Leading Digits, Factor Complexity, and Flows on Homogeneous Spaces: A Dynamical Approach

Mehdi Golafshan
Julien Cassaigne
Ivan Mitrofanov
Margarita Shevtsova

University of Liège

Uniform Distribution of Sequences-ESI
April, 24

A Snapshot of its Historical Milestones

How much faster the first pages [of logarithmic tables] wear out than the last ones (1881)

Simon Newcomb: American astronomer

He deduced the logarithmic probabilities shown in the first two row for the first and second digits [111]

d	0	1	2	3	4	5	6	7	8	9
$\mathbb{P}(D_1 = d)$	0	30.10	17.60	12.49	9.69	7.91	6.69	5.79	5.11	4.57
$\mathbb{P}(D_2 = d)$	11.96	11.38	10.88	10.43	10.03	9.66	9.33	9.03	8.75	8.49
$\mathbb{P}(D_3 = d)$	10.17	10.13	10.09	10.05	10.01	9.97	9.94	9.90	9.86	9.82
$\mathbb{P}(D_4 = d)$	10.01	10.01	10.00	10.00	10.00	9.99	9.99	9.99	9.98	9.98

Probabilities (in percent) of the frst four signifcant decimal digits

Fifty-seven Years Later

American physicist (1937)

PERCENTAGE OF TIMES THE NATURAL NUMBERS 1 TO 9 ARE USED AS FIRST DIGITS IN NUMBERS, AS DETERMINED BY 20,229 OBSERVATIONS

Group	Title	First Digit									Count
		1	2	3	4	5	6	7	8	9	
A	Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
B	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
C	Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
D	Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
E	Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
F	Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
G	H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
H	Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
I	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
J	Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
K	$n^{-1}, \sqrt{n}, \cdots$	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
L	Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
M	*Digest*	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
N	Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
O	X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
P	Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Q	Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
R	Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
S	$n^1, n^2 \cdots n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
T	Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
	Average.......	30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
	Probable Error	±0.8	±0.4	±0.4	±0.3	±0.2	±0.2	±0.2	±0.2	±0.3	—

American Philosophical Society

Frank Benford

WHY Benford's Law?

Benford's Law for arithmetic sequences.

(I) Sequences such as the squares that grow at linear or polynomial rate.

(II) Sequences such as (2^n), (2^{n^2}), or (n^n) that grow at faster than polynomial rate, but whose logarithms grow at polynomial rate.

(III) Sequences whose logarithms grow at faster than polynomial rate; e.g, (2^{2^n})

Main Definition

- Let $x \in \mathbb{R}^+$; and $m \in \mathbb{N}_{\geqslant 2}$.

Definition (1)

The **most significant digit** of x, is the unique $j \in [\![1, 9]\!]$ satisfying

$$10^k j \leqslant |x| < 10^k (j+1), \quad \exists!\ k \in \mathbb{Z}.$$

- $D_1(x) := j$

Definition (2)

The m^{th} **significant digit** of x, is defined inductively as the unique $[\![0, 9]\!]$ such that

$$10^k \left(\sum_{i=1}^{m-1} D_i(x)\, 10^{\,m-i} + j \right) \leqslant |x| < 10^k \left(\sum_{i=1}^{m-1} D_i(x)\, 10^{\,m-i} + j + 1 \right), \quad \exists!\ k \in \mathbb{Z}$$

- $D_m(x) := j$

Observation

(2^n)	2481361251 2481361251 2481361251 2481361251 2481371251
(2^{n^2})	2156365121 2271519342 5412132118 1169511474 1146399353
$(n!)$	1262175433 3468123612 5126141382 8282131528 3162152163
$(p(n))$	1235711234 5711122346 7111123345 6811122333 4567811112

From Benford back to equidistribution

World of numbers X	World of logs $Y = \{\log_{10} X\}$
Leading-digit pattern	Interval length in $[0, 1)$
Benford probabilities	Uniform (equidistributed) density
Base change / Scaling $X \mapsto cX$	Shift $Y \mapsto Y + \log_{10} c \pmod 1$

Homogeneous–space point of view

A homogeneous space is a quotient G/Γ of a Lie group G by a lattice Γ.

