

SUPERLINEAR COMPLEXITY OF THE $(3/2)^n$ STEERING WORD

RALF STEPHAN

ABSTRACT. Write $(3/2)^n = m_n + \varepsilon_n$ with m_n the nearest integer and $\varepsilon_n \in [-\frac{1}{2}, \frac{1}{2})$, and let $T = (t_n)$, $t_n = 2m_{n+1} - 3m_n$, be the resulting *steering word*: the step-by-step record of the map $x \mapsto \frac{3}{2}x$ on the orbit of 1, coded by nearest-integer rounding. We prove that the subword complexity $p_T(k)$ of T is superlinear, $p_T(k)/k \rightarrow \infty$. A pigeonhole argument reduces this to exponential pair repulsion of the orbit modulo one, which we establish by two applications of the p -adic Subspace Theorem: once in three variables, which handles all gaps between the two orbit points at once, and once in two variables (Ridout's theorem) for the single degenerate family of constant gap. The argument is completely formalized in Lean 4, together with the Subspace Theorem it uses.

1. INTRODUCTION

For a real number x we write $\|x\| = \min_{n \in \mathbb{Z}} |x - n|$ for its distance to the nearest integer. The distribution of the sequence $(\|\alpha^n\|)$ for a fixed real $\alpha > 1$ is poorly understood; already for $\alpha = 3/2$ it is open whether $\|(3/2)^n\| \rightarrow 0$ can fail to hold densely. Mahler [Mah57] proved that for rational $\alpha > 1$ that is not an integer, and any $0 < \ell < 1$, the fractional part of α^n exceeds ℓ^n for all but finitely many n ; Corvaja and Zannier [CZ04] recast and extended this through the Schmidt subspace theorem. A version of Mahler's theorem with a rational multiplier, a special case of Ridout's theorem [Rid57] and of the Main Theorem of [CZ04], is one of the two Diophantine inputs below.

This paper concerns not the sizes $\|(3/2)^n\|$ themselves but the *symbolic dynamics* they generate. Writing $(3/2)^n = m_n + \varepsilon_n$ with $m_n \in \mathbb{Z}$ and $\varepsilon_n \in [-\frac{1}{2}, \frac{1}{2})$, the increments

$$t_n := 2m_{n+1} - 3m_n \in \{-2, -1, 0, 1, 2\}$$

record, at each step, which integer translate the multiplication-by- $3/2$ map selects. The word $T = (t_n)_{n \geq 0}$ is the full itinerary of that autonomous dynamics; we call it the *steering word*. It is the round-to-nearest analogue of the itineraries of the $\frac{3}{2}$ -dynamics studied through rational base number systems [AFS08]. Its *subword complexity* $p_T(k)$ is the number of distinct length- k factors it contains. Our main result is the following.

Institute for Globally Distributed Open Research and Education (IGDORE) .

Theorem 1.1 (Main theorem). *The subword complexity of the steering word T is superlinear: for every C there is a K with $p_T(k) > Ck$ for all $k \geq K$; equivalently $p_T(k)/k \rightarrow \infty$.*

The proof has the shape of a transfer of the Adamczewski–Bugeaud method [AB07] for complexity lower bounds of algebraic numbers to the present dynamical word, and it simplifies in one respect: because the $\frac{3}{2}$ -dynamics is autonomous, the bookkeeping is position-uniform, and every repeated factor of T compresses into a single clean Diophantine inequality, with no prefix-anchoring. Concretely the argument runs in three stages.

Stage 0 (Section 2) is an exact identity: a length- k factor of T occurring at two positions $a < c$ forces $3^k(\varepsilon_c - \varepsilon_a) = 2^k(\varepsilon_{c+k} - \varepsilon_{a+k})$ and the divisibility $2^k \mid m_c - m_a$. This already gives two unconditional theorems — T is not eventually periodic (Theorem 2.9), and $p_T(k) \geq (41/24)k - 3$ (Theorem 2.10), strictly above the universal Morse–Hedlund floor $k + 1$.

Stage 1 (Section 3) reduces Theorem 1.1 by pigeonhole to a single statement:

(K) for every $\theta < 1$, the inequality $\|(3/2)^c - (3/2)^a\| \leq \theta^c$ has only finitely many solutions $2 \leq a < c$.

This is *exponential pair repulsion* for the orbit — an improvement of the trivial repulsion floor $\|(3/2)^c - (3/2)^a\| \geq 2^{-c}$ from an exponential base 2^{-1} to an arbitrary base $\theta < 1$.

Stage 2 (Section 4) establishes (K). To a violator (a, c) we attach the integer point $(3^c, 2^{c-a}3^a, 2^cM)$, where M is the integer nearest to $(3/2)^c - (3/2)^a$, and apply the p -adic Subspace Theorem in three variables with the primes 2 and 3. The 2- and 3-adic smallness of the coordinates exactly pays for their size, and the theorem confines all violators to finitely many planes, whatever the gap $c - a$. On each plane elementary arithmetic finishes the job, except on planes of constant gap. There the problem has only one parameter left, and it falls to the two-variable case of the same theorem, Ridout’s theorem [Rid57].

The elementary consequences of Stage 0 and the Stage 1 reduction are unconditional. Stage 2 rests on the p -adic subspace theorem [Sch77, BG06], used in three variables and in two.

The whole argument is formalized in Lean 4 with Mathlib, in the directory `Stephan2026a` of the repository <https://github.com/rwst/Subspace-Theorems>. That repository also formalizes the subspace theorem with finite places, so the formal proof of Theorem 1.1 uses no axioms beyond those of Lean’s logic. The two formal statements of Theorem 1.1 are checked by the `comparator` tool against a short statement of record, which imports only Mathlib and defines only the steering word and the subword complexity.

1.1. Relation to the complexity landscape. The conclusion $p_T(k)/k \rightarrow \infty$ is precisely the shape of the Adamczewski–Bugeaud theorem [AB07] for

the b -ary expansions of irrational algebraic numbers, and the analogy extends to the gap that theorem leaves open: there Borel's conjecture predicts normality, that is $p(k) = b^k$, so superlinearity is the unconditional shadow of an expected exponential truth. The same is the case here, and it can be made precise.

Superlinearity already places T outside every classical low-complexity family. Automatic sequences satisfy $p(k) = O(k)$ [Cob72, AS03], as do linearly recurrent words [DHS99], Sturmian words and the codings of rotations and of interval exchanges, and fixed points of primitive substitutions; Theorem 1.1 excludes T from all of them. For the general theory of the factor complexity function we refer to [CN10], and to [Fer99, Puz25] for surveys.

What superlinearity does *not* decide is whether T is morphic, and here a rigidity theorem of Pansiot [Pan84] enters: the complexity of a pure morphic word is one of $\Theta(1)$, $\Theta(k)$, $\Theta(k \log \log k)$, $\Theta(k \log k)$ and $\Theta(k^2)$ — in particular it is polynomially bounded. Hilion and Levitt [HL25] have recently proved the analogous four-class theorem for right-infinite words representing attracting fixed points of an automorphism of a finitely generated free group. The following two statements turn this rigidity into a dichotomy for the steering word; both are consequences of Lemma 2.7, read backwards.

