$ equal to $\delta^{(l)}(v_0; \mathbf{e})$.

We will now identify the vertex v_1. As for $\mathbf{a} = (a_\omega)_\omega$, we notice that

$$a_\omega(n) = \pm r(2^l n + (1,\omega)\cdot \mathbf{e}) = \pm r(n) r((1,\omega)\cdot \mathbf{e}) = r(n). \tag{VI.30}$$

Above, we use that fact that $(1,\omega)\cdot \mathbf{e} < 2^{l-1}$. Next, $\mathbf{r} = (r_\omega)_\omega$ is given by

$$r_\omega = \left\lfloor \frac{(1,\omega)\cdot \mathbf{e}}{2^l} \right\rfloor = 0,$$

by virtue of the same estimate as before. Finally, σ is the product of the ± 1 factors implicit in (VI.30), hence

$$\sigma = \prod_{\omega\in\{0,1\}^s} r((1,\omega)\cdot \mathbf{e}) = \prod_{m=0}^{2^s-1} r(m+e_0).$$

Thus, v_1 is equal to $\mathcal{R}(v_0)$, provided that $\sigma = \sigma(e_0)$ defined above is equal to -1, and $v_1 = v_0$ otherwise. It remains to find e_0 for which $\sigma(e_0) = -1$. In fact, it will suffice to show that $\sigma(e_0)$ is not constant with respect to e_0. Since $\sigma(e_0+1)/\sigma(e_0) = r(2^s+e_0)/r(e_0)$, we have $\sigma(e_0+1) = -\sigma(e_0)$ for any choice $2^{s-1} \le e_0 < 2^s$, which finishes the argument. □

Corollary VI.3.3. *There exists a constant $c > 0$ such that $\|t\|_{U^s[2^L]} = O(2^{-cL})$.*

Proof. Direct adaptation of the argument in Corollary VI.2.4. □

Proof of Theorem VI.1.2. We begin by splitting the interval $[N]$ into a disjoint union of intervals I_j of the form $[m_j 2^{L_j+1}, m_j 2^{L_j+1} + 2^{L_j})$, and a remainder part J so that $|J| \ll \log N$ and each exponent L appears among L_j's at most $\log N$ times. This can be accomplished by first splitting $[N]$ into dyadic intervals $[m'_j 2^{L'_j}, (m'_j+1)2^{L'_j})$ (with L'_j all distinct), and then splitting each of these further into intervals $[(m'_j+1)2^{L'_j} - 2^k, (m'_j+1)2^{L'_j} - 2^k + 2^{k-2})$ for $2 \le k < L'_j$, and the singleton of $(m'_j+1)2^{L'_j} - 1$ (which we put in J).

For each of the intervals I_j and for any $n + m_j 2^{L_j+1} \in I_j$, $n \in [2^j]$, we have $r(n + m_j 2^{L_j+1}) = r(n) r(m_j)$, whence $\left\|1_{I_j} r\right\|_{U^s[N]} \ll \frac{2^{L_j}}{N} \|r\|_{U^s[2^{L_j}]}$. Using Corollary VI.3.3 and the bound $\|1_J\|_{U^s[N]} \ll (|J|/N)^{1/2^s}$, we may now estimate that

$$\begin{aligned} \|t\|_{U^s[N]} &\ll \sum_j \frac{2^{L_j(1-c)}}{N} + N^{-1/2^s+o(1)} \\ &\ll \frac{\log N}{N} \sum_{L\le \log N} 2^{L(1-c)} + N^{-1/2^s+o(1)} \ll N^{-c'}, \end{aligned}$$

where $0 < c' < \min(c, 1/2^s)$ is arbitrary. □

VI.4 Closing remarks

It is natural to ask if the conclusions of Theorems VI.1.1 and VI.1.2 can be extended to a wider class of automatic sequences. Alternatively, one can hope to obtain a decomposition of an arbitrary automatic sequence into a structured part and a Gowers uniform part.

Both the Thue-Morse and the Rudin-Shapiro sequences can be construed as the number of times a certain pattern π appears in the binary expansion of the input; $\pi = 1$ for Thue-Morse and $\pi = 11$ for Rudin-Shapiro.

More generally, let us fix $k \in \mathbb{N}$, and for a pattern $\pi \in \Sigma_k^* \setminus \{\epsilon\}$ denote by $f_\pi(n)$ the number of times π appears in the base k expansion $(n)_k$. For more background on such pattern-counting sequences, we refer to [4, Chpt. 3]. To avoid technical difficulties, we will assume from this point that π begins with a non-zero digit.

For any $t \in \mathbb{R}$, we may consider the sequence $e(tf_\pi(n))$. If $t \in \mathbb{Q}$, thus obtained sequence is automatic, but one may work in greater generality. In analogy with the examples of Thue-Morse and Rudin-Shapiro, we expect these sequences to be Gowers uniform.

Conjecture VI.4.1. *Let $k \in \mathbb{N}$, $t \in \mathbb{R} \setminus \{0\}$, let $\pi \in \Sigma_k^*$ be a pattern starting with a non-zero digit, and let the sequence $b\colon \mathbb{N}_0 \to \mathbb{C}$ be given by $b(n) = e(tf_\pi(n))$. Then, for any $s \in \mathbb{N}$, there exists a constant c (dependent on s, t and π) such that $\|b\|_{U^s[N]} \ll N^{-c}$.*

In fact, in the case when $t \in \mathbb{Q}$, an argument reminiscent of the proof of Theorem VI.1.2 can be applied. For general $t \in \mathbb{R} \setminus \mathbb{Q}$, the question remains open.

