

PREVIOUS-COPY COMPRESSION AND ALGEBRAIC DIGIT EXPANSIONS

LUCA BLANCHI

ABSTRACT. Let $b \geq 2$ be an integer and let

$$\alpha = \sum_{n \geq 1} a_n b^{-n} \in (0, 1) \cap (\overline{\mathbb{Q}} \setminus \mathbb{Q})$$

be algebraic irrational, with $a_n \in \{0, \dots, b-1\}$. Write

$$w_N(\alpha) = a_1 a_2 \cdots a_N$$

for the prefix of length N of the base- b expansion of α .

We study the compressibility of $w_N(\alpha)$ under a left-to-right previous-copy model. A phrase is either a literal symbol or an exact copy of an earlier substring; self-overlap is allowed, as in self-referential variants of LZ77. Let $\text{oc}(W)$ be the minimum number of phrases in such a parsing of a finite word W . We prove

$$\text{oc}(w_N(\alpha)) = \omega(\log N).$$

Consequently, every standard LZ77-type previous-factor parsing, self-referential or not, uses $\omega(\log N)$ phrases on $w_N(\alpha)$.

The proof combines a simple multiplicative-growth lemma with Diophantine approximation. A logarithmic online previous-copy parsing forces a copied phrase whose endpoint is a fixed multiplicative factor larger than its starting boundary. If the copy displacement tends to infinity, this gives a repetition forbidden by the Adamczewski–Bugeaud Diophantine exponent criterion. If the displacement is bounded, it gives too good an approximation by rationals with denominators supported on a fixed finite set of primes, contradicting Ridout’s theorem.

We also prove a structured-noise variant using Nguyen’s refined Diophantine exponent theorem. If copied phrases may differ from their sources on a vanishing proportion of positions contained in a bounded number of intervals, logarithmically many phrases still do not suffice.

Finally, we add an effective finite-prefix component. We introduce the linear extension complexity $\lambda(W)$, the least C such that W extends to an infinite word of factor complexity at most Cn . Using Bugeaud’s effective theorem for long prefixes of algebraic numbers, we prove

$$\lambda(w_N(\alpha)) \gg_{\alpha} \frac{(\log N)^{1/11}}{(\log \log N)^{4/11}}.$$

Since small online-copy parsings and small string attractors both imply small linear extension complexity, this gives the same effective lower bound for the minimum string attractor size of $w_N(\alpha)$, and hence for all compression measures which induce string attractors of comparable size. This part is weaker than the superlogarithmic LZ77 bound, but applies to offline repetitiveness measures outside the reach of the online argument.

1. INTRODUCTION

The base- b expansion of an algebraic irrational number is expected to be highly complex. This expectation cannot be usefully expressed through unrestricted Kolmogorov complexity of finite prefixes: algebraic numbers are computable, and hence their first N digits have descriptions of length $O(\log N)$. One must instead work with restricted models of description or compression.

A fundamental theorem of Adamczewski and Bugeaud says that the base- b expansion of an algebraic irrational number cannot have low factor complexity. More precisely, if $p(n, \alpha, b)$ denotes the number of distinct blocks of length n occurring in the base- b expansion of an algebraic irrational α , then

$$\frac{p(n, \alpha, b)}{n} \rightarrow +\infty.$$

Their method rests on a Diophantine principle: sufficiently strong repetitions in the digit expansion produce rational approximations that are too good for algebraic irrational numbers.

The purpose of this note is to apply that principle to a concrete compression model. We consider finite prefixes under a left-to-right previous-copy parsing. This model is deliberately generous. Copy sources and lengths are not charged. A copied phrase is allowed to overlap its target. Therefore a lower bound in this model is not an artefact of bit-level encoding conventions.

The first main result is the following.

Theorem 1.1 (Exact online-copy lower bound). *Let $b \geq 2$, and let*

$$\alpha = \sum_{n \geq 1} a_n b^{-n} \in (0, 1) \cap (\overline{\mathbb{Q}} \setminus \mathbb{Q}).$$

Then

$$\text{oc}(w_N(\alpha)) = \omega(\log N).$$

Equivalently,

$$\lim_{N \rightarrow \infty} \frac{\text{oc}(w_N(\alpha))}{\log N} = +\infty.$$

As a direct consequence, every standard LZ77-type previous-factor parsing of $w_N(\alpha)$, self-referential or not, has $\omega(\log N)$ phrases.

The result should be read carefully. It is not a lower bound for arbitrary straight-line programs, unrestricted grammar compression, bidirectional macro schemes, or offline copy systems. The online condition is essential in the proof. However, a second and independent part of the note gives effective lower bounds for several offline repetitiveness measures through a different route.