Proposition 1.2 (Packing bound). *Let $k \in \mathbb{N}$ and let $A \subset \mathbb{N}$ be a finite set of positions whose orbit points are pairwise more than $(2/3)^k$ apart, that is $|\varepsilon_c - \varepsilon_a| > (2/3)^k$ for all $a \neq c$ in A . Then $p_T(k) \geq |A|$. Equivalently $p_T(k) \geq N((2/3)^k)$, where $N(\delta)$ is the largest cardinality of a δ -separated subset of $\{\varepsilon_n\}$.*

Proof. By Lemma 2.7, two positions carrying the same length- k factor satisfy $|\varepsilon_c - \varepsilon_a| \leq (2/3)^k$. Hence the length- k factors at the positions of A are pairwise distinct. $\square$

Corollary 1.3 (Density forces exponential complexity). *If $((3/2)^n)_{n \geq 0}$ is dense modulo one, then $p_T(k) \geq \lceil (3/2)^k \rceil - 1$ for every k .*

Proof. Fix k and a positive integer $M < (3/2)^k$, so that $(2/3)^k < 1/M$. The M points $-\frac{1}{2} + j/M$, $0 \leq j < M$, lie in $[-\frac{1}{2}, \frac{1}{2})$ and are $1/M$ -separated. Approximating each of them within $\delta = \frac{1}{3}(1/M - (2/3)^k) > 0$ by a point of the orbit produces M positions whose ε -values are pairwise more than $1/M - 2\delta = \frac{1}{3} \cdot \frac{1}{M} + \frac{2}{3}(2/3)^k > (2/3)^k$ apart, so $p_T(k) \geq M$ by Proposition 1.2. Since $\lceil (3/2)^k \rceil - 1 < (3/2)^k$, the claim follows. $\square$

Two readings of Corollary 1.3 place the main theorem. First, it identifies Theorem 1.1 as the weakest nontrivial instance of what the density conjecture predicts: the expected truth is exponential, and the difficulty of the subject lies in the gap between $k \cdot \omega(1)$ and $(3/2)^k$. Second, read contrapositively, *any* subexponential upper bound $p_T(k) = o((3/2)^k)$ would refute density of $((3/2)^n)$ modulo one. This is why no upper bound of that

strength is offered here: beyond the trivial alphabet ceiling $p_T(k) \leq 5^k$, the sharp known environment is Kopra's trace subshift [Kop21], which gives $p_T(k) \leq 4 \cdot 3^k - 3 \cdot 2^k$ — of base 3 rather than $3/2$.

Combining Corollary 1.3 with Pansiot's theorem: if $((3/2)^n)$ is dense modulo one, then T is not morphic, and by [HL25] it does not represent an attracting fixed point of a free group automorphism either; equivalently, exhibiting a morphism that generates T would refute the density conjecture. It is Theorem 1.1 that makes the dichotomy non-vacuous, by eliminating the classes $\Theta(1)$ and $\Theta(k)$ outright.

The implication runs one way only. Proposition 1.2 bounds the complexity below by a packing number, not conversely: the letter t_n is a function of $(\varepsilon_n, m_n \bmod 2)$, but $m_{n+1} \bmod 2$ already requires $m_n \bmod 4$, so a length- k window is governed by $(\varepsilon_a, m_a \bmod 2^k)$ and distinct factors need not arise from separated orbit points. We claim no converse to Lemma 2.7, and Theorem 1.1 accordingly transfers nothing back to the density question. The complementary unconditional information is due to Flatto, Lagarias and Pollington [FLP95], whose theorem gives the set of limit points of $\{\xi(3/2)^n\}$ a diameter of at least $1/3$: the orbit closure is known to be spread out, but is not known to have positive box dimension — which by Proposition 1.2 would already force $p_T(k) \geq c^k$ for some $c > 1$.

Finally, on the shape of p_T as a function of k . The region bracketed by the bounds above is genuinely populated: Cassaigne [Cas03] constructs words whose complexity grows faster than every polynomial and slower than every exponential; and in the linear regime the pairs (α, β) realizable as $\alpha = \liminf_k p_T(k)/k$ and $\beta = \limsup_k p_T(k)/k$ constitute the Heine spectrum, recently shown by Erazo and Moreira [EM25] to have non-empty interior.

Setup and notation. Throughout, $\|\cdot\|$ is the distance to the nearest integer, computed exactly in $\mathbb{Q}$ for all the rationals that occur.

Definition 1.4 (The orbit decomposition and the steering word). For $n \in \mathbb{N}$ put

$$m_n := \text{round}((3/2)^n), \quad \varepsilon_n := (3/2)^n - m_n \in [-\tfrac{1}{2}, \tfrac{1}{2}), \quad R_n := 3^n - 2^n m_n,$$

so that $\varepsilon_n = R_n/2^n$; here round rounds half-integers up, placing ε_n in the half-open window $[-\frac{1}{2}, \frac{1}{2})$. The *steering letter* is $t_n := 2m_{n+1} - 3m_n$, the *steering word* is $T = (t_n)_{n \geq 0}$, and the *parity word* is $b = (b_n)$, $b_n := m_n \bmod 2$.

Elementary manipulations show $t_n = 3\varepsilon_n - 2\varepsilon_{n+1}$, whence $|t_n| \leq 2$ (a five-letter alphabet) and $t_n \equiv m_n \pmod{2}$, so that b is the letter-to-letter reduction of T modulo 2. Moreover R_n is odd for $n \geq 1$; this parity is the source of the trivial repulsion floor. The sequence begins $m = 1, 2, 2, 3, 5, 8, \dots$ and $t = 1, -2, 0, 1, 1, \dots$. All objects live in $\mathbb{Z}$ or $\mathbb{Q}$; no real analysis enters before Section 4.

Definition 1.5 (Circuit sum). For $a, k \in \mathbb{N}$ put

$$W(a, k) := \sum_{i < k} 3^{k-1-i} 2^i t_{a+i} \in \mathbb{Z}.$$

2. THE REPETITION IDENTITY

The whole analysis rests on one exact cancellation. Iterating the step relation $2m_{n+1} = 3m_n + t_n$ over a window telescopes the circuit sum into a closed form.

Lemma 2.1 (Closed form of the circuit sum). *For all $a, k \in \mathbb{N}$,*

$$W(a, k) = 3^k \varepsilon_a - 2^k \varepsilon_{a+k}, \quad \text{equivalently} \quad 2^k m_{a+k} = 3^k m_a + W(a, k).$$

Proof. Induction on k , using the recurrence $W(a, k+1) = 3W(a, k) + 2^k t_{a+k}$ and $t_n = 3\varepsilon_n - 2\varepsilon_{n+1}$; the main terms of the orbit cancel exactly. $\square$

Definition 2.2 (Repetition). A *repetition* of length k at positions $a < c$ is an equality of factors, $t_{a+i} = t_{c+i}$ for all $i < k$. (Occurrences may overlap; nothing anchors them to a prefix.)

Equal factors accumulate equal circuit sums, so subtracting two instances of the closed form gives the central identity.

Theorem 2.3 (The repetition identity). *If there is a repetition of length k at $a < c$, then*

$$3^k(\varepsilon_c - \varepsilon_a) = 2^k(\varepsilon_{c+k} - \varepsilon_{a+k}), \quad 3^k(m_c - m_a) = 2^k(m_{c+k} - m_{a+k}).$$

Proof. By Definition 2.2 the circuit sums agree, $W(a, k) = W(c, k)$. Substituting the two forms of Lemma 2.1 and cancelling yields both identities (the fractional form from the ε -closed form, the integer form from its integer companion). $\square$

Corollary 2.4 (Divisibilities). *Under the hypothesis of Theorem 2.3,*

$$2^k \mid m_c - m_a \quad \text{and} \quad 3^k \mid m_{c+k} - m_{a+k}.$$

Proof. 2^k divides $3^k(m_c - m_a)$ by the integer identity, and $\gcd(2^k, 3^k) = 1$; symmetrically for 3^k . $\square$

The divisibility upgrades to size once m is strictly increasing, which it is from index 2 on. This is the mechanism that forbids long repetitions late in the word.

Proposition 2.5 (Growth ceiling). *If there is a repetition of length k at $2 \leq a < c$, then*

$$2^{k+c+1} \leq 3^{c+1},$$

that is, $k \leq (c+1) \log_2(3/2) \approx 0.585(c+1)$.