- flows are given by one-parameter subgroups of G.
- follow the point Γ under such a flow and write down in which piece of a fixed finite partition it lands $\implies$ an infinite word in a finite alphabet

Arithmetic sequence	Log map	Orbit lives on	Flow
a^n	$n \log_b a \bmod 1$	$\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$	an irrational rotation
a^{n^2}, $n!$, $\Gamma(cn), \ldots$	polynomial in n	higher-dimensional torus or nil-manifold	polynomial or diagonal
general $a^{P(n)}$ (P polynomial)	polynomial in n	nilmanifold G/Γ	unipotent or diagonal

Factor Complexity

Morse and Hedlund (1938)

- under the name block growth
- subword complexity (1975)

Definition (4)

The **factor complexity** of $\mathbf{w}$ is the map $p_{\mathbf{w}} : \mathbb{N} \to \mathbb{N}$ defined as follows:

$$n \mapsto \#\mathrm{Fac}_n(\mathbf{w})$$

Outline of This Talk

- F.C. is stand for factor complexity.

F.C. of sequences of the most significant digits of the decimal expansion of a^n.

Outline of This Talk

- F.C. is stand for factor complexity.

F.C. of sequences of the most significant digits of the decimal expansion of a^n.

F.C. of sequences of the most significant digits of the b-expansion of a^n.

Outline of This Talk

- F.C. is stand for factor complexity.

F.C. of sequences of the most significant digits of the decimal expansion of a^n.

F.C. of sequences of the most significant digits of the b-expansion of a^n.

Beyond of F.C. of sequences of the most significant digits of the b-expansion of a^n.

Outline of This Talk

- F.C. is stand for factor complexity.

F.C. of sequences of the most significant digits of the decimal expansion of a^n.

F.C. of sequences of the most significant digits of the b-expansion of a^n.

Beyond of F.C. of sequences of the most significant digits of the b-expansion of a^n.

F.C. of sequences of the most significant digits of the b-expansion of a^{n^d}

Outline of This Talk

- F.C. is stand for factor complexity.

F.C. of sequences of the most significant digits of the decimal expansion of a^n.

F.C. of sequences of the most significant digits of the b-expansion of a^n.

Beyond of F.C. of sequences of the most significant digits of the b-expansion of a^n.

F.C. of sequences of the most significant digits of the b-expansion of a^{n^d}

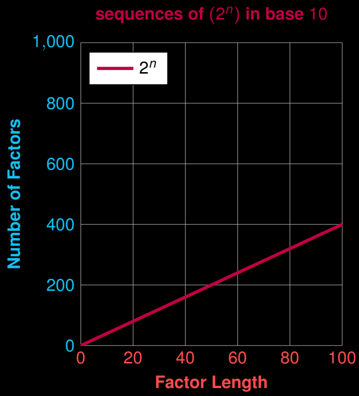
F.C. of sequences of the most significant digits of the decimal expansion of $n!$.

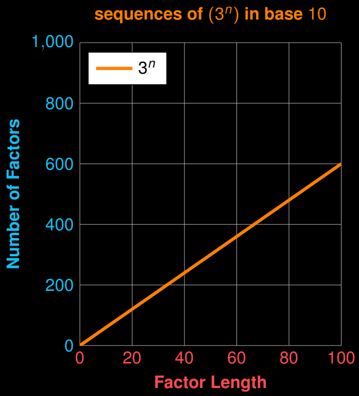
The Most Significant Digits of First 50 Terms (Concatenated)

Sequence	The most significant digits of first 50 terms (concatenated)
(2^n)	2481361251 2481361251 2481361251 2481361251 2481371251
(3^n)	3928272615 1514141313 1392827262 6151514141 3139282727
(4^n)	4162141621 4162141621 4172141731 4172141731 4173141731
(5^n)	5216317319 4216317319 4215217319 4215217319 4215217318
(6^n)	6321742116 3217421163 2174211632 1742116321 8421163218
(7^n)	7432118542 1196432117 5321196432 1175321196 4321175321
(8^n)	8654322111 8654322111 9754322111 9765432211 1865432211
(9^n)	9876554433 3222211111 1987765544 3332222111 1119877655

Table: Leading digits (in base 10) of the first 50 terms of the sequences (a^n), $a \in [\![2, 9]\!]$

Empirical Factor Complexities Based on the First $100,000$ Terms





Admissible Pairs

Definition

A pair (a, b) is called admissible if

i) a is a positive rational number;.

ii) b is a squarefree integer $\geqslant 5$;

iii) a and b are multiplicative independent: $\log_b(a) \notin \mathbb{Q}$.

- A pair (a, b) is called strong admissible if

I) a is a positive real number;.

II) b is a squarefree integer $\geqslant 5$;

III) a and b are multiplicative independent: $\log_b(a) \notin \mathbb{Q}$.

- A pair (a, b) is called weak admissible if

A) a is a positive real number;.