To state that route, define the linear extension complexity $\lambda(W)$ of a finite word W to be the least C such that W is the prefix of an infinite word

x satisfying

$$p_x(n) \leq Cn \quad (n \geq 1),$$

where $p_x(n)$ is the number of distinct factors of length n of x . We prove that both online-copy parsings and string attractors control $\lambda(W)$:

$$\lambda(W) \leq \text{oc}(W) + 1, \quad \lambda(W) \leq \gamma(W) + 1,$$

where $\gamma(W)$ is the minimum string attractor size of W . Combining this with Bugeaud's effective finite-prefix theorem gives the second main result.

Theorem 1.2 (Effective finite-prefix lower bound). *Let $b \geq 2$, and let*

$$\alpha \in (0, 1) \cap (\overline{\mathbb{Q}} \setminus \mathbb{Q}).$$

Then, for all sufficiently large N ,

$$\lambda(w_N(\alpha)) \gg_\alpha \frac{(\log N)^{1/11}}{(\log \log N)^{4/11}}.$$

Consequently,

$$\gamma(w_N(\alpha)) \gg_\alpha \frac{(\log N)^{1/11}}{(\log \log N)^{4/11}},$$

and the same lower bound holds for every compression measure which is known to induce a string attractor of size bounded linearly by that measure.

This effective bound is weaker than $\text{oc}(w_N(\alpha)) = \omega(\log N)$, but it applies to offline measures for which the online multiplicative-growth argument does not apply.

We also prove a structured noisy version of the online theorem. There, a copied phrase may differ from its source on a small exceptional set, provided the exceptional set is contained in a bounded number of intervals. The unbounded-displacement case is handled by Nguyen's refined Diophantine exponent theorem; the bounded-displacement case is handled directly by Ridout's theorem.

The final section isolates a quantitative profile

$$\Theta_\alpha(T)$$

measuring the strongest exact repetition in the digit sequence beyond scale T . The exact theorem implies qualitatively that

$$\Theta_\alpha(T) \rightarrow 0.$$

Moreover, any explicit decay bound on $\Theta_\alpha(T)$ would immediately imply an explicit superlogarithmic lower bound for $\text{oc}(w_N(\alpha))$. This reduces the main quantitative breakthrough problem to a finite Diophantine problem, closely related to a moving-period version of Ridout's theorem.

Throughout the note, logarithms are natural. All implicit constants may depend on fixed parameters such as α , its degree and height, and, where relevant, b , although the effective theorem quoted from Bugeaud is itself independent of the base.

2. DIOPHANTINE INPUTS

Let

$$\mathbf{a} = a_1 a_2 a_3 \cdots$$

be an infinite word over a finite set of integers. For integers $i \leq j$, write

$$\mathbf{a}[i, j] = a_i a_{i+1} \cdots a_j.$$

For an integer $b \geq 2$, put

$$\xi_{\mathbf{a}, b} = \sum_{n \geq 1} a_n b^{-n}.$$

We use three Diophantine inputs: the Adamczewski–Bugeaud repetition criterion, Ridout’s theorem, and Bugeaud’s effective finite-prefix theorem. For the noisy section we also use Nguyen’s refined repetition criterion.

2.1. Exact repetitions. The following is the form of the Adamczewski–Bugeaud Diophantine exponent criterion needed below.

Theorem 2.1 (Exact repetition criterion). *Let $b \geq 2$. Suppose that there exist $\rho > 1$ and sequences of integers*

$$0 \leq r_m < s_m < t_m$$

such that

$$s_m - r_m \rightarrow \infty, \quad t_m \geq \rho s_m,$$

and

$$\mathbf{a}[r_m + 1, r_m + t_m - s_m] = \mathbf{a}[s_m + 1, t_m]$$

for every m . Then

$$\xi_{\mathbf{a}, b}$$

is rational or transcendental.

Indeed, the prefix $\mathbf{a}[1, t_m]$ contains a repetition beginning after a preperiod of length r_m , with period $s_m - r_m$, and with total length at least a fixed multiple of s_m . In the language of Adamczewski and Bugeaud, the Diophantine exponent of \mathbf{a} is greater than 1. The stated conclusion is precisely the corresponding transcendence criterion.

2.2. Ridout’s theorem. We shall use Ridout’s theorem in the following form.

Theorem 2.2 (Ridout). *Let S be a finite set of prime numbers, let θ be a real algebraic number, and let $\varepsilon > 0$. Then there are only finitely many rational numbers P/Q , written in lowest terms, such that every prime divisor of Q belongs to S and*

$$\left| \theta - \frac{P}{Q} \right| < Q^{-1-\varepsilon}.$$

2.3. Structured noisy repetitions. Let U and V be two words of the same length L . For $\eta > 0$ and an integer $\Delta \geq 0$, say that U and V are (η, Δ) -close if the set

$$\{1 \leq i \leq L : U_i \neq V_i\}$$

is contained in a union of at most Δ intervals of $\{1, \dots, L\}$ whose total cardinality is at most ηL .

We use the following consequence of Nguyen’s refined Diophantine exponent theorem.

