ACTA ARITHMETICA
215.3 (2024)

On asymptotically automatic sequences

by

JAKUB KONIECZNY (Lyon and Oxford)

Abstract. We study the notion of an asymptotically automatic sequence, which generalises the notion of an automatic sequence. While k-automatic sequences are characterised by finiteness of k-kernels, the k-kernels of asymptotically k-automatic sequences are only required to be finite up to equality almost everywhere. We prove basic closure properties and a linear bound on asymptotic subword complexity, show that results concerning frequencies of symbols are no longer true for the asymptotic analogue, and discuss some classification problems.

1. Introduction. Automatic sequences are a widely-studied class of sequences, which can be defined in many equivalent ways, each shedding light from a different perspective. Perhaps the shortest definition uses the notion of k-kernel. For a sequence $\mathbf{a} = (a_n)_{n=0}^{\infty}$ and a base $k \geq 2$, the *k-kernel* of $\mathbf{a}$ is the set

$$\mathcal{N}_k(\mathbf{a}) := \{(a_{k^i n+r})_{n=0}^{\infty} \mid i, r \in \mathbb{N}_0,\ r < k^i\}.$$

A sequence $\mathbf{a}$ is *k-automatic* if its k-kernel is finite, $\#\mathcal{N}_k(\mathbf{a}) < \infty$.

We will say that two sequences $\mathbf{a} = (a_n)_{n=0}^{\infty}$ and $\mathbf{b} = (b_n)_{n=0}^{\infty}$ are *asymptotically equal*, denoted $\mathbf{a} \simeq \mathbf{b}$, if $a_n = b_n$ for almost all $n \in \mathbb{N}_0$, i.e.,

$$\frac{1}{N}\#\{0 \leq n < N \mid a_n \neq b_n\} \to 0 \quad \text{as } N \to \infty.$$

In [Kon23], the author introduced the notion of an *asymptotically k-automatic sequence*, which is a sequence $\mathbf{a}$, over a finite alphabet, whose k-kernel is finite up to asymptotic equality. In other words, we require that the set

$$\widetilde{\mathcal{N}}_k(\mathbf{a}) := \mathcal{N}_k(\mathbf{a})/{\simeq}$$

should be finite, $\#\widetilde{\mathcal{N}}_k(\mathbf{a}) < \infty$, or equivalently that there exists a finite number of sequences $\mathbf{a}^{(0)}, \mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(d-1)}$ such that for each $\mathbf{b} \in \mathcal{N}_k(\mathbf{a})$ there exists $0 \leq j < d$ such that $\mathbf{b} \simeq \mathbf{a}^{(j)}$. Asymptotically k-automatic sequences

2020 *Mathematics Subject Classification*: Primary 11B85; Secondary 68Q45.
Key words and phrases: automatic sequences.
Received 19 June 2023; revised 11 April 2024.
Published online 12 August 2024 in Open Access (under CC-BY license).

include k-automatic sequences, their modifications on density zero sets of positions, as well as sequences $\mathbf{a} = (a_n)_{n=0}^{\infty}$ of the form $a_n = F(\{\log_k n\})$, where $F\colon [0,1) \to \Omega$ is a Riemann-measurable map taking values in some finite set Ω. More examples are also constructed in [Kon23].

The purpose of the present paper is to systematically investigate asymptotically automatic sequences, and to establish analogues of various well-known facts about automatic sequences—or to disprove them. There are several sources of motivation behind this endeavour. Firstly, we believe that the class of asymptotically automatic sequences is interesting in its own right and worthy of investigation. Secondly, some of the new results obtained here immediately imply new density versions of results expressed in terms of automatic sequences. In particular, this is the case for classification results obtained in Section 6. Finally, we hope that the arguments presented here will better illuminate the theory of automatic sequences.

In Section 2 we discuss basic properties of asymptotically k-automatic sequences. In particular, we prove closure under Cartesian products and codings, and under passing to an arithmetic progression. We also discuss the connection with automata, obtained in [Kon23].

In Section 3 we discuss the dependence of the notion of a k-automatic sequence on the base k. We show that for multiplicatively dependent bases k and ℓ, asymptotically k-automatic sequences are the same as asymptotically ℓ-automatic sequences. For multiplicatively independent bases k and ℓ, a celebrated theorem of Cobham asserts that the only sequences that are automatic in both bases are the eventually periodic ones. In contrast, as shown in [Kon23], there exist sequences that are asymptotically automatic in two multiplicatively independent bases without being asymptomatically equal to a periodic sequence; we can only show a considerably weaker property of asymptotic invariance under a shift. We investigate the set of bases with respect to which a given sequence is asymptotically automatic and show that it has a particularly structured form.

In Section 4 we discuss frequencies of symbols and subwords in asymptotically automatic sequences. In the case of automatic sequences, frequencies are not guaranteed to exist, but logarithmic frequencies are. Additionally, if frequencies do exist then they are rational. In contrast, we construct examples which show that for asymptotically automatic sequences logarithmic frequencies are not guaranteed to exist, and if frequencies exist then they are not guaranteed to be rational.

In Section 5 we discuss subword complexity and its newly introduced asymptotic analogue. Automatic sequences are notable for having linear subword complexity. We argue that subword complexity is not the correct notion to consider in the asymptotic regime, and show that asymptotically automatic sequences have linear asymptotic subword complexity. We also prove several basic facts about asymptotic subword complexity. In particular, we classify sequences whose asymptotic subword complexity is especially low, which is analogous to the classification of periodic and Sturmian sequences in terms of subword complexity.

In Section 6 we discuss classification problems. Specifically, we fully classify bracket words that are asymptotically automatic, in particular showing that Sturmian words are never asymptotically automatic. We also partially classify multiplicative sequences that are asymptotically automatic.

In Section 7 we introduce the notion of an asymptotically k-regular sequence, and we prove a variant of Cobham's theorem.

Notation. We let $\mathbb{N} = \{1, 2, \dots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For $N \in \mathbb{N}_0$, we let $[N] = \{0, 1, \dots, N-1\}$.

For a set Ω we let Ω^* denote the set of all finite words $w = w_0 w_1 \dots w_{\ell-1}$ ($\ell \in \mathbb{N}_0$, $w_0, w_1, \dots, w_{\ell-1} \in \Omega$) over Ω, including the empty word ϵ. The length of a word w is denoted by $|w|$. We let Ω^∞ denote the set of all sequences $\mathbf{w} = w_0 w_1 w_2 \dots$ over Ω. Given a sequence $\mathbf{w}$, we let w_i denote its ith entry (with $i \geq 0$) and for $i \leq j$ we let $w_{[i,j)} = w_i w_{i+1} \dots w_{j-1}$ denote the corresponding subword.

For a word $w \in \Omega^*$ and $n \in \mathbb{N}_0$, we let $w^n = w \dots w$ denote the n-fold repetition of w, and $w^\infty = ww \cdots \in \Omega^\infty$ denote the periodic sequence with period w. More generally, for $t \in \mathbb{R}_+$ we let w^t denote the prefix of w^∞ of length $\lfloor t|w| \rfloor$. (We caution that, in general, $(w^t)^s \neq w^{ts}$.)

For an integer $k \geq 2$, we let $\Sigma_k = \{0, 1, \dots, k-1\}$ denote the set of base-k digits. For an integer $n \in \mathbb{N}_0$, we let $(n)_k$ denote its base-k expansion (without leading zeros) and for $u \in \Sigma_k^*$ (possibly with leading zeros) we let $[u]_k$ denote the integer encoded by $[u]$. For $i \in \mathbb{N}_0$ we also let $(n)_k^i$ denote the unique word $u \in \Sigma_k^i$ such that $n \equiv [u]_k \bmod k^i$ (this is the length-i suffix of $(n)_k$ or the result of padding $(n)_k$ with leading zeros, depending on the length of $(n)_k$).

We use standard asymptotic notation. For two quantities X, Y, we write $X = O(Y)$ or $X \ll Y$ if $|X| \leq CY$ for a constant C. If C is additionally allowed to depend on a parameter A, we write $X = O_A(Y)$ or $X \ll_A Y$. Assuming that X and Y depend on a parameter N, we write $X = o_{N\to\infty}(Y)$ if $X/Y \to 0$ and $X = \omega_{N\to\infty}(Y)$ if $X/Y \to \infty$ as $N \to \infty$; if N is clear from the context and there is no risk of confusion, we simply write $o(Y)$ and $\omega(Y)$ instead.

We say that a statement $\varphi(n)$ holds *for almost all* $n \in \mathbb{N}$ if the set of $n \in \mathbb{N}$ for which $\varphi(n)$ is false has density zero, i.e., $\#\{n < N \mid \neg\varphi(n)\} = o(N)$ as $N \to \infty$.

2. Basic properties. In this section, we review some basic properties of asymptotically automatic sequences. Most of them are direct analogues of similar properties of automatic sequences and can be established using essentially the same argument.

LEMMA 2.1. *Let $k \geq 2$ and let $\mathbf{a}, \mathbf{b}$ be asymptotically k-automatic sequences. Then the Cartesian product $\mathbf{a} \times \mathbf{b} = ((a_n, b_n))_{n=0}^{\infty}$ is an asymptotically k-automatic sequence.*

Proof. It is enough to notice that $\mathcal{N}_k(\mathbf{a} \times \mathbf{b}) \subseteq \mathcal{N}_k(\mathbf{a}) \times \mathcal{N}_k(\mathbf{b})$. ∎

LEMMA 2.2. *Let $k \geq 2$, let $\mathbf{a}$ be an asymptotically k-automatic sequence over a finite alphabet Ω and let $\rho\colon \Omega \to \Omega'$ be a map. Then the coding $\rho(\mathbf{a}) = (\rho(a_n))_{n=0}^{\infty}$ is an asymptotically k-automatic sequence.*

Proof. It is enough to notice that $\mathcal{N}_k(\rho(\mathbf{a})) \subseteq \{\rho(\mathbf{b}) \mid \mathbf{b} \in \mathcal{N}_k(\mathbf{a})\}$. ∎

COROLLARY 2.3. *Let $k \geq 2$ and let $(S, \cdot)$ be a semigroup. Then asymptotically k-automatic sequences over S form a semigroup, with the operation defined pointwise, i.e. by $\mathbf{a} \cdot \mathbf{b} = (a_n \cdot b_n)_{n=0}^{\infty}$.*

Proof. Apply Lemmas 2.1 and 2.2 with $\rho\colon S \times S \to S$ given by $\rho(x, y) = x \cdot y$. ∎

In particular, the family of complex-valued asymptotically k-automatic sequences is closed under addition and multiplication, i.e. it constitutes a ring. Similarly, the multiplicative inverse of an asymptotically k-automatic sequence, if it exists, is asymptotically k-automatic.

LEMMA 2.4. *Let $k \geq 2$, let $\mathbf{a}$ be an asymptotically k-automatic sequence and let $\mathbf{a}'$ be a sequence with $\mathbf{a}' \simeq \mathbf{a}$. Then $\mathbf{a}'$ is asymptotically k-automatic.*

Proof. It is enough to notice that $\widetilde{\mathcal{N}}_k(\mathbf{a}) = \widetilde{\mathcal{N}}_k(\mathbf{a}')$. ∎

As a consequence of Lemma 2.4, it makes sense to speak of asymptotic k-automaticity of a sequence that is only defined almost everywhere. For instance, in the following result we consider the restriction $\mathbf{a}' = (a_{qn+r})_{n=0}^{\infty}$ of an asymptotically k-automatic sequence $\mathbf{a}$ to an arithmetic progression $(qn + r)_{n=0}^{\infty}$. When $r < 0$, a few initial terms of $\mathbf{a}'$ are not well-defined. However, we can assign arbitrary values to those entries, and whether or not the resulting sequence is asymptotically k-automatic does not depend on the choices made.

LEMMA 2.5. *Let $k \geq 2$, let $\mathbf{a}$ be a sequence over a finite alphabet, and let $q \in \mathbb{N}$. For $r \in \mathbb{Z}$, let $\mathbf{a}^{(r)} := (a_{qn+r})_{n=0}^{\infty}$, where we let $a_n := a_0$ for $n < 0$.*

(i) *If $\mathbf{a}$ is asymptotically k-automatic then so is $\mathbf{a}^{(r)}$ for each $r \in \mathbb{Z}$.*
(ii) *If for all r with $0 \leq r < q$ the sequence $\mathbf{a}^{(r)}$ is asymptotically k-automatic then $\mathbf{a}$ is also asymptotically k-automatic.*

Proof. (i) It will suffice to consider the special cases where $q = 1$ and $r = \pm 1$, and where $r = 0$. The general case can be recovered by repeated applications of these two cases. For $r = 0$, the proof follows the argument of [AS03b, Theorem 6.8.1], except that we replace $\mathcal{N}_k$ with $\widetilde{\mathcal{N}}_k$. The key idea is that the set

$$\{(b_{qn+s})_{n=0}^{\infty} \mid \mathbf{b} \in \mathcal{N}_k(\mathbf{a}),\ 0 \le s < q\}$$

is finite up to equality almost everywhere and contains $\mathcal{N}_k(\mathbf{a}^{(0)})$. In the case where $q = 1$ and $r = \pm 1$, similarly, the proof follows the argument for [AS03b, Theorem 6.8.3]. The key idea this time is that the set

$$\mathcal{N}_k(\mathbf{a}) \cup \{\mathbf{b}^{(r)} \mid \mathbf{b} \in \mathcal{N}_k(\mathbf{a})\},$$

is finite up to equality almost everywhere and contains $\mathcal{N}_k(\mathbf{a}^{(r)})$. (Here, $\mathbf{b}^{(r)}$ is defined by $b_n^{(r)} := b_{qn+r} = b_{n\pm 1}$ if $qn + r \ge 0$ and $b_n^{(r)} := b_0$ otherwise.)

(ii) Pick any $\mathbf{b} \in \mathcal{N}_k(\mathbf{a})$, say $b_n = a_{k^i n+m}$ with $m, i \in \mathbb{N}_0$, $m < k^i$. Then $\mathbf{b}$ is uniquely determined by the sequences $\mathbf{b}^{(s)}$ for s with $0 \le s < q$. Pick any $0 \le s < q$ and put $r := k^i s + m \bmod q$ and $h := \lfloor (k^i s + m)/q \rfloor$. Then $h < k^i$ and

$$b_n^{(s)} = a_{qk^i n + k^i s + m} = a^{(r)}_{k^i n + h}.$$

Thus, $\mathbf{b}^{(s)} \in \mathcal{N}_k(\mathbf{a}^{(r)})$. Since $\widetilde{\mathcal{N}}_k(\mathbf{a}^{(r)})$ is finite and r is chosen from a finite set, we conclude that $\widetilde{\mathcal{N}}_k(\mathbf{a})$ is finite. ∎

For a sequence $\mathbf{a}$ taking values in a commutative monoid we may consider the sequence of partial sums $\Sigma\mathbf{a}$ given by $(\Sigma a)_n = \sum_{m=0}^{n-1} a_m$ (in particular, $a_0 = 0$). The following lemma is an analogue of [AS03b, Corollary 6.9.3].

Lemma 2.6. *Let $\mathbf{a}$ be an asymptotically k-automatic sequence over a finite abelian monoid $(S, +)$. Then the sequence $\Sigma\mathbf{a}$ is asymptotically k-automatic.*

Proof. For any sequence $\mathbf{b}$ and $0 \le i < k$, let $\mathbf{b}^{(i)}$ be the sequence given by $b^{(i)} := b_{kn+i}$. For each $0 \le j \le k$ we have

$$\begin{aligned}(\Sigma b)_{kn+j} &= \sum_{m=0}^{n-1}\sum_{i=0}^{k-1} b_{km+i} + \sum_{i=0}^{j-1} b_{kn+i}\\ &= \sum_{i=0}^{k-1}(\Sigma b^{(i)})_n + \sum_{i=0}^{j-1} b_n^{(i)} = \sum_{i=0}^{j-1}(\Sigma b^{(i)})_{n+1} + \sum_{i=j}^{k-1}(\Sigma b^{(i)})_n.\end{aligned}$$

Let $\mathcal{S} \subseteq S^\infty$ be the submonoid of S^∞ generated by $\Sigma\mathbf{b}$ and their shifts by 1 (i.e., $((\Sigma b)_{n+1})_{n=0}^{\infty}$) for $\mathbf{b} \in \mathcal{N}_k(\mathbf{a})$, with the operation defined pointwise. Since S and $\widetilde{\mathcal{N}}_k(\mathbf{a})$ are finite, $\mathcal{S}/{\simeq}$ is finite. Applying the identity obtained above, we see that $\mathcal{S}$ is closed under the operation $\mathbf{c} \mapsto (c_{kn+i})_{n=0}^{\infty}$ for each $0 \le i < k$. It follows that $\mathcal{N}_k(\Sigma\mathbf{a}) \subseteq \mathcal{S}$. Thus, $\widetilde{\mathcal{N}}_k(\Sigma\mathbf{a})$ is finite. ∎

We close this section by citing a result from [Kon23] which elucidates the connection between automata and asymptotically automatic sequences. For a map $\phi\colon \Sigma_k^* \to \Omega$, we define the *k-kernel* (or simply the *kernel*, since there is no risk of ambiguity)

$$\mathcal{N}_k(\phi) = \{\phi_v \mid v \in \Sigma_k^*\}, \qquad \phi_v(u) := \phi(uv) \quad \text{for } u, v \in \Sigma_k^*.$$

The map ϕ is automatic if and only if its kernel is finite, $\#\mathcal{N}_k(\phi) < \infty$.