Sketch of Proof of Conjecture VI.4.1 for $t \in \mathbb{Q}$. Fix $s \in \mathbb{N}$. Suppose that $t = p/q$ where $q \geq 2$ and p, q are coprime. For each $a \in \mathcal{N}_k(b)$ and each $j \in \mathbb{N}_0$, the sequence $e(j/q)a$ belongs to $\mathcal{N}_k(b)$. Let $\mathcal{N}_k^+(b)$ consist of the sequences $a \in \mathcal{N}_k(b)$ such that $a(0) = +1$, so that $\mathcal{N}_k(b) = \bigcup_{j=0}^{q-1} e(j/q)\mathcal{N}_k^+(b)$, and the union is disjoint.

We introduce the vertex set

$$V = \mathcal{N}_k^+(b)^{2^s} \times \mathbb{Z}^{2^s} \times \{z \in \mathbb{C} \mid z^q = 1\},$$

together with the initial vertex $v_0 = (\mathbf{b}, \mathbf{0}, 1)$. To each $v = (\mathbf{a}, \mathbf{r}, \sigma) \in V$ and $L \in \mathbb{N}_0$ we associate the averages

$$A(L; v) := \mathop{\mathbb{E}}_{n,\mathbf{h}} \prod_{\omega \in \{0,1\}^s} a_\omega(n + \omega \cdot \mathbf{h} + r_\omega), \tag{VI.31}$$

where the average is taken over $n, \mathbf{h}$ with $\{n + \omega \cdot \mathbf{h} \mid \omega \in \{0,1\}^s\} \subset [2^L]$. It will suffice to prove that $A(L, v_0) \ll e^{-cL}$ for a constant $c > 0$.

For any l, for any vertex $v = (\mathbf{a}, \mathbf{r}, \sigma) \in V$ and any $\mathbf{e} \in [k^l]^{s+1}$ we introduce the transitions $\delta^{(l)}(v; \mathbf{e}) = (\mathbf{a}', \mathbf{r}', \sigma')$, where $\mathbf{a}'$ is chosen so that

$$a'_\omega(n) = \lambda_\omega a_\omega \left(k^l n + \left(r_\omega + (1, \omega) \cdot \mathbf{e} \bmod k^l\right)\right), \qquad (n \in \mathbb{N}_0) \tag{VI.32}$$

where $\lambda_\omega \in \mathbb{C}$ is chosen so that $a'_\omega \in \mathcal{N}_k^+(b)$; $\mathbf{r}'$ is given by

$$r_\omega = \left\lfloor \frac{r_\omega + (1, \omega) \cdot \mathbf{e}}{k^l} \right\rfloor; \tag{VI.33}$$

and finally $\sigma' = \sigma \prod_\omega \lambda_\omega$. With this notation, we introduce the random walk on V so that the probability of transition from v to v' in l steps is given by

$$P^{(l)}(v, v') = \mathbb{P}_{\mathbf{e} \in [k^l]^{s+1}} \left(\delta^{(l)}(v, \mathbf{e}) = v'\right).$$

A directed edge is present between v and v' if $P^{(1)}(v, v') > 0$. We have the symmetry $\mathcal{R} \colon V \to V$ given by $(\mathbf{a}, \mathbf{r}, \sigma) \mapsto (\mathbf{a}, \mathbf{r}, e(1/q)\sigma)$, which preserves the transition probabilities and has $\mathcal{R}^q = \mathrm{id}$.

For any $l < L$, we have in analogy with Lemma VI.3.1

$$A(L; v) = \sum_{v'} P^{(l)}(v, v') A(L; v') + O(k^{-(L-l)}).$$

Hence, making use of Perron-Frobenius Theorem and the symmetry of the graph, we are able to finish the argument once we show that the graph supported on the set V_0 of vertices reachable from v_0 is finite, strongly connected, aperiodic and preserved by $\mathcal{R}$. This, in turn, will follow once we show that $\mathcal{R}(v_0) \in V_0$.

Denote by $\Lambda \subset \mathbb{C}$ the set of these $\lambda = e(j/q)$ that $\mathcal{R}^j(v_0) = (\mathbf{b}, \mathbf{0}, \lambda) \in V_0$. By construction, $1 \in \Lambda$ and Λ is a group. Let us consider, for sufficiently large l, the path from v_0 corresponding to the label $\mathbf{e}$ where $e_i = 2^{i-1}$ for $1 \le i \le s$ and we leave e_0 unspecified (but small in terms of l). Just like in the case of Rudin-Shapiro, we find that

$$\delta^{(l)}(v_0, \mathbf{e}) = (\mathbf{b}, \mathbf{0}, \sigma(e_0)),$$

where $\sigma(e_0) = \prod_{m=0}^{2^s-1} b(m + e_0) \in \Lambda$. Hence, $\sigma(e_0 + 1)/\sigma(e_0) = b(2^s + e_0)/b(e_0) \in \Lambda$. Pick $e_0 = [\pi]_k k^{l_0} - 2^s$ where l_0 is chosen so that $k^{l_0} > 2^s$, and let e'_0 be obtained by discarding the leading digit of e_0. Note that by construction $b(e_0 + 2^s) = e(p/q)$, $b(e'_0 + 2^s) = 0$, and $b(e_0) = b(e'_0)$. If follows that

$$b(2^s + e_0) b(e'_0) / b(e_0) b(2^s + e'_0) = e(p/q) \in \Lambda,$$

whence $e(1/q) \in \Lambda$, $\mathcal{R}(v_0) \in V_0$, and we are done. □

We now turn to general automatic sequences. There are several easy to describe classes of k-automatic sequences which fail to have small uniformity norms. It will be more convenient to speak of sequences whose domain is Σ_k^* rather than $\mathbb{N}_0$. Recall from Chapter V that these approaches are essentially equivalent. We restrict our attention to k-automatic sequences taking values in $\mathbb{R}$.