Theorem 2.3 (Refined repetition criterion). *Let $b \geq 2$. Suppose that there exist $\rho > 1$, an integer $\Delta \geq 0$, a sequence $\eta_m \rightarrow 0$, and integers*

$$0 \leq r_m < s_m < t_m$$

such that

$$s_m - r_m \rightarrow \infty, \quad t_m \geq \rho s_m,$$

and such that the two words

$$\mathbf{a}[r_m + 1, r_m + t_m - s_m], \quad \mathbf{a}[s_m + 1, t_m]$$

are (η_m, Δ) -close for every m . Then

$$\xi_{\mathbf{a}, b}$$

is rational or transcendental.

Only the exact criterion and Ridout’s theorem are needed for the exact online-copy lower bound. The refined criterion is used solely in the noisy section.

2.4. Bugeaud’s effective finite-prefix theorem. We also use the following theorem of Bugeaud.

Theorem 2.4 (Bugeaud). *Let $b \geq 2$, and let ξ be a real algebraic irrational number of degree D and height at most H , with $H \geq e^e$. Let a be the base- b expansion of ξ . Let w be an infinite word satisfying*

$$p_w(n) \leq Cn \quad (n \geq 1)$$

for some integer $C \geq 2$. If the first L digits of a coincide with the first L digits of w , then either

$$H \geq \exp \left\{ 10^{-2} C^{-1} L^{1/(8 \log(4C))} \right\},$$

or

$$D \geq \exp \left\{ 10^{-100} C^{-11/2} (\log C)^{-1} (\log L)^{1/2} (\log \log L)^{-1} \right\}.$$

The numerical constants play no role in this note. What matters is that, for fixed algebraic ξ , the theorem forbids arbitrarily long prefixes from coinciding with infinite words of complexity Cn unless C grows at least like

$$\frac{(\log L)^{1/11}}{(\log \log L)^{4/11}}$$

up to a constant depending on ξ .

3. ONLINE PREVIOUS-COPY PARSINGS

Let $W = W[1]W[2] \cdots W[N]$ be a finite word.

Definition 3.1. An online previous-copy parsing of W is a factorization

$$W = F_1 F_2 \cdots F_z$$

with boundaries

$$0 = n_0 < n_1 < \cdots < n_z = N,$$

where

$$F_j = W[n_{j-1} + 1, n_j].$$

Each phrase F_j is of one of two types.

First, F_j may be a literal, in which case $|F_j| = 1$.

Second, F_j may be a copied phrase. Writing

$$s = n_{j-1}, \quad t = n_j, \quad L = t - s,$$

this means that there is a source position p with

$$1 \leq p \leq s$$

such that

$$W[p + h] = W[s + 1 + h] \quad (0 \leq h < L).$$

The source starts strictly to the left of the target phrase. It may overlap the target.

Define

$$\text{oc}(W)$$

to be the minimum number of phrases in an online previous-copy parsing of W .

This is a structural parsing measure, not a bit-level encoding length. In particular, source positions and phrase lengths are not charged.

Remark 3.2. The first phrase in every online previous-copy parsing is necessarily a literal, since no previous source is available at position 1.

4. A MULTIPLICATIVE-GROWTH LEMMA

The following elementary lemma is the combinatorial core of the exact lower bound.

Lemma 4.1 (Multiplicative-growth phrase). *Let $C > 0$. Suppose that, for infinitely many N , a word W_N of length N has an online previous-copy parsing with at most $C \log N$ phrases. Then there exist $\rho > 1$ and infinitely many copied phrases, arising in such parsings, with boundaries $s < t$ such that*

$$t \geq \rho s$$

and

$$t \rightarrow \infty.$$

One may take

$$\rho = \exp\left(\frac{1}{3C}\right).$$

Proof. Fix such a parsing of W_N , and write its boundaries as

$$0 = n_0 < n_1 < \cdots < n_z = N, \quad z \leq C \log N.$$

Since the first phrase is a literal, $n_1 = 1$. Thus

$$N = \frac{n_z}{n_1} = \prod_{j=2}^z \frac{n_j}{n_{j-1}}.$$

Set

$$\rho = \exp\left(\frac{1}{3C}\right) > 1.$$

Suppose, for contradiction, that there is a constant T such that every phrase satisfying

$$\frac{n_j}{n_{j-1}} \geq \rho$$

has endpoint $n_j \leq T$, along the infinite family of parsings under consideration. After the last such phrase, all remaining ratios are $< \rho$. Therefore, for all sufficiently large N ,

$$N \leq T\rho^z \leq T \exp\left(\frac{1}{3C}C \log N\right) = TN^{1/3},$$

which is impossible.