PROPOSITION 2.7 ([Kon23]). *Let* $\mathbf{a}$ *be an asymptotically* k*-automatic sequence. Then there exist* $d \in \mathbb{N}$ *and an automatic map* $\phi\colon \Sigma_k^* \to \Sigma_d$ *as well as sequences* $\mathbf{a}^{(0)}, \mathbf{a}^{(1)}, \dots, \mathbf{a}^{(d-1)}$ *such that for each* $u \in \Sigma_k^*$ *we have*

$$(a_{k^{|u|}n+[u]_k})_{n=0}^{\infty} \simeq (a_n^{(\phi(n))})_{n=0}^{\infty}.$$

In applications, we will always assume that the value of d is minimal, meaning that no pair of sequences $\mathbf{a}^{(0)}, \mathbf{a}^{(1)}, \dots, \mathbf{a}^{(d-1)}$ is asymptotically equal. The following can be construed as an analogue of the pumping lemma for automatic sequences [AS03b, Lem. 4.2.1].

LEMMA 2.8. *Let* $\mathbf{a}$ *be an asymptotically* k*-automatic sequence. Then there exists* $p \in \mathbb{N}$ *such that for each word* $w \in \Sigma_k^*$ *with* $|w| \geq p$ *there exist words* $u, v, v' \in \Sigma_k^*$ *such that* $w = vuv'$, $|w| - p \leq |u| < |w|$ *and, letting* $w(t) := vu^tv'$, *for all* $t \in \mathbb{N}_0$ *we have*

$$(a_{k^{|w|}n+[w]_k})_{n=0}^{\infty} \simeq (a_{k^{|w(t)|}n+[w(t)]_k})_{n=0}^{\infty}.$$

Proof. Apply Proposition 2.7 and then apply the pumping lemma to ϕ. ■

3. Bases. A classical theorem due to Cobham asserts that for an automatic sequence there is, essentially, only one base with respect to which it is automatic. To be more precise, it is not hard to show that each eventually periodic sequence is automatic in every base and that for any pair of multiplicatively dependent integers $k, \ell \geq 2$ (i.e., $\log k/\log \ell \in \mathbb{Q}$), k-automatic sequences are the same as ℓ-automatic sequences. Cobham's theorem asserts that for each automatic but not eventually periodic sequence $\mathbf{a}$ there exists precisely one integer $k \geq 2$ that is not a perfect power and such that $\mathbf{a}$ is k-automatic.

In contrast, an example from [Kon23] shows that there exists a sequence $\mathbf{a}$ that is asymptotically 2- and 3-automatic, but not asymptotically periodic. (It also seems very plausible that this sequence is not k-automatic for those k that have at least one prime factor $p \geq 5$, but the proof of this remains elusive.) Thus, for a given sequence over a finite alphabet, it makes sense to inquire into the set of bases with respect to which it is asymptotically automatic.

We first note that, in analogy with automatic sequences, asymptotically automatic sequences in multiplicatively dependent bases are the same. The argument is virtually identical to the one for automatic sequences, but we include it here for the sake of completeness.

LEMMA 3.1. *Let* $\mathbf{a}$ *be a sequence over a finite alphabet* Ω *and let* $k, \ell \geq 2$ *be multiplicatively dependent integers. Then* $\mathbf{a}$ *is asymptotically* k*-automatic if and only if* $\mathbf{a}$ *is asymptotically* ℓ*-automatic.*

Proof. It is enough to consider the case where $\ell = k^m$ is an integer power of k with $m \geq 2$. In this case we have $\mathcal{N}_\ell(\mathbf{a}) \subseteq \mathcal{N}_k(\mathbf{a})$ and hence also $\widetilde{\mathcal{N}}_\ell(\mathbf{a}) \subseteq \widetilde{\mathcal{N}}_k(\mathbf{a})$. Thus, if $\mathbf{a}$ is k-automatic then it is also ℓ-automatic. Conversely, suppose that $\mathbf{a}$ is ℓ-automatic, and consider a sequence $\mathbf{b} \in \widetilde{\mathcal{N}}_k(\mathbf{a})$, say $b_n = a_{k^i n + r}$ where $0 \leq r < k^i$. Put $j = i \bmod m$ and $s = \lfloor r/\ell^{\lfloor i/m \rfloor} \rfloor$. Bearing in mind the identity

$$k^i n + r = \ell^{\lfloor i/m \rfloor}(k^j n + s) + (r \bmod \ell^{\lfloor i/m \rfloor}),$$

we see that $\mathbf{b} \in \widetilde{\mathcal{N}}_\ell(\mathbf{a}')$, where $a'_n = a_{k^j n + s}$. By Lemma 2.5, $\mathbf{a}'$ is asymptotically ℓ-automatic, and hence $\widetilde{\mathcal{N}}_\ell(\mathbf{a}')$ is finite. Since $\mathbf{b} \in \widetilde{\mathcal{N}}_k(\mathbf{a})$ was arbitrary and j and s are uniformly bounded with respect to $\mathbf{b}$, it follows that $\widetilde{\mathcal{N}}_k(\mathbf{a})$ is finite. ∎

Next, we show some closure properties of the set of bases with respect to which a given sequence is asymptotically automatic.

LEMMA 3.2. *Let* $k, \ell \geq 2$ *be multiplicatively independent integers and let* $\mathbf{a}$ *be an asymptotically* k*-automatic sequence. Then* $\mathbf{a}$ *is asymptotically* ℓ*-automatic if and only if it is asymptotically* $k\ell$*-automatic.*

Proof. Suppose first that $\mathbf{a}$ is asymptotically ℓ-automatic. To show that $\mathbf{a}$ is asymptotically $k\ell$-automatic it is enough to notice that

$$\mathcal{N}_{k\ell}(\mathbf{a}) \subseteq \bigcup_{\mathbf{b} \in \mathcal{N}_\ell(\mathbf{a})} \mathcal{N}_k(\mathbf{b}),$$

and each sequence $\mathbf{b} \in \mathcal{N}_\ell(\mathbf{a})$ is asymptotically k-automatic by Lemma 2.5.

Suppose next that $\mathbf{a}$ is asymptotically $k\ell$-automatic. Let $I \in \mathbb{N}_0$ be sufficiently large that for each integer R with $0 \leq R < k^I$ there exist infinitely many pairs $(i, r) \in \mathbb{N}_0^2$ with $0 \leq r < k^i$ such that $(a_{k^i n + r})_{n=0}^\infty \simeq (a_{k^I n + R})_{n=0}^\infty$. We will show that for each $0 \leq R < k^I$ the sequence $(a_{k^I n + R})_{n=0}^\infty$ is asymptotically ℓ-automatic, which will finish the argument by Lemma 2.5. Each sequence in $\mathcal{N}_\ell((a_{k^I n + R})_{n=0}^\infty)$ takes the form $(a_{k^I \ell^j n + t})_{n=0}^\infty$ for some $j, t \in \mathbb{N}_0$ with $0 \leq t < k^I \ell^j$ and $t \equiv R \bmod k^I$. By the definition of I, we can find arbitrarily large $i \in \mathbb{N}_0$ and $0 \leq r < k^i$ such that

$$\begin{aligned}(a_{k^I \ell^j n + t})_{n=0}^\infty &= (a_{k^I(\ell^j n + \lfloor t/k^I \rfloor) + R})_{n=0}^\infty \\ &\simeq (a_{k^i(\ell^j n + \lfloor t/k^I \rfloor) + r})_{n=0}^\infty = (a_{k^i \ell^j n + u})_{n=0}^\infty,\end{aligned}$$

where $u = k^i\lfloor t/k^I \rfloor + r < k^i \ell^j$. Since $\mathbf{a}$ is asymptotically $k\ell$-automatic, we can find $J, T \in \mathbb{N}_0$ with J bounded by a constant independent of j and $0 \le T < (k\ell)^J$ such that

$$(a_{k^i\ell^j n+u})_{n=0}^{\infty} = (a_{(k\ell)^j(k^{i-j}n+\lfloor u/(k\ell)^j\rfloor)+(u \bmod (k\ell)^j)})_{n=0}^{\infty}$$
$$\simeq (a_{(k\ell)^J(k^{i-j}n+\lfloor u/(k\ell)^j\rfloor)+T})_{n=0}^{\infty} = (a_{k^{i-j+J}\ell^J n+v})_{n=0}^{\infty},$$

where $v = (k\ell)^J\lfloor u/(k\ell)^j\rfloor + T < k^{i-j+J}\ell^J$. Finally, using the fact that $\mathbf{a}$ is asymptotically k-automatic, we can find $M, V \in \mathbb{N}_0$ with M bounded by a constant independent of j and $0 \le V < k^M$ such that

$$(a_{k^{i-j+J}\ell^J n+v})_{n=0}^{\infty} = (a_{k^{i-j+J}(\ell^J n+\lfloor v/k^{i-j+J}\rfloor)+(v \bmod k^{i-j+J})})_{n=0}^{\infty}$$
$$\simeq (a_{k^M(\ell^J n+\lfloor v/k^{i-j+J}\rfloor)+V})_{n=0}^{\infty} = (a_{k^M\ell^J n+S})_{n=0}^{\infty},$$

where $S = k^M\lfloor v/k^{i-j+J}\rfloor + V < k^M\ell^J$. Ultimately, $(a_{k^I\ell^j n+t})_{n=0}^{\infty}$ is asymptotically equal to the sequence $(a_{k^M\ell^J n+S})_{n=0}^{\infty}$, which is taken from a finite set independent of j and t. This shows $\widetilde{\mathcal{N}}_\ell((a_{k^I n+R})_{n=0}^{\infty})$ is finite, as needed. ■

As a consequence of Lemma 3.2, the set of bases with respect to which $\mathbf{a}$ is asymptotically automatic has a particularly structured form. Before we proceed, we need the following simple lemma on vector spaces.

LEMMA 3.3. *Let $S \subseteq \bigoplus_{i=1}^{\infty}\mathbb{Q}$ be a set satisfying the following properties:*

(i) *for each $u \in S$ and $i \in \mathbb{N}$, $u_i \in \mathbb{N}_0$;*
(ii) *for each $u \in S$ and $m \in \mathbb{N}$, if $m \mid u_i$ for all $i \in \mathbb{N}$ then $(1/m)u \in S$;*
(iii) *for all $u, v \in S$, $u + v \in S$;*
(iv) *for all $u, v \in S$, if $u_i \ge v_i$ for all $i \in \mathbb{N}$ then $u - v \in S$.*

Then there exists a vector space $V < \bigoplus_{i=1}^{\infty}\mathbb{Q}$ such that

$$S = V \cap \mathbb{N}_0^{\infty} = \{v \in V \mid v_i \in \mathbb{N}_0 \quad \textit{for all } i \in \mathbb{N}\}.$$

Proof. Let V denote the set of all $v \in \bigoplus_{i=1}^{\infty}\mathbb{Q}$ such that there exist $m \in \mathbb{N}$ and $u \in S$ with $mv + u \in S$. It is routine to check that if $v, v' \in V$ and $n \in \mathbb{N}$ then $v + v' \in V$, $(1/n)v \in V$ and $-v \in V$. Thus, V is a vector space. Directly by definition, for all $v \in S$ we have $v \in V$ and $v_i \in \mathbb{N}_0$ for all $i \in \mathbb{N}$. Suppose conversely that $v \in V$ and $v_i \in \mathbb{N}_0$ for all $i \in \mathbb{N}_0$. Then $mv + u \in S$ for some $u \in S$ and $m \in \mathbb{N}$. We then infer from the closure properties of S that $mv \in S$ and $v \in S$. ■

For a prime p, we let ν_p denote the p-adic valuation, meaning that $p^{\nu_p(n)}$ is the largest power of p that divides n. Let also $p_1, p_2, \ldots$ be the increasing enumeration of the primes.

COROLLARY 3.4. *Let $\mathbf{a}$ be a sequence over a finite alphabet Ω. There exists a vector subspace $V < \bigoplus_{i=1}^{\infty}\mathbb{Q}$ such that for $k \in \mathbb{N}$ with $k > 1$, $\mathbf{a}$ is asymptotically k-automatic if and only if $(\nu_{p_1}(k), \nu_{p_2}(k), \ldots) \in V$.*

Proof. By Lemma 3.2, the set

$$S = \{(0, 0, \dots)\} \cup \{(\nu_{p_1}(k), \nu_{p_2}(k), \dots) \mid \mathbf{a} \text{ is asymptotically } k\text{-automatic}\}$$

satisfies the assumptions of Lemma 3.3. ∎

It seems plausible that the following converse to Corollary 3.4 is also true.

CONJECTURE 3.5. Let $V < \bigoplus_{i=1}^{\infty} \mathbb{Q}$ be a vector subspace. Then there exists a sequence $\mathbf{a}$ over the alphabet $\{0, 1\}$ such that for $k \in \mathbb{N}$ with $k > 1$, $\mathbf{a}$ is asymptotically k-automatic if and only if $(\nu_{p_1}(k), \nu_{p_2}(k), \dots) \in V$.

Let us briefly explain the rationale behind the conjecture above. Pick a set P of primes. We will construct a candidate for a sequence that is asymptotically k-automatic if and only if k is a product of primes from P. (The fact that we are dealing specifically with primes does not seem to be crucial, but it makes the discussion somewhat simpler.)

Assume first that $P = \{p_1, \dots, p_r\}$ is finite. Let $\mathcal{D} \subseteq \mathbb{N}$ be the set of all products of primes in P,

$$\mathcal{D} = \{p_1^{i_1} \dots p_r^{i_r} \mid i_1, \dots, i_r \in \mathbb{N}_0\}.$$

We use $\mathcal{D}$ to partition $\mathbb{N}$ into intervals $I = [N, M)$, where $N < M$ are consecutive elements of $\mathcal{D}$. Let us say that an interval I as mentioned above is *exceptional* if there exists a prime $p \in P$ such that $p \nmid N$ or $p \nmid M$. If I is not exceptional then for any $p \in P$ the integers N/p M/p are again consecutive elements of $\mathcal{D}$. This motivates us to consider a sequence $\mathbf{a}$ that is constant between consecutive elements of $\mathcal{D}$. For concreteness, writing $N = p_1^{i_1} \dots p^{i_r}$ and $M = p_1^{j_1} \dots p^{j_r}$, we can set

$$a_n = i_1 + \dots + i_r + j_1 + \dots + j_r \bmod 1 \quad \text{for } n \in I.$$

If I is not exceptional then for $n \in I$ we have $a_n = a_{\lfloor n/p \rfloor}$. In other words, letting E denote the union of all exceptional intervals, we have $a_n = a_{\lfloor n/p \rfloor}$ for all $n \in \mathbb{N} \setminus E$. We conjecture that E has density zero; this is trivial when $r = 1$, a simple lemma proved in [Kon23] when $r = 2$, and open for $r \geq 3$. If this conjecture is true then the sequence $\mathbf{a}$ defined above is asymptotically p-automatic for all $p \in P$, and hence also for all k that are products of primes in P. (Indeed, $\widetilde{\mathcal{N}}_p(\mathbf{a})$ has only one element.) On the other hand, when k has prime factors not belonging to P, there seems to be no particular reason why $\mathbf{a}$ should be asymptotically k-automatic—hence, we can reasonably conjecture that $\mathbf{a}$ is not asymptotically k-automatic.

In the case where $P = \{p_1, p_2, \dots\}$ is infinite, we modify the construction above by setting

$$\mathcal{D} = \{N = p_1^{i_1} \dots p_r^{i_r} \mid i_1, \dots, i_r \in \mathbb{N}_0,\ r \leq \ell(N)\},$$

where $\ell \colon \mathbb{N} \to \mathbb{N}$ is non-decreasing and $\ell(N)$ tends to infinity very slowly as $N \to \infty$. We then proceed as above. Assuming that the finite case goes through (meaning that the two conjectures we put forward are indeed true), we can derive the infinite case by a limiting argument. (Since the discussion is largely conjectural, we skip further details.)

4. Frequencies of symbols

4.1. Logarithmic frequencies. For a sequence $\mathbf{a}$ over a finite alphabet Ω and for a symbol $\omega \in \Omega$, we define the *lower* and *upper* (*asymptotic*) *frequencies* of ω in $\mathbf{a}$:

$$\overline{\mathrm{freq}}_{\mathbf{a}}(\omega) := \overline{d}(\{n \in \mathbb{N}_0 \mid a_n = \omega\}) = \limsup_{N\to\infty} \frac{\#\{n \in [N] \mid a_n = \omega\}}{N}, \tag{1}$$

$$\underline{\mathrm{freq}}_{\mathbf{a}}(\omega) := \underline{d}(\{n \in \mathbb{N}_0 \mid a_n = \omega\}) = \liminf_{N\to\infty} \frac{\#\{n \in [N] \mid a_n = \omega\}}{N}. \tag{2}$$

More generally, for a word $w \in \Omega^\ell$, we define

$$\overline{\mathrm{freq}}_{\mathbf{a}}(w) := \overline{d}(\{n \in \mathbb{N}_0 \mid a_{[n,n+\ell)} = w\}), \tag{3}$$

$$\underline{\mathrm{freq}}_{\mathbf{a}}(w) := \underline{d}(\{n \in \mathbb{N}_0 \mid a_{[n,n+\ell)} = w\}). \tag{4}$$

When $\overline{\mathrm{freq}}_{\mathbf{a}}(w) = \underline{\mathrm{freq}}_{\mathbf{a}}(w)$, the common value is called the *frequency* of w in $\mathbf{a}$ and is denoted by $\mathrm{freq}_{\mathbf{a}}(w)$. Replacing $\overline{d}$ and $\underline{d}$ with upper and lower logarithmic densities given by

$$\overline{d}^{\log}(A) = \limsup_{N\to\infty} \frac{1}{\log N} \sum_{n=0}^{N-1} \frac{1_A(n)}{n+1}, \qquad \underline{d}^{\log}(A) = \liminf_{N\to\infty} \frac{1}{\log N} \sum_{n=0}^{N-1} \frac{1_A(n)}{n+1},$$

we similarly obtain the notions of *upper* and *lower logarithmic frequencies*:

$$\overline{\mathrm{freq}}^{\log}_{\mathbf{a}}(w) := \overline{d}^{\log}(\{n \in \mathbb{N}_0 \mid a_{[n,n+\ell)} = w\}), \tag{5}$$

$$\underline{\mathrm{freq}}^{\log}_{\mathbf{a}}(w) := \underline{d}^{\log}(\{n \in \mathbb{N}_0 \mid a_{[n,n+\ell)} = w\}), \tag{6}$$

as well as the notion of *logarithmic frequency*,

$$\mathrm{freq}^{\log}_{\mathbf{a}}(w) = \overline{\mathrm{freq}}^{\log}_{\mathbf{a}}(w) = \underline{\mathrm{freq}}^{\log}_{\mathbf{a}}(w).$$

When $\mathbf{a}$ is automatic, the frequencies of symbols are not guaranteed to exist, as shown by the basic example $a_n = |(n)_2| \bmod 2$. However, the logarithmic frequencies exist for all symbols, and more generally for all words [AS03b, Thm. 8.4.9]. Thus, it may come as a surprise that there exist asymptotically automatic sequences such that logarithmic frequencies of symbols do not exist.