Proof. Corollary 2.4 gives $2^k \mid m_c - m_a$, and $m_a < m_c$ (strict monotonicity of m for indices ≥ 2), so $2^k \leq m_c - m_a \leq m_c$. Combined with the a priori bound $2(2^c m_c) \leq 2 \cdot 3^c + 2^c$ — itself a restatement of $\varepsilon_c \geq -\frac{1}{2}$ — this yields $2^{k+c+1} \leq 3^{c+1}$ as an inequality of integers. $\square$

Remark 2.6. The same computation bounds a $(j+1)$ -fold repetition of a length- p block starting at $a \geq 2$ by $2^{jp+(a+p)+1} \leq 3^{a+p+1}$: long periodic windows cannot occur late in the word.

The identity has two further quantitative shadows, the interface to the Diophantine kernel of Section 3.

Lemma 2.7 (Contraction). *A repetition of length k at $a < c$ forces $|\varepsilon_c - \varepsilon_a| \leq (2/3)^k$.*

Proof. Rewrite the fractional identity of Theorem 2.3 as

$$\varepsilon_c - \varepsilon_a = (2/3)^k (\varepsilon_{c+k} - \varepsilon_{a+k})$$

and bound the difference of two centered fractional parts by 1. $\square$

Lemma 2.8 (Trivial repulsion floor). *For $1 \leq a < c$,*

$$2^c \|(3/2)^c - (3/2)^a\| \geq 1.$$

Proof. The orbit difference is

$$3^a(3^{c-a} - 2^{c-a})/2^c,$$

a fraction with odd numerator over 2^c ; hence its distance to the nearest integer is at least 2^{-c} . (Equivalently $2^c(\varepsilon_c - \varepsilon_a)$ is an odd integer.) $\square$

Lemma 2.8 is the estimate the kernel (K) will improve: from the exponential base 2^{-1} to an arbitrary base $\theta < 1$. Two unconditional theorems already drop out of Stage 0.

Theorem 2.9 (Aperiodicity). *The steering word T is not eventually periodic.*

Proof. If $t_{n+p} = t_n$ for all $n \geq N$ with $p \geq 1$, then at $a = \max(N, 2)$ and $c = a+p$ there is a repetition of *every* length k . Taking $k = 3^{c+1}$ contradicts the growth ceiling $2^{k+c+1} \leq 3^{c+1}$ of Proposition 2.5, since $3^{c+1} < 2^{3^{c+1}}$. (Aperiodicity of this family is also a special case of [DN05, Lemma 1].) $\square$

Theorem 2.10 (Complexity floor). *For all k ,*

$$41k \leq 24p_T(k) + 72, \quad \text{i.e.} \quad p_T(k) \geq \frac{41}{24}k - 3 \approx 1.708k.$$

Proof. By Proposition 2.5 and the certified estimate $3^{41} \leq 2^{65}$ (which encodes $\log_2 3 \leq 65/41$), a repetition of length k at $2 \leq a < c$ forces $41k \leq 24c+24$. Hence the windows starting at $2, 3, \dots, C$ with $C = \lfloor (41k-25)/24 \rfloor$ are pairwise distinct as factors, giving at least $C-1$ distinct length- k factors; comparing with the total count $p_T(k)$ and simplifying yields the stated inequality. $\square$

Remark 2.11. The constant intrinsic to the method is $1/\log_2(3/2) \approx 1.7095$; the rational slope $41/24 \approx 1.7083$ formalized here is the certified rounding of it obtained from the integer certificate $3^{41} \leq 2^{65}$. In either form the bound lies strictly above the Morse–Hedlund floor $k + 1$ valid for every aperiodic word [MH38]. A lower bound of exactly this shape, with the constant $1/\log_2(3/2)$, is due to Dubickas [Dub09] for the digit words of $[(p/q)^n]$; the present statement is its transfer to the nearest-integer steering word of the orbit of 1.

The parity word. The primary target is the five-letter word T , not its binary reduction b . Since b is a letter-to-letter image of T we have $p_T(k) \geq p_b(k)$, so Theorem 2.10 says nothing about p_b . What survives is a dichotomy.

Proposition 2.12 (Parity dichotomy). *A length- k repetition of the parity word b at positions a, c either lifts to a repetition of T (so Theorem 2.10 applies), or the two windows first disagree at some step $j < k$ at which the steering letters differ by a nonzero even amount.*

Proof. If the T -windows agree, we are in the first case. Otherwise let j be the least index of disagreement; there $b_{a+j} = b_{c+j}$ forces $t_{c+j} - t_{a+j}$ even, and it is nonzero by choice of j . $\square$

Remark 2.13. While the two orbit windows remain T -synchronized their ε -difference expands by exactly $3/2$ per step,

$$\varepsilon_{c+j} - \varepsilon_{a+j} = (3/2)^j (\varepsilon_c - \varepsilon_a).$$

Consequently a *maximal* desynchronization $|t_{c+j} - t_{a+j}| = 4$ after j synchronized steps certifies $|\varepsilon_c - \varepsilon_a| \geq (2/3)^{j+1}$, whereas a minimal one $|t_{c+j} - t_{a+j}| = 2$ carries no such constraint. This is the precise sense in which superlinearity for b is harder; we claim no unconditional linear bound for p_b .

3. REDUCTION TO A DIOPHANTINE KERNEL

Definition 3.1 (The kernel). For a rational scale $\theta \in (0, 1)$ let

$$V(\theta) := \{(a, c) : 2 \leq a < c, \|(3/2)^c - (3/2)^a\| \leq \theta^c\}.$$

We say *pair repulsion* holds at θ if $V(\theta)$ is finite, and we call the assertion

$$(K) \quad V(\theta) \text{ is finite for every rational } \theta \in (0, 1)$$

the *Diophantine kernel*. The target property is that p_T be *superlinear*: for every C there is a K with $p_T(k) > Ck$ for all $k \geq K$.

Restricting θ to the rationals loses nothing: $V(\theta)$ is monotone in θ and $\mathbb{Q}$ is dense, so the rational and real formulations of (K) are equivalent. This keeps the reduction inside $\mathbb{Q}$.

Theorem 3.2 (Reduction). *The kernel (K) implies that p_T is superlinear.*

Proof. Fix C and suppose $p_T(k) \leq Ck$ for some large k . Among the $Ck + 1$ windows starting at positions $2, \dots, Ck + 2$, pigeonhole produces two equal factors, i.e. a repetition of length k at some $2 \leq a < c \leq (C + 2)k$. Choose once and for all a rational scale $\theta < 1$ with $\theta^{C+2} \geq 2/3$ (a Bernoulli estimate provides one). By Lemma 2.7,

$$\|(3/2)^c - (3/2)^a\| \leq |\varepsilon_c - \varepsilon_a| \leq (2/3)^k \leq \theta^{(C+2)k} \leq \theta^c,$$

so $(a, c) \in V(\theta)$. As k grows, the growth ceiling (Proposition 2.5, in the form $41k \leq 24c + 24$) forces $c \rightarrow \infty$, contradicting the finiteness of $V(\theta)$ granted by (K). Hence no such C bounds $p_T(k)/k$. $\square$

The kernel is thus the sole remaining obstacle, and it is unconditional as a *reduction*: (K) enters Theorem 3.2 only as a hypothesis. The rest of the paper establishes it.

4. THE KERNEL VIA THE SUBSPACE THEOREM

The kernel (K) is a statement in two parameters, the position a and the gap $s = c - a$. We prove it by applying the Subspace Theorem once in three variables, which confines all violators to finitely many planes at once, whatever their gaps. Elementary arithmetic disposes of every plane except those on which the gap is constant. On such a plane only one parameter is left, and the two-variable case of the same theorem (Ridout's theorem) finishes the proof.

4.1. The Subspace Theorem. It helps to say informally what the theorem does before the formulas begin. Its ancestor is Roth's theorem: an algebraic irrational admits no rational approximations appreciably better than the trivial ones. Schmidt's subspace theorem is the higher-dimensional form of the same phenomenon. If several linear forms take simultaneously much smaller values at an integer point than the size of that point warrants, then the point is not free to lie anywhere: all such points fall into finitely many proper subspaces, that is, they satisfy one of finitely many fixed linear relations with fixed coefficients. Unusually good approximation remains possible, but only along degenerate families.