Thus there are phrases with

$$\frac{n_j}{n_{j-1}} \geq \rho$$

and unbounded endpoints. A literal phrase with starting boundary n_{j-1} has ratio

$$\frac{n_j}{n_{j-1}} = 1 + \frac{1}{n_{j-1}},$$

which is $< \rho$ for all sufficiently large n_{j-1} . Hence the high-growth phrases with unbounded endpoints are copied phrases. \square

5. EXACT ONLINE-COPY LOWER BOUND

Theorem 5.1 (Compression criterion). *Let $b \geq 2$, and let*

$$\xi_{\mathbf{a},b} = \sum_{n \geq 1} a_n b^{-n}.$$

If

$$\text{oc}(a_1 a_2 \cdots a_N) = O(\log N)$$

for infinitely many N , then $\xi_{\mathbf{a},b}$ is rational or transcendental.

Proof. It is enough to prove the contrapositive. Suppose that

$$\xi_{\mathbf{a},b} = \alpha \in (0, 1) \cap (\overline{\mathbb{Q}} \setminus \mathbb{Q}).$$

Assume, for contradiction, that there are $C > 0$ and infinitely many N such that

$$\text{oc}(w_N(\alpha)) \leq C \log N.$$

By the multiplicative-growth lemma, after passing to an infinite subsequence, we obtain copied phrases

$$w_N(\alpha)[s+1, t]$$

such that

$$t \geq \rho s$$

for some fixed $\rho > 1$, and $t \rightarrow \infty$.

For each such copied phrase, put $L = t - s$. There is a source position p , with $1 \leq p \leq s$, such that

$$a_{p+h} = a_{s+1+h} \quad (0 \leq h < L).$$

Set

$$r = p - 1, \quad d = s - r = s - p + 1.$$

Then

$$0 \leq r < s < t,$$

and

$$\mathbf{a}[r+1, r+t-s] = \mathbf{a}[s+1, t].$$

We split into two cases.

First suppose that

$$d = s - r \rightarrow \infty$$

along an infinite subsequence. Then the triples $r < s < t$ satisfy

$$s - r \rightarrow \infty, \quad t \geq \rho s,$$

and the exact repetition condition above. By the exact repetition criterion, α is rational or transcendental, contradicting the assumption that α is algebraic irrational.

It remains to consider the case in which $d = s - r$ is bounded along an infinite subsequence. Passing to a further subsequence, we may assume that d is fixed.

The equality

$$a_{r+1+h} = a_{s+1+h} \quad (0 \leq h < t - s)$$

can be rewritten as

$$a_k = a_{k+d} \quad (r+1 \leq k \leq t-d).$$

Thus the block

$$a_{r+1}a_{r+2} \cdots a_t$$

is periodic with period d .

Define the rational number

$$\beta = \sum_{n=1}^r a_n b^{-n} + \sum_{q \geq 0} \sum_{i=1}^d a_{r+i} b^{-(r+qd+i)}.$$

Then β has a base- b expansion agreeing with that of α through at least the first t digits. Hence

$$|\alpha - \beta| \ll b^{-t}.$$

Write $\beta = P/Q$ in lowest terms. Since β has preperiod r and period d in base b , its denominator divides

$$b^r(b^d - 1).$$

Therefore every prime divisor of Q belongs to the fixed finite set

$$S = \{q \text{ prime} : q \mid b(b^d - 1)\},$$

and

$$Q \ll_d b^r.$$

If r is bounded along an infinite subsequence, then only finitely many such rationals β can occur. Since the agreement length t tends to infinity, one of them agrees with α to arbitrarily many digits, and hence equals α . This contradicts the irrationality of α .

Thus $r \rightarrow \infty$. Since

$$s = r + d$$

and

$$t \geq \rho s,$$

there is $\rho' > 1$ such that, for all sufficiently large terms of the subsequence,

$$t \geq \rho' r.$$

Choose $\varepsilon > 0$ with $1 + \varepsilon < \rho'$. Since $Q \ll_d b^r$, we obtain

$$b^{-t} \ll Q^{-1-\varepsilon}$$

along the subsequence. Therefore

$$|\alpha - \beta| < Q^{-1-\varepsilon}$$

for infinitely many distinct rationals P/Q whose denominators have all prime factors in the fixed finite set S . This contradicts Ridout's theorem.

Both cases are impossible. Hence an algebraic irrational α cannot have

$$\text{oc}(w_N(\alpha)) = O(\log N)$$

along an infinite sequence of N . □

Corollary 5.2 (Exact online-copy lower bound). *Let $b \geq 2$, and let*

$$\alpha = \sum_{n \geq 1} a_n b^{-n} \in (0, 1) \cap (\overline{\mathbb{Q}} \setminus \mathbb{Q}).$$

Then

$$\text{oc}(w_N(\alpha)) = \omega(\log N).$$

6. CONSEQUENCES FOR LZ77-TYPE PARSINGS

The online previous-copy model is at least as permissive as the usual left-to-right previous-factor models.

Let $z(W)$ denote the number of phrases in a standard LZ77-type parsing of W , under any convention in which each phrase is either a literal, or a previous copy, possibly followed by one trailing literal. The previous copy may be self-referential or non-self-referential.