PROPOSITION 4.1. *There exists an asymptotically* 2*-automatic sequence* $\mathbf{a}$ *over the alphabet* $\{0,1\}$ *such that* $\overline{\mathrm{freq}}^{\log}_{\mathbf{a}}(1) = 1$ *and* $\underline{\mathrm{freq}}^{\log}_{\mathbf{a}}(1) = 0$*. In particular,* $\mathrm{freq}^{\log}_{\mathbf{a}}(1)$ *does not exist.*

Proof. Following [Kon23, Sec. 4], let $(H_i)_{i=0}^{\infty} = (2^{\alpha_i}3^{\beta_i})_{i=0}^{\infty}$ be the increasing enumeration of the set

$$\mathcal{H} = \{2^{\alpha}3^{\beta} \mid \alpha, \beta \in \mathbb{N}_0\} = \{1, 2, 3, 4, 6, 8, 9, 12, \dots\}. \tag{7}$$

Let $(\gamma_j)_{j=0}^{\infty}$ be an increasing sequence of integers with $\gamma_0 = 0$, which will be specified in the course of the argument. We define the sequence $\mathbf{a}$ by putting $a_0 = 0$ and for each $j \in \mathbb{N}_0$ and each $i \in \mathbb{N}_0$ such that $\beta_i \in [\gamma_j, \gamma_{j+1})$ letting

$$a_n = j \bmod 2 \in \{0, 1\} \quad \text{for } n \in [H_i, H_{i+1}). \tag{8}$$

It follows by the same argument as in [Kon23, Lem. 4.2] that

$$a_n = a_{n+1} = a_{2n} \quad \text{for almost all } n \in \mathbb{N}_0, \tag{9}$$

and as a consequence, $\mathbf{a}$ is asymptotically 2-automatic.

We plan to show that for each $j \in \mathbb{N}$, letting $N = 3^{\gamma_j}$, we have

$$\frac{1}{\log N} \sum_{n=0}^{N-1} \frac{a_n}{n+1} \in \begin{cases} [0, 2^{-j}) & \text{if } j \equiv 1 \bmod 2, \\ (1 - 2^{-j}, 1] & \text{if } j \equiv 0 \bmod 2. \end{cases} \tag{10}$$

Once (10) has been proved, letting $j \to \infty$ we conclude that $\overline{\mathrm{freq}}_{\mathbf{a}}^{\log}(1) = 1$ and $\underline{\mathrm{freq}}_{\mathbf{a}}^{\log}(1) = 0$, as needed. We also note that (10) only depends on $\gamma_0, \gamma_1, \dots, \gamma_j$ and not on $\gamma_{j+1}, \gamma_{j+2}, \dots$ (this is because each $n \in [1, N) = [1, 3^{\gamma_j})$ belongs to an interval $[H_i, H_{i+1}) \subseteq [1, 3^{\gamma_j})$ which satisfies $\beta_i < \gamma_j$). Our plan is to let γ_j grow very rapidly with j. Thus, it will suffice to show that, given any values of $\gamma_0, \gamma_1, \dots, \gamma_{j-1}$, condition (10) holds for all sufficiently large values of γ_j. For concreteness, suppose that $j \equiv 1 \bmod 2$. For each interval $[H_i, H_{i+1}) \subseteq [1, 3^{\gamma_j})$, if $\beta_i \geq \gamma_{j-1}$ then we have $a_n = 0$ for all $n \in [H_i, H_{i+1})$. Using the rough estimate $a_n \leq 1$ for all remaining n and applying Lemma 4.2 below concludes the argument. ∎

LEMMA 4.2. *For each $\beta \in \mathbb{N}_0$, we have*

$$\lim_{k \to \infty} \frac{1}{\log H_k} \sum_{i<k,\, \beta_i = \beta} \sum_{n=H_i}^{H_{i+1}-1} \frac{1}{n+1} = 0. \tag{11}$$

Proof. Let S_k denote the sum in (11) and pick $\varepsilon > 0$. By [Kon23, Lem. 4.1] we have $H_{i+1}/H_i \to 0$ as $i \to \infty$. Hence, we can pick i_1 such that $H_{i+1} \leq (1+\varepsilon)H_i$ for all $i \geq i_1$. Splitting the outer sum in (11) at i_1 and estimating $1/(n+1)$ from above by $1/H_i$ for $n \in [H_i, H_{i+1})$, we obtain

$$S_k \leq O_\varepsilon(1) + \varepsilon \cdot \#\{i \mid i_1 \leq i < k,\ \beta_i = \beta\} \leq O_\varepsilon(1) + \varepsilon \log_2 H_k. \tag{12}$$

It follows that

$$\limsup_{k \to \infty} \frac{S_k}{\log H_k} \leq \frac{\varepsilon}{\log 2}. \tag{13}$$

Letting $\varepsilon \to 0$ concludes the proof. ∎

4.2. Rationality of frequencies. Another notable property of automatic sequences is that frequencies of symbols, if they exist, are rational [AS03b, Thm. 8.4.5(b)]. We prove that also this property does not extend to asymptotically automatic sequences, which can have arbitrary frequencies of symbols. In particular, this shows that there are uncountably many distinct asymptotically automatic sequences (up to equality almost everywhere and renaming symbols), in contrast with automatic sequences, which are easily seen to be countable.

PROPOSITION 4.3. *For each $\theta \in [0,1]$, there exists an asymptotically 2-automatic sequence $\mathbf{a}$ over the alphabet $\{0,1\}$ such that $\mathrm{freq}_{\mathbf{a}}(1) = \theta$.*

Proof. For each $n \in \mathbb{N}_0$, there is a unique decomposition of the binary expansion

$$(n)_2 = u_n^{(1)} u_n^{(2)} \dots u_n^{(r(n))} v_n, \tag{14}$$

where $r(n) \in \mathbb{N}_0$, $u_n^{(i)} \in \Sigma_2^* 1$ and $|u_n^{(i)}|_1 = i$ for $i = 1, 2, \dots, r(n)$, and $v_n \in \Sigma_2^*$ and $|v_n|_1 \le r(n)$. We will take $\mathbf{a}$ of the form

$$a_n = f(u_n^{(r(n))}), \tag{15}$$

where $f \colon \Sigma_2^* \to \{0,1\}$ remains to be specified.

CLAIM 1. *We have $\#\widetilde{\mathcal{N}}_2(\mathbf{a}) = 1$. In particular, $\mathbf{a}$ is asymptotically 2-automatic.*

Proof. We need to prove that for almost all $n \in \mathbb{N}_0$ we have

$$a_n = a_{2n} = a_{2n+1}. \tag{16}$$

The decomposition of $(2n)_2$ as in (14) takes the form

$$(n)_2 = u_n^{(1)} \dots u_n^{(r(n))} (v_n 0), \tag{17}$$

and thus $u_{2n}^{(r(2n))} = u_n^{(r(n))}$ and $a_n = a_{2n}$ for all n.

To prove the second equality in (16), we begin by noting that, similarly to (17), the decomposition of $(2n+1)_2$ as in (14) takes the form

$$(n)_2 = u_n^{(1)} \dots u_n^{(r(n))} (v_n 1), \tag{18}$$

unless $|v_n|_1 = r(n)$, in which case

$$s_2(n) = |(n)_2|_1 = 1 + \dots + r(n) + r(n) = (r(n)^2 + 3r(n))/2. \tag{19}$$

Let $R = \{(r^2+3r)/2 \mid r \in \mathbb{N}_0\}$. It will suffice to show that $s_2(n) \notin R$ for almost all $n \in \mathbb{N}_0$.

Let L be a large integer and consider a random variable $\mathbf{n}$ uniformly distributed on $[2^L]$. Then $s_2(\mathbf{n})$ has binomial distribution $\mathrm{Bin}(L, 1/2)$. By Hoeffding's inequality, we have

$$\mathbb{P}(|s_2(\mathbf{n}) - L/2| > \sqrt{L \log L}) \le 2/L^2. \tag{20}$$

A standard estimate on binomial coefficients, which is easily derived from Stirling's formula, asserts that

$$\max_{x\in\mathbb{Z}} \mathbb{P}\big(s_2(\mathbf{n}) = x\big) = 2^{-L}\binom{L}{\lfloor L/2\rfloor} \ll \frac{1}{\sqrt{L}}. \tag{21}$$

Thus, bearing in mind $\#R\cap[N] = \sqrt{2N} + O(1)$, we can estimate

$$\begin{aligned}\mathbb{P}\big(s_2(\mathbf{n}) \in R\big) &\le \sum_{|x-L/2|\le\sqrt{L\log L}} 1_R(x)\mathbb{P}\big(s_2(\mathbf{n}) = x\big) \\ &\quad + \mathbb{P}(|s_2(\mathbf{n}) - L/2| > \sqrt{L\log L}) \\ &\ll \sqrt{\frac{\log L}{L}} + \frac{1}{L^2} \to 0 \text{ as } L\to\infty.\end{aligned}$$

It follows that $s_2(n) \notin R$ for almost all $n \in \mathbb{N}_0$, as needed. △

We let f be any map from Σ_2^* to $\{0,1\}$ with the property that for any $r, m \in \mathbb{N}_0$ with $r \le m$ we have

$$\#\{w \in \Sigma_2^m \mid f(w) = 1,\ |w|_1 = r\} = \left\lfloor \theta\binom{m}{r}\right\rfloor. \tag{22}$$

Such a map can easily be constructed by picking, for all $r, m \in \mathbb{N}_0$ with $r \le m$, a partition $\{w \in \Sigma_2^m \mid |w|_1 = r\} = X_0 \cup X_1$ into two sets with cardinalities $\#X_1 = \lfloor\theta\binom{m}{r}\rfloor$ and $\#X_0 = \lceil(1-\theta)\binom{m}{r}\rceil$, and then putting $f(w) = 1$ for $w \in X_1$ and $f(w) = 0$ for $w \in X_0$.

Claim 2. *We have* $\mathrm{freq}_{\mathbf{a}}(1) = \theta$.

Proof. Let N be a large integer, and let $L = \lceil\log_2 N\rceil$. We need to approximate $\#\{n \in [N] \mid a_n = 1\}$. Towards this end, we partition $[N]$ into cells

$$E = E(r, m, u^{(1)}, \dots, u^{(r-1)}, v), \tag{23}$$

where $r, m \in \mathbb{N}_0,\ u^{(1)}, \dots, u^{(r-1)}, v \in \Sigma_2^*$, defined as the set of those $n \in [N]$ for which $r(n) = r$, $|u_n^{(r)}| = m$, $u_n^{(1)} = u^{(1)}$, $u_n^{(2)} = u^{(2)}$, …, $u_n^{(r-1)} = u^{(r-1)}$, and $v_n = v$. We will say that the cell E given by (23) is *good* if all of the following conditions hold, and *bad* otherwise:

(i) $|r - \sqrt{L}| \le \sqrt{2\log L}$;
(ii) $|m - 2r| \le \sqrt{20r\log L}$;
(iii) $[u^{(1)}\dots u^{(r-1)}1^m v]_2 < N$.

Let also U_{g} and U_{b} denote the union of all good and bad cells, respectively. Our plan is to show that f is well-behaved on U_{g}, and U_{b} is negligibly small.

We begin by estimating the cardinality of U_{b}. Let Z denote the set of those $n \in [N]$ whose binary expansion $(n)_2^L = 0\dots0(n)_2$ (padded with 0s to length L) contains a subword w with

$$\big||w|_1 - \tfrac{1}{2}|w|\big| > \sqrt{\tfrac{3}{2}|w| \log L}. \tag{24}$$

It follows from Hoeffding's inequality that if we fix the positions where w begins and ends and we choose $n \in [2^L]$ uniformly at random, the probability that (24) holds is at most $2\exp(-3\log L) = 2L^{-3}$. Since there are $\binom{L+1}{2}$ ways to pick a non-empty subword of a length-L word, the union bound yields

$$\#Z \le 2^L \cdot \binom{L+1}{2} \cdot 2L^{-3} \ll \frac{N}{L}.$$

Consider a bad cell E, given by (23), and assume that N (and hence also L) is sufficiently large. If (i) is false then $E \subseteq Z$, since for any $n \in E$, condition (24) holds for $w = (n)_2^L$, which can easily be derived from the fact that $|w| = L$ and $(r^2+r)/2 \le |w|_1 \le (r^2+3r)/2$. Similarly, if (i) is true but (ii) is false then $E \subseteq Z$, since for any $n \in E$, condition (24) holds for $w = u_n^{(r(n))}$. Finally, if (i) is true but (iii) is false then the word $u^{(1)} \dots u^{(r-1)}$ is a prefix of $(N)_2$, and its length is at least $r(r-1)/2 \ge L/3$. Thus, E is contained in an interval of length $2^{2/3L} \ll N^{2/3} < N/L$. Combining the estimates obtained above, we conclude that

$$\#U_{\mathrm{b}} \ll N/L = o_{N\to\infty}(N). \tag{25}$$

Consider now a good cell E, given by (23). It follows from (iii) that

$$E = \{[u^{(1)} \dots u^{(r-1)} wv]_2 \mid w \in \Sigma_2^m,\ |w|_1 = r\}. \tag{26}$$

Thus, it follows from (22) that

$$\#\{n \in E \mid a_n = 1\} = \left\lfloor \theta \binom{m}{r} \right\rfloor = \lfloor \theta \cdot \#E \rfloor. \tag{27}$$

Assuming that L is sufficiently large, we have

$$\binom{m}{r} \ge \left(\frac{m}{r}\right)^r \ge \left(\frac{3}{2}\right)^{\sqrt{L}/2} = \omega_{N\to\infty}(1). \tag{28}$$

(We recall that $\omega_{N\to\infty}(1)$ denotes an expression that tends to ∞ with N.) Thus, the expression in (27) is $(\theta - o_{N\to\infty}(1))\#E$.

We are now ready to estimate

$$\begin{aligned}\#\{n \in [N] \mid a_n = 1\} \ge \#\{n \in U_{\mathrm{g}} \mid a_n = 1\} &= \sum_{E \text{ good}} (\theta - o_{N\to\infty}(1)) \cdot \#E \\ &= (\theta - o_{N\to\infty}(1))(N - \#U_{\mathrm{b}}) = \theta N - o_{N\to\infty}(N),\end{aligned}$$

where the sum is taken over all the cells E that are good. Dividing by N and letting $N \to \infty$, we conclude that $\underline{\mathrm{freq}}_{\mathbf{a}}(1) \ge \theta$. Using analogous reasoning, we obtain the complementary estimate $\overline{\mathrm{freq}}_{\mathbf{a}}(1) \le \theta$. △

Combining Claims 1 and 2 completes the argument for Proposition 4.3. ■

5. Subword complexity. Subword complexity of a sequence $\mathbf{a}$ over a finite alphabet Ω is the function $p_{\mathbf{a}}$ which assigns to each $\ell \in \mathbb{N}$ the number of length-ℓ subwords of $\mathbf{a}$:

$$p_{\mathbf{a}}(\ell) := \#\{w \in \Omega^\ell \mid (\exists n \in \mathbb{N}_0)\ w = a_{[n,n+\ell)}\}. \tag{29}$$

A notable feature of automatic sequences is that their subword complexity is linear: $p_{\mathbf{a}}(\ell) = O(\ell)$. This property does not directly extend to asymptotically automatic sequences. In fact, given a finite alphabet Ω it is not hard to construct a sequence over Ω that is asymptotically constant (and thus asymptotically k-automatic for each $k \geq 2$) and which contains each finite word as a subword (and thus has maximal possible subword complexity, $p_{\mathbf{a}}(\ell) = \#\Omega^\ell$). In particular, we see that two asymptotically equal sequences can have diametrically different subword complexities.

The discussion above suggests that subword complexity is not the right notion to consider in the asymptotic regime. Instead, we introduce a new notion of *asymptotic subword complexity*, defined as the function $\widetilde{p}_{\mathbf{a}}$ which assigns to each $\ell \in \mathbb{N}$ the number of length-ℓ subwords of $\mathbf{a}$ which appear with positive asymptotic frequency:

$$\widetilde{p}_{\mathbf{a}}(\ell) := \#\{w \in \Omega^\ell \mid \overline{\mathrm{freq}}_{\mathbf{a}}(w) > 0\}. \tag{30}$$

It follows directly from the definition that for any sequence $\mathbf{a}$ and any $\ell \in \mathbb{N}$ we have $\widetilde{p}_{\mathbf{a}}(\ell) \leq p_{\mathbf{a}}(\ell)$. Moreover, $p_{\mathbf{a}}(\ell)$ only depends on the equivalence class of $\mathbf{a}$ up to asymptotic equality, that is, if $\mathbf{a} \simeq \mathbf{b}$ then $\widetilde{p}_{\mathbf{a}}(\ell) = \widetilde{p}_{\mathbf{b}}(\ell)$ for all $\ell \in \mathbb{N}$.