(i) *Constant sequences.* Clearly, for any $c \in \mathbb{R} \setminus \{0\}$, $\|c\|_{U^s[N]} = |c| > 0$.

(ii) *Periodic sequences.* Let $a\colon \mathbb{N}_0 \to \mathbb{R}$ be a periodic, non-constant sequence. Then $\|a\|_{U^s[N]} \gg 1$, since $a(n)$ correlates with a Fourier character.

(iii) *Length-dependent sequences.* Let $a\colon \mathbb{N}_0 \to \mathbb{R}$ take the form $a(n) = b(\lfloor \log_k n \rfloor)$ where b is periodic and non-zero. Then $a(n)$ takes a constant, non-zero value on any interval $[k^l, k^{l+1} - 1)$, whence the triangle inequality for Gowers norms implies that $\|a\|_{U^s[N]} \not\to 0$ as $N \to \infty$.

(iv) *Almost periodic sequences.* Let us call a sequence $a\colon \Sigma_k^* \to \mathbb{R}$ *almost periodic* if there exists a word $u \in \Sigma_k^*$ such that for any word $v \in \Sigma_k^*$, the sequence $w \mapsto a(wuv)$ is constant. If at least one of thus obtained constant sequences is non-zero, then a correlates with a periodic sequence, and so $\|a\|_{U^s[N]} \not\to 0$ as $N \to \infty$..

(v) *Almost constant sequences.* Let us call a sequence $a\colon \Sigma_k^* \to \mathbb{R}$ *almost constant* if there exists a word $u \in \Sigma_k^*$ such that for any word $v \in \Sigma_k^*$, the sequence $w \mapsto a(vuw)$ is constant. If at least one thus obtained sequence is non-zero, then a is constant on long intervals and so $\|a\|_{U^s[N]} \gg 1$.

Combining the above examples, for a given $k \in \mathbb{N}$ let us call an automatic sequence $a\colon \Sigma_k^* \to \mathbb{R}$ *structured* if there exist words $u_1, u_2 \in \Sigma_k^*$ and integers $L, M \in \mathbb{N}$ such that for any words v_1, v_2, the value of $a(v_1u_1wu_2v_2)$ is a function of v_1, v_2, $([w]_k \bmod M)$ and $(|w| \bmod L)$. In absence of other known examples, and in light of some preliminary (unpublished) results with Jakub Byszewski and Clemens Müllner, we pose the following conjecture.

Conjecture VI.4.2. *Let $k \in \mathbb{N}$, and let $a\colon \mathbb{N}_0 \to \mathbb{R}$ be an automatic sequence. Then, there exists a decomposition $a = a_{\mathrm{st}} + a_{\mathrm{uni}}$, where a_{st} is structured (as defined above) and a_{uni} is Gowers uniform, in the sense that for each $s \in \mathbb{N}$ there exists $c > 0$ such that $\|a_{\mathrm{uni}}\|_{U^s[N]} \ll N^{-c}$.*

Appendix A

Continued fractions

In this appendix we recall some fairly standard facts concerning continued fractions. Because the results are standard, we do not provide proofs, merely references.

A.1 Basic definitions

A continued fraction is an expression of the form:

$$[a_0; a_1, a_2, \dots] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \dots}}},$$

where $a_0 \in \mathbb{Z}$ and $a_i \in \mathbb{N}$ for $i > 0$. This can be either finite or infinite; we focus mostly on the infinite case.

A standard way to make sense of infinite fractions of this form is to consider consecutive finite approximations, which we typically denote as $\frac{p_n}{q_n}$, given by

$$\frac{p_n}{q_n} = [a_0; a_1, a_2, \dots, a_n] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}.$$

In particular, $p_0 = a_0,\ q_0 = 1$. It is also convenient to define $p_{-1} = 1$ and $q_{-1} = 0$.

We list some basic properties of the partial approximations. Throughout, a_i denote integers, and p_i, q_i are defined as above. Perhaps the most fundamental fact that we shall use is the following.

Fact A.1.1 ([70, Thm. 5])**.** We have the relation $[a_0, a_1, \dots, a_n, x] = \frac{xp_n + p_{n-1}}{xq_n + q_{n-1}}$.

It is not difficult to derive the following consequences.

Fact A.1.2 ([70, Thm. 1, 12])**.** Sequences p_n, q_n are given recursively by

$$\begin{aligned} p_{n+2} &= a_{n+2}p_{n+1} + p_n, & p_{-1} &= 1,\ p_0 = a_0, \\ q_{n+2} &= a_{n+2}q_{n+1} + q_n, & q_{-1} &= 0,\ q_0 = 1. \end{aligned}$$

In particular, $p_{n+m} \geq 2^{\frac{m-1}{2}} p_n$ and $q_{n+m} \geq 2^{\frac{m-1}{2}} q_n$ for each n, m.

Fact A.1.3 ([70, Thm. 2])**.** We have

$$q_n p_{n+1} - q_{n+1}p_n = (-1)^n, \quad \text{or equivalently:} \quad \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n q_{n+1}}.$$

The sequence $\frac{p_n}{q_n}$ converges rather rapidly. We shall denote

$$\alpha := \lim_{n\to\infty} \frac{p_n}{q_n} = [a_0; a_1, a_2, \dots].$$

and refer to a_i as the continued fraction expansion of α.