B) $b \geqslant 5$;

C) a and b are multiplicative independent: $\log_b(a) \notin \mathbb{Q}$.

Complexity of Leading Digit Sequences

Notation

$p_{a,b}(n)$: factor complexity of the most significant digits of the sequence (a^n) in base b.

	(2^n)	(3^n)	(4^n)	(5^n)	(6^n)	(7^n)	(8^n)	(9^n)
b = 5	**2n + 2**	**3n + 1**	**3n + 1**		**4n**	**4n**	**4n**	**3n + 2**
b = 6	**2n + 3**	**2n + 3**	**3n + 2**	**4n + 1**		**5n**	**4n + 1**	**3n + 2**
b = 7	**3n + 3**	**4n + 2**	**5n + 1**	**5n + 1**	**5n + 1**		**6n**	**6n**
b = 10	**4n + 5**	**6n + 3**	**6n + 3**	**4n + 5**	**7n + 2**	**8n + 1**	**7n + 2**	**8n + 1**

Table: Formulas for sequences in different bases

Complexity of Leading Digit Sequences

Notation

$\mathtt{p}_{a,b}(n)$: factor complexity of the most significant digits of the sequence (a^n) in base b.

	(2^n)	(3^n)	(4^n)	(5^n)	(6^n)	(7^n)	(8^n)	(9^n)
$b = 5$	$2n + 2$	$3n + 1$	$3n + 1$		$4n$	$4n$	$4n$	$3n + 2$
$b = 6$	$2n + 3$	$2n + 3$	$3n + 2$	$4n + 1$		$5n$	$4n + 1$	$3n + 2$
$b = 7$	$3n + 3$	$4n + 2$	$5n + 1$	$5n + 1$	$5n + 1$		$6n$	$6n$
$b = 10$	$4n + 5$	$6n + 3$	$6n + 3$	$4n + 5$	$7n + 2$	$8n + 1$	$7n + 2$	$8n + 1$

Table: Formulas for sequences in different bases

Theorem 1. (He & Hildebrand & ... 2020)

- Let (a, b) be an admissible pair. Then $\mathtt{p}_{a,b}(n)$ is an affine function for $n \geqslant 1$.

Complexity of Leading Digit Sequences

Notation

$\mathrm{p}_{a,b}(n)$**:** factor complexity of the most significant digits of the sequence (a^n) in base b.

	(2^n)	(3^n)	(4^n)	(5^n)	(6^n)	(7^n)	(8^n)	(9^n)
$\mathbf{b = 5}$	$\mathbf{2n + 2}$	$\mathbf{3n + 1}$	$\mathbf{3n + 1}$		$\mathbf{4n}$	$\mathbf{4n}$	$\mathbf{4n}$	$\mathbf{3n + 2}$
$\mathbf{b = 6}$	$\mathbf{2n + 3}$	$\mathbf{2n + 3}$	$\mathbf{3n + 2}$	$\mathbf{4n + 1}$		$\mathbf{5n}$	$\mathbf{4n + 1}$	$\mathbf{3n + 2}$
$\mathbf{b = 7}$	$\mathbf{3n + 3}$	$\mathbf{4n + 2}$	$\mathbf{5n + 1}$	$\mathbf{5n + 1}$	$\mathbf{5n + 1}$		$\mathbf{6n}$	$\mathbf{6n}$
$\mathbf{b = 10}$	$\mathbf{4n + 5}$	$\mathbf{6n + 3}$	$\mathbf{6n + 3}$	$\mathbf{4n + 5}$	$\mathbf{7n + 2}$	$\mathbf{8n + 1}$	$\mathbf{7n + 2}$	$\mathbf{8n + 1}$

Table: Formulas for sequences in different bases

Theorem 1. (He & Hildebrand & ... 2020)

- Let (a, b) be an admissible pair. Then $\mathrm{p}_{a,b}(n)$ is an affine function for $n \geqslant 1$.

Theorem 2. (P. Alessandri. PhD thesis, 1996)

- A coding of an irrational rotation, the complexity has the form $\mathrm{p}(n) = cn + d$, for n large enough.