This bites here because the quantities we must control, $(3/2)^a$ and $(3/2)^c$, are built from the two primes 2 and 3 alone. They are huge numbers, of height growing exponentially in the exponent, and yet they are described by very few integer parameters. To say that a small integer combination of such numbers lies unexpectedly close to an integer is to say that a linear form is unexpectedly small at a large integer point. The S -arithmetic version of the theorem, due to Schlickewei [Sch77], measures smallness at the primes 2 and 3 as well as at the real place. This is what makes the sparse multiplicative structure visible to the machinery.

For a prime p let $|x|_p = p^{-v_p(x)}$ denote the p -adic absolute value on $\mathbb{Q}$, and write $|x|_\infty = |x|$. A nonzero integer n satisfies the product formula

$|n| \prod_p |n|_p = 1$, and hence

$$(1) \quad |n| \cdot |n|_2 \cdot |n|_3 \geq 1, \quad \text{with equality if and only if } n = \pm 2^i 3^j.$$

In words: measured at all three places together, an integer of the form $\pm 2^i 3^j$ (an S -unit, for $S = \{2, 3\}$) has size exactly 1, however large it is at the real place, and a general integer $n = 2^i 3^j n'$ with $\gcd(n', 6) = 1$ has size $|n'|$. This accounting principle is behind both applications below.

Theorem 4.1 (Subspace Theorem, p -adic form [Sch77]; see [BG06, Ch. 7]). *Let $n \geq 2$, let S be a finite set of primes, and for each $v \in S \cup \{\infty\}$ let $L_{v,1}, \dots, L_{v,n}$ be linearly independent linear forms in n variables with rational coefficients. For every $\epsilon > 0$ the solutions $x \in \mathbb{Z}^n \setminus \{0\}$ of*

$$\prod_{v \in S \cup \{\infty\}} \prod_{i=1}^n |L_{v,i}(x)|_v \leq \|x\|^{-\epsilon}, \quad \|x\| := \max_i |x_i|,$$

lie in finitely many proper linear subspaces of $\mathbb{Q}^n$.

Throughout, $S = \{2, 3\}$, and the forms at 2 and at 3 are the coordinate forms $X_1, \dots, X_n$; only the forms at ∞ vary. With this choice, by (1), a coordinate of x that is of the form $\pm 2^i 3^j$ contributes nothing to the double product as long as the coordinate form X_i is also one of the forms at ∞ . The product can therefore only be small because of the remaining forms at ∞ , which measure the near-coincidence one wants to rule out.

Two features of Theorem 4.1 shape everything that follows. First, it does not say that there are only finitely many solutions, and it could not: a one-parameter family of solutions can lie in a single subspace. Each of the finitely many subspaces must therefore be examined by hand. Second, it is ineffective: it neither bounds the solutions nor identifies the subspaces. All results of this section inherit this ineffectivity.

4.2. Constant gap: two variables.

Lemma 4.2 (Mahler's theorem with a rational multiplier). *Let $\xi \in \mathbb{Q}_{>0}$ and $0 < \theta < 1$. Then only finitely many $a \in \mathbb{N}$ satisfy*

$$0 < \|\xi(3/2)^a\| \leq \theta^a.$$

The meaning of the lemma is best seen by dividing through by $(3/2)^a$. If M is the integer nearest to $\xi(3/2)^a$, the inequality says $|\xi - M2^a/3^a| \leq \theta^a(2/3)^a$. This is an approximation to the fixed rational ξ by fractions whose denominator is a power of 3 and whose numerator is divisible by a large power of 2. Ridout's theorem [Rid57] says that the p -adic constraints on numerator and denominator are paid for in exactly the way (1) describes, so that such approximations cannot be appreciably better than trivial. The proof makes this precise.

Proof. Write $\xi = u/v$ with $u, v \in \mathbb{N}$, and let M be the integer nearest to $\xi(3/2)^a$; for a large, $M \geq 1$. Put

$$y = (y_1, y_2) := (u3^a, vM2^a) \in \mathbb{Z}^2, \quad N' := y_1 - y_2 = v2^a(\xi(3/2)^a - M),$$

so that $0 < |N'| \leq v2^a\theta^a$. Since $|N'| \leq u3^a$ for a large, we have $\|y\| \leq u3^a + |N'| \leq 2u3^a$. Take the forms $Y_1 - Y_2$ and Y_2 at ∞ . Since $|y_1|_2 \leq 1$, $|y_2|_2 \leq 2^{-a}$, $|y_1|_3 \leq 3^{-a}$, $|y_2|_3 \leq 1$ and $M \leq \xi(3/2)^a + 1$,

$$|N'| \cdot |y_2| \cdot |y_1|_2 |y_2|_2 \cdot |y_1|_3 |y_2|_3 \leq |N'| \cdot vM2^a \cdot 2^{-a} \cdot 3^{-a} \leq v^2(\xi + 1)\theta^a.$$

With $\epsilon := \log(1/\theta)/(2 \log 3)$ we have $\|y\|^{-\epsilon} \geq (2u)^{-\epsilon}\theta^{a/2}$, so the right-hand side is at most $\|y\|^{-\epsilon}$ for $a \geq a_0$. By Theorem 4.1 with $n = 2$, the points y lie on finitely many lines through the origin; none of them is the line $Y_1 = 0$, as $y_1 \neq 0$, so each has the form $Y_2 = \rho Y_1$. On such a line $N' = (1 - \rho)u3^a$. The value $\rho = 1$ is excluded by $N' \neq 0$, and for $\rho \neq 1$ the inequality $|1 - \rho|u3^a \leq v2^a\theta^a$ bounds a . $\square$

For $\xi = 1$ Lemma 4.2 is Mahler's theorem [Mah57]. It is also the simplest instance (group $\langle 2, 3 \rangle$, degree $d = 1$, integer multiplier $q = 1$) of the Main Theorem of Corvaja and Zannier [CZ04]. The formalization follows the proof above, through Theorem 4.1 with $n = 2$. We need the lemma only for one family of multipliers.

Corollary 4.3 (Gap-bounded slice). *For every $S \in \mathbb{N}$ and every rational $\theta \in (0, 1)$, the set of $(a, c) \in V(\theta)$ with $c \leq a + S$ is finite.*

Proof. Fix a gap $1 \leq s_0 \leq S$. For $c = a + s_0$,

$$(3/2)^c - (3/2)^a = \xi(3/2)^a, \quad \xi := (3/2)^{s_0} - 1 \in \mathbb{Q}_{>0}.$$

The distance of this number to $\mathbb{Z}$ is positive by Lemma 2.8, and it is at most $\theta^c \leq \theta^a$. Lemma 4.2 leaves finitely many a . Taking the union over the finitely many gaps proves the claim. $\square$

Remark 4.4. The same input shows that fixed-gap repetitions are short in a strong sense: for each $s_0 \geq 1$ and each rational slope $\mu > 0$, only finitely many pairs (a, k) satisfy $k \geq \mu a$ together with a repetition of length k at $(a, a + s_0)$.

4.3. Exponential pair repulsion: three variables.

Theorem 4.5 (Exponential pair repulsion; the kernel (K) holds). *For every rational $\theta \in (0, 1)$ the set $V(\theta)$ is finite: only finitely many pairs $2 \leq a < c$ satisfy*

$$\|(3/2)^c - (3/2)^a\| \leq \theta^c.$$

The idea is to treat the unknown integer nearest to $(3/2)^c - (3/2)^a$ as a third coordinate. The near-integer condition then becomes the smallness of a linear form at an integer point whose other coordinates are built from 2 and 3.