Fact A.1.4 ([70, Thm. 9])**.** The speed of convergence above in $\alpha = \lim_{n\to\infty} \frac{p_n}{q_n}$ is described by

$$\alpha - \frac{p_n}{q_n} = \sum_{i=n}^{\infty} \frac{(-1)^i}{q_i q_{i+1}},$$

and in particular $\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n q_{n+1}} < \frac{1}{a_{n+1} q_n^2}$.

As a consequence, it is always easy to compare two continued fraction approximations.

Fact A.1.5 ([70, Thm. 4])**.** We have the following ordering:

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \frac{p_4}{q_4} < \cdots < \alpha < \cdots < \frac{p_5}{q_5} < \frac{p_3}{q_3} < \frac{p_1}{q_1}.$$

It is often useful to have a good understanding of the ratio $\frac{q_{n+1}}{q_n}$. Fortunately, this quantity has a simple description.

Fact A.1.6 ([70, Thm. 6])**.** We have $\frac{q_{n+1}}{q_n} = [a_{n+1}; a_n, \dots, a_1]$.

A.2 Ergodic perspective

We refer to the sequence a_i as the continued fractions expansion of α above. Every irrational number has precisely one (infinite) expansion. (A similar statement is true for rational numbers, except one needs to be careful with uniqueness.) More precisely, we have the following.

Fact A.2.1 ([35, Lem. 3.4])**.** The map $\mathbb{Z}\times\mathbb{N}^{\mathbb{N}}\colon (a_0, a_1, \dots) \mapsto [a_0; a_1, a_2, \dots,] \in \mathbb{R}\setminus\mathbb{Q}$ is a bijection. Likewise, the map $\mathbb{N}^{\mathbb{N}}\colon (a_1, \dots) \mapsto [0; a_1, a_2, \dots,] \in [0,1] \setminus \mathbb{Q}$ is a bijection.

Definition A.2.2 (Continued fraction transformation)**.** Define the transformation $T\colon [0,1] \setminus \mathbb{Q} \to [0,1] \setminus \mathbb{Q}$ by $T\alpha = \left\{\frac{1}{\alpha}\right\}$, where $\{x\}$ denotes the fractional part of x. (One may extend the definition to $\mathbb{Q}$ by setting $T\alpha = 0$ for $x \in \mathbb{Q}$ if one wishes to have a map $T\colon [0,1] \to [0,1]$.)

Define the measure μ on $[0,1]$ by $\mu(A) = \frac{1}{\log 2}\int_A \frac{dx}{x+1}$ for $A \in \mathcal{B}([0,1])$, where $\mathcal{B}([0,1])$ denotes the Borel σ-algebra.

We refer to the transformation T as the *continued fraction transformation* and to μ as the *Gauss measure*.

Fact A.2.3 ([35, Chpt. 3])**.** The transformation T acts on $\mathbb{N}^{\mathbb{N}}$ by a shift, given by

$$T([0; a_1, a_2, \dots]) = T([0; a_2, a_3, \dots]).$$

Fact A.2.4 ([35, Chpt. 3])**.** The measure μ is equivalent to the Lebesgue measure. The transformation T is measurable and piecewise continuous.

Fact A.2.5 ([35, Chpt. 3])**.** The transformation T is μ-invariant, in the sense that for each $A \in \mathcal{B}([0,1])$ we have $\mu(T^{-1}(A)) = \mu(A)$.

Thus, $([0,1], T, \mathcal{B}([0,1]), \mu)$ is a measure preserving system (for introduction to measure preserving systems, see e.g. [35, Chapter 1]).

Fact A.2.6 ([35, Thm. 3.7])**.** The measure preserving system $([0,1], T, \mathcal{B}([0,1]), \mu)$ is ergodic.

A.3 Good rational approximations

Essentially all good rational approximations of a number $\alpha = [a_0; a_1, a_2, \dots]$ come from continued fractions.

Fact A.3.1 (Legendre, [70, Thm. 4])**.** If $\left|\alpha - \frac{p}{q}\right| < \frac{1}{2q^2}$ for some $\frac{p}{q}$, then there exists some i with $\frac{p}{q} = \frac{p_i}{q_i}$.

For context, we also mention a result which we don't use, even implicitly.

Fact A.3.2 (Hurwitz)**.** For every α there are p, q such that $\left|\alpha - \frac{p}{q}\right| < \frac{1}{\sqrt{5}q^2}$.

Call a fraction $\frac{p}{q}$ a *best rational approximation* (of the second kind, in terminology of [70]) of α if $|q\alpha - p| < |q'\alpha - p'|$ for any $\frac{p'}{q'} \neq \frac{p}{q}$ with $q' \leq q$.

Fact A.3.3 ([70, Thm. 16, 17])**.** If $\frac{p}{q}$ is a best rational approximation of α, then there exists some i with $\frac{p}{q} = \frac{p_i}{q_i}$. Conversely, if $i \geq 1$ then $\frac{p_i}{q_i}$ is a best rational approximation of α. In particular, if $\left|\alpha - \frac{p}{q}\right| \leq \left|\alpha - \frac{p_i}{q_i}\right|$ then $q \geq q_i$.

See also Chapter 6 of [70] for different notions of a best rational approximation and more similar results.

Badly approximable numbers can be characterised in terms of their continued fraction expansion. Recall that α is badly approximable precisely when $\left|\alpha - \frac{p}{q}\right| \gg \frac{1}{q^2}$ for all $\frac{p}{q}$.

Fact A.3.4 ([35, Prop. 3.10])**.** The number α is badly approximable if and only if the sequence $(a_i)_{i=1}^{\infty}$ is bounded.

A particularly important class of badly approximable are the quadratic irrationals.