PROPOSITION 5.1. *Let $\mathbf{a}$ be an asymptotically automatic sequence. Then $\widetilde{p}_{\mathbf{a}}(\ell) = O(\ell)$.*

Proof. Assume that $\mathbf{a}$ is asymptotically k-automatic and takes values in a finite alphabet Ω. Pick d, $\mathbf{a}^{(0)}, \mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(d-1)}$ and $\phi\colon \Sigma_k^* \to \Sigma_d$ as in Proposition 2.7.

Let $\ell \in \mathbb{N}$ and pick $i \in \mathbb{N}$ with $k^{i-1} \leq \ell < k^i$. Consider the sequence $\mathbf{a}'$ given by

$$a'_n = a^{(\phi((n)_k^i))}_{\lfloor n/k^i \rfloor}.$$

Since $\mathbf{a} \simeq \mathbf{a}'$, we have $\widetilde{p}_{\mathbf{a}'}(\ell) = \widetilde{p}_{\mathbf{a}}(\ell)$. Thus, it suffices to show that there are $O(\ell)$ length-ℓ words which appear in $\mathbf{a}'$ with positive upper frequency. Pick one such word $w \in \Omega^\ell$. Applying the pigeonhole principle, we can find $0 \leq r < k^i$ such that

$$\overline{d}(\{n \in \mathbb{N}_0 \mid a'_{[k^i n+r, k^i n+r+\ell)} = w\}) > 0.$$

Recall that for $h \in [0, 2k^i)$ we have

$$a'_{k^i n+h} = a^{(\phi((h)_k^i))}_{n+\lfloor h/k^i \rfloor},$$

where $\lfloor h/k^i \rfloor \in \{0,1\}$. Thus, the word $a'_{[k^i n+r, k^i n+r+\ell)}$ is completely determined by $\big(a_n^{(j)}\big)_{j=0}^{d-1}$, $\big(a_{n+1}^{(j)}\big)_{j=0}^{d-1}$ and r. It follows that

$$\widetilde{p}_{\mathbf{a}'}(\ell) \le \#\Omega^d \cdot \#\Omega^d \cdot k^i \le C\ell,$$

where $C := k \cdot \#\Omega^{2d}$. In particular, $\widetilde{p}_{\mathbf{a}'}(\ell) = O(\ell)$, as needed. ∎

Although the notion of asymptotic subword complexity is quite natural, it does not seem to have been systematically studied in the literature. To justify interest in this notion, we prove several basic properties, which are analogues of standard facts about subword complexity. Given a sequence $\mathbf{a} \in \Omega^\infty$, it will be convenient to consider the set given by

$$\widetilde{X}_{\mathbf{a}} := \{\mathbf{x} \in \Omega^\infty \mid \overline{\mathrm{freq}}_{\mathbf{a}}(w) > 0 \text{ for each subword } w \text{ of } \mathbf{x}\}. \tag{31}$$

It is routine to verify that for a word $w \in \Omega^*$ we have $\overline{\mathrm{freq}}_{\mathbf{a}}(w) > 0$ if and only if w is a subword of some $\mathbf{x} \in \widetilde{X}_{\mathbf{a}}$. In particular, $\widetilde{p}_{\mathbf{a}}(\ell) \ge p_{\mathbf{x}}(\ell)$ for all $\mathbf{x} \in \widetilde{X}_{\mathbf{a}}$ and $\ell \in \mathbb{N}$.

Lemma 5.2. *Let* $\mathbf{a}$ *be a sequence over a finite alphabet* Ω *such that* $\overline{\mathrm{freq}}_{\mathbf{a}}(\omega) > 0$ *for all* $\omega \in \Omega$. *Then exactly one of the following holds:*

(†) *there exist* $w \in \Omega^*$ *and* $\mathbf{x} \in \widetilde{X}_{\mathbf{a}}$ *such that* $\mathbf{a} = w\mathbf{x}$;
(‡) *for all* $i \in \mathbb{N}_0$ *and* $\mathbf{a} = w_0 w_1 w_2 \ldots$ *there exists* $w_i \in \Omega^*$ *such that* $\overline{\mathrm{freq}}_{\mathbf{a}}(w_i) > 0 = \overline{\mathrm{freq}}_{\mathbf{a}}(w_i (w_{i+1})_1)$.

Proof. First note that if $\mathbf{a}$ admits a representation as in (†) then $\overline{\mathrm{freq}}_{\mathbf{a}}(u) > 0$ for each subword u of $\mathbf{x}$ and consequently $\mathbf{a}$ cannot be represented as in (‡).

Assume next that $\mathbf{a}$ cannot be represented as in (†). We greedily construct a representation as in (‡). Let α_0 denote the initial symbol of $\mathbf{a}$. Note that $\mathbf{a}$ has a prefix w with $\overline{\mathrm{freq}}_{\mathbf{a}}(w) = 0$ since otherwise we would have $\mathbf{a} \in \widetilde{X}_{\mathbf{a}}$. Since $\overline{\mathrm{freq}}_{\mathbf{a}}(\omega) > 0$ for $\omega \in \Omega$, we have $|w| \ge 2$. Pick the shortest possible w and write it in the form $w = w_0 \alpha_1$, where $w_0 \in \Omega^*$ and $\alpha_1 \in \Omega$. By definition, $\overline{\mathrm{freq}}_{\mathbf{a}}(w_0) > 0$. We may write $\mathbf{a}$ in the form $\mathbf{a} = w_0 \mathbf{a}'$, where $\mathbf{a}' \in \Omega^\infty$ begins with α_1. Then $\mathbf{a}'$ has a minimal prefix, which we will denote by $w_1 \alpha_2$, with $\overline{\mathrm{freq}}_{\mathbf{a}}(w_1 \alpha_2) = 0$, since otherwise we would have $\mathbf{a} \in w_0 \widetilde{X}_{\mathbf{a}}$. Thus, we may write $\mathbf{a}' = w_1 \mathbf{a}''$. Iterating this reasoning, we find $w_0, w_1, w_2, \ldots \in \Omega^*$ and $\alpha_0, \alpha_1, \alpha_2, \ldots \in \Omega$ such that $\mathbf{a} = w_0 w_1 w_2 \ldots$ and for each $i \in \mathbb{N}_0$, α_i is the initial symbol of w_i, $\overline{\mathrm{freq}}_{\mathbf{a}}(w_i) > 0$ and $\overline{\mathrm{freq}}_{\mathbf{a}}(w_i \alpha_{i+1}) = 0$. ∎

We point out that the assumptions of Lemma 5.2 are easy to satisfy. Indeed, if it is not the case that $\overline{\mathrm{freq}}_{\mathbf{a}}(\omega) > 0$ for all $\omega \in \Omega$ then we can find a smaller alphabet $\Omega' \subseteq \Omega$ and a sequence $\mathbf{a}' \in (\Omega')^\infty$ with $\mathbf{a}' \simeq \mathbf{a}$ and $\overline{\mathrm{freq}}_{\mathbf{a}'}(\omega) > 0$ for all $\omega \in \Omega'$.

LEMMA 5.3. *Let* $\mathbf{a}$ *be a sequence over a finite alphabet* Ω *such that* $\overline{\mathrm{freq}}_{\mathbf{a}}(\omega) > 0$ *for all* $\omega \in \Omega$, *and assume that* $\mathbf{a}$ *admits a representation as in* (‡). *Then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |w_n| = \infty. \tag{32}$$

Proof. For a contradiction, suppose that

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |w_n| < L \tag{33}$$

for some integer L. Let N be a large integer such that

$$\sum_{n=0}^{N-1} |w_n| < NL, \tag{34}$$

and let

$$\mathcal{M}(N) = \{n \in [N] \mid |w_n| \le 2L\}. \tag{35}$$

It follows from Markov's inequality that $\#\mathcal{M}(N) \ge N/2$. For $w \in \Omega^*$ with $|w| \le 2L$ and $\alpha \in \Omega$ such that $\overline{\mathrm{freq}}_{\mathbf{a}}(w\alpha) = 0$, consider

$$\mathcal{M}(N; w, \alpha) = \{n \in \mathcal{M}(N) \mid w_n = w \text{ and } w_{n+1} \text{ begins with } \alpha\}.$$

Note that $[N] = \bigcup_{w,\alpha} \mathcal{M}(N; w, \alpha)$, where the union is taken over all pairs w, α which satisfy the conditions mentioned above. If $\#\Omega = 1$, there is nothing to prove, so we may freely assume that $\#\Omega \ge 2$. The number of possible pairs w, α is at most $\#\Omega \cdot (\#\Omega^{2L} + \#\Omega^{2L-1} + \cdots + 1) < \#\Omega^{2L+2}$, so we can find a pair w, α with $\#\mathcal{M}(N; w, \alpha) \ge N/2\#\Omega^{2L+2}$.

Since there are finitely many possible pairs w, α, we conclude that for one of them we have $\#\mathcal{M}(N; w, \alpha) \ge N/2\#\Omega^{2L+2}$ for infinitely many values of N. In particular,

$$\overline{\mathrm{freq}}_{\mathbf{a}}(w\alpha) \ge \limsup_{N\to\infty} \frac{\#\mathcal{M}(N; w, \alpha)}{\sum_{n=0}^{N-1} |w_n|} \ge \frac{1}{2L\#\Omega^{2L+2}} > 0, \tag{36}$$

contradicting an earlier assumption. ∎

We are now ready to classify sequences with particularly low asymptotic subword complexity. Recall that each sequence $\mathbf{a}$ such that $p_{\mathbf{a}}(\ell) \le \ell$ for at least one $\ell \in \mathbb{N}$ is ultimately periodic, in which case $p_{\mathbf{a}}(\ell)$ is bounded as $\ell \to \infty$ (see [CH73]). A similar result is true in the asymptotic regime. Recall that we say that a sequence $\mathbf{a}$ is *asymptotically invariant under a shift* if there exists $m \in \mathbb{N}$ such that $(a_n)_{n=0}^{\infty} \simeq (a_{n+m})_{n=0}^{\infty}$.

PROPOSITION 5.4. *Let* $\mathbf{a}$ *be a sequence over a finite alphabet* Ω. *If there exists* $\ell \in \mathbb{N}$ *such that* $\widetilde{p}_{\mathbf{a}}(\ell) \le \ell$ *then* $\mathbf{a}$ *is asymptotically invariant under a* *shift. Conversely, if* $\mathbf{a}$ *is asymptotically invariant under a shift then* $\widetilde{p}_{\mathbf{a}}(\ell)$ *is bounded as* $\ell \to \infty$.

Proof. We may assume without loss of generality that $\overline{\mathrm{freq}}_{\mathbf{a}}(\omega) > 0$ for all $\omega \in \Omega$. For each $\mathbf{x} \in \widetilde{X}_{\mathbf{a}}$ we have $p_{\mathbf{x}}(\ell) \le \widetilde{p}_{\mathbf{a}}(\ell) \le \ell$, and hence $\mathbf{x} = vu^\infty$ is eventually periodic with the pre-period $|v|$ and the period $|u|$ bounded as a function of ℓ and $\#\Omega$, say $|v|, |u| \le L$ (the bound on the period can be inferred e.g. from the argument given in [PF02, Prop. 1.1.1]). It now follows from Lemma 5.3 that $\mathbf{a}$ is asymptotically invariant under the shift by $L!$. Conversely, if $\mathbf{a}$ is asymptotically invariant under the shift by some $m \in \mathbb{N}$ then $\widetilde{p}_{\mathbf{a}}(\ell) \le \#\Omega^m$ for all $\ell \in \mathbb{N}$. ∎

It follows from the remark above the proposition that for each sequence $\mathbf{a}$ that is not eventually periodic we have $p_{\mathbf{a}}(\ell) \ge \ell + 1$ for all $\ell \in \mathbb{N}$. Sequences for which $p_{\mathbf{a}}(\ell) = \ell + 1$ for all $\ell \in \mathbb{N}$ are known as *Sturmian sequences* and have been extensively studied; see e.g. [PF02, Chap. 5] for an extensive discussion. There are several alternative characterisations of Sturmian sequences. Specifically, a sequence $\mathbf{a}$ is Sturmian if and only if it takes one of the following forms:

(37)
$$a_n = \lfloor \theta(n+1) + \rho \rfloor - \lfloor \theta n + \rho \rfloor \quad \text{or} \quad a_n = \lceil \theta(n+1) + \rho \rceil - \lceil \theta n + \rho \rceil,$$

where $\theta \in (0,1) \setminus \mathbb{Q}$ and $\rho \in [0,1)$. We have the following asymptotic analogue. (Recall that by "for almost all" we mean "for all outside of a set with density zero".)

PROPOSITION 5.5. *Let* $\mathbf{a}$ *be a sequence over* $\{0,1\}$ *and assume that* $\widetilde{p}_{\mathbf{a}}(\ell) = \ell+1$ *for all* $\ell \in \mathbb{N}$. *Then there exist* $\theta \in [0,1)\setminus\mathbb{Q}$ *and* $\boldsymbol{\rho} \in [0,1)^\infty$ *such that* $\rho_{n+1} = \rho_n$ *for almost all* $n \in \mathbb{N}_0$ *and* $a_n = \lfloor (n+1)\theta + \rho_n \rfloor - \lfloor n\theta + \rho_n \rfloor$ *for all* $n \in \mathbb{N}_0$.

Proof. We may assume without loss of generality that $\overline{\mathrm{freq}}_{\mathbf{a}}(\omega) > 0$ for $\omega \in \{0,1\}$. If $\mathbf{a}$ admits a decomposition $\mathbf{a} = w\mathbf{x}$ as in (†) then $\widetilde{p}_{\mathbf{a}}(\ell) = p_{\mathbf{x}}(\ell) = \ell+1$ for all $\ell \in \mathbb{N}$, and thus $\mathbf{x}$ is a Sturmian word. As a consequence, we can find the required representation of $\mathbf{a}$ with $\boldsymbol{\rho}$ eventually constant. Thus, we may assume that $\mathbf{a}$ does not admit a decomposition (†) and hence it has a decomposition (‡) for some words $w_0, w_1, w_2, \ldots$.

For all $\mathbf{x} \in \widetilde{X}_{\mathbf{a}}$ we have $p_{\mathbf{x}}(\ell) \le \widetilde{p}_{\mathbf{a}}(\ell) = \ell + 1$ for all $\ell \in \mathbb{N}$. Thus, all $\mathbf{x} \in \widetilde{X}_{\mathbf{a}}$ are either Sturmian or eventually periodic. Suppose first that $\widetilde{X}_{\mathbf{a}}$ contains at least one Sturmian word $\mathbf{x}$. Let $\theta \in [0,1) \setminus \mathbb{Q}$ denote the slope of $\mathbf{x}$; thus, $\mathbf{x}$ is given by $x_n = \lfloor (n+1)\theta + \rho \rfloor - \lfloor n\theta + \rho \rfloor$ ($n \in \mathbb{N}_0$) for some $\rho \in [0,1)$ (possibly with the exception of one value of n). For each $\ell \in \mathbb{N}$ we have $\widetilde{p}_{\mathbf{a}}(\ell) = p_{\mathbf{x}}(\ell) = \ell + 1$. Hence, each word w with $\overline{\mathrm{freq}}_{\mathbf{a}}(w) > 0$ is a subword of $\mathbf{x}$. In particular, for each $i \in \mathbb{N}_0$, there is some $\sigma_i \in [0,1)$ such that $w_{i,m} = \lfloor (m+1)\theta + \sigma_i \rfloor - \lfloor m\theta + \sigma_i \rfloor$ for $0 \leq m < |w_i|$. It follows that we can pick $\boldsymbol{\rho}$ of the form

$$\boldsymbol{\rho} = \underbrace{\rho_0'\rho_0' \dots \rho_0'}_{|w_0| \text{ times}} \underbrace{\rho_1'\rho_1' \dots \rho_1'}_{|w_1| \text{ times}} \underbrace{\rho_2'\rho_2' \dots \rho_2'}_{|w_2| \text{ times}} \dots$$

for some $\rho_0', \rho_1', \rho_2', \dots \in [0,1)$. As a consequence of Lemma 5.3 we find that $\rho_{n+1} = \rho_n$ for almost all $n \in \mathbb{N}_0$.

Now, suppose that all $\mathbf{x} \in \widetilde{X}_{\mathbf{a}}$ are eventually periodic, and hence can be written in the form $\mathbf{x} = vu^\infty$ for some $v \in \Omega^*$ and $u \in \Omega^*$ with $u \neq \epsilon$ which cannot be written as a power of a shorter word. Our next goal is to show that the lengths $|u|$ of the periods are uniformly bounded.

For a contradiction, assume that for each ℓ_0 there is a word $\mathbf{x} = vu^\infty \in \widetilde{X}_{\mathbf{a}}$ as above with $\ell := |u| \geq \ell_0$. Note that $p_{\mathbf{x}}(\ell - 1) = \ell$. Indeed, $p_{\mathbf{x}}(\ell-1)$ cannot be larger since $p_{\mathbf{x}}(\ell-1) \leq \widetilde{p}_{\mathbf{a}}(\ell-1) = \ell$, and it cannot be smaller since this would imply that u^∞ has a period strictly less than ℓ. Thus, each word $w \in \Omega^{\ell-1}$ with $\overline{\mathrm{freq}}_{\mathbf{a}}(w) > 0$ is a subword of u^∞. Let us next consider a pair $v_1u_1^\infty, v_2u_2^\infty \in \widetilde{X}_{\mathbf{a}}$ of sequences with $|u_1| =: \ell_1 < \ell_2 := |u_2|$. Since $\overline{\mathrm{freq}}_{\mathbf{a}}(u_1^{\ell_2/\ell_1}) > 0$, up to a periodic shift and up to changing at most one symbol, the words $u_1^{\ell_2/\ell_1}$ and u_2 are equal. Consider now a third sequence $v_3u_3^\infty \in \widetilde{X}_{\mathbf{a}}$ with $\ell_3 := |u_3| > 100\ell_1\ell_2$. Applying the earlier argument to the pairs $v_1u_1^\infty, v_3u_3^\infty$ and $v_2u_2^\infty, v_3u_3^\infty$ we conclude that there exist periodic shifts $\widetilde{u}_1$ and $\widetilde{u}_2$ of u_1 and u_2 such that $\widetilde{u}_1^{\ell_3/\ell_1}$ and $\widetilde{u}_2^{\ell_3/\ell_2}$ differ from u_3 in at most one position. Bearing in mind that $\ell_1\ell_2$ is a period of both $\widetilde{u}_1^\infty$ and $\widetilde{u}_2^\infty$, we conclude that $\widetilde{u}_1^\infty = \widetilde{u}_2^\infty$, contradicting the assumption that $\ell_1 < \ell_2$.