Proof. Let $(a, c) \in V(\theta)$, put $s := c - a \geq 1$, and let M be the integer nearest to $(3/2)^c - (3/2)^a$. Since $a \geq 2$, this number is at least $(3/2)^a \cdot \frac{1}{2} \geq \frac{9}{8}$, so $M \geq 1$. Put

$$x = (x_1, x_2, x_3) := (3^c, 2^s 3^a, 2^c M) \in \mathbb{Z}^3, \quad N := x_1 - x_2 - x_3 = 2^c((3/2)^c - (3/2)^a - M).$$

We collect four facts about x .

- N is odd, since $x_1 = 3^c$ is odd and x_2, x_3 are even; in particular $N \neq 0$. (This is Lemma 2.8.)
- $|N| = 2^c \|(3/2)^c - (3/2)^a\| \leq 2^c \theta^c$.
- $M \leq (3/2)^c - (3/2)^a + \frac{1}{2} \leq (3/2)^c$.
- $\|x\| = x_1 = 3^c$. Indeed $x_2 = 2^s 3^a < 3^s 3^a = 3^c$; and $x_3 = 2^c M \leq 3^c - 2^s 3^a + 2^{c-1} < 3^c$, because $2^s 3^a > 2^s 2^{a-1} = 2^{c-1}$.

Take the forms $X_1 - X_2 - X_3$, X_2 , X_3 at ∞ ; they are linearly independent. Using $|x_1|_2 = 1$, $|x_2|_2 = 2^{-s}$, $|x_3|_2 \leq 2^{-c}$, $|x_1|_3 = 3^{-c}$, $|x_2|_3 = 3^{-a}$ and $|x_3|_3 \leq 1$,

$$\prod_v \prod_i |L_{v,i}(x)|_v \leq \underbrace{|N| \cdot 2^s 3^a \cdot 2^c M}_{v=\infty} \cdot \underbrace{2^{-s} 2^{-c}}_{v=2} \cdot \underbrace{3^{-c} 3^{-a}}_{v=3} = |N| M 3^{-c} \leq \theta^c = \|x\|^{-\epsilon}, \quad \epsilon := \frac{\log(1/\theta)}{\log 3}$$

By Theorem 4.1 with $n = 3$, the points x lie in finitely many proper subspaces $\alpha X_1 + \beta X_2 + \gamma X_3 = 0$ of $\mathbb{Q}^3$. Since x determines (a, c) (from x_1 we read off c , and then a from x_2), it suffices to show that each of these subspaces contains only finitely many of the points. Fix one.

Case $\gamma = 0$. Then $\alpha 3^c = -\beta 2^s 3^a$ with $(\alpha, \beta) \neq (0, 0)$, so both are nonzero and $(3/2)^s = -\beta/\alpha$. The gap $s = s_0$ is therefore constant on this subspace, and Corollary 4.3 applies.

Case $\gamma \neq 0$. Normalise $\gamma = -1$. Then $2^c M = \alpha 3^c + \beta 2^s 3^a$, that is, $M = \alpha(3/2)^c + \beta(3/2)^a$. The nearest integer is thus a fixed rational combination of the two orbit points, and therefore so is the distance:

$$\|(3/2)^c - (3/2)^a\| = |(3/2)^c - (3/2)^a - M| = |\beta_1(3/2)^c + \beta_2(3/2)^a|, \quad \beta_1 := 1 - \alpha, \beta_2 := -1 - \beta.$$

This distance is at most θ^c , and it is positive because $N \neq 0$.

- If $\beta_1 = \beta_2 = 0$, the distance vanishes, which is impossible.
- If $\beta_1 = 0 \neq \beta_2$, the distance is $|\beta_2|(3/2)^a \geq |\beta_2|$, so $\theta^c \geq |\beta_2|$, which bounds c .
- If $\beta_1 \neq 0$, choose s_1 with $|\beta_1|(3/2)^{s_1} \geq 2|\beta_2|$. For gaps $s \geq s_1$,

$$|\beta_1(3/2)^c + \beta_2(3/2)^a| \geq (3/2)^a (|\beta_1|(3/2)^s - |\beta_2|) \geq \frac{1}{2} |\beta_1|(3/2)^c \geq \frac{1}{2} |\beta_1|,$$

so $\theta^c \geq \frac{1}{2} |\beta_1|$, which bounds c . The pairs with gap $s < s_1$ lie in a gap-bounded slice, finite by Corollary 4.3.

A bound on c leaves finitely many pairs $a < c$. So in every case the subspace contains only finitely many of the points x . $\square$

Remark 4.6 (What the proof does). Two steps in the proof carry the weight, and it may help to see why each works.

The collapse of the double product. Of the nine factors in the product, those belonging to the S -unit coordinates $x_1 = 3^c$ and $x_2 = 2^s 3^a$ cancel exactly, by (1). The coordinate $x_3 = 2^c M$ is an S -unit times the unknown integer M ; its 2-adic smallness pays for the factor 2^c , and only $M \leq (3/2)^c$ survives. What is left is $|N| M 3^{-c}$, with N the near-miss itself. Without the hypothesis on (a, c) one only knows $|N| \leq 2^{c-1}$, and the product is then at most $\frac{1}{2}$, which is no saving at all. The hypothesis $\|(3/2)^c - (3/2)^a\| \leq \theta^c$ brings it down to $\theta^c = \|x\|^{-\epsilon}$, a fixed power of the size of the point, and a power saving is exactly what Theorem 4.1 needs. The gap s enters only through $x_2 = 2^s 3^a$, whose archimedean factor 2^s is repaid by its 2-adic factor 2^{-s} . This is why a single application covers all gaps at once, and why no case distinction in the gap is needed until the subspaces are examined.

The exceptional subspaces. The constant-gap points $(3^c, 2^{s_0} 3^{c-s_0}, 2^c M)$ all lie on the plane $X_2 = (2/3)^{s_0} X_1$, whatever M is. Theorem 4.1 in three variables therefore cannot rule them out: they form a one-parameter family inside one of its subspaces. Inside that plane only a two-variable problem remains, and it is solved by Lemma 4.2. Every other subspace is harmless because the ratio of the two orbit points is $(3/2)^c / (3/2)^a = (3/2)^s$, which tends to infinity with the gap. A fixed combination $\beta_1 (3/2)^c + \beta_2 (3/2)^a$ can be small only if this ratio is close to $-\beta_2 / \beta_1$. Since the ratio cannot accumulate, a small combination forces the gap into a finite set, and there is nothing further to control.

Theorem 4.7 (Main theorem). *The subword complexity of the steering word is superlinear, $p_T(k)/k \rightarrow \infty$.*

Proof. Theorem 4.5 supplies the kernel (K), and the reduction (Theorem 3.2) turns it into superlinearity. $\square$

Remark 4.8 (Dependencies). The results of Sections 2 and 3 are unconditional. The results of this section, and hence the main theorem, depend on one deep input: the p -adic subspace theorem, Theorem 4.1 ([Sch77]; see [BG06, Ch. 7] for the form used). It is used twice, with $n = 3$ in Theorem 4.5 and with $n = 2$ in Lemma 4.2. In the formalization it is not an assumption: it is proved there over number fields and with finite places, and specialized to the form of Theorem 4.1. Every dependency is ineffective; no bound on the exceptional sets is obtained.

5. IMPLICATIONS AND FURTHER DIRECTIONS

The results above sit at the junction of three programmes and leave a different residue in each.

5.1. Diophantine approximation. Detached from its dynamical origin, the kernel of Theorem 4.5 reads

$$\#\{(a, c) : 2 \leq a < c, \|(3/2)^c - (3/2)^a\| \leq \theta^c\} < \infty \quad \text{for every } \theta < 1.$$

Mahler's theorem, in the form used by Zudilin [Zud07], says that $\|(3/2)^k\| \leq \theta^k$ has finitely many solutions for every $\theta < 1$; the kernel is exactly its two-parameter companion. The orbit of 1 under $x \mapsto \frac{3}{2}x$ does not merely repel the integers: its points repel *each other* modulo one at every exponential rate. Only the base $\theta = \frac{1}{2}$ of that statement is elementary (Lemma 2.8); everything above it is Subspace.