Fact A.3.5 ([70, Thm. 28])**.** The expansion $(a_i)_{i=1}^{\infty}$ of α is eventually periodic if and only if α is a quadratic irrational.

In particular, if α is quadratic irrational then α is badly approximable.

Appendix B

Ultrafilters and limits

In this appendix we introduce some basic facts about ultrafilters and associated limits. The material is basic, but may be somewhat non-standard. The references to all statements can be found in many papers, such as [11]; for a thorough introduction see [64]; for a simple-minded survey see [72].

Definition B.0.1 (Ultrafilter)**.** An *ultrafilter* on a ground set X is a collection of sets $p \subset \mathcal{P}(X)$ obeying the following properties:

(i) if $A \in p$ and $A \subset B \subset X$, then $B \in p$;

(ii) if $A, B \in p$ then $A \cap B \in p$;

(iii) $\emptyset \notin p$ and $X \in p$;

(iv) for any A, either $A \in p$ or $X \setminus A \in p$.

The set of ultrafilters on X is denoted βX.

Trivial examples of ultrafilters include the *principal* ultrafilters, which take the form $p_x = \{A \in \mathcal{P}(X) \mid x \in X\}$. Ultrafilters not of this form are called non-principal. Under the identification $x \leftrightarrow p_x$, the set of non-principal ultrafilters is $\beta X \setminus X$.

Existence of non-trivial ultrafilters already requires the axiom of choice.

Fact B.0.2. If X is an infinite set, then there exist non-principal ultrafilters on X.

Our main application for ultrafilters is the convenient notion of limit.

Definition B.0.3. Let p be an ultrafilter on a set X, and let $f : X \to Z$ be a map to a Hausdorff topological space Z. Then a point $z \in Z$ is the limit of f along p, if for each open neighbourhood $U \ni z$, we have $f^{-1}(U) \in p$. The point z with the above property, if it exists, is denoted $p\text{–}\lim_x f(x)$.

Likewise, if $S(x)$ is a statement which may be true or false for $x \in X$, then $S(x)$ is said to hold *p-almost always*, or *for p-almost all x* if $\{x \mid S(x) \text{ holds}\} \in p$.

If S is identified with a map $X \to \{\text{True}, \text{False}\}$ then this $S(x)$ holds for p-almost all x if and only if $p\text{-}\lim_x S(x) = \text{True}$. Conversely, $p\text{-}\lim_x f(x) = z$ precisely when for any $U \ni z$, open, $f(x) \in U$ for p-almost all x.

Fact B.0.4. Let $f\colon X \to Z$ be a map to a Hausdorff space, and let $p \in \beta X$. The there exists at most one limit of f along p. Moreover, if Z is compact, then the limit $p\text{-}\lim_x f(x)$ exists.

We list some basic properties of ultralimits. Because of the applications we have in mind, we specialise to ultrafilters on $\mathbb{N}$.

Fact B.0.5. Let $p \in \beta\mathbb{N} \setminus \mathbb{N}$, and let $f\colon \mathbb{N} \to Z$, where Z is a compact Hausdorff space. If the limit $\lim_{n\to\infty} f(n)$ exists, then $p\text{-}\lim_n f(n) = \lim_{n\to\infty} f(x)$.

In particular, if $f(x) = z$ is constant, then $p\text{-}\lim_n f(n) = z$.

Ultralimits behave nicely with respect to continuous functions, and hence also with arithmetic operations.

Fact B.0.6. Let $p \in \beta\mathbb{N} \setminus \mathbb{N}$, and $f\colon \mathbb{N} \to Z$, and let $g\colon Z \to Y$ be continuous, where Y, Z are Hausdorff spaces. Then $p\text{-}\lim_n g \circ f(n) = g\,(p\text{-}\lim_n f(n))$, provided that $p\text{-}\lim_n f(n)$ exists.

In particular, if $f_1, f_2\colon \mathbb{N} \to \mathbb{T}$, then $p\text{-}\lim_n(f_1(n) + f_2(n)) = p\text{-}\lim_n f_1(n) + p\text{-}\lim_n f_2(n)$. For functions $\mathbb{N} \to \mathbb{R}$, the usual formulas hold for limits of sums, products, etc.

For the convenience of the reader, we spell out a characterisation of iterated limits.

Fact B.0.7. Let $p_i \in \beta\mathbb{N}$ for $1 \le i \le r$, and $f\colon \mathbb{N}^r \to Z$, where Z is a Hausdorff compact space. Then, the statement $p_1\text{-}\lim_{n_1} p_2\text{-}\lim_{n_2} \dots p_r\text{-}\lim_{n_r} f(n_1, \dots, n_r) = z$ is equivalent the statement that for any open neighbourhood U of z, there exist families of sets $A_1 \in p_1,\ A_2(n_1) \in p_2,\ A_3(n_1, n_2) \in p_3,\ \dots, A_r(n_1, \dots, n_{r-1}) \in p_r$ such that for any n_i which satisfy the condition $n_j \in A_j(n_1, \dots, n_{j-1})$ it holds that $f(n_1, \dots, n_r) \in U$.

A key fact behind many applications of ultrafilters is that many interesting combinatorial properties have a natural characterisation in terms of ultrafilters. One may extend semigroup structure to from $\beta\mathbb{N}$, declaring $p + q = p\text{-}\lim_n q\text{-}\lim_m(n + m)$. An ultrafilter $p \in \beta\mathbb{N}$ is *idempotent* if $p + p = p$.