Admissibility Requirement

Admissibility Requirement

Multiplicative independent: $\log_b(a) \notin \mathbb{Q}$

$\log_b a \in \mathbb{Q} \implies$ the sequence of the most significant digits: periodic $\implies \mathtt{p}_{a,b}(n)$: *bounded*

Admissibility Requirement

Multiplicative independent: $\log_b(a) \notin \mathbb{Q}$

$\log_b a \in \mathbb{Q} \implies$ the sequence of the most significant digits: periodic $\implies \mathtt{p}_{a,b}(n)$: *bounded*

Lower bound of base: $b \geqslant 5$

$b = 3 \implies$ interval $I = [\log_b 1, \log_b 2) \implies \ell(I) > \frac{1}{2}$

• the intersection of the interval $[0, 2/3]$ with its translate by $1/2$ consists of the two disjoint intervals $[0, 1/6]$ and $[1/2, 2/3]$.

Admissibility Requirement

Multiplicative independent: $\log_b(a) \notin \mathbb{Q}$

$\log_b a \in \mathbb{Q} \implies$ the sequence of the most significant digits: periodic $\implies \mathtt{p}_{a,b}(n)$: *bounded*

Lower bound of base: $b \geqslant 5$

$b = 3 \implies$ interval $I = [\log_b 1, \log_b 2) \implies \ell(I) > \frac{1}{2}$

• the intersection of the interval $[0, 2/3]$ with its translate by $1/2$ consists of the two disjoint intervals $[0, 1/6]$ and $[1/2, 2/3]$.

Squarefree bases

$b \geqslant 5$: non-squarefree integer, q: prime, $q^2 \mid b$, $a = q$

Case I: $b \neq q^n \implies \mathtt{p}_{a,b}(n)$: *NOT affine*

Case II: $b = q^n \implies$ *periodic* $\implies \mathtt{p}_{a,b}(n)$: *bounded*

Transcendental or Algebraic

Transcendental or Algebraic

Question: The following number made by the most significant digits of 2^n should be transcendental?

$$A := 0.124813612512\cdots$$

Answer:

Transcendental or Algebraic

Question: The following number made by the most significant digits of 2^n should be transcendental?

$$A := 0.124813612512\cdots$$

Answer:

$$\boxed{\textit{YES}!}$$

Transcendental or Algebraic

Question: The following number made by the most significant digits of 2^n should be transcendental?

$$A := 0.124813612512\cdots$$

Answer:

$$\boxed{YES!}$$

Theorem 3. (Adamczewski & Bugeaud. 2004)

- Let $b \geqslant 2$ be an integer. The factor complexity of the b-ary expansion of every irrational algebraic number satisfies

$$\liminf_{n\to\infty} \frac{\mathrm{p}(n)}{n} = \infty$$

Transcendental or Algebraic

Question: The following number made by the most significant digits of 2^n should be transcendental?

$$A := 0.124813612512\cdots$$

Answer:

$$\boxed{\mathit{YES}!}$$

Theorem 3. (Adamczewski & Bugeaud. 2004)

- Let $b \geqslant 2$ be an integer. The factor complexity of the b-ary expansion of every irrational algebraic number satisfies

$$\liminf_{n\to\infty} \frac{\mathrm{p}(n)}{n} = \infty$$

Example

Let $a = 2$ and $b = 10$. Then $\mathrm{p}_{2,10}(n) = 4n + 5$. Hence, A is a transcendental number.

Leading Digit Complexity Function

Leading Digit Complexity Function

Theorem 4. (Cassaigne. 1997)

- Any function of the form $cn + d$, where c and d are positive integers, is the complexity function of some word **w** for all $n \in \mathbb{N}$.

Leading Digit Complexity Function

Theorem 4. (Cassaigne. 1997)

- Any function of the form $cn + d$, where c and d are positive integers, is the complexity function of some word $\mathbf{w}$ for all $n \in \mathbb{N}$.

Definition (5)

(i) A function $cn + d$ is called a **leading digit complexity function** if there exists an admissible pair (a, b) such that

$$cn + d \;=\; \mathrm{p}_{a,b}(n); \quad \forall n.$$

(ii) A pair (c, d) of integers is called **good** if it is the pair of coefficients of a leading digit complexity function $cn + d$.