In the self-referential case, the copied part is an online previous-copy phrase. In the non-self-referential case, it is a fortiori an online previous-copy phrase. If a convention attaches one trailing literal to a copied phrase, split that phrase into at most two online previous-copy phrases. Hence there is an absolute constant A , depending only on the LZ77 convention, such that

$$\text{oc}(W) \leq Az(W) + A$$

for every finite word W .

Corollary 6.1 (LZ77 lower bound). *Let $b \geq 2$, and let*

$$\alpha \in (0, 1) \cap (\overline{\mathbb{Q}} \setminus \mathbb{Q}).$$

Then every standard LZ77-type previous-factor parsing of

$$w_N(\alpha)$$

has

$$\omega(\log N)$$

phrases.

Proof. If some such parsing had $O(\log N)$ phrases along an infinite sequence of N , then the comparison above would give

$$\text{oc}(w_N(\alpha)) = O(\log N)$$

along the same sequence, contradicting the exact online-copy lower bound. \square

7. LINEAR EXTENSION COMPLEXITY

The online argument gives a strong superlogarithmic lower bound, but it relies essentially on left-to-right growth. We now introduce a weaker measure which also interacts with offline repetitiveness notions such as string attractors.

Let \mathcal{A} be a finite alphabet. For an infinite word $x \in \mathcal{A}^{\mathbb{N}}$, let

$$p_x(k)$$

be the number of distinct factors of length k occurring in x .

Definition 7.1. For a finite word $W \in \mathcal{A}^*$, define its linear extension complexity by

$$\lambda(W) = \inf \left\{ C \geq 1 : \begin{array}{l} \text{there exists } x \in \mathcal{A}^{\mathbb{N}} \text{ with } x[1, |W|] = W, \\ \text{and } p_x(k) \leq Ck \text{ for every } k \geq 1 \end{array} \right\}.$$

Thus $\lambda(W)$ measures whether W can be embedded as a prefix of an infinite word of uniformly linear factor complexity.

Lemma 7.2 (Online-copy parsings give linear extensions). *For every finite word W ,*

$$\lambda(W) \leq \text{oc}(W) + 1.$$

Proof. Let

$$W = F_1 F_2 \cdots F_z$$

be an online previous-copy parsing with $z = \text{oc}(W)$ phrases. Let

$$0 = n_0 < n_1 < \cdots < n_z = |W|$$

be its phrase boundaries. Fix a symbol $c \in \mathcal{A}$, and extend W to the infinite word

$$x = Wccc \cdots .$$

Let $k \geq 1$, and consider any factor U of x of length k . Choose the leftmost occurrence of U in x . We claim that this occurrence crosses either the first position of a phrase, or the first position of the infinite constant tail.

Indeed, suppose first that the occurrence lies entirely inside W . If it is contained in a copied phrase

$$W[s+1, t]$$

and does not cross the first position $s+1$ of that phrase, then it starts at some position $i > s+1$ and ends at most at t . Let $p \leq s$ be the source position of the copied phrase. By the copy equality, the same length- k factor occurs starting at

$$p + (i - s - 1) < i,$$

contradicting the leftmost choice. Thus the occurrence crosses the beginning of a phrase, unless it lies in a literal phrase, in which case it also crosses the beginning of that phrase.

If the occurrence meets the infinite tail, then it crosses the first position $|W| + 1$ of the tail, unless it is entirely contained in the tail. In the latter case $U = c^k$, which also occurs starting at $|W| + 1$.

There are z phrase-start positions and one tail-start position. For a fixed position q , at most k factors of length k cross q . Therefore

$$p_x(k) \leq (z+1)k \quad (k \geq 1).$$

Hence $\lambda(W) \leq z + 1$. □

8. EFFECTIVE LOWER BOUNDS FROM BUGEAUD'S THEOREM

The previous lemma allows us to convert Bugeaud's finite-prefix theorem into an effective lower bound for $\lambda(w_N(\alpha))$.

Theorem 8.1 (Effective lower bound for linear extension complexity). *Let $b \geq 2$, and let*

$$\alpha \in (0, 1) \cap (\overline{\mathbb{Q}} \setminus \mathbb{Q}).$$

Then there is a constant $c_\alpha > 0$ such that, for all sufficiently large N ,

$$\lambda(w_N(\alpha)) \geq c_\alpha \frac{(\log N)^{1/11}}{(\log \log N)^{4/11}}.$$

Proof. Let D be the degree of α , and let $H \geq e^e$ be an upper bound for its height. Put

$$L = N.$$

Let $C_0 > \lambda(w_N(\alpha))$. By the definition of λ , there is an infinite word x whose first N symbols are $w_N(\alpha)$ and such that

$$p_x(k) \leq C_0 k \quad (k \geq 1).$$

Choose an integer

$$C \geq \max\{2, C_0\}$$

with $C \leq C_0 + 3$. Then

$$p_x(k) \leq Ck \quad (k \geq 1).$$

Bugeaud's theorem applies to α , x , C , and $L = N$. Hence either

$$H \geq \exp \left\{ 10^{-2} C^{-1} N^{1/(8 \log(4C))} \right\},$$

or

$$D \geq \exp \left\{ 10^{-100} C^{-11/2} (\log C)^{-1} (\log N)^{1/2} (\log \log N)^{-1} \right\}.$$

We show that both alternatives are impossible if

$$C \leq c \frac{(\log N)^{1/11}}{(\log \log N)^{4/11}}$$

with $c > 0$ sufficiently small and N sufficiently large.