Theorem B.0.8. *Let $A \subset \mathbb{N}$. Then A is an* IP *set if and only if $A \in p$ for some idempotent $p \in \beta\mathbb{N}$. Likewise, A is* IP* *if $A \in p$ for any idempotent $p \in \beta\mathbb{N}$.*

Bibliography

[1] B. Adamczewski and J. Bell. Function fields in positive characteristic: expansions and Cobham's theorem. *J. Algebra*, 319(6):2337–2350, 2008.

[2] S. Akiyama. Cubic Pisot units with finite beta expansions. In *Algebraic number theory and Diophantine analysis (Graz, 1998)*, pages 11–26. de Gruyter, Berlin, 2000.

[3] J.-P. Allouche and J. Shallit. The ring of k-regular sequences. *Theoret. Comput. Sci.*, 98(2):163–197, 1992.

[4] J.-P. Allouche and J. Shallit. *Automatic sequences*. Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations.

[5] J.-P. Allouche and J. Shallit. The ring of k-regular sequences. II. *Theoret. Comput. Sci.*, 307(1):3–29, 2003. Words.

[6] L. E. Baum and M. M. Sweet. Continued fractions of algebraic power series in characteristic 2. *Ann. of Math. (2)*, 103(3):593–610, 1976.

[7] J. E. Baumgartner. A short proof of Hindman's theorem. *J. Combinatorial Theory Ser. A*, 17:384–386, 1974.

[8] J. P. Bell, M. Coons, and K. G. Hare. The minimal growth of a k-regular sequence. *Bull. Aust. Math. Soc.*, 90(2):195–203, 2014.

[9] J. P. Bell, M. Coons, and K. G. Hare. Growth degree classification for finitely generated semigroups of integer matrices. *Semigroup Forum*, 92(1):23–44, 2016.

[10] V. Bergelson. Weakly mixing PET. *Ergodic Theory Dynam. Systems*, 7(3):337–349, 1987.

[11] V. Bergelson. Ultrafilters, IP sets, dynamics, and combinatorial number theory. In *Ultrafilters across mathematics*, volume 530 of *Contemp. Math.*, pages 23–47. Amer. Math. Soc., Providence, RI, 2010.

[12] V. Bergelson, H. Furstenberg, and R. McCutcheon. IP-sets and polynomial recurrence. *Ergodic Theory Dynam. Systems*, 16(5):963–974, 1996.

[13] V. Bergelson, H. Furstenberg, and B. Weiss. Piecewise-Bohr sets of integers and combinatorial number theory. In *Topics in discrete mathematics*, volume 26 of *Algorithms Combin.*, pages 13–37. Springer, Berlin, 2006.

[14] V. Bergelson and A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.*, 9(3):725–753, 1996.

[15] V. Bergelson and A. Leibman. Topological multiple recurrence for polynomial configurations in nilpotent groups. *Adv. Math.*, 175(2):271–296, 2003.

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

[17] V. Bergelson and A. Leibman. $\mathrm{IP}_{\mathrm{r}}^{*}$ recurrence and nilsystems. Preprint. arXiv:1604.02489 [math.DS], 2017.

[18] V. Bergelson, A. Leibman, and E. Lesigne. Intersective polynomials and the polynomial Szemerédi theorem. *Adv. Math.*, 219(1):369–388, 2008.

[19] V. Bergelson and R. McCutcheon. Idempotent ultrafilters, multiple weak mixing and Szemerédi's theorem for generalized polynomials. *J. Anal. Math.*, 111:77–130, 2010.

[20] J. Berstel and P. Séébold. *A characterization of Sturmian morphisms*, pages 281–290. Springer Berlin Heidelberg, Berlin, Heidelberg, 1993.

[21] V. Berthé, H. Ei, S. Ito, and H. Rao. On substitution invariant Sturmian words: an application of Rauzy fractals. *Theor. Inform. Appl.*, 41(3):329–349, 2007.

[22] Y. Bugeaud. Automatic continued fractions are transcendental or quadratic. *Ann. Sci. Éc. Norm. Supér. (4)*, 46(6):1005–1022, 2013.

[23] J. Byszewski and J. Konieczny. Automatic sequences and generalised polynomials. Preprint, 2017.

[24] J. Byszewski and J. Konieczny. Sparse generalised polynomials. Preprint. arXiv:1612.00073 [math.NT]. To appear in Trans. Amer. Math. Soc., 2017.

[25] O. A. Camarena and B. Szegedy. Nilspaces, nilmanifolds and their morphisms. Preprint. arXiv:1009.3825 [math.DS], 2012.

[26] N. Chekhova, P. Hubert, and A. Messaoudi. Propriétés combinatoires, ergodiques et arithmétiques de la substitution de Tribonacci. *J. Théor. Nombres Bordeaux*, 13(2):371–394, 2001.

[27] N. Chevallier. Best simultaneous Diophantine approximations and multidimensional continued fraction expansions. *Mosc. J. Comb. Number Theory*, 3(1):3–56, 2013.

[28] A. Cobham. On the base-dependence of sets of numbers recognizable by finite automata. *Math. Systems Theory*, 3:186–192, 1969.

[29] H. Cohen. *A Course in Computational Algebraic Number Theory*. Graduate Texts in Mathematics. Springer Berlin Heidelberg, 2013.

[30] H. Derksen. A Skolem-Mahler-Lech theorem in positive characteristic and finite automata. *Invent. Math.*, 168(1):175–224, 2007.

[31] 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):1837–1866, 2012.

[32] M. Drmota. Subsequences of automatic sequences and uniform distribution. In *Uniform distribution and quasi-Monte Carlo methods*, volume 15 of *Radon Ser. Comput. Appl. Math.*, pages 87–104. De Gruyter, Berlin, 2014.