Good Pairs

Good Pairs

Notation

$$G := \{(c,d) \mid cn + d \text{ is a leading digit complexity function}\},$$

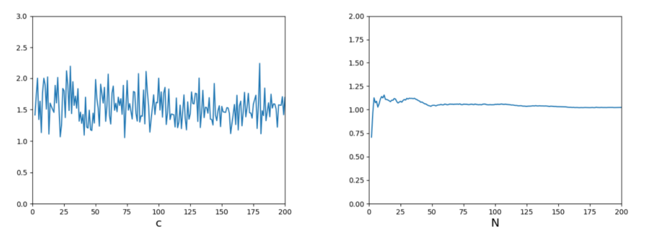
$$G(c) := \{d \mid (c,d) \in G\}.$$

Good Pairs

Notation

$$G := \{(c,d) \mid cn+d \text{ is a leading digit complexity function}\},$$

$$G(c) := \{d \mid (c,d) \in G\}.$$

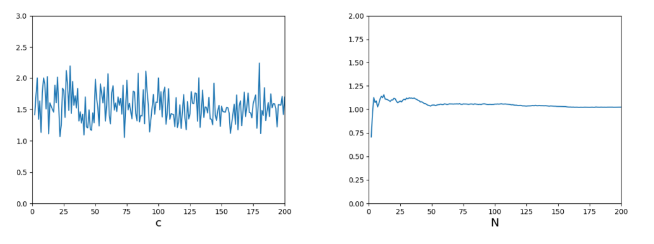


$$F_1(c) := \frac{\#G(c)}{\sqrt{c}} \qquad F_2(N) := \left(\sum_{c \leqslant N} (\#G(c)\right) N^{\frac{-3}{2}}$$

Good Pairs

Notation

$$G := \{(c,d) \mid cn+d \text{ is a leading digit complexity function}\},$$
$$G(c) := \{d \mid (c,d) \in G\}.$$



$$F_1(c) := \frac{\#G(c)}{\sqrt{c}} \qquad F_2(N) := \left(\sum_{c \leqslant N} (\#G(c)\right) N^{\frac{-3}{2}}$$

- bounded above and below by positive constants
- $F_1(c)$: not converge to a limit
- $F_2(N)$: appears to converge to a limit

Number of Good Pairs

Number of Good Pairs

Theorem 5. (M. G. 2025$^+$)

1. There exist positive constants k_1 and k_2 such that
$$k_1\sqrt{c} \leqslant \#G(c) \leqslant k_2\sqrt{c}$$
for all sufficiently large c, but the limit
$$\lim_{c\to\infty} \frac{\#G(c)}{\sqrt{c}} := \textbf{NOT EXIST}$$
2. There exists a positive constant k such that
$$\sum_{c\leq N} \#G(c) \sim k\,N^{3/2} \quad (N\to\infty).$$

Arithmetic Complexity

Arithmetic Complexity

- Let $\mathbf{w} \in \mathcal{A}^{\mathbb{N}}$. An *n-arithmetic factor* in $\mathbf{w}$ is any subwords of the form

$$w_k \; w_{k+d} \; w_{k+2d} \; \cdots , \; w_{k+(n-1)d}, \quad k \geqslant 0, \; d \geqslant 1.$$

Arithmetic Complexity

- Let $\mathbf{w} \in \mathcal{A}^{\mathbb{N}}$. An *n-arithmetic factor* in $\mathbf{w}$ is any subwords of the form

$$w_k \; w_{k+d} \; w_{k+2d} \; \cdots \; , \; w_{k+(n-1)d}, \quad k \geqslant 0, \; d \geqslant 1.$$

Definition (6)

The **arithmetical complexity** of $\mathbf{w}$ is given by:

$$\mathrm{a}_w(n) \;=\; \#\Big\{ w_k \; w_{k+d} \; \cdots \; w_{k+(n-1)d} \;\Big|\; k \geqslant 0, \; d \geqslant 1 \Big\}.$$

- counts how many distinct arithmetic n-factors of appear in w.

Arithmetic Complexity

- Let $\mathbf{w} \in \mathcal{A}^{\mathbb{N}}$. An *n-arithmetic factor* in $\mathbf{w}$ is any subwords of the form

$$w_k\ w_{k+d}\ w_{k+2d}\ \cdots,\ w_{k+(n-1)d}, \quad k \geqslant 0,\ d \geqslant 1.$$

Definition (6)

The **arithmetical complexity** of $\mathbf{w}$ is given by:

$$\mathrm{a}_w(n) = \#\left\{ w_k\ w_{k+d}\ \cdots\ w_{k+(n-1)d} \;\middle|\; k \geqslant 0,\ d \geqslant 1 \right\}.$$

- counts how many distinct arithmetic n-factors of appear in w.

Theorem 6. (Cassaign & Frid. 2007)

- Let $\mathbf{s}$ be a sturmain word. Then $\mathrm{a}_{\mathbf{s}}(k) = \mathcal{O}(k^3)$

Notation

$\mathrm{a}_{a,b}(n)$**:** arthmetic complexity of the most significant digits of the sequence (a^n) in base b.