Effectivity. Every result of Section 4 is ineffective, so no explicit K can be extracted in Theorem 1.1. For the one-parameter problem the effective theory is in better shape, though it barely clears the trivial base: after Baker and Coates, and Beukers, Zudilin [Zud07] proved $\|(3/2)^k\| > 0.5803^k$ for all k beyond an effectively computable bound, against the trivial 2^{-k} . Nothing comparable is known for pairs. The reduction of Theorem 3.2 converts any effective slice of the kernel directly into an effective complexity bound, and the exchange rate is explicit.

Problem 5.1. Exhibit a $\theta \in (0, 1)$ and an effectively computable $A(\theta)$ such that every pair $(a, c) \in V(\theta)$ satisfies $c \leq A(\theta)$.

Such a θ yields, by the proof of Theorem 3.2, the effective bound $p_T(k) > Ck$ for all explicitly large k , with $C = \log(2/3)/\log \theta - 2$. The threshold to beat is therefore $\theta = (2/3)^{24/89} \approx 0.896$: below it the certified slope $41/24$ of Theorem 2.10 is already better, and at θ near Zudilin's 0.5803 the exchange rate returns nothing at all. The whole difficulty of Problem 5.1 lies in that gap.

From qualitative to quantitative. The unconditional bound $p(n)/n \rightarrow \infty$ of [AB07] was later sharpened by Bugeaud and Evertse [BE08], who replaced the Subspace Theorem by its quantitative form and obtained $p(n, \xi, b) \geq n(\log n)^{0.09}$ for infinitely many n . The corresponding upgrade here is not automatic, and the obstruction can be named. A quantitative Subspace Theorem bounds the number of *subspaces* carrying the large solutions, but not the coefficients of those subspaces, and not the number of solutions in each. In the proof of Theorem 4.5 each subspace is disposed of either by a floor $|\beta_1|/2$ or $|\beta_2|$ that depends on its coefficients, or by an appeal to Lemma 4.2 on a fixed gap, and both steps are qualitative.

Problem 5.2. Make the analysis of the exceptional subspaces in Theorem 4.5 quantitative — by controlling the coefficients β_1, β_2 and counting the solutions of Lemma 4.2 on each fixed gap — and deduce a bound of the shape $p_T(k) \geq k(\log k)^\delta$ for infinitely many k .

Generality. Only two features of $3/2$ were used: that the orbit points are built from the finitely many primes 2 and 3, and that the numerator of $(3/2)^n$ is odd — the source of every positivity step (Lemma 2.8, the oddness of N in Theorem 4.5). For coprime $p > q \geq 2$ the second becomes $N \equiv p^c \not\equiv 0 \pmod{q}$, and the proof of Theorem 4.5 goes through with S the set of primes dividing pq ; so does the proof of Lemma 4.2, which already

allows a rational multiplier. We have not written out the remaining details for general ξ , nor formalized any of this.

Problem 5.3. Prove superlinearity of the steering word of $\xi(p/q)^n$ for every $\xi \in \mathbb{Q}^\times$ and all coprime $p > q \geq 2$. The growth ceiling becomes $k \leq (c+1)\log_q(p/q)$, so the analogue of Theorem 2.10 carries the constant $\log q / \log(p/q)$.

Finally, the Subspace Theorem in more variables suggests *higher-order* repulsion: finiteness of near-integer coincidences among three or more orbit points. The analysis of the exceptional subspaces would then no longer be elementary, because two ratios such as $(3/2)^{c-b}$ and $(3/2)^{b-a}$ can stay bounded independently of each other. Exploiting such a statement would also require a combinatorial step the present argument lacks: the pigeonhole of Theorem 3.2 extracts coincidences two at a time.

5.2. Symbolic dynamics. The sharpest consequence for this field is the dichotomy of Section 1.1: if $((3/2)^n)$ is dense modulo one then T is not morphic, so a morphism generating T would refute the density conjecture. Theorem 1.1 is what makes this non-vacuous, having removed the classes $\Theta(1)$ and $\Theta(k)$ from Pansiot’s list; the remaining candidates are $\Theta(k \log \log k)$, $\Theta(k \log k)$ and $\Theta(k^2)$. A systematic search for a morphism generating the observed prefix of T is therefore a well-posed experiment with an asymmetric payoff: a hit would settle a hard number-theoretic conjecture in the negative, while failure accumulates evidence for density.

The asymptotics are best phrased through the topological entropy of the subshift generated by T ,

$$h(T) = \lim_{k \rightarrow \infty} \frac{\log p_T(k)}{k},$$

which exists by subadditivity. Our results say nothing about it: superlinearity is compatible with $h(T) = 0$. What is known is the bracket $h(T) \leq \log 3$ from Kopra’s trace subshift [Kop21], while Corollary 1.3 gives $h(T) \geq \log(3/2)$ under density. Contrapositively, *a proof that the steering subshift has zero entropy would refute the density of $((3/2)^n)$ modulo one* — a purely symbolic-dynamical route into the problem, and one on which the trivial ceiling $p_T(k) \leq 5^k$ is silent.

Problem 5.4. Decide whether $h(T) > 0$. More modestly, close the gap between the base 3 of Kopra’s ceiling and the base 3/2 that density would demand.

Two structural obstacles are worth recording for anyone who takes this up. First, Proposition 1.2 is one-way: a length- k factor is governed not by ε_a alone but by the pair $(\varepsilon_a, m_a \bmod 2^k)$, so the natural phase space is $[-\frac{1}{2}, \frac{1}{2}) \times \mathbb{Z}_2$ rather than the interval, and it is there that a converse — factors corresponding to separated points — would have to be sought. Establishing one would make the transfer between complexity and equidistribution

two-way, which it currently is not. Second, the parity word b resists the method for a reason isolated in the remark after Proposition 2.12: a minimal desynchronization $|t_{c+j} - t_{a+j}| = 2$ carries no constraint on $|\varepsilon_c - \varepsilon_a|$, so Lemma 2.7 has no b -analogue. Any unconditional superlinear bound for p_b needs a substitute — plausibly a statistical bound on how often minimal desynchronization can occur along a window.

6. ACKNOWLEDGEMENTS

A previous version of this paper closed the kernel (K) by a longer route: a gap dichotomy whose infinite branch used arguments from the paper [NKR25] of Nair, Kumar and Rout. We thank Yann Bugeaud for the simplification of the argument.

The author utilized Claude Code as an AI coding assistant to aid in the Lean 4 formalization of the proofs presented in this paper. The author directed and reviewed all generated code and takes full responsibility for the mathematical integrity and final content of the work.

APPENDIX A. NUMERICAL EXPLORATION

The steering word is exactly computable, and it repays a look. Everything in this appendix was produced from the division-free integer recurrence

$$m_{n+1} = \frac{3m_n + t_n}{2}, \quad R_{n+1} = 3R_n - 2^n t_n, \quad \varepsilon_n = \frac{R_n}{2^n},$$

in which t_n is the unique integer with $\varepsilon_{n+1} \in [-\frac{1}{2}, \frac{1}{2})$ and $t_n \equiv m_n \pmod{2}$; no floating-point arithmetic enters the generation of T . We use the first $N = 10^6$ letters. Write $p_T^{(N)}(k)$ for the number of distinct length- k factors of the prefix $t_0 t_1 \cdots t_{N-1}$. Two properties of this statistic govern how it may be read: it is a *lower* bound for $p_T(k)$, every factor of a prefix being a factor of T ; and it can never exceed $N - k + 1$, so it stops being informative as soon as $p_T(k)$ approaches N . Nothing in this appendix belongs to the formal development, and Conjecture A.2 is a conjecture, not a theorem.