Let L be the maximal period of $\mathbf{x} \in \widetilde{X}_{\mathbf{a}}$ and put $M = L!$. Note that it follows from Proposition 5.4 that $\mathbf{a}$ is not asymptotically invariant under the shift by M. Hence, there exists $\varepsilon > 0$ such that we can find arbitrarily long words $w \in \Omega^*$ with $\overline{\mathrm{freq}}_{\mathbf{a}}(w) > 0$ such that there are at least $\varepsilon|w|$ positions $0 \leq n < |w| - M$ such that $w_n \neq w_{n+M}$ (this follows readily from Lemmas 5.2 and 5.3). For each word $w \in \Omega^*$ with $\overline{\mathrm{freq}}_{\mathbf{a}}(w) > 0$ and $|w| \geq M$, let $\overline{w}$ denote the word over $\{0,1\}$ with length $|\overline{w}| = |w| - M$ given by $\overline{w}_n = 0$ if $w_n = w_{n+M}$ and $w_n = 1$ otherwise, and let $\mathcal{L}$ be the family of all words $\overline{w}$ that arise this way. With the same $\varepsilon > 0$ as above, we see that there exist arbitrarily long words $\overline{w} \in \mathcal{L}$ such that $|\overline{w}|_1 \geq \varepsilon|w|$. Thus, it follows from Lemma 5.6 below that there exists $\overline{\mathbf{x}} \in \{0,1\}^\infty$ such that each subword of $\overline{\mathbf{x}}$ belongs to $\mathcal{L}$ and $\overline{\mathrm{freq}}_{\overline{\mathbf{x}}}(1) > 0$. A compactness argument shows that there exists $\mathbf{x} \in \widetilde{X}_{\mathbf{a}}$ such that for each $n \in \mathbb{N}_0$ we have $\overline{x}_n = 0$ if $\overline{x}_n = x_{n+M}$ and $x_n = 1$ otherwise. This contradicts the earlier observation that each $\mathbf{x} \in \widetilde{X}_{\mathbf{a}}$ is eventually periodic with period at most L, and finishes the argument. ∎

Lemma 5.6. *Let $\mathcal{L} \subseteq \{0,1\}^*$ be an infinite set of words closed under taking subwords. Suppose that there exists $\varepsilon > 0$ such that for infinitely many words $w \in \mathcal{L}$ we have $|w|_1 \geq \varepsilon|w|$. Then there exists $\mathbf{x} \in \{0,1\}^\infty$ such that $\overline{\mathrm{freq}}_{\mathbf{x}}(1) > 0$ and each subword of $\mathbf{x}$ belongs to $\mathcal{L}$.*

Proof. Let α denote the infimum of all real numbers $\varepsilon > 0$ such that for each sufficiently long $w \in \mathcal{L}$ we have $|w|_1 < \varepsilon|w|$. By assumption, $\alpha > 0$. Pick a decreasing sequence $\delta_\ell > 0$ such that $\delta_\ell \to 0$ as $\ell \to \infty$ and $|w|_1 < (\alpha + \delta_\ell)|w|$ for all $w \in \Sigma_k^\ell$. Let $(m_i)_{i=1}^\infty$ be a rapidly increasing sequence of positive integers, to be determined in the course of the argument, and let $(\ell_i)_{i=0}^\infty$ be the sequence given by $\ell_0 = 1$ and $\ell_{i+1} = m_{i+1}\ell_i$.

Consider, for some $i \in \mathbb{N}$, a word $w \in \mathcal{L}$ of length ℓ_i satisfying $|w|_1 \geq (\alpha - \delta'_{\ell_i})\ell_i$, where $\delta'_{\ell_i} := (4/\alpha - 1)\delta_{\ell_i}$. We may decompose $w = w_0 w_1 \ldots w_{m_i - 1}$, where $|w_r| = \ell_{i-1}$ for all $0 \leq r < m_i$. Expressing $|w|_1$ as $\sum_{r=0}^{m_i-1} |w_r|_1$, we see that the proportion of words w_r $(0 \leq r < m_i)$ such that $|w_r|_1 < (\alpha - \delta'_{\ell_{i-1}})\ell_{i-1}$ is at most

$$\frac{\delta_{\ell_{i-1}} + \delta'_{\ell_i}}{\delta'_{\ell_{i-1}} + \delta_{\ell_{i-1}}} = \frac{\alpha}{4} + \delta_{\ell_i} \cdot \frac{4/\alpha - 1}{\delta'_{\ell_{i-1}} + \delta_{\ell_{i-1}}}. \tag{38}$$

Picking m_i sufficiently large as a function of ℓ_{i-1}, we can ensure that the expression in (38) is less than $\alpha/2$. As a consequence we can decompose w as $w = uw'v$, where $|w'| = \ell_{i-1}$, $|w'|_1 \geq (\alpha - \delta'_{\ell_{i-1}})\ell_{i-1}$ and $|v|_1 \geq (\alpha/2)\ell_i$ (we take $w' = w_r$ for the minimal admissible value of r). Iterating this construction, we can find words $w^{(j)}, u^{(j)}, v^{(j)} \in \mathcal{L}$ for $0 \leq j \leq i$ such that $w^{(i)} = w$ and $w^{(j+1)} = u^{(j)} w^{(j)} v^{(j)}$, $|w^{(j)}| = \ell_j$, $|w^{(j)}|_1 \geq (\alpha - \delta'_{\ell_j})\ell_j$ and $\left|v^{(j)}\right|_1 \geq (\alpha/2)\ell_{j+1}$ for $0 \leq j < i$. Applying a compactness argument, we can find infinite sequences $w^{(j)}, u^{(j)}, v^{(j)} \in \mathcal{L}$ which satisfy the properties mentioned above for all $j \in \mathbb{N}_0$.

Consider the sequence $\mathbf{x} = v^{(0)} v^{(1)} v^{(2)} \ldots$. It is clear from the construction that all subwords of $\mathbf{x}$ belong to $\mathcal{L}$. Since $|v^{(j)}|_1 \geq (\alpha/2)\ell_{j+1}$ and $\sum_{h=0}^{j} |v^{(h)}| < \ell_{j+1}$, we have $\overline{\mathrm{freq}}_{\mathbf{x}}(1) \geq \alpha/2 > 0$, as needed. ∎

6. Classification problems. Given a class of sequences $\mathcal{C}$, a natural question in the theory of automatic sequences is: Which members of $\mathcal{C}$ are automatic? Here, we discuss some problems of this type in the asymptotic regime.

6.1. Bracket sequences. A generalised polynomial is an expression constructed using ordinary polynomials, the floor function, addition and multiplication. In contrast with ordinary polynomials, their generalised counterparts can be bounded or even finitely-valued but non-constant. Bounded generalised polynomials have been extensively investigated by many authors, including Håland Knutson, Bergelson, Leibman and others; see e.g. [Hål93, Hål94, HK95, BL07, Lei12, BHKS21]. In [AK23] we undertook a systematic study of letter-to-letter codings of finitely-valued generalised polynomial sequences, which we dub *bracket sequences.* In a series of papers [BK18, BK20, Kon22], we obtained a complete classification of automatic bracket sequences: they are precisely the eventually periodic sequences. The vast bulk of the difficulty in the aforementioned problem stems from the need to consider asymptotically constant sequences. This issue does not arise when we consider asymptotically automatic sequences, leading to a much shorter argument.

PROPOSITION 6.1. *Let* $\mathbf{a}$ *be a bracket word over a finite alphabet. Then the following conditions are equivalent:*

(i) $\mathbf{a}$ *is asymptotically* k*-automatic for at least one* $k \geq 2$*;*
(ii) $\mathbf{a}$ *is asymptotically* k*-automatic for all* $k \geq 2$*;*
(iii) $\mathbf{a}$ *is asymptotically equal to a periodic sequence.*

Proof. It is clear that we have the chain of implications (iii)⇒(ii)⇒(i), so we only need to prove (i)⇒(iii). Since the argument closely follows the proof of [BK20, Theorem B] we will only provide a sketch with emphasis on the points where the two arguments differ, and refer to [BK20] for details.

In [BK20] we use the notion of a weakly periodic function $f\colon \mathbb{N}_0 \to \Omega$ (Ω being any set). A function f is *weakly periodic* if for any restriction f' of f to an arithmetic sequence (given by $f'(n) = f(an + b)$ with $a \in \mathbb{N}$ and $b \in \mathbb{N}_0$) there exist $q \in \mathbb{N}$ and distinct $r, r' \in \mathbb{N}_0$ such that

$$f'(qn + r) = f'(qn + r') \quad \text{for all } n \in \mathbb{N}_0. \tag{39}$$

Here, we will use an even weaker notion of an *asymptotically weakly periodic* function, which is defined in the same way with the exception that in (39) we only require the equality to hold for almost all $n \in \mathbb{N}_0$.

Lemma 2.1 in [BK20] asserts that each automatic sequence is weakly periodic. With essentially the same argument we conclude that each asymptotically automatic sequence is asymptotically weakly periodic. (The only difference is that in the asymptotic variant we identify sequences that are equal almost everywhere.)

We will show, in analogy with [BK20, Theorem 2.6], that if $f\colon \mathbb{N}_0 \to \mathbb{R}$ is a finitely-valued generalised polynomial sequence that is asymptotically weakly periodic then f is asymptotically equal to a periodic sequence. Combined with the aforementioned analogue of [BK20, Lemma 2.1], this will complete the proof of the implication (i)⇒(iii).

We rely on the structure theorem due to Bergelson and Leibman [BL07] (cf. [BK20, Theorem 1.13]) which implies that there exists a minimal nilsystem (X,T), a point $z \in X$, a semialgebraic partition $X = \bigcup_{j=1}^{r} S_j$ and constants $c_j \in \mathbb{R}$ such that $g(n) = c_j$ if and only if $T^n(z) \in S_j$ $(1 \le j \le r)$. (A nilsystem is a special type of a topological dynamical system, i.e., a compact space X equipped with a homeomorphism $T\colon X \to X$. Minimality means that each orbit $\{T^n(x) \mid n \in \mathbb{N}\}$ is dense in X. There is a natural choice of a T-invariant measure μ_X on X, namely the corresponding Haar measure. For our purposes, all that we need to know about the sets S_j is that $\mu_X(\partial S_j) = 0$, where $\partial S = \mathrm{cl}\, S \setminus \mathrm{int}\, S$ denotes the boundary of a set S.) Passing to an arithmetic progression (cf. [BK20, Remark 1.14]) we may assume that (X,T) is totally minimal, meaning that (X,T^a) is minimal for each $a \in \mathbb{N}$. (We note that this reduction is possible because we are working with nilsystems rather than more general topological dynamical systems.)

Replacing c_j with 0s and 1s, we may freely reduce the problem to the simpler situation where $f\colon \mathbb{N}_0 \to \{0,1\}$ and $f(n) = 1$ if and only if $T^n(z) \in S$ for some semialgebraic set S, which is reminiscent of [BK20, Lemma 2.4]. Our goal is to show that f is asymptotically equal to a constant sequence. Replacing S with its interior, we may freely assume that S is open. (Corollary 1.12 in [BK20] guarantees that this operation only changes f on a density zero set of positions, and hence does not affect asymptotic weak periodicity nor the property of being asymptotically constant.) If S is empty then there is nothing to prove, so assume that this is not the case.

Since f is asymptotically weakly periodic, we can find $q, r, r' \in \mathbb{N}_0$ with $q \neq 0$ and $r < r'$ such that $f(qn + r) = f(qn + r')$ for almost all n. Let $d = r' - r$. We claim that $T^d(S) \subseteq S$. Pick $y \in S$ and an open neighbourhood V of $T^d(y)$; it will suffice to show that $V \cap S \neq \emptyset$. Let $U \subseteq S$ be an open neighbourhood of y with $T^d(U) \subseteq V$, and let $\mathcal{N} = \{n \in \mathbb{N}_0 \mid T^{qn+r}(z) \in U\}$. Since (X,T) is minimal, the set $\mathcal{N}$ is syndetic. For almost all $n \in \mathcal{N}$ (thus, in particular, for at least one n) we have $f(qn + r') = f(qn + r) = 1$ and consequently $T^{qn+r'}(z) = T^d(T^{qn+r}(z)) \in U$. Thus, $V \cap S \supseteq T^d(U) \cap S \neq \emptyset$, as needed.

Since $T^d(S) \subseteq S$, the orbit of any point from S under T^d is contained in S. Since (X,T^d) is minimal, it follows that S is dense in X. Recalling that $\mu_X(\partial S) = 0$, we conclude that $\mu_X(S) = 1$ and hence $f(n) = 1$ for almost all $n \in \mathbb{N}_0$ (cf. [BK20, Corollary 1.12]). ∎

As an application, we can consider bracket sequences of a particularly simple form and classify the ones that are asymptotically equal to an automatic sequence; cf. [AS03a, Thm. 6.2].

Corollary 6.2. *Let $\alpha, \beta \in \mathbb{R}$, $m \in \mathbb{N}_{\ge 2}$, and let $\mathbf{a}$ be the sequence given by $a_n = \lfloor \alpha n + \beta \rfloor \bmod m$. If $\mathbf{a}$ is asymptotically equal to an automatic sequence then $\alpha \in \mathbb{Q}$.*

6.2. Multiplicative sequences. A sequence $\mathbf{a} = (a_n)_{n=1}^{\infty}$ of complex numbers is *multiplicative* if $a_{nm} = a_n a_m$ for all $n, m \in \mathbb{N}$ with $\gcd(n, m) = 1$, and *completely multiplicative* if the requirement that $\gcd(n, m) = 1$ can be dropped. The problem of classifying multiplicative automatic sequences ([1]) has been considered in many papers, including [Yaz01, SP03, BC09, BCC10, Coo10, BBC12, BCH14, AG18, KK19, KK20, Kon20, Li20], culminating in a complete resolution in [KLM22]:

Theorem 6.3. *Let* $\mathbf{a}$ *be a finitely-valued multiplicative sequence. Then* $\mathbf{a}$ *is automatic if and only if it is* p*-automatic for a prime* p *and takes the form*

$$a_{p^i n} = c_i b_n \quad \text{for } i \in \mathbb{N}_0,\ n \in \mathbb{N},\ p \nmid n,$$

where $\mathbf{b}$ *and* $\mathbf{c}$ *are eventually periodic sequences.*

In this section we will show that each asymptotically automatic completely multiplicative sequence behaves in a periodic manner on all prime powers, except possibly for some more complicated behaviour on small primes. It it very likely that one can obtain a complete classification, in the spirit of [KLM22], but for the sake of exposition we prove a somewhat weaker result:

Proposition 6.4. *Let* $\mathbf{a}$ *be a sequence that is both multiplicative and asymptotically* k*-automatic. Then* $\mathbf{a} \simeq 0$ *or there exists a Dirichlet character* χ *such that* $a_{p^i} = \chi(p^i)$ *for each sufficiently large prime* p *and each* $i \in \mathbb{N}$.

Before we approach the proof, we discuss some corollaries and preliminaries. First, since any modification of an automatic sequence on a set with density zero is asymptotically automatic, we have the following consequence, expressed purely in terms of automatic and multiplicative sequences.

Corollary 6.5. *Let* $\mathbf{a}$ *be a multiplicative sequence that is asymptotically equal to an automatic sequence. Then* $\mathbf{a} \simeq 0$ *or there exists a Dirichlet character* χ *such that* $a_{p^i} = \chi(p^i)$ *for each sufficiently large prime* p *and each* $i \in \mathbb{N}$.

In particular, we see that some of the classical multiplicative sequences are not asymptotically equal to an automatic sequence.

Corollary 6.6. *There is no automatic sequence* $\mathbf{a}$ *with* $\mathbf{a} \simeq \boldsymbol{\mu}$*, where* $\boldsymbol{\mu}$ *denotes the Möbius sequence given by*

$$\mu_n = \begin{cases} (-1)^k & \text{if } n = p_1 \dots p_k \text{ is the product of } k \text{ distinct primes,} \\ 0 & \text{if } n \text{ is divisible by a square.} \end{cases}$$

([1]) A slight notational inconvenience stems from the fact that automatic sequences are most naturally defined on $\mathbb{N}_0$ and multiplicative sequences—on $\mathbb{N}$. In order to avoid this issue we adopt the convention that each multiplicative sequence $\mathbf{a} = (a_n)_{n=1}^{\infty}$ indexed by $\mathbb{N}$ is extended to $\mathbb{N}_0$ by setting $a_0 = 0$.

Likewise, there is no automatic sequence $\mathbf{a}$ *with* $\mathbf{a} \simeq \boldsymbol{\lambda}$*, where* $\boldsymbol{\lambda}$ *denotes the Liouville sequence given by*

$$\lambda_n = (-1)^k \quad \text{if } n = p_1 \dots p_k \text{ is the product of } k \text{ primes.}$$

The fact that there exist asymptotically shift-invariant sequences that are not asymptotically equal to a periodic sequence leads to some technical difficulties. In order to overcome them, we will need the following result, which is essentially contained in [Klu17]. To avoid breaking the flow of discussion, we delegate the proof to the Appendix. We call a sequence 1-*bounded* if it takes values in $\{z \in \mathbb{C} \mid |z| \leq 1\}$.