First, under this hypothesis,

$$\log(4C) \ll \log \log N.$$

Therefore

$$N^{1/(8 \log(4C))} = \exp \left(\frac{\log N}{8 \log(4C)} \right)$$

grows faster than every fixed power of $\log N$. Since C is only a fixed power of $\log N$, we have

$$C^{-1} N^{1/(8 \log(4C))} \rightarrow +\infty.$$

Thus the first alternative would force $H \rightarrow +\infty$, impossible because H is fixed.

Second, using

$$C \leq c \frac{(\log N)^{1/11}}{(\log \log N)^{4/11}}$$

and

$$\log C \leq \log \log N$$

for large N , we obtain

$$\begin{aligned} & C^{-11/2}(\log C)^{-1}(\log N)^{1/2}(\log \log N)^{-1} \\ & \geq c^{-11/2}(\log \log N)^2(\log C)^{-1}(\log \log N)^{-1}. \end{aligned}$$

Since $\log C \leq \log \log N$, the right-hand side is bounded below by a positive constant multiple of $c^{-11/2}$. Choosing c sufficiently small, this lower bound exceeds $\log D$. Thus the second alternative would force

$$D > \deg(\alpha),$$

again impossible.

Consequently,

$$C \gg_{\alpha} \frac{(\log N)^{1/11}}{(\log \log N)^{4/11}}.$$

Since $C \leq C_0 + 3$, and since $C_0 > \lambda(w_N(\alpha))$ was arbitrary, the theorem follows after adjusting the implicit constant. \square

Corollary 8.2 (Effective online-copy lower bound). *Let $b \geq 2$, and let*

$$\alpha \in (0, 1) \cap (\overline{\mathbb{Q}} \setminus \mathbb{Q}).$$

Then, for all sufficiently large N ,

$$\text{oc}(w_N(\alpha)) \gg_{\alpha} \frac{(\log N)^{1/11}}{(\log \log N)^{4/11}}.$$

Proof. This follows from

$$\lambda(W) \leq \text{oc}(W) + 1$$

and the preceding theorem. \square

Remark 8.3. This effective bound is much weaker than

$$\text{oc}(w_N(\alpha)) = \omega(\log N),$$

which was proved by the exact Diophantine exponent argument. Its value is that it comes from a finite-prefix low-complexity obstruction and will also apply to offline repetitiveness measures.

9. STRING ATTRACTORS AND OFFLINE COMPRESSION

We now apply the same finite-prefix method to string attractors.

Definition 9.1. Let W be a finite word of length N . A set

$$\Gamma \subseteq \{1, \dots, N\}$$

is a string attractor for W if every distinct factor of W has an occurrence in W crossing at least one position of Γ . Let

$$\gamma(W)$$

be the minimum size of a string attractor for W .

Lemma 9.2 (Attractors give linear extensions). *For every finite word W ,*

$$\lambda(W) \leq \gamma(W) + 1.$$

Proof. Let Γ be a string attractor for W , with

$$|\Gamma| = \gamma(W).$$

Fix a symbol c in the alphabet and extend W to

$$x = Wccc\cdots.$$

Let $k \geq 1$. Every factor of length k of x either is a factor of W , crosses the boundary between W and the infinite tail, or is c^k .

Every distinct factor of length k of W has an occurrence crossing a position of Γ . For each fixed position $q \in \Gamma$, at most k length- k factors cross q . Hence factors coming from W contribute at most

$$\gamma(W)k$$

distinct words.

The factors crossing the tail boundary contribute at most k additional words, and the factor c^k is included among them by considering the occurrence starting at the first tail position. Therefore

$$p_x(k) \leq (\gamma(W) + 1)k \quad (k \geq 1),$$

and so

$$\lambda(W) \leq \gamma(W) + 1. \quad \square$$

Theorem 9.3 (Effective string-attractor lower bound). *Let $b \geq 2$, and let*

$$\alpha \in (0, 1) \cap (\overline{\mathbb{Q}} \setminus \mathbb{Q}).$$

Then, for all sufficiently large N ,

$$\gamma(w_N(\alpha)) \gg_{\alpha} \frac{(\log N)^{1/11}}{(\log \log N)^{4/11}}.$$