Letters and pairs. Suppose ε_n equidistributes in $[-\frac{1}{2}, \frac{1}{2})$ and the parity $m_n \bmod 2$ behaves like an independent fair coin. Since t_n is one of the two integers in $(3\varepsilon_n - 1, 3\varepsilon_n + 1]$, selected between them by the parity constraint, the model predicts the letter 0 with probability $\frac{1}{3}$, the letters ± 1 with probability $\frac{1}{4}$ and the letters ± 2 with probability $\frac{1}{12}$. Measured over 10^6 letters:

letter	-2	-1	0	1	2
measured	0.08302	0.25017	0.33362	0.25028	0.08292
model	1/12	1/4	1/3	1/4	1/12

Each cell agrees to within $5 \cdot 10^{-4}$. The equidistribution the model assumes is visible directly in the orbit: a 50-bin histogram of ε_n , $n < 10^6$, has occupancies between 19 732 and 20 305 against an expected 20 000. Pairs of letters, by contrast, are constrained outright.

Proposition A.1. *For every n , $|t_n + t_{n+1}| \leq 2$. Consequently the six pairs $(\pm 2, \pm 2)$, $(\pm 2, \pm 1)$, $(\pm 1, \pm 2)$ with equal signs never occur, and $p_T(2) = 19$.*

Proof. Applying $t_n = 3\varepsilon_n - 2\varepsilon_{n+1}$ twice,

$$t_n + t_{n+1} = 3\varepsilon_n + \varepsilon_{n+1} - 2\varepsilon_{n+2}.$$

Since $-\frac{1}{2} \leq \varepsilon_j < \frac{1}{2}$ for every j , the right-hand side is strictly below $\frac{3}{2} + \frac{1}{2} + 1 = 3$ and strictly above $-\frac{3}{2} - \frac{1}{2} - 1 = -3$; being an integer, it lies in $[-2, 2]$. The excluded pairs are exactly the six elements of $\{-2, \dots, 2\}^2$ with $|t_n + t_{n+1}| \geq 3$, and $25 - 6 = 19$ is the measured value. $\square$

The rule is not the whole story. The subshift of finite type it defines has 75 words of length 3, whereas $p_T(3) = 65$, so constraints of range 3 are present as well; and it yields no better ceiling than Section 1.1 already has, its entropy being $\log 3.935 \dots$ against Kopra's $\log 3$.

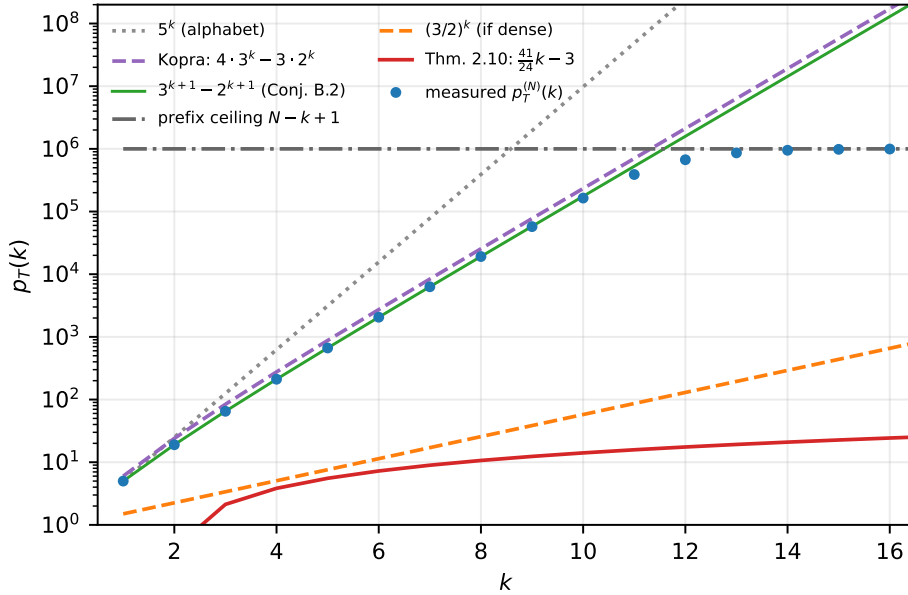


FIGURE 1. Measured $p_T^{(N)}(k)$ at $N = 10^6$ against the proven and conjectural environment. The data follow the curve $3^{k+1} - 2^{k+1}$ for as long as 10^6 letters suffice to exhibit every factor, fall away from it as sampling fails, and finally saturate at the prefix ceiling $N - k + 1$; both effects are artefacts of the finite sample. Note the distance to the certified floor of Theorem 2.10 and to the bound $(3/2)^k$ that density would force.

The complexity profile. The measured profile is, for $k \leq 9$,

k	1	2	3	4	5	6	7	8	9
$p_T^{(N)}(k)$	5	19	65	211	665	2059	6304	19138	57447
$3^{k+1} - 2^{k+1}$	5	19	65	211	665	2059	6305	19171	58025

Conjecture A.2. $p_T(k) = 3^{k+1} - 2^{k+1}$ for every $k \geq 1$.

Agreement is exact for $1 \leq k \leq 6$. At $k = 7$ and $k = 8$ the prefix falls short by 1 and by 33, which is the expected effect of rarity rather than evidence against the formula: the letter frequencies are far from uniform, so a factor built from extreme letters has probability about 12^{-k} per position — at $k = 7$, some three hundredths of an occurrence in 10^6 positions. Figure 2 exhibits the effect: for each N the ratio $p_T^{(N)}(k)/(3^{k+1} - 2^{k+1})$ sits at exactly 1 up to a threshold that moves right as N grows, and collapses beyond it.

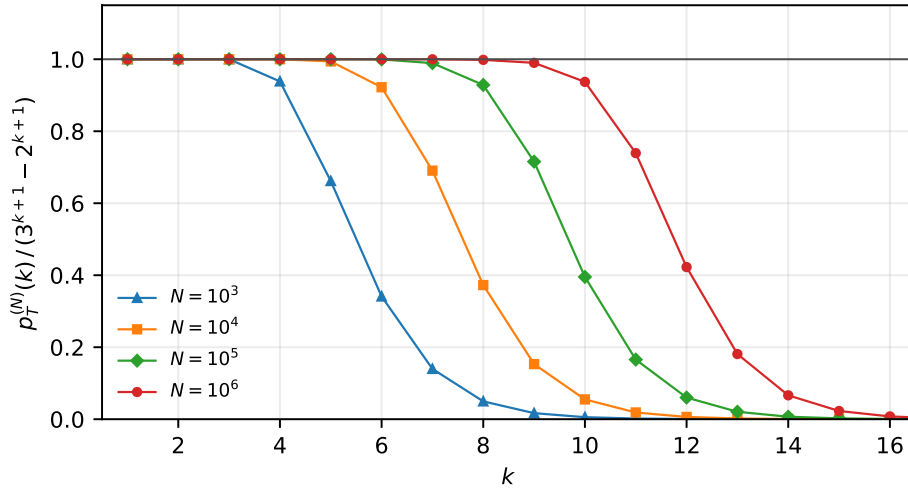


FIGURE 2. $p_T^{(N)}(k)$ normalized by the conjectured value, for four prefix lengths. The plateau at 1 lengthens with N ; the collapse to its right is under-sampling, not a failure of the formula.

Three things would follow from Conjecture A.2. The topological entropy of the steering subshift would be exactly $\log 3$, so the base of Kopra's ceiling is sharp and only its constant (4 against 3) is not; this would settle Problem 5.4 in both its parts, with $h(T) = \log 3 > 0$. The complexity would exceed by the exponential factor 2^k the bound $\lceil (3/2)^k \rceil - 1$ that density forces through Corollary 1.3; the $(3/2)^k$ there counts the ε -coordinate alone, and the missing 2^k is exactly the parity coordinate $m_a \bmod 2^k$, their product 3^k being one more reason to regard $[-\frac{1}{2}, \frac{1}{2}) \times \mathbb{Z}_2$ as the right phase space (Section 5). And the distance to what is proved is stark: at $k = 8$ the certified floor of Theorem 2.10 reads 10.7 against a measured 19 138.