[33] M. Drmota, C. Mauduit, and J. Rivat. The Thue-Morse sequence along squares is normal. Preprint available at www.dmg.tuwien.ac.at/drmota/, 2013.

[34] M. Einsiedler, A. Katok, and E. Lindenstrauss. Invariant measures and the set of exceptions to Littlewood's conjecture. *Ann. of Math. (2)*, 164(2):513–560, 2006.

[35] M. Einsiedler and T. Ward. *Ergodic theory with a view towards number theory*, volume 259 of *Graduate Texts in Mathematics*. Springer-Verlag London, Ltd., London, 2011.

[36] G. Everest, A. van der Poorten, I. Shparlinski, and T. Ward. *Recurrence sequences*, volume 104 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2003.

[37] J.-H. Evertse. On sums of S-units and linear recurrences. *Compositio Math.*, 53(2):225–244, 1984.

[38] I. Fagnot. A little more about morphic Sturmian words. *Theor. Inform. Appl.*, 40(3):511–518, 2006.

[39] A. Fan. Oscillating sequences of higher order and topological systems of quasi-discrete spectrum. Preprint, 2017.

[40] A. Fish. Polynomial largeness of sumsets and totally ergodic sets. *Online J. Anal. Comb.*, 0(5):19, 2010.

[41] A. Fish. Solvability of linear equations within weak mixing sets. *Israel J. Math.*, 184:477–504, 2011.

[42] H. Furstenberg. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Analyse Math.*, 31:204–256, 1977.

[43] H. Furstenberg. *Recurrence in ergodic theory and combinatorial number theory.* Princeton University Press, Princeton, N.J., 1981. M. B. Porter Lectures.

[44] H. Furstenberg and Y. Katznelson. An ergodic Szemerédi theorem for commuting transformations. *J. Analyse Math.*, 34:275–291 (1979), 1978.

[45] H. Furstenberg and Y. Katznelson. An ergodic Szemerédi theorem for IP-systems and combinatorial theory. *J. Analyse Math.*, 45:117–168, 1985.

[46] H. Furstenberg and B. Weiss. Topological dynamics and combinatorial number theory. *J. Analyse Math.*, 34:61–85 (1979), 1978.

[47] A. O. Gel'fond. Sur les nombres qui ont des propriétés additives et multiplicatives données. *Acta Arith.*, 13:259–265, 1967/1968.

[48] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[49] W. T. Gowers and J. Wolf. The true complexity of a system of linear equations. *Proc. Lond. Math. Soc. (3)*, 100(1):155–176, 2010.

[50] B. Green. Higher-Order Fourier Analysis, I. (Notes available from the author).

[51] B. Green. *Additive combinatorics* [book review of [104]. *Bull. Amer. Math. Soc. (N.S.)*, 46(3):489–497, 2009.

[52] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167(2):481–547, 2008.

[53] B. Green and T. Tao. Linear equations in primes. *Ann. of Math. (2)*, 171(3):1753–1850, 2010.

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

[55] B. Green, T. Tao, and T. Ziegler. An inverse theorem for the Gowers $U^{s+1}[N]$-norm. *Ann. of Math. (2)*, 176(2):1231–1372, 2012.

[56] Y. Gutman, F. Manners, and P. P. Varjú. The structure theory of Nilspaces, I. Preprint. arXiv:1605.08945 [math.DS], 2016.

[57] Y. Gutman, F. Manners, and P. P. Varjú. The structure theory of Nilspaces, II: Representaion as nilmanifolds. Preprint. arXiv:1605.08948 [math.DS], 2016.

[58] Y. Gutman, F. Manners, and P. P. Varjú. The structure theory of Nilspaces, III: Inverse limit representations and topological dynamics. Preprint. arXiv:1605.08950 [math.DS], 2016.

[59] I. J. Håland. Uniform distribution of generalized polynomials. *J. Number Theory*, 45(3):327–366, 1993.

[60] I. J. Håland. Uniform distribution of generalized polynomials of the product type. *Acta Arith.*, 67(1):13–27, 1994.

[61] I. J. Håland and D. E. Knuth. Polynomials involving the floor function. *Math. Scand.*, 76(2):194–200, 1995.

[62] P. R. Halmos. *Lectures on ergodic theory*. Chelsea Publishing Co., New York, 1960.

[63] N. Hindman. Finite sums from sequences within cells of a partition of N. *J. Combinatorial Theory Ser. A*, 17:1–11, 1974.

[64] N. Hindman and D. Strauss. *Algebra in the Stone-Čech compactification*. de Gruyter Textbook. Walter de Gruyter & Co., Berlin, second edition, 2012.

[65] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math. (2)*, 161(1):397–488, 2005.

[66] B. Host and B. Kra. Uniformity seminorms on ℓ^∞ and applications. *J. Anal. Math.*, 108:219–276, 2009.

[67] B. Host and B. Kra. Nil-Bohr sets of integers. *Ergodic Theory Dynam. Systems*, 31(1):113–142, 2011.

[68] W. Huang, S. Shao, and X. Ye. Nil Bohr-sets and almost automorphy of higher order. *Mem. Amer. Math. Soc.*, 241(1143):v+83, 2016.

[69] P. Hubert and A. Messaoudi. Best simultaneous Diophantine approximations of Pisot numbers and Rauzy fractals. *Acta Arith.*, 124(1):1–15, 2006.

[70] A. Y. Khinchin. *Continued fractions*. Dover Publications, Inc., Mineola, NY, russian edition, 1997. With a preface by B. V. Gnedenko, Reprint of the 1964 translation.