Proof. By the preceding lemma,

$$\lambda(w_N(\alpha)) \leq \gamma(w_N(\alpha)) + 1.$$

The result follows from the effective lower bound for $\lambda(w_N(\alpha))$. \square

Corollary 9.4 (Offline dictionary-compression consequences). *Let $M(W)$ be any compression measure for which every representation of size $M(W)$ induces a string attractor of size $O(M(W))$, under the corresponding standard size convention. Then, for every algebraic irrational*

$$\alpha \in (0, 1)$$

and every integer base $b \geq 2$,

$$M(w_N(\alpha)) \gg_{\alpha} \frac{(\log N)^{1/11}}{(\log \log N)^{4/11}}.$$

Proof. If a representation of size $M(W)$ induces a string attractor of size $O(M(W))$, then

$$\gamma(W) \ll M(W).$$

Apply this to $W = w_N(\alpha)$ and use the preceding theorem. \square

Remark 9.5. This corollary applies to the standard repetitiveness measures for which such a linear attractor extraction is known, under the corresponding size conventions. These include, among others, LZ77, straight-line programs under usual grammar-size conventions, collage systems, macro schemes, and related measures discussed in the string-attractor literature. The principle is uniform: once a compressor induces an attractor of size $O(M)$, the lower bound for γ transfers to M .

Remark 9.6. The string-attractor result is not a substitute for the exact online-copy theorem. It gives only

$$\gamma(w_N(\alpha)) \gg_{\alpha} \frac{(\log N)^{1/11}}{(\log \log N)^{4/11}},$$

not $\gamma(w_N(\alpha)) = \omega(\log N)$. Proving a superlogarithmic lower bound for string attractors or straight-line programs would require a new idea beyond the online multiplicative-growth argument.

10. NOISY ONLINE-COPY PARSINGS

We now return to online parsings and introduce a structured-noise version.

Let U, V be words of the same length L . Recall that U and V are (η, Δ) -close if their mismatch set is contained in a union of at most Δ intervals of total cardinality at most ηL .

Definition 10.1. Let W be a finite word. An (η, Δ) -noisy online-copy parsing of W is a factorization

$$W = F_1 F_2 \cdots F_z$$

with boundaries

$$0 = n_0 < n_1 < \cdots < n_z = |W|,$$

such that each phrase is either a literal or an approximate previous copy in the following sense. If

$$F_j = W[s + 1, t], \quad L = t - s,$$

then there exists a source position p , with

$$1 \leq p \leq s,$$

such that

$$W[p, p + L - 1] \quad \text{and} \quad W[s + 1, t]$$

are (η, Δ) -close.

Let

$$\text{oc}_{\eta, \Delta}(W)$$

be the minimum number of phrases in such a parsing.

This is again a structural measure. The exceptional intervals, their locations, and their contents are not charged.

11. NOISY ONLINE-COPY LOWER BOUND

Theorem 11.1 (Noisy online-copy lower bound). *Let $b \geq 2$, and let*

$$\alpha = \sum_{n \geq 1} a_n b^{-n} \in (0, 1) \cap (\overline{\mathbb{Q}} \setminus \mathbb{Q}).$$

Fix an integer $\Delta \geq 0$, and let

$$\eta_N \rightarrow 0.$$

Then

$$\text{oc}_{\eta_N, \Delta}(w_N(\alpha)) = \omega(\log N).$$

Proof. Assume, for contradiction, that there exist $C > 0$ and infinitely many N such that

$$\text{oc}_{\eta_N, \Delta}(w_N(\alpha)) \leq C \log N.$$

The multiplicative-growth lemma applies verbatim, since it uses only the phrase boundaries. Thus, after passing to an infinite subsequence, there are approximate copied phrases

$$w_N(\alpha)[s+1, t]$$

such that

$$t \geq \rho s, \quad t \rightarrow \infty$$

for some fixed $\rho > 1$.

For each such phrase, put $L = t - s$. There exists p , with $1 \leq p \leq s$, such that

$$a_p a_{p+1} \cdots a_{p+L-1}$$

and

$$a_{s+1} a_{s+2} \cdots a_t$$

are (η_N, Δ) -close. Set

$$r = p - 1, \quad d = s - r.$$

Then the two words

$$\mathbf{a}[r+1, r+t-s], \quad \mathbf{a}[s+1, t]$$

are (η_N, Δ) -close.

We split into two cases.

First suppose that

$$d = s - r \rightarrow \infty$$

along an infinite subsequence. Since $\eta_N \rightarrow 0$, the refined repetition criterion applies to the triples $r < s < t$. It follows that α is rational or transcendental, contradicting the assumption that α is algebraic irrational.

It remains to treat the case where $d = s - r$ is bounded. Passing to a further subsequence, assume that d is fixed.