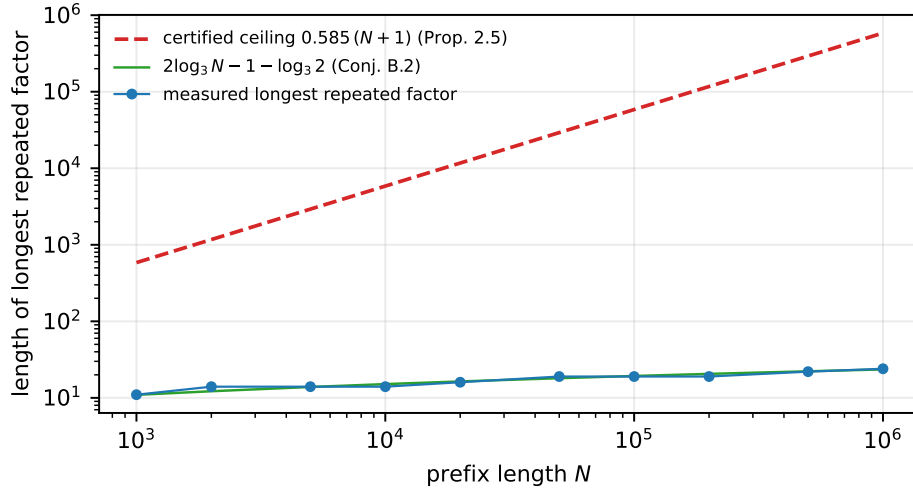


FIGURE 3. Longest repeated factor in a prefix of length N , against the ceiling certified by Proposition 2.5 and against the birthday prediction of Conjecture A.2. The certified ceiling is linear in N ; the repetitions actually present die out logarithmically.

Repetitions. Proposition 2.5 bounds repetition lengths linearly in the position: inside a prefix of length N it permits repeated factors of length up to $0.585(N+1)$, that is 585 000 at $N = 10^6$. The longest repeated factor actually present there has length 24. What the measurements track instead is

$$2 \log_3 N - \log_3 2 - 1,$$

the length at which, for a word of complexity 3^{k+1} , the expected number of coinciding pairs among N windows falls below one; the fit is within one letter over three decades. This tests Conjecture A.2 at $k \approx 24$, far outside the range in which $p_T(k)$ can be counted directly — though it tests the exponential base only, since replacing 3^{k+1} by Kopra’s $4 \cdot 3^k$ would move the prediction by $\log_3(4/3) \approx 0.26$, well below the resolution of the data.

The four orders of magnitude between the two curves of Figure 3 are the visual form of the gap that Sections 3–4 do not close. The mechanism of Section 2 forbids only very long repetitions, whereas in the word itself repetitions die out logarithmically; an argument that saw the true decay rate would deliver an exponential complexity bound, and the Diophantine input used here sees only the linear ceiling.

REFERENCES

- [AB07] B. Adamczewski and Y. Bugeaud, *On the complexity of algebraic numbers I. Expansions in integer bases*, Ann. of Math. (2) **165** (2007), no. 2, 547–565.

- [AS03] J.-P. Allouche and J. Shallit, *Automatic sequences: theory, applications, generalizations*, Cambridge University Press, Cambridge, 2003.
- [BE08] Y. Bugeaud and J.-H. Evertse, *On two notions of complexity of algebraic numbers*, Acta Arith. **133** (2008), no. 3, 221–250.
- [BG06] E. Bombieri and W. Gubler, *Heights in Diophantine geometry*, New Mathematical Monographs **4**, Cambridge University Press, Cambridge, 2006.
- [Cas03] J. Cassaigne, *Constructing infinite words of intermediate complexity*, Developments in Language Theory (DLT 2002), Lecture Notes in Comput. Sci. **2450**, Springer, Berlin, 2003, 173–184.
- [CN10] J. Cassaigne and F. Nicolas, *Factor complexity*, in: V. Berthé and M. Rigo (eds.), Combinatorics, automata and number theory, Encyclopedia Math. Appl. **135**, Cambridge University Press, Cambridge, 2010, 163–247.
- [Cob72] A. Cobham, *Uniform tag sequences*, Math. Systems Theory **6** (1972), 164–192.
- [CZ04] P. Corvaja and U. Zannier, *On the rational approximations to the powers of an algebraic number: solution of two problems of Mahler and Mendès France*, Acta Math. **193** (2004), no. 2, 175–191. arXiv:math/0403522.
- [Dub09] A. Dubickas, *On integer sequences generated by linear maps*, Glasg. Math. J. **51** (2009), no. 2, 243–252.
- [DN05] A. Dubickas and A. Novikas, *Integer parts of powers of rational numbers*, Math. Z. **251** (2005), no. 3, 635–648.
- [AFS08] S. Akiyama, C. Frougny, and J. Sakarovitch, *Powers of rationals modulo 1 and rational base number systems*, Israel J. Math. **168** (2008), 53–91.
- [DHS99] F. Durand, B. Host, and C. Skau, *Substitution dynamical systems, Bratteli diagrams and dimension groups*, Ergodic Theory Dynam. Systems **19** (1999), no. 4, 953–993.
- [EM25] H. Erazo and C. G. Moreira, *The Heins spectrum has non-empty interior*, Combinatorics on Words (WORDS 2025), Lecture Notes in Comput. Sci., vol 15729. Springer, Cham, 2025.
- [Fer99] S. Ferenczi, *Complexity of sequences and dynamical systems*, Discrete Math. **206** (1999), no. 1–3, 145–154.
- [FLP95] L. Flatto, J. C. Lagarias, and A. D. Pollington, *On the range of fractional parts $\{\xi(p/q)^n\}$* , Acta Arith. **70** (1995), no. 2, 125–147.
- [HL25] A. Hilion and G. Levitt, *A Pansiot-type subword complexity theorem for automorphisms of free groups*, Israel J. Math. **270**, 59–93 (2025). arXiv:2208.00676.
- [Kop21] J. Kopra, *On the trace subshifts of fractional multiplication automata*, Theoret. Comput. Sci. **851** (2021), 92–110.
- [Mah57] K. Mahler, *On the fractional parts of the powers of a rational number II*, Mathematika **4** (1957), 122–124.
- [MH38] M. Morse and G. A. Hedlund, *Symbolic dynamics*, Amer. J. Math. **60** (1938), no. 4, 815–866.
- [Pan84] J.-J. Pansiot, *Complexité des facteurs des mots infinis engendrés par morphismes itérés*, Automata, Languages and Programming (ICALP 1984), Lecture Notes in Comput. Sci. **172**, Springer, Berlin, 1984, 380–389.
- [Puz25] S. Puzynina, *Complexity of infinite words*, Machines, Computations, and Universality (MCU 2024), Lecture Notes in Comput. Sci. **15270**, Springer, Cham, 2025, 1–16.
- [NKR25] P. S. Nair, V. Kumar, and S. S. Rout, *Algebraic approximations to linear combinations of S -units*, arXiv:2506.02898 (v3, 18 Nov 2025).
- [Rid57] D. Ridout, *Rational approximations to algebraic numbers*, Mathematika **4** (1957), 125–131.
- [Sch77] H. P. Schlickewei, *The p -adic Thue–Siegel–Roth–Schmidt theorem*, Arch. Math. (Basel) **29** (1977), no. 3, 267–270.

- [Zud07] W. Zudilin, *A new lower bound for $\|(3/2)^k\|$* , J. Théor. Nombres Bordeaux **19** (2007), no. 1, 311–323.