$\mathrm{a}_{\mathbf{r}}(n)$**:** arthmetic complexity of the rotation word $\mathbf{r}$

Notation

$\mathrm{a}_{a,b}(n)$**:** arthmetic complexity of the most significant digits of the sequence (a^n) in base b.

$\mathrm{a}_{\mathbf{r}}(n)$**:** arthmetic complexity of the rotation word **r**

Conclusion (admissible pair)

$$\mathrm{a}_{\mathbf{s}}(k) \quad \Longrightarrow \quad \mathrm{a}_{\mathbf{r}}(k) \quad \Longrightarrow \quad \mathrm{a}_{a,b}(k) = \mathcal{O}(k^3)$$

Polynomial Complexity

Polynomial Complexity

Let $\mathbf{w} \in \mathcal{A}^{\mathbb{N}}$. For any polynomial

$$P(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x].$$

Polynomial Complexity

Let $\mathbf{w} \in \mathcal{A}^{\mathbb{N}}$. For any polynomial

$$P(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x].$$

- The *polynomial subword*:

$$w_{k+P(0)}\, w_{k+P(1)}\, w_{k+P(2)}\, \ldots, \quad d \geqslant 1,\ a_d > 0,\ k >\!\!> 0.$$

Polynomial Complexity

Let $\mathbf{w} \in \mathcal{A}^{\mathbb{N}}$. For any polynomial

$$P(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x].$$

- The *polynomial subword*:

$$w_{k+P(0)}\, w_{k+P(1)}\, w_{k+P(2)}\, \ldots, \quad d \geqslant 1,\ a_d > 0,\ k >> 0.$$

- The polynomial closure:

$$P_d(\mathbf{w}) = \bigcup_{k \geqslant 0} \bigcup_{\substack{P \in \mathbb{Z}[x] \\ \deg(P)=d,\, P(n) \geqslant 0}} \mathrm{Fac}\Big(w_{k+P(0)}\, w_{k+P(1)}\, w_{k+P(2)}\, \ldots \Big).$$

Polynomial Complexity

Let $\mathbf{w} \in \mathcal{A}^{\mathbb{N}}$. For any polynomial

$$P(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x].$$

- The *polynomial subword*:

$$w_{k+P(0)}\, w_{k+P(1)}\, w_{k+P(2)}\, \ldots, \quad d \geqslant 1,\ a_d > 0,\ k >\!\!> 0.$$

- The polynomial closure:

$$P_d(\mathbf{w}) = \bigcup_{k \geqslant 0} \bigcup_{\substack{P \in \mathbb{Z}[x] \\ \deg(P)=d,\, P(n)\geqslant 0}} \mathrm{Fac}\Big(w_{k+P(0)}\, w_{k+P(1)}\, w_{k+P(2)}\, \ldots \Big).$$

Definition (7)

The **polynomial complexity** of **w** is given by:

$$\mathrm{P}_d^{\mathbf{w}}(n) = \#\Big(P_d(\mathbf{w}) \cap \mathcal{A}^n \Big),$$

- counts the number of distinct n-factors that appear in $P_d(\mathbf{w})$.

Theorem 7. (M.G. & Cassaign 2025^{+})

Let **s** be a Sturmian word. Then

$$\mathrm{P}_d^{\mathbf{s}}(k) = \mathcal{O}\left(k^{\frac{(d+1)(d+2)}{2}}\right).$$

Theorem 7. (M.G. & Cassaign 2025^{+})

Let **s** be a Sturmian word. Then

$$\mathrm{P}_d^{\mathbf{s}}(k) = \mathcal{O}\left(k^{\frac{(d+1)(d+2)}{2}}\right).$$

Notation

Theorem 7. (M.G. & Cassaign 2025^{+})

Let $\mathbf{s}$ be a Sturmian word. Then

$$\mathrm{P}_d^{\mathbf{s}}(k) = \mathcal{O}\left(k^{\frac{(d+1)(d+2)}{2}}\right).$$

Notation

$\mathrm{P}_d^{a,b}(n)$**:** polynomial complexity of the most significant digits of the sequence (a^n) in base b.

$\mathrm{P}_d^{\mathbf{r}}(n)$**:** polynomial complexity of the rotation word $\mathbf{r}$

Theorem 7. (M.G. & Cassaign 2025$^+$)

Let **s** be a Sturmian word. Then

$$\mathrm{P}_d^{\mathbf{s}}(k) = \mathcal{O}\left(k^{\frac{(d+1)(d+2)}{2}}\right).$$

Notation

$\mathrm{P}_d^{a,b}(n)$**:** polynomial complexity of the most significant digits of the sequence (a^n) in base b.