[71] A. N. Khovanskii. *The application of continued fractions and their generalizations to problems in approximation theory*. Translated by Peter Wynn. P. Noordhoff N. V., Groningen, 1963.

[72] J. Konieczny. Applications of ultrafilters in ergodic theory and combinatorial number theory. Master's thesis, Jagiellonian University & Vrije Universiteit Amsterdam, 2013. arXiv:1310.1056 [math.DS].

[73] J. Konieczny. Sets of recurrence as bases for the positive integers. *Acta Arith.*, 174(4):309–338, 2016.

[74] J. Konieczny. Combinatorial properties of Nil–Bohr sets. *Israel J. Math.*, 220(1):333–385, 2017.

[75] J. Konieczny. Gowers norms for the Thue-Morse and Rudin-Shapiro sequences. Preprint. arXiv:1611.09985 [math.NT], 2017.

[76] J. Konieczny. Weakly mixing sets of integers and polynomial equations. *The Quarterly Journal of Mathematics*, 68(1):141, 2017.

[77] L. Kronecker. Zwei Sätze über Gleichungen mit ganzzahligen Coefficienten. *J. Reine Angew. Math.*, 53:173–175, 1857.

[78] J. C. Lagarias. Best simultaneous Diophantine approximations. II. Behavior of consecutive best approximations. *Pacific J. Math.*, 102(1):61–88, 1982.

[79] M. Lazard. Sur les groupes nilpotents et les anneaux de Lie. *Ann. Sci. Ecole Norm. Sup. (3)*, 71:101–190, 1954.

[80] A. Leibman. Polynomial sequences in groups. *J. Algebra*, 201(1):189–206, 1998.

[81] A. Leibman. Polynomial mappings of groups. *Israel J. Math.*, 129:29–60, 2002.

[82] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Israel J. Math.*, 146:303–315, 2005.

[83] A. Leibman. Pointwise convergence of ergodic averages for polynomial actions of $\mathbb{Z}^d$ by translations on a nilmanifold. *Ergodic Theory Dynam. Systems*, 25(1):215–225, 2005.

[84] A. Leibman. Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold. *Ergodic Theory Dynam. Systems*, 25(1):201–213, 2005.

[85] A. Leibman. A canonical form and the distribution of values of generalized polynomials. *Israel J. Math.*, 188:131–176, 2012.

[86] E. Lindenstrauss. Pointwise theorems for amenable groups. *Invent. Math.*, 146(2):259–295, 2001.

[87] M. Lothaire. *Algebraic combinatorics on words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002.

[88] A. I. Malcev. On a class of homogeneous spaces. *Amer. Math. Soc. Translation*, 1951(39):33, 1951.

[89] C. Mauduit and J. Rivat. La somme des chiffres des carrés. *Acta Math.*, 203(1):107–148, 2009.

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

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

[92] C. Mauduit and A. Sárközy. On finite pseudorandom binary sequences. II. The Champernowne, Rudin-Shapiro, and Thue-Morse sequences, a further construction. *J. Number Theory*, 73(2):256–276, 1998.

[93] R. McCutcheon. An infinitary polynomial van der Waerden theorem. *J. Combin. Theory Ser. A*, 86(2):214–231, 1999.

[94] C. Müllner and L. Spiegelhofer. Normality of the Thue-Morse sequence along Piatetski-Shapiro sequences, II. Preprint. To appear in Israel J. Math. arXiv:1511.01671 [math.NT], 2015.

[95] W. Parry. Ergodic properties of affine transformations and flows on nilmanifolds. *Amer. J. Math.*, 91:757–771, 1969.

[96] M. Pollicott and M. Yuri. *Dynamical systems and ergodic theory*, volume 40 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1998.

[97] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

[98] K. F. Roth. Rational approximations to algebraic numbers. *Mathematika*, 2:1–20; corrigendum, 168, 1955.

[99] A. Sárkőzy. On difference sets of sequences of integers. I. *Acta Math. Acad. Sci. Hungar.*, 31(1–2):125–149, 1978.

[100] W. M. Schmidt. Norm form equations. *Ann. of Math. (2)*, 96:526–551, 1972.

[101] J. Shallit. A generalization of automatic sequences. *Theoret. Comput. Sci.*, 61(1):1–16, 1988.

[102] E. Szemerédi. On sets of integers containing no four elements in arithmetic progression. *Acta Math. Acad. Sci. Hungar.*, 20:89–104, 1969.

[103] T. Tao. *Higher order Fourier analysis*, volume 142 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[104] T. Tao and V. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

[105] S. Tu. $\mathrm{Nil}\,\mathrm{Bohr}_0$-sets and polynomial recurrence. *J. Math. Anal. Appl.*, 409(2):890–898, 2014.

[106] A. J. van der Poorten and H. P. Schlickewei. Additive relations in fields. *J. Austral. Math. Soc. Ser. A*, 51(1):154–170, 1991.

[107] R. C. Vaughan. *The Hardy-Littlewood method*, volume 125 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, second edition, 1997.

[108] S.-I. Yasutomi. On Sturmian sequences which are invariant under some substitutions. In *Number theory and its applications (Kyoto, 1997)*, volume 2 of *Dev. Math.*, pages 347–373. Kluwer Acad. Publ., Dordrecht, 1999.

[109] T. Ziegler. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.*, 20(1):53–97, 2007.

[110] P. Zorin-Kranich. *Ergodic theorems for polynomials in nilpotent groups*. PhD thesis, Universiteit van Amsterdam, Sept. 2013.

[111] P. Zorin-Kranich. A nilpotent IP polynomial multiple recurrence theorem. *J. Anal. Math.*, 123:183–225, 2014.