THEOREM 6.7. *Let* $\mathbf{a}$ *be a* 1*-bounded, finitely-valued multiplicative sequence that is asymptotically invariant under a shift. Then* $\mathbf{a} \simeq \mathbf{0}$ *or* $\mathbf{a}$ *is periodic.*

We will also need the following basic lemma.

LEMMA 6.8. *Let* $\mathbf{a}$ *be a multiplicative sequence. Suppose that there exist* $Q, r \in \mathbb{N}$ *with* $\gcd(r, Q) = 0$ *such that* $a_n = 0$ *for almost all* $n \equiv r \bmod Q$*. Then* $\mathbf{a} \simeq 0$.

Proof. Let $\mathcal{Z}$ be the set of all primes p such that $a_p = 0$. There are two cases to consider, depending on the size of $\mathcal{Z}$. Suppose first $\sum_{p \in \mathcal{Z}} 1/p = \infty$. Since $a_n = 0$ for each n with $p \mid n$ and $p^2 \nmid n$ for some $p \in \mathcal{Z}$, for each finite set $\mathcal{F} \subseteq \mathcal{Z}$ we have

$$\overline{d}(\{n \mid a_n \neq 0\}) \leq \prod_{p \in \mathcal{F}} \left(1 - \frac{p-1}{p^2}\right) \ll \exp\left(-\sum_{p \in \mathcal{F}} \frac{1}{p}\right).$$

Since $\mathcal{F} \subseteq \mathcal{Z}$ was arbitrary, we conclude that $\mathbf{a} \simeq \mathbf{0}$, as needed.

Suppose next that $\sum_{p \in \mathcal{Z}} 1/p < \infty$ and hence (bearing in mind the prime number theorem in arithmetic progressions) in each residue class $s + Q\mathbb{Z}$ with $\gcd(s, Q) = 1$ we can find infinitely many primes p with $a_p \neq 0$. Since for each $n \in \mathbb{N}_0$ with $a_n \neq 0$ that is not divisible by p we have $a_{pn} = a_p a_n \neq 0$, it follows that

$$\begin{aligned} 0 &= d\left(\left\{ n \in \mathbb{N}_0 \;\middle|\; \begin{array}{l} a_n \neq 0, \\ n \equiv r \bmod Q \end{array} \right\}\right) \\ &\geq \frac{1}{p}\left(\overline{d}\left(\left\{ n \in \mathbb{N}_0 \;\middle|\; \begin{array}{l} a_n \neq 0, \\ n \equiv s^{-1} r \bmod Q \end{array} \right\}\right) - \frac{1}{p}\right). \end{aligned}$$

Letting s vary and passing to the limit $p \to \infty$, we conclude that $a_n = 0$ for almost all $n \in \mathbb{N}_0$ coprime to Q, and hence also for almost all $n \in \mathbb{N}_0$. ∎

Before moving on to the proof of Proposition 6.4, let us briefly discuss the strategy. We begin with a relatively quick reduction to the case where the sequence $\mathbf{a}$ is completely multiplicative. Next, using Proposition 2.7, we find an automatic sequence ϕ which encodes much of the behaviour of $\mathbf{a}$. (For technical reasons, we will work with a slightly different automatic sequence ψ defined in the body of the argument, but the difference between the two is not relevant for now.) The benefit of working with ϕ as opposed to $\mathbf{a}$ is that ϕ is genuinely automatic and hence we can apply existing results on automatic sequences to it. Our plan is to exploit the multiplicative behaviour of ϕ. Of course, ϕ takes values in some set Σ_d ($d \in \mathbb{N}$) which does not have multiplicative structure ([2]) so in order to talk about multiplicative behaviour of ϕ we need to do more work. The idea is, essentially, to define multiplication on Σ_d in a way that is compatible with ϕ. For technical reasons, we are marginally more careful: We pick a multiplicatively rich integer M, a set $\mathcal{G} \subseteq \Sigma_d$ and define a map $f\colon \mathbb{N}_0 \to \mathcal{G}_0 := \mathcal{G} \cup \{0_{\mathcal{G}}\}$ by setting $f(n) = \phi(n)$ if n is coprime to M and $f(n) = 0_{\mathcal{G}}$ otherwise. (Formally, $\mathcal{G}$ is simply the set of all values of ϕ on integers coprime to M; the symbol $\mathcal{G}$ is chosen to emphasise that we think of $\mathcal{G}$ as a group.) We define an operation $\odot$ on $\mathcal{G}$ by setting $f(n) \odot f(m) := f(nm)$ for all $n, m \in \mathbb{N}_0$. Verification that this definition is well-posed and makes $\mathcal{G}$ into a group occupies the bulk of the argument. Once this is accomplished, we can use the classification of multiplicative automatic sequences as a black box to conclude that f is periodic. Letting L denote a period of f, it is relatively easy to show that $\mathbf{a}$ is asymptotically invariant under the shift by L. Combined with the results discussed earlier, this finishes the argument.

Proof of Proposition 6.4. If $\mathbf{a} \simeq \mathbf{0}$, there is nothing to prove, so assume that this is not the case. As a preliminary simplification, we will use an argument similar to one employed in [Kon20] to reduce to the case where $\mathbf{a}$ is completely multiplicative. Let p be a large prime. It follows from the pigeonhole principle that there exist $i, r, r' \in \mathbb{N}_0$ with $0 \le r, r' < k^i/p$ and $\gcd(k, r) = \gcd(k, r') = 1$ such that $r \not\equiv r' \bmod p$ and

$$(a_{k^i n + r})_{n=0}^{\infty} \simeq (a_{k^i n + r'})_{n=0}^{\infty} \quad \text{and} \quad (a_{k^i n + pr})_{n=0}^{\infty} \simeq (a_{k^i n + pr'})_{n=0}^{\infty}. \tag{40}$$

As a consequence, for almost all $n \in \mathbb{N}_0$ with $k^i n + r' \not\equiv 0 \bmod p$ we have

$$a_{k^i pn + pr} = a_{k^i pn + pr'} = a_p a_{k^i n + r'} = a_p a_{k^i n + r}. \tag{41}$$

Let $j \in \mathbb{N}$. By (41), for almost all $n \in \mathbb{N}_0$ with $k^i n + r \equiv p^j \bmod p^{j+1}$ we have

$$a_{p^{j+1}} a_{(k^i n + r)/p^j} = a_p a_{p^j} a_{(k^i n + r)/p^j}. \tag{42}$$

Using Lemma 6.8 we conclude that $a_{p^{j+1}} = a_p a_{p^j}$. Thus, applying induction with respect to j we conclude that $a_{p^j} = a_p^j$. Pick a threshold p_0 such that

([2]) The fact that Σ_d is a subset of the multiplicative semigroup $(\mathbb{N}_0, \cdot)$ is a red herring: Σ_d is only used to index the elements of $\widetilde{\mathcal{N}}_k(\mathbf{a})$ and any other d-element set would do equally well.

$a_{p^j} = a_p^j$ for all primes $p \geq p_0$ and all $j \in \mathbb{N}$, and consider the sequence $\bar{\mathbf{a}}$ given by

$$\bar{a}_n = \begin{cases} 0 & \text{if } n \text{ is divisible by a prime } p < p_0, \\ a_n & \text{otherwise.} \end{cases} \tag{43}$$

Then $\bar{\mathbf{a}}$ is completely multiplicative, asymptotically k-automatic, and takes the same value as $\mathbf{a}$ on all powers of sufficiently large primes. Hence, replacing $\mathbf{a}$ with $\bar{\mathbf{a}}$ if necessary, we may assume that $\mathbf{a}$ is completely multiplicative.

Pick $d, \mathbf{a}^{(0)}, \mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(d-1)}$ and $\phi\colon \Sigma_k^* \to \Sigma_d$ as in Proposition 2.7. Let H be an integer that is sufficiently multiplicatively rich that $\phi(0^H u) = \phi(0^{2H} u)$ for all $u \in \Sigma_k^*$, and put $K := k^H$. Define a map $\psi\colon \mathbb{N}_0 \to \Sigma_d$ by $\psi(m) = \phi(0^h (m)_k)$, where h is the unique integer satisfying $H \leq h < 2H$ and $h + |(m)_k| \equiv 0 \bmod H$. This construction ensures that for each $m \in \mathbb{N}_0$ and for each $i \in \mathbb{N}_0$ that is sufficiently large in terms of m (more precisely, $K^{i-1} > m$) we have

$$\big(a_{K^i n+m}\big)_{n=0}^{\infty} \simeq \big(a_n^{(\psi(m))}\big)_{n=0}^{\infty}. \tag{44}$$

We note that automaticity of ϕ implies that ψ is k-automatic.

Claim 1. *For each $q \in \mathbb{N}$ and $m, m' \in \mathbb{N}_0$, if $a_q \neq 0$ and $\psi(qm) = \psi(qm')$ then $\psi(m) = \psi(m')$.*

Proof. For each sufficiently large i, for almost all $n \in \mathbb{N}_0$ we have

$$\begin{aligned} a_n^{(\psi(m))} &= a_{K^i n+m} = a_q^{-1} a_{K^i qn+qm} = a_q^{-1} a_{qn}^{(\psi(qm))} \\ &= a_q^{-1} a_{qn}^{(\psi(qm'))} = a_q^{-1} a_{K^i qn+qm'} = a_{K^i n+m'} = a_n^{(\psi(m'))}. \end{aligned}$$

It follows that $\mathbf{a}^{(\psi(m))} \simeq \mathbf{a}^{(\psi(m'))}$ and hence $\psi(m) = \psi(m')$. △

We will need the following standard technical result, which is closely related to [Kon20, Lem. 3.4] and is fully proved e.g. in [KK24].

Claim 2. *There exists a positive integer M such that for each $j \in \Sigma_d$ with*

$$d_{\log}(\{n \in \mathbb{N}_0 \mid \gcd(n, M) = 1,\ \psi(n) = j\}) > 0,$$

for each $q \in \mathbb{N}$ coprime to M and each $r \in \mathbb{Z}/q\mathbb{Z}$ we also have

$$d_{\log}(\{n \in \mathbb{N}_0 \mid \gcd(n, M) = 1,\ \psi(n) = j,\ n \equiv r \bmod q\}) > 0.$$

For the rest of the argument, we let M denote the integer introduced above. We assume that M is divisible by k (otherwise we can freely replace M with kM).

Claim 3. *We have $d_{\log}(\{n \in \mathbb{N} \mid \gcd(n, M) = 1,\ \psi(n) = \psi(1)\}) > 0$.*

Proof. Pick any j with $d_{\log}(\{n \in \mathbb{N} \mid \gcd(n, M) = 1,\ \psi(n) = j\}) > 0$. Pick any q belonging to that set, that is, $\psi(q) = j$ and $\gcd(q, M) = 1$.

Consider the set

$$A = \{n \in \mathbb{N} \mid \gcd(n, M) = 1,\ \psi(qn) = j\}.$$

By Claim 2 we have $d_{\log}(A) > 0$, as

$$d_{\log}(qA) = d_{\log}(\{n \in \mathbb{N} \mid \gcd(n, M) = 1,\ \psi(n) = j,\ n \equiv 0 \bmod q\}) > 0.$$

For each $n \in A$ we have $\psi(qn) = \psi(q) = j$. If $a_q \neq 0$ then it follows from Claim 1 that $\psi(n) = \psi(1)$. Since $n \in A$ was arbitrary, we conclude that

$$d_{\log}(\{n \in \mathbb{N} \mid \gcd(n, M) = 1,\ \psi(n) = \psi(1)\}) \geq d_{\log}(A) > 0,$$

and the proof is complete. It remains to consider the case where $a_q = 0$ for all q as above. Inspecting the argument, we see that this is only possible if $a_n = 0$ for almost all $n \in \mathbb{N}$ that are coprime to M, and hence (since $\mathbf{a}$ is multiplicative) $\mathbf{a} \simeq \mathbf{0}$, contradicting an earlier assumption. △

We will routinely use the following observation in conjunction with Claim 1 to remove the assumption that $a_q \neq 0$.

CLAIM 4. *For each $q \in \mathbb{N}$ coprime to M we have $a_q \neq 0$.*

Proof. For a contradiction, suppose that $a_q = 0$. By Claims 2 and 3, there exists $r \in \mathbb{N}$ such that $\psi(qr) = \psi(1)$. Let i be a sufficiently large integer. Then

$$\mathbf{0} = (a_{K^i qn+qr})_{n=0}^{\infty} \simeq (a_{K^i qn+1})_{n=0}^{\infty}.$$

Hence, it follows from Lemma 6.8 that $\mathbf{a} \simeq 0$, contradicting an earlier assumption. △

CLAIM 5. *Let $q, r \in \mathbb{N}$ be coprime to M. If $\psi(q) = \psi(r) = \psi(1)$ then $\psi(qr) = \psi(1)$.*

Proof. By Claim 2, there exists $u \in \mathbb{N}$ coprime to M and such that

$$\psi(qru) = \psi(1) = \psi(q).$$

Hence, by Claim 1, we have

$$\psi(ru) = \psi(1) = \psi(r).$$

Using Claim 1 again, we obtain

$$\psi(u) = \psi(1) = \psi(qru).$$

Yet another application of Claim 1 yields

$$\psi(1) = \psi(qr),$$

as needed. △

CLAIM 6. *Let $q, q', r \in \mathbb{N}$ be coprime to M. If $\psi(q) = \psi(q')$ and $\psi(qr) = \psi(1)$ then $\psi(q'r) = \psi(1)$.*

Proof. By Claim 2, there exists $u \in \mathbb{N}$ coprime to M such that

$$\psi(q'ru) = \psi(1) = \psi(qr).$$

Hence, by repeated applications of Claim 1 we obtain

$$\psi(q'u) = \psi(q) = \psi(q'), \quad \psi(u) = \psi(1) = \psi(q'ru), \quad \psi(1) = \psi(q'r),$$

as needed. △

Claim 7. *Let $q, q', r, r' \in \mathbb{N}$ be coprime to M. If $\psi(q) = \psi(q')$ and $\psi(r) = \psi(r')$ then $\psi(qr) = \psi(q'r')$.*

Proof. By Claim 2, there exist $\hat{q}, \hat{r} \in \mathbb{N}$ coprime to M and such that

$$\psi(q\hat{q}) = \psi(r\hat{r}) = \psi(1).$$

By Claim 6 we also have

$$\psi(q'\hat{q}) = \psi(r'\hat{r}) = \psi(1).$$

By Claim 5, it follows that

$$\psi(qr\hat{q}\hat{r}) = \psi(1) = \psi(q'r'\hat{q}\hat{r}).$$

Thus, by Claim 1 we conclude that $\psi(qr) = \psi(q'r')$, as needed. △

We are now ready to use ψ to construct a multiplicative sequence that is genuinely automatic (as opposed to merely asymptotically automatic). Define

$$\mathcal{G} := \{\psi(q) \mid q \in \mathbb{N}, \gcd(q, M) = 1\},$$

and $\mathcal{G}_0 = \mathcal{G} \cup \{0_{\mathcal{G}}\}$, where $0_{\mathcal{G}} \notin \mathcal{G}$. Let $f : \mathbb{N}_0 \to \mathcal{G}$ be given by

$$f(n) = \begin{cases} \psi(n) & \text{if } \gcd(n, M) = 1, \\ 0_{\mathcal{G}} & \text{if } \gcd(n, M) > 1. \end{cases}$$

We define a binary operation $\odot$ on $\mathcal{G}$ by

$$\psi(q) \odot \psi(r) = \psi(qr)$$

for $q, r \in \mathbb{N}$ coprime to M; this definition is well-posed thanks to Claim 7. We extend $\odot$ to $\mathcal{G}_0$ by setting $g \odot 0_{\mathcal{G}} = 0_{\mathcal{G}} \odot g = 0_{\mathcal{G}}$ for all $g \in \mathcal{G}_0$. It is routine to verify for all $n, m \in \mathbb{N}_0$ we have

$$f(nm) = f(n) \odot f(m).$$

It follows that $(\mathcal{G}_0, \odot)$ is a finite abelian semigroup, which is the image of $(\mathbb{N}_0, \cdot)$ under the semigroup homomorphism f. Moreover, $\psi(1)$ is the identity element in $\mathcal{G}$, and by Claim 2, each element of $\mathcal{G}$ has an inverse. Hence, $(\mathcal{G}, \odot)$ is a group.

Claim 8. *The map f is periodic.*

Proof. If $(\mathcal{G}, \odot)$ is isomorphic to the cyclic group C_n for some $n \in \mathbb{N}$ then $\mathcal{G}_0$ can be identified with the subset of complex numbers $\{z \in \mathbb{C} \mid z^n = 1$ or $z = 0\}$, under which identification f can be construed as a multiplicative map $f\colon \mathbb{N}_0 \to \mathbb{C}$. Hence, the conclusion follows from Theorem 6.3. The general case can readily be derived by considering characters of $(\mathcal{G}, \odot)$. $\triangle$

Let L be a period of f that is also a multiple of M. Consider the complex-valued multiplicative sequence $\mathbf{a}'$ given by

$$a'_n := \begin{cases} a_n & \text{if } \gcd(n,k) = 1, \\ 0 & \text{if } \gcd(n,k) > 1. \end{cases}$$

Claim 9. *The sequence* $\mathbf{a}'$ *is asymptotically invariant under the shift by* L.