$\mathrm{P}_d^{\mathbf{r}}(n)$**:** polynomial complexity of the rotation word **r**

Conclusion (admissible pair)

$$\mathrm{P}_d^{\mathbf{s}}(k) \quad \Longrightarrow \quad \mathrm{P}_d^{\mathbf{r}}(k) \quad \Longrightarrow \quad \mathrm{P}_d^{a,b}(k) = \mathcal{O}\left(k^{\frac{(d+1)(d+2)}{2}}\right)$$

Beyond

Beyond

Theorem 8. (M. G., Kanel-Belov, Kondakov, & Mitrofanov. 2021)

Let $\mathrm{P}(n)$ be a polynomial with an irrational leading coefficient. Let $\mathbf{w}$ be an infinite word where

$$\mathbf{w}_n = [2\{P(n)\}]$$

Then there is a polynomial $\mathrm{Q}(k)$ that depends only on $\operatorname{deg}(\mathrm{P})$, such that

$$\mathrm{Q}(k) = \mathrm{p}_{\mathbf{w}}(k)$$

for all sufficiently large k.

Beyond

Theorem 8. (M. G., Kanel-Belov, Kondakov, & Mitrofanov. 2021)

Let $\mathrm{P}(n)$ be a polynomial with an irrational leading coefficient. Let $\mathbf{w}$ be an infinite word where

$$\mathbf{w}_n = [2\{P(n)\}]$$

Then there is a polynomial $\mathrm{Q}(k)$ that depends only on $\deg(\mathrm{P})$, such that

$$\mathrm{Q}(k) = \mathrm{p}_{\mathbf{w}}(k)$$

for all sufficiently large k.

Theorem 9. (M. G. & Mitrofanov. 2024)

Let $d \in \mathbb{Z}_{>0}$, let $b \geqslant 5$ be an integer, and let $a > 0$ be a real number such that a and b are multiplicatively independent. Consider the sequence w, where w_n is the most significant digit of a^{n^d} when expressed in base b.

Then, there exists a polynomial $\mathrm{P}(k)$ of degree $\frac{d(d+1)}{2}$ such that:

$$\mathrm{P}(k) = \mathrm{p}_{\mathbf{w}}(k) \quad \text{for large enough } k.$$

Bounding of Significant Factor Complexity

Bounding of Significant Factor Complexity

Definition (8)

A **significant factor** in base b is a contiguous subword that appears in the most significant digits word of $(c \cdot a^n)_{n\in\mathbb{N}}$, where $c, a > 0$.

Bounding of Significant Factor Complexity

Definition (8)

A **significant factor** in base b is a contiguous subword that appears in the most significant digits word of $(c \cdot a^n)_{n\in\mathbb{N}}$, where $c, a > 0$.

Theorem 10. (M. G. & Shevtsova. 2025^+)

Let $k \in \mathbb{N}$. Then

$$\frac{(k-1)k(k+1)}{6} \leqslant \mathrm{p}_{\mathbf{d}}(k) \leqslant 3(k-1)k(k+1).$$

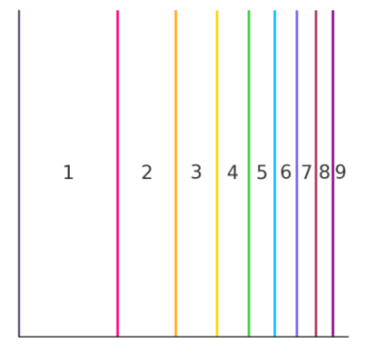


Figure: Lines for $k = 1$

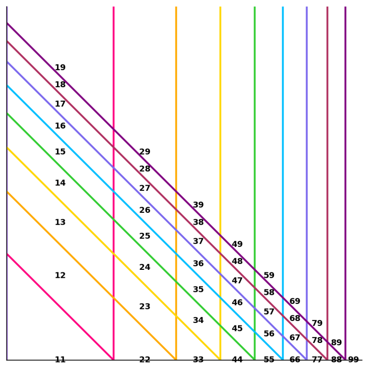


Figure: Lines for $k = 2$

Sketch of proof

Sketch of proof

- Counting Faces: Exactly $9k$ lines partition $\mathbb{T}^2$; each face in this partition corresponds uniquely to a length-k factor in $\mathbf{d}$.

Sketch of proof

- Counting Faces: Exactly $9k$ lines partition $\mathbb{T}^2$; each face in this partition corresponds uniquely to a length-k factor in $\mathbf{d}$.
- Lower Bound: Coloring lines cleverly ensures many intersection points (vertices), giving at least $\frac{(k-1)k(k+1)}{6}$ faces.