Proof. For each n with $\gcd(k,n) > 1$ we have $a'_n = a'_{n+L} = 0$, so we only need to verify that $a'_n = a'_{n+L}$ for almost all n coprime to k. Let $P_{i,r}$ be the arithmetic progression given by $P_{i,r} = K^i\mathbb{N}_0 + r$, where $i, r \in \mathbb{N}_0$, $r < K^{i-1} - L$, and $\gcd(r, M) = 1$. For almost all $n \in \mathbb{N}_0$ we have

$$a'_{K^i n+r} = a_{K^i n+r} = a_n^{(\psi(r))} = a_n^{(\psi(r+L))} = a_{K^i n+r+L} = a'_{K^i n+r+L}.$$

Hence, for almost all $n \in P_{i,r}$ we have $a'_{n+L} = a'_n$. Thus, letting

$$B_j := \{n \in \mathbb{N} \mid \gcd(n,k) = 1\} \setminus \bigcup_{i=1}^{j} \bigcup_{r} P_{i,r},$$

where the inner union runs over all integers r with $0 \le r < K^{i-1} - L$ and $\gcd(r, M) = 1$, it will suffice to show that $d(B_j) \to 0$ as $j \to \infty$. It is not hard to construct a word $w \in \Sigma_K^*$ that is not contained in the last j digits of the base-K expansion of any $n \in B_j$. (For instance, one can pick $w = (0^h1)^M$, where h is a sufficiently large and multiplicatively rich integer; the key idea is that if w appears in $(n)_K$ and $\gcd(n,k) = 1$ then n belongs to M different arithmetic progressions of the form $K^i\mathbb{N}_0 + r$ with $r < K^{i-h} \le K^{i-1} - L$ and at least one of those progressions satisfies $\gcd(r, M) = 1$.) It follows that

$$d(B_j) \le K^{-j} \cdot \#\{n \in [K^j] \mid (n)_K \text{ does not contain } w\} \to 0 \quad \text{as } j \to \infty,$$

which together with earlier considerations proves the claim. $\triangle$

Combining Claim 9 with Theorem 6.7 finishes the proof of Proposition 6.4. ■

7. Regular sequences. The fact that automatic sequences are necessarily finitely-valued poses a significant limitation. This motivated Allouche and Shallit to introduce a closely related notion of a *regular sequence* [AS92, AS03a]. Let R be an integral domain (e.g. $R = \mathbb{Z}$ or $R = \mathbb{F}_p$) and let K be the field of fractions of R (e.g. $K = \mathbb{Q}$ or $K = \mathbb{F}_p$). A sequence $\mathbf{a} \in R^\infty$ is k-*regular* if the R-module generated by $\mathcal{N}_k(\mathbf{a})$,

$$\mathcal{M}_k(\mathbf{a}) := \operatorname{span}_R \mathcal{N}_k(\mathbf{a}),$$

is finitely generated. In the case where $R = \mathbb{Z}$, this is equivalent to the assertion that $\mathcal{N}_k(\mathbf{a})$ spans a finite-dimensional space over $\mathbb{Q}$. In particular, k-regular sequences include k-automatic sequences and polynomial sequences. Conversely, one can show that each finitely-valued k-regular sequence is k-automatic.

In analogy with asymptotically k-automatic sequences, we can define asymptotically k-regular sequences. Thus, a sequence $\mathbf{a} \in R^\infty$ is *asymptotically k-regular* if $\widetilde{\mathcal{M}}_k(\mathbf{a}) := \mathcal{M}_k(\mathbf{a})/\simeq$ is a finitely-generated R-module, or equivalently, if there exist a finite number of sequences $\mathbf{a}^{(0)}, \mathbf{a}^{(1)}, \dots, \mathbf{a}^{(d-1)}$ such that for each $\mathbf{b} \in \mathcal{M}_k(\mathbf{a})$ there exist $t_0, t_1, \dots, t_{d-1} \in R$ with

$$\mathbf{b} \simeq t_0\mathbf{a}^{(0)} + t_1\mathbf{a}^{(1)} + \dots + t_{d-1}\mathbf{a}^{(d-1)}.$$

Clearly, the asymptotically k-regular sequences include the k-regular sequences and the asymptotically k-automatic sequences. It also remains true, with essentially the same argument, that each finitely-valued asymptotically k-regular sequence is asymptotically k-automatic.

An important feature of real-valued k-regular sequences is that their rate of growth is at most polynomial, meaning that if $\mathbf{a} \in \mathbb{R}^\infty$ is k-regular then

$$\limsup_{n\to\infty} \frac{\log |a_n|}{\log n} < \infty.$$

This is no longer true for asymptotically k-regular sequences, which, as we will see below, can grow arbitrarily fast. This suggests that asymptotically k-regular sequences exhibit significantly less structured behaviour than k-regular sequences.

PROPOSITION 7.1. *For each $f\colon \mathbb{N}_0 \to \mathbb{N}_0$, there exists a k-regular sequence $\mathbf{a} \in \mathbb{Z}^\infty$ such that $a_n \geq f(n)$ for all $n \in \mathbb{N}_0$.*

Proof. We rely (once more) on the construction in [Kon23, Sec. 4]; cf. Proposition 4.1 above. Let $\mathcal{H} = \{H_i\}_{i=0}^\infty = \{2^{\alpha_i}3^{\beta_i}\}_{i=0}^\infty$ be as in Proposition 4.1, and let $\mathbf{a}$ be given by

$$a_n = F(\beta_i) \quad \text{for } n \in [H_i, H_{i+1}) \text{ and } i \in \mathbb{N}_0, \tag{45}$$

where $F\colon \mathbb{N}_0 \to \mathbb{N}_0$ is an increasing function that remains to be specified. Following the same argument as in [Kon23, Lem. 4.2], we get $\#\mathcal{N}_2(\mathbf{a}) = 1$, which in particular implies that $\mathbf{a}$ is asymptotically 2-regular.

Since for each $\gamma \in \mathbb{N}$, the union of all intervals $[H_i, H_{i+1})$ such that $\beta_i \leq \gamma$ has density zero, we can construct a non-decreasing sequence $\boldsymbol{\gamma} \in \mathbb{N}_0^\infty$ such that $\gamma_n \to \infty$ as $n \to \infty$ and for almost all n, letting i denote the unique index such that $n \in [H_i, H_{i+1})$, we have $\gamma_n \leq \beta_i$. Thus, $a_n \geq F(\gamma_n)$ for almost all n. Picking $F(\gamma) := \max\{f(n) \mid \gamma_n \leq \gamma\}$ we conclude that $a_n \geq f(n)$ for almost all n.

Finally, let $\mathbf{a}'$ be given by $a'_n = \max(a_n, f(n))$. Then directly from definition we have $a'_n \geq f(n)$ for all $n \in \mathbb{N}_0$ and $\mathbf{a}'$ is asymptotically 2-regular because $\mathbf{a}' \simeq \mathbf{a}$. ∎

In [Bel06], Bell proved the analogue of Cobham's theorem for k-regular sequences. With a similar argument, we can obtain a slightly stronger version, which is analogous to [Kon23, Thm. B]. We will say that a sequence $\mathbf{a} \in R^\infty$ satisfies a linear recurrence (over R) if there exist $d \in \mathbb{N}$ and $t_1, t_2, \ldots, t_d \in R$ such that

$$a_{n+d} = t_1 a_{n+d-1} + t_2 a_{n+d-2} + \cdots + t_d a_n \quad \text{for all } n \in \mathbb{N}_0.$$

If $R = \mathbb{Z}$ and $\mathbf{a}$ is additionally k-regular for some $k \geq 2$ then this is equivalent to the requirement that there exist $Q \in \mathbb{N}$ and polynomial sequences p_m for $0 \leq m < Q$ such that $a_n = p_m(n)$ for each sufficiently large $n \equiv m \bmod Q$. For the sake of simplicity, we only consider the case where $R = \mathbb{Z}$ and the bases k, ℓ are coprime.

Theorem 7.2. *Let $k, \ell \in \mathbb{N}$ be coprime, and let $\mathbf{a} \in \mathbb{Z}^\infty$ be a sequence that is both k-regular and asymptotically ℓ-regular. Then $\mathbf{a} \simeq \mathbf{a}'$ for a k-regular sequence $\mathbf{a}' \in \mathbb{Z}^\infty$ that satisfies a linear recurrence.*

Before we approach the proof, it will be convenient to introduce some preliminaries. For $u \in \Sigma_k$ we define a linear map $T_u \colon \mathbb{Q}^\infty \to \mathbb{Q}^\infty$, given by

$$T_u(\mathbf{a}) := (a_{k^{|u|}n+[u]_k})_{n=0}^\infty, \qquad \mathbf{a} \in \mathbb{Q}^\infty. \tag{46}$$

The operators T_u follow a familiar composition rule: $T_{uv} = T_v T_u$ for $u, v \in \Sigma_k^*$. We note that, for a k-regular sequence $\mathbf{a}$, the module $\mathcal{M}_k(\mathbf{a})$ and the space $\operatorname{span}_{\mathbb{Q}} \mathcal{N}_k(\mathbf{a})$ are invariant under T_u. Since the space of sequences that are almost everywhere zero is closed under T_u, for each asymptotically k-regular $\mathbf{a}$ we also have a map $T_u \colon \widetilde{\mathcal{M}}_k(\mathbf{a}) \to \widetilde{\mathcal{M}}_k(\mathbf{a})$. The same formula (46) defines a map $T_u \colon \mathbb{F}_p^\infty \to \mathbb{F}_p^\infty$ for prime p.

Lemma 7.3. *Let $k \in \mathbb{N}$, let p be a prime and let $\mathbf{a}, \mathbf{a}' \in \mathbb{F}_p^\infty$ be k-regular sequences such that for all $i \in \Sigma_k$ the maps $T_i \colon \mathcal{M}_k(\mathbf{a}) \to \mathcal{M}_k(\mathbf{a})$ and $T_i \colon \mathcal{M}_k(\mathbf{a}') \to \mathcal{M}_k(\mathbf{a}')$ are invertible. If $\mathbf{a} \simeq \mathbf{a}'$ then $\mathbf{a} = \mathbf{a}'$.*

Proof. The sequence $\mathbf{b} \in \mathbb{F}_p^\infty$ given by

$$b_n = \begin{cases} 0 & \text{if } a_n = a'_n, \\ 1 & \text{if } a_n \neq a'_n \end{cases}$$

is almost everywhere zero and k-regular, and hence k-automatic, since $\mathbb{F}_p$ is finite. It follows that there exists a word $w \in \Sigma_k^*$ such that $a_n = a'_n$ for all $n \in \mathbb{N}_0$ such that $(n)_k$ contains w as a subword. In particular, we have $T_w(\mathbf{a}) = T_w(\mathbf{a}')$. Since T_w is invertible (on either $\mathcal{M}_k(\mathbf{a})$ or $\mathcal{M}_k(\mathbf{a}')$), we conclude that $\mathbf{a} = \mathbf{a}'$. ∎

Lemma 7.4. *Let $k \in \mathbb{N}$ and let $\mathbf{a} \in \mathbb{Q}^\infty$ be a k-regular sequence. Then there exists $\theta > 0$ such that for each $\mathbf{a}' \in \mathcal{M}_k(\mathbf{a})$ we have either $\mathbf{a}' \simeq \mathbf{0}$ or $\overline{d}(\{n \in \mathbb{N}_0 \mid a'_n \neq 0\}) \geq \theta$.*

Proof. Suppose that, on the contrary, we have a family $\mathbf{a}^{(i)} \in \mathcal{M}_k(\mathbf{a})$ ($i \in \mathbb{N}_0$) with

$$0 < \overline{d}(\{n \in \mathbb{N}_0 \mid a^{(i)}_n \neq 0\}) \to 0 \quad \text{ as } i \to \infty.$$

Passing to a subsequence we can assume without loss of generality that

$$\overline{d}(\{n \in \mathbb{N}_0 \mid a^{(i)}_n \neq 0\}) \geq 2\overline{d}(\{n \in \mathbb{N}_0 \mid a^{(i+1)}_n \neq 0\}) \quad \text{ for each } i \in \mathbb{N}_0.$$

As a consequence, for each linear combination

$$\mathbf{b} = \sum_{i=I}^{I+J} t_i \mathbf{a}^{(i)}$$

with $I, J \in \mathbb{N}_0$, $t_i \in \mathbb{Q}$ and $t_I \neq 0$ we have

$$\overline{d}(\{n \in \mathbb{N}_0 \mid b_n \neq 0\}) \geq \overline{d}(\{n \in \mathbb{N}_0 \mid a^{(i)}_n \neq 0\}) \cdot \left(1 - \frac{1}{2} - \cdots - \frac{1}{2^J}\right) > 0.$$

Therefore, the sequences $\mathbf{a}^{(i)}$ are linearly independent, contradicting the assumption of regularity. ∎

Proof of Theorem 7.2. We first consider the case where for each $i \in \Sigma_k^*$ the map T_i is invertible on $\operatorname{span}_{\mathbb{Q}} \mathcal{N}_k(\mathbf{a})$. For each prime p the sequence $\mathbf{a} \bmod p := (a_n \bmod p)_{n=0}^\infty$ is both k-automatic and asymptotically ℓ-automatic. Hence, it follows from [Kon23, Thm. B] that $\mathbf{a} \bmod p$ is asymptotically periodic. With the same argument as in [Bel06, Prop. 3.10], we see that $T_u(\mathbf{a}) \bmod p$ is asymptotically periodic with period coprime to k for each $u \in \Sigma_k^*$ such that $|u| \geq \dim \operatorname{span}_{\mathbb{Q}} \mathcal{N}_k(\mathbf{a})$. Note that the maps $T_i \bmod p$: $\mathcal{M}_k(\mathbf{a} \bmod p) \to \mathcal{M}_k(\mathbf{a} \bmod p)$ are invertible for all primes p sufficiently large. Thus, it follows from Lemma 7.3 that $T_u(\mathbf{a}) \bmod p$ is (exactly) periodic. Hence, we deduce from [Bel06, Thm. 3.4] that $T_u(\mathbf{a})$ satisfies a linear recurrence. Applying this result to all $u \in \Sigma_k^*$ with $|u| = \dim \operatorname{span}_{\mathbb{Q}} \mathcal{N}_k(\mathbf{a})$, we conclude that $\mathbf{a}$ satisfies a linear recurrence (cf. [Bel06, Thm. 3.11]). Additionally, inspecting the proof of [Bel06, Thm. 3.4] we see that the length of the said linear recurrence is bounded by a quantity depending only on $\dim \operatorname{span}_{\mathbb{Q}} \mathcal{N}_k(\mathbf{a})$.

We now proceed to consider the general case. Let $w \in \Sigma_k^*$ be such that $\dim T_w(\operatorname{span}_{\mathbb{Q}} \mathcal{N}_k(\mathbf{a}))$ is least possible. Then for each $u \in \Sigma_k^*$ such that w is a subword of u, the sequence $\mathbf{a}' = T_u(\mathbf{a})$ has the property that for each $i \in \Sigma_k$, the restriction of T_i to $\mathcal{M}_k(\mathbf{a}')$ is invertible. Hence, we conclude from the special case considered above that the restriction of $\mathbf{a}'$ to each residue class modulo some $Q = O(1)$ coincides with a polynomial sequence of degree at most $d = O(1)$. Thus, we have constructed a large family of arithmetic progressions of the form $k^i Q\mathbb{N}_0 + r$ $(i, r \in \mathbb{N}_0)$ where $\mathbf{a}$ behaves in the expected way; our remaining task is to understand how the restrictions of $\mathbf{a}$ to different progressions are interrelated.

For $i \in \mathbb{N}_0$, let $\mathbf{b}^{(i)}$ denote the sequence given by

$$b_n^{(i)} = \sum_{m=0}^{d+1} (-1)^m \binom{d+1}{m} a_{n+k^i Qm} \quad \text{for } n \in \mathbb{N}_0. \tag{47}$$

Then $b_n^{(i)} = 0$ for each $n \in \mathbb{N}_0$ such that $(n)_k^i$ contains w as a subword. For $j, s \in \mathbb{N}_0$ such that $s + k^i Q(d+1) < \ell^j$, consider further the sequence $\mathbf{c}^{(i,j,s)}$ given by

$$c_n^{(i,j,s)} = b_{\ell^j n+s}^{(i)} \quad \text{for } n \in \mathbb{N}_0. \tag{48}$$

Inspecting the definitions, we see that $\mathbf{c}^{(i,j,s)} \in \mathcal{M}_\ell(\mathbf{a})$. Moreover, since k and ℓ are coprime, we have

$$\begin{aligned}&\overline{d}(\{n \in \mathbb{N}_0 \mid c_n^{(i,j,s)} \neq 0\}) \\ &\qquad \leq \overline{d}(\{n \in \mathbb{N}_0 \mid w \text{ does not appear in } (n)_k^i\}) \to 0 \quad \text{as } i \to \infty.\end{aligned}$$

Thus, it follows from Lemma 7.4 that there exists $i \in \mathbb{N}_0$ such that for all j, s like above we have $\mathbf{c}^{(i,j,s)} \simeq 0$. In other words, $b_n^{(i)} = 0$ for almost all $n \in \ell^j \mathbb{N}_0 + s$. Note that almost all positive integers belong to $\ell^j \mathbb{N}_0 + s$ for some admissible j, s; more precisely, we have

$$\underline{d}\Big(\bigcup_{j=0}^{J} \bigcup_{s=0}^{\ell^j - k^i Q(d+1)} (\ell^j \mathbb{N}_0 + s)\Big) \to 1 \quad \text{as } J \to \infty.$$

It follows that $\mathbf{b}^{(i)} \simeq 0$. Thus, we may partition $\mathbb{N}_0$ into a union of pairwise disjoint arithmetic progressions $(I_m)_{m=1}^{\infty}$ with step $k^i Q$ such that $\sum_{m=1}^{M} |I_m|/M \to \infty$ as $M \to \infty$ and for each m, $(a_n)_{n \in I_m}$ coincides with a polynomial sequence of degree at most d. Consider any pair of infinite arithmetic progressions $P_1 = k^{i_1} Q\mathbb{N}_0 + r_1$ and $P_2 = k^{i_2} Q\mathbb{N}_0 + r_2$ with $0 \leq r_1 < k^{i_1}$, $0 \leq r_2 < k^{i_2}$ such that the restriction of $\mathbf{a}$ to either progression coincides coincides with a polynomial of degree at most d, as constructed earlier. Suppose additionally that $i_1, i_2 \geq i$ and $r_1 \equiv r_2 \bmod k^i$. Then we can find m such that $\#(P_1 \cap I_m), \#(P_2 \cap I_m) > d$ and as a consequence the restrictions of $\mathbf{a}$ to P_1 and P_2 (and I_m) are given by the same polynomial. Varying the choice of P_1 and P_2 we conclude that $\mathbf{a}$ coincides with a polynomial of degree at most d almost everywhere on each residue class modulo $k^i Q$. ∎

We close this section with a conjecture, which appears to be the correct analogue of Cobham's theorem (cf. [Bel06, Thm. 1.4] and [Kon23, Thm. A]).