Sketch of proof

- Counting Faces: Exactly $9k$ lines partition $\mathbb{T}^2$; each face in this partition corresponds uniquely to a length-k factor in $\mathbf{d}$.
- Lower Bound: Coloring lines cleverly ensures many intersection points (vertices), giving at least $\frac{(k-1)k(k+1)}{6}$ faces.
- Upper Bound: One also cannot exceed $3\,(k-1)\,k\,(k+1)$ faces by topological constraints (Euler characteristic on the torus).

A bit terminology

A bit terminology

Definition (9)

A **log-factorial sequence** is any sequence of the form $(\lg(c \cdot n!))_{n\in\mathbb{N}}$, $c > 0$.

- Let $(s_n)_{n\in\mathbb{N}}$ be a sequence of nonzero real numbers.

Definition (10)

The **leading decimal signature** of this sequence is the infinite word formed by the most significant digits of the sequence $\left(10^{s_n}\right)_{n\in\mathbb{N}}$ in base 10, which we denote by $\mathfrak{D}$.

Definition (11)

A **factorial signature** is the leading decimal signature associated with a log-factorial sequence, and it is denoted by $\mathfrak{F}$.

Recurrence of Significant Factors in Factorial Signature

Recurrence of Significant Factors in Factorial Signature

Theorem 11. (M. G. & Shevtsova. 2025^{+})

Any significant factor in the factorial signature $\mathfrak{F}$ appears infinitely often.

Factor Complexity of Residual Words

Factor Complexity of Residual Words

Definition (12)

Any recurrent factor of a factorial signature that is not a significant factor is called a **residual factor**.

Factor Complexity of Residual Words

Definition (12)

Any recurrent factor of a factorial signature that is not a significant factor is called a **residual factor**.

Theorem 12. (M. G. & Shevtsova. 2025^{+})

The number of distinct residual factors of length k in a factorial signature is $\Theta\left(k^2 \ln k\right)$.

Measure-Zero Exceptions

Measure-Zero Exceptions

Theorem 13. (M. G. & Shevtsova. 2025^{+})

For any fixed k, the set of real $c > 0$ whose factorial signature carries a length-k *residual* factor infinitely often has Lebesgue measure zero.

On the Recurrence of Gamma Words

On the Recurrence of Gamma Words

We call $\left(c_1 \cdot \Gamma(c_2 \cdot n)\right)_{n \in \mathbb{N}}$ is *Gamma sequence* if $c_1, c_2 > 0$.

On the Recurrence of Gamma Words

We call $\left(c_1 \cdot \Gamma(c_2 \cdot n)\right)_{n\in\mathbb{N}}$ is *Gamma sequence* if $c_1, c_2 > 0$.

Definition (13)

A **Gamma word** is the leading decimal signature associated with associated with a gamma sequence, and it is denoted by $\mathfrak{G}$.

On the Recurrence of Gamma Words

We call $\left(c_1 \cdot \Gamma(c_2 \cdot n)\right)_{n \in \mathbb{N}}$ is *Gamma sequence* if $c_1, c_2 > 0$.

Definition (13)

A **Gamma word** is the leading decimal signature associated with associated with a gamma sequence, and it is denoted by $\mathfrak{G}$.

Theorem 14. (M. G. & Shevtsova. 2025^{+})

For every residual factor of a factorial signature, there exists a gamma word in which that same factor appears infinitely many times.

References

- P. Alessandri. “Codages de rotations et basses complexités.” Ph.D. Thesis, Aix-Marseille University II, Marseille, France (1996).
- X. He, A. J. Hildebrand, Y. Li, and Y. Zhang. “Complexity of Leading Digit Sequences.” *Discrete Math. Theor. Comput. Sci.* **1**, no. 1 (2020).
- M. Golafshan and I. Mitofanov. “Factor complexity of the most significant digits of a^{n^d}.” *arXiv:2402.16210v2*
- A. Kanel-Belov, G. V. Kondakov, and I. V. Mitrofanov. “Inverse problems of symbolic dynamics.” *Banach Cent. Publ.* **94**, no. 1 (2011): 43–60.
- L. Kuipers and H. Niderreiter. “Uniform Distribution of Sequences.” *Pure and applied mathematics, A Wiley-Interscience publication.* New York (1974).
- H. Weyl. “Über die Gleichverteilung von Zahlen mod. Eins.” *Math. Ann.* **77** (1996): 313–352.