Conjecture 7.5. Let $k, \ell \in \mathbb{N}$ be multiplcatively independent, and let $\mathbf{a} \in \mathbb{Z}^\infty$ be a sequence that is both asymptotically k-regular and asymptotically ℓ-regular. Then $\mathbf{a}$ satisfies a linear recurrence almost everywhere, i.e., there exist $d \in \mathbb{N}$ and $t_1, t_2, \dots, t_d \in \mathbb{Z}$ such that

$$a_{n+d} = t_1 a_{n+d-1} + t_2 a_{n+d-2} + \dots + t_d a_n \quad \text{for almost all } n \in \mathbb{N}_0.$$

Appendix. Shift-invariant multiplicative sequences. Proof of Theorem 6.7. Suppose first that $\mathbf{a}$ is invariant under the shift by 1. Then in particular

$$\frac{1}{N}\sum_{n=1}^{N} |a_{n+1} - a_n| = 0. \tag{49}$$

Hence, Kátai's conjecture proved by Klurman [Klu17, Thm. 1.8] implies that either

$$\frac{1}{N}\sum_{n=1}^{N} |a_n| = 0, \tag{50}$$

or there exists $s \in \mathbb{C}$ with $\mathrm{Re}(s) < 1$ such that $a_n = n^s$ for all $n \in \mathbb{N}$. In the former case, since the set $\{|a_n| \mid n \in \mathbb{N},\ a_n \neq 0\}$ is bounded away from 0, we have $\mathbf{a} \simeq \mathbf{0}$. In the latter case, since a_n is finitely-valued, we have $s = 0$, meaning that $\mathbf{a} = \mathbf{1}$.

Let us now consider the general case, where $\mathbf{a}$ is invariant under the shift by some $d \in \mathbb{N}$, not necessarily equal to 1. We will heavily rely on the material in [Klu17]. If $\mathbf{a} \simeq \mathbf{0}$ then there is nothing left to prove, so assume that this is not the case. It will be convenient to decompose $\mathbf{a}$ as the product of multiplicative sequences $\mathbf{a}^{(p)}$ corresponding to different primes p. If p is a prime then the sequence $\mathbf{a}^{(p)}$ is specified for primes r and exponents i by

$$a^{(p)}_{r^i} = \begin{cases} a_{r^i} & \text{if } r = p, \\ 1 & \text{otherwise,} \end{cases}$$

and extended to $\mathbb{N}$ by multiplicativity. Clearly, $a_n = \prod_p a^{(p)}_n$ for all $n \in \mathbb{N}$. We will also use the sequences $\mathbf{c}$ and $\mathbf{c}^{(p)}$ given by $c_n = |a_n|^2$ and $c^{(p)}_n = |a^{(p)}_n|^2$.

Let us consider the limit

$$A := \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} a_n \overline{a}_{n+d}. \tag{51}$$

On the one hand, because $\mathbf{a}$ is asymptotically invariant under the shift by d, we have

$$A = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} a_n \overline{a}_n = \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} c_n. \tag{52}$$

By Halász's theorem [Hal71] (cf. [Klu17, eq. (1)]), the mean value of $\mathbf{c}$ exists (thus, A is well-defined) and is the product of contributions coming from the primes:

$$A = \prod_p C_p, \quad \text{where} \quad C_p := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} c_n^{(p)}. \tag{53}$$

Note that A is real and $A > 0$ because the limit in (52) exists and $\mathbf{a}$ is finitely-valued and not asymptotically equal to $\mathbf{0}$.

The "distance" between $\mathbf{a}$ and another 1-bounded multiplicative sequence $\mathbf{b}$ is defined by

$$\mathbb{D}(\mathbf{a}, \mathbf{b}) := \sqrt{\sum_{p \text{ prime}} \frac{1 - \operatorname{Re}(a_p \bar{b}_p)}{p}} \in [0, \infty].$$

Claim 1. *If $\mathbb{D}(\mathbf{a}, \mathbf{1}) < \infty$ then for each prime p there exists a root of unity λ such that $a_{n+d}^{(p)} = \lambda a_n^{(p)}$ for almost all $n \in \mathbb{N}$.*

Proof. By [Klu17, Theorem 1.3], the limit in (51) admits a factorisation:

$$A = \prod_p B_p, \quad \text{where} \quad B_p := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} a_n^{(p)} \overline{a}_{n+d}^{(p)}. \tag{54}$$

An application of the Cauchy–Schwarz inequality shows that $|B_p| \leq C_p$, and hence comparing (53) and (54) we conclude that $|B_p| = C_p$ for all primes p. Let p be a prime and pick $\lambda \in \mathbb{C}$ with $|\lambda|$ such that $\lambda B_p \in \mathbb{R}_+$. Then

$$0 \leq \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} |a_{n+d}^{(p)} - a_n^{(p)}|^2 = 2C_p - 2|B_p| = 0. \tag{55}$$

Since $\mathbf{a}$ is finitely-valued, it follows that $a_{n+d}^{(p)} = \lambda a_n^{(p)}$ for almost all n. For each $z \in \mathbb{C}$ that appears in $\mathbf{a}$ with positive upper frequency, the same is true of $\lambda z, \lambda^2 z, \ldots$, and consequently λ is a root of unity. $\triangle$

Claim 2. *Let p be a prime and let λ be a root of unity such that $a_{n+d}^{(p)} = \lambda a_n^{(p)}$ for almost all $n \in \mathbb{N}$. Then $\mathbf{a}^{(p)}$ is periodic. If $p \nmid 2d$ then $\mathbf{a}^{(p)} = \mathbf{1}$.*

Proof. Let $h := \nu_p(d)$ and for $i \in \mathbb{N}_0$ let $z_i := a_{p^i}$. For each $n \in \mathbb{N}$ with $\nu_p(n) =: i$ we have

$$\nu_p(n+d) = \begin{cases} i & \text{if } i < h, \\ h & \text{if } i > h, \\ h + \nu_p(n+d)/p^h & \text{if } i = h. \end{cases} \tag{56}$$

If $p \neq 2$ then $\nu_p((n+d)/p^h)$ can take any value in $\mathbb{N}_0$, and if $p = 2$ then $\nu_p((n+d)/p^h)$ can take any value in $\mathbb{N}$. We also note that each possible pair of values $\nu_p(n), \nu_p(n+d)$ occurs for a periodic set of values of n, and hence with positive frequency. Thus, assuming for a moment that $p \neq 2$, from (56) that

$$\lambda z_i = z_i \quad \text{for all } i < h, \tag{57}$$

$$\lambda z_i = z_h \quad \text{for all } i > h, \tag{58}$$

$$\lambda z_h = z_i \quad \text{for all } i \geq h. \tag{59}$$

If $p = 2$ then (59) only holds for $i > h$.

It follows from (58) that $(z_i)_{i=0}^{\infty}$ is eventually constant: $z_i = \lambda^{-1} z_h$ for all $i > h$. As a consequence, $\mathbf{a}^{(p)}$ is periodic with period p^{h+1}. This proves the first part of the claim. For the second part, taking $i = h = 0$ in (59) we conclude that $\lambda = 1$. Hence, (58) implies that $z_i = z_0 = 1$ for all $i \geq 1$ and consequently $\mathbf{a}^{(p)} = \mathbf{1}$. △

In the case where $\mathbb{D}(\mathbf{a}, \mathbf{1}) < \infty$, we are now ready to finish the argument. Indeed, in this case each of the sequences $\mathbf{a}^{(p)}$ is periodic, and each except for finitely many of them is the constant sequence $\mathbf{1}$, and hence $\mathbf{a}$ is periodic. In general, it is not necessarily the case that $\mathbb{D}(\mathbf{a}, \mathbf{1}) < \infty$ but we have the following slightly weaker fact.

Claim 3. *There exists $t \in \mathbb{R}$ and a primitive Dirichelet character χ with some conductor q such that for the sequence $\mathbf{b}$ given by $b_n = \chi(n) n^{it}$ we have $\mathbb{D}(\mathbf{a}, \mathbf{b}) < \infty$.*

Proof. Since the limit defining A exists, we also have

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{n=1}^{N} \frac{a_n \overline{a}_{n+d}}{n} = A > 0.$$

This puts us in the same position as in [Klu17, second half of proof of Lemma 4.3]. The claim follows by combining Tao's result on two-point correlations of multiplicative functions [Tao16], [Klu17, Thm. 4.1] with a technical lemma due to Elliott [Ell10] (see [Klu17, Lem. 4.2]) and repeating the reasoning at the end of proof of Lemma 4.3 verbatim. △

Claim 4. *We have $t = 0$.*

Proof. For a contradiction, suppose that $t \neq 0$ and let $\tau := t/\pi$. Since $\mathbf{a}$ and χ are finitely-valued, there exist $\theta \in [0, 1)$ and $\varepsilon, \delta > 0$ such that for each sufficiently large prime p with $|\{\tau \log p\} - \theta| < \delta$ we have $\mathrm{Re}(a_p \overline{\chi}(p) p^{-it}) < 1 - \varepsilon$. It follows from the prime number theorem that for each sufficiently large n the number of primes p with $|\tau \log p - n - \theta| < \delta$ is

$$\frac{\exp\left(\frac{n+\theta+\delta}{\tau}\right)}{(n+\theta+\delta)/\tau} - \frac{\exp\left(\frac{n+\theta-\delta}{\tau}\right)}{(n+\theta-\delta)/\tau} + o\left(\frac{\exp(\frac{n}{\tau})}{n}\right) = (c + o(1)) \frac{\exp\left(\frac{n}{\tau}\right)}{n},$$

where $c > 0$ is a constant depending on δ, τ and θ. It follows that

$$\mathbb{D}(\mathbf{a},\mathbf{b})^2 \geq \varepsilon \sum_{n=1}^{\infty} \sum_{\substack{p \text{ prime} \\ |\tau \log p - \theta - n| < \delta}} \frac{1}{p} \gg \varepsilon \sum_{n=1}^{\infty} \frac{c+o(1)}{n} = \infty,$$

contrary to the defining property of $\mathbf{b}$. $\triangle$

Claim 5. *The sequence* $\mathbf{a}\overline{\mathbf{b}} = (a_n\overline{\chi}(n))_{n=1}^{\infty}$ *is periodic.*

Proof. The sequence $\mathbf{a}\overline{\mathbf{b}}$ is multiplicative, asymptotically invariant under the shift by qd, and satisfies $\mathbb{D}(\mathbf{a}\overline{\mathbf{b}},\mathbf{1}) = \mathbb{D}(\mathbf{a},\mathbf{b}) < \infty$. Thus, applying the earlier reasoning with $\mathbf{a}\overline{\mathbf{b}}$ in place of $\mathbf{a}$ we conclude that $\mathbf{a}\overline{\mathbf{b}}$ is periodic. $\triangle$

Claim 6. *The sequence* $\mathbf{a}$ *is periodic.*

Proof. It follows from Claim 5 that the restriction of $\mathbf{a}$ to integers coprime to qd is periodic with period qd. For integers $m, m', r \in \mathbb{N}_0$ such that m, m' divide a power of qd and that $0 \leq r < qd\,\mathrm{lcm}(m,m')$, consider the set P of integers n such that (1) n/m is an integer coprime to qd; (2) $(n+d)/m'$ is an integer coprime to qd; (3) $n \equiv r \bmod qd\,\mathrm{lcm}(m,m')$. Clearly, P is periodic. For $n \in P$ we have

$$a_n = a_{n/m}a_m = a_{r/n}a_m, \qquad a_{n+d} = a_{(n+d)/m'}a_{m'} = a_{(r+d)/m'}a_{m'},$$

and in particular a_n and a_{n+d} are constant on P. Since $a_{n+d} = a_n$ for almost all $n \in \mathbb{N}$ and P has positive density or is empty, we conclude that $a_{n+d} = a_n$ for all $n \in P$. Taking the union over all possible choices of m, m', r we conclude that $a_{n+d} = a_n$ for all $n \in \mathbb{N}$. $\triangle$ ∎

Acknowledgements. The author is grateful to Boris Adamczewski, Jakub Byszewski, Daniel Krenn, Oleksiy Klurman, Clemens Müllner, and Jeffrey Shallit for helpful comments and discussions.

During a significant part of the work on this project, the author worked within the framework of the LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). The author currently works at the University of Oxford and is supported by UKRI Fellowship EP/X033813/1.

References

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

[AG18] J.-P. Allouche and L. Goldmakher, *Mock characters and the Kronecker symbol*, J. Number Theory 192 (2018), 356–372.

[AS92] J.-P. Allouche and J. Shallit, *The ring of k-regular sequences*, Theoret. Comput. Sci. 98 (1992), 163–197.

[AS03a] J.-P. Allouche and J. Shallit, *The ring of k-regular sequences, II*, Theoret. Comput. Sci. 307 (2003), 3–29.

[AS03b] J.-P. Allouche and J. Shallit, *Automatic Sequences*, Cambridge Univ. Press, Cambridge, 2003.

[Bel06] J. P. Bell, *A generalization of Cobham's theorem for regular sequences*, Sém. Lothar. Combin. 54A (2006), art. B54Ap, 15 pp.

[BBC12] J. P. Bell, N. Bruin, and M. Coons, *Transcendence of generating functions whose coefficients are multiplicative*, Trans. Amer. Math. Soc. 364 (2012), 933–959.

[BCH14] J. P. Bell, M. Coons, and K. G. Hare, *The minimal growth of a k-regular sequence*, Bull. Austral. Math. Soc. 90 (2014), 195–203.

[BHKS21] V. Bergelson, I. J. Håland Knutson, and Y. Son, *An extension of Weyl's equidistribution theorem to generalized polynomials and applications*, Int. Math. Res. Notices 2021, 14965–15018.

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

[BCC10] P. Borwein, S. K. K. Choi, and M. Coons, *Completely multiplicative functions taking values in $\{-1, 1\}$*, Trans. Amer. Math. Soc. 362 (2010), 6279–6291.

[BC09] P. Borwein and M. Coons, *Transcendence of power series for some number theoretic functions*, Proc. Amer. Math. Soc. 137 (2009), 1303–1305.

[BK18] J. Byszewski and J. Konieczny, *Sparse generalised polynomials*, Trans. Amer. Math. Soc. 370 (2018), 8081–8109.

[BK20] J. Byszewski and J. Konieczny, *Automatic sequences and generalised polynomials*, Canad. J. Math. 72 (2020), 392–426.

[Coo10] M. Coons, *(Non)Automaticity of number theoretic functions*, J. Théor. Nombres Bordeaux 22 (2010), 339–352.

[CH73] E. M. Coven and G. A. Hedlund, *Sequences with minimal block growth*, Math. Systems Theory 7 (1973), 138–153.

[Ell10] P. D. T. A. Elliott, *The value distribution of additive arithmetic functions on a line*, J. Reine Angew. Math. 642 (2010), 57–108.

[Hål93] I. J. Håland, *Uniform distribution of generalized polynomials*, J. Number Theory 45 (1993), 327–366.

[Hål94] I. J. Håland, *Uniform distribution of generalized polynomials of the product type*, Acta Arith. 67 (1994), 13–27.

[HK95] I. J. Håland and D. E. Knuth, *Polynomials involving the floor function*, Math. Scand. 76 (1995), 194–200.

[Hal71] G. Halász, *On the distribution of additive and the mean values of multiplicative arithmetic functions*, Studia Sci. Math. Hungar. 6 (1971), 211–233.

[Klu17] O. Klurman, *Correlations of multiplicative functions and applications*, Compos. Math. 153 (2017), 1622–1657.

[KK24] O. Klurman and J. Konieczny, *Classification of automatic semigroups and related problems*, 2024, in preparation.

[KK19] O. Klurman and P. Kurlberg, *A note on multiplicative automatic sequences*, C. R. Math. Acad. Sci. Paris 357 (2019), 752–755.

[KK20] O. Klurman and P. Kurlberg, *A note on multiplicative automatic sequences on II*, Bull. London Math. Soc. 52 (2020), 185–188.

[Kon20] J. Konieczny, *On multiplicative automatic sequences*, Bull. London Math. Soc. 52 (2020), 175–184.

[Kon22] J. Konieczny, *Generalised polynomials and integer powers*, J. London Math. Soc. 105 (2022), 154–219.

[Kon23] J. Konieczny, *An asymptotic version of Cobham's theorem*, Acta Arith. 211 (2023), 323–343.

[KLM22] J. Konieczny, M. Lemańczyk, and C. Müllner, *Multiplicative automatic sequences*, Math. Z. 300 (2022), 1297–1318.

[Lei12] A. Leibman, *A canonical form and the distribution of values of generalized polynomials*, Israel J. Math. 188 (2012), 131–176.

[Li20] S. Li, *On completely multiplicative automatic sequences*, J. Number Theory 213 (2020), 388–399.

[PF02] N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, Lecture Notes in Math. 1794, Springer, Berlin, 2002.

[SP03] J.-C. Schlage-Puchta, *A criterion for non-automaticity of sequences*, J. Integer Seq. 6 (2003), art. 03.3.8, 5 pp.

[Tao16] T. Tao, *The logarithmically averaged Chowla and Elliott conjectures for two-point correlations*, Forum Math. Pi 4 (2016), art. e8, 36 pp.

[Yaz01] S. Yazdani, *Multiplicative functions and k-automatic sequences*, J. Théor. Nombres Bordeaux 13 (2001), 651–658.

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

Current address:
Department of Computer Science
University of Oxford
Oxford OX1 3QD, UK

E-mail: jakub.konieczny@gmail.com