The approximate copy condition says that the equalities

$$a_{r+1+h} = a_{s+1+h} \quad (0 \leq h < L)$$

fail only for values of h contained in at most Δ intervals of total cardinality at most $\eta_N L$. Equivalently, up to enlarging these intervals by $O_\Delta(1)$ endpoints, the period- d relations

$$a_k = a_{k+d}$$

fail only on a subset of

$$r + 1 \leq k \leq r + L$$

covered by at most Δ intervals of total cardinality at most $\eta_N L + O_\Delta(1)$.

Remove these bad intervals. The remaining good set is a union of at most $\Delta + 1$ intervals and has cardinality at least

$$(1 - \eta_N)L - O_\Delta(1).$$

Thus one good interval has length

$$H \geq \frac{(1 - \eta_N)L - O_\Delta(1)}{\Delta + 1}.$$

Since

$$L = t - s \geq \left(1 - \frac{1}{\rho}\right)t,$$

there is a constant $c > 0$, depending only on ρ and Δ , such that

$$H \geq ct$$

for all sufficiently large N .

Let this good interval of edge positions begin at A . Then

$$a_k = a_{k+d} \quad (A \leq k \leq A + H - 1).$$

It follows that the digit block

$$a_A a_{A+1} \cdots a_{A+H+d-1}$$

is periodic with period d .

Set

$$m = A - 1.$$

Since $A \leq r + L \leq t$, we have $m \leq t$. Hence

$$H \geq cm$$

unless m is bounded, in which case the argument below is even stronger.

Define the rational number

$$\beta = \sum_{n=1}^m a_n b^{-n} + \sum_{q \geq 0} \sum_{i=1}^d a_{m+i} b^{-(m+qd+i)}.$$

The periodicity just obtained implies that α and β agree through at least position $m + H$. Therefore

$$|\alpha - \beta| \ll b^{-(m+H)}.$$

Writing $\beta = P/Q$ in lowest terms, we have

$$Q \mid b^m(b^d - 1)$$

up to cancellation. Thus all prime divisors of Q lie in the fixed finite set

$$S = \{q \text{ prime} : q \mid b(b^d - 1)\},$$

and

$$Q \ll_d b^m.$$

If m is bounded along an infinite subsequence, then only finitely many rationals β occur. Since $H \rightarrow \infty$, one of them agrees with α to arbitrarily many digits, and hence equals α , contradicting irrationality.

Thus $m \rightarrow \infty$. Since $H \geq cm$, choose $\varepsilon > 0$ with $\varepsilon < c$. Then, for all sufficiently large terms,

$$b^{-(m+H)} \ll Q^{-1-\varepsilon}.$$

Therefore

$$|\alpha - \beta| < Q^{-1-\varepsilon}$$

for infinitely many distinct rationals P/Q whose denominators have all prime factors in S . This contradicts Ridout's theorem.

Both cases are impossible. Hence

$$\text{oc}_{\eta_N, \Delta}(w_N(\alpha)) = \omega(\log N).$$

□

12. A QUANTITATIVE REPETITION PROFILE

The exact online-copy theorem is qualitative. It proves superlogarithmic divergence but gives no explicit rate. The next definitions isolate the precise finite Diophantine obstruction that controls possible quantitative improvements.

Let $\mathbf{a} = a_1 a_2 \dots$ be the digit sequence of α .

Definition 12.1. For $T \geq 1$, define

$$\Theta_{\mathbf{a}}(T) = \sup \left\{ \frac{t}{s} - 1 : \begin{array}{l} 0 \leq r < s < t, \ t \geq T, \\ \mathbf{a}[r+1, r+t-s] = \mathbf{a}[s+1, t] \end{array} \right\},$$

with the convention that the supremum is 0 if no such triple exists.

When \mathbf{a} is the base- b digit sequence of α , we write

$$\Theta_{\alpha}(T)$$

instead of $\Theta_{\mathbf{a}}(T)$.

Proposition 12.2 (Qualitative decay of the repetition profile). *Let $b \geq 2$, and let*

$$\alpha \in (0, 1) \cap (\overline{\mathbb{Q}} \setminus \mathbb{Q}).$$

Then

$$\Theta_\alpha(T) \rightarrow 0 \quad (T \rightarrow \infty).$$

Proof. Suppose not. Then there exist $\varepsilon > 0$ and triples

$$0 \leq r_m < s_m < t_m, \quad t_m \rightarrow \infty,$$

such that

$$t_m \geq (1 + \varepsilon)s_m$$

and

$$\mathbf{a}[r_m + 1, r_m + t_m - s_m] = \mathbf{a}[s_m + 1, t_m].$$

Put

$$d_m = s_m - r_m.$$

If $d_m \rightarrow \infty$ along an infinite subsequence, the exact repetition criterion implies that α is rational or transcendental, contradiction.

Thus d_m is bounded along an infinite subsequence. Passing to a further subsequence, assume $d_m = d$ is fixed. The equality above gives a period- d block of length tending to infinity. Repeating the bounded-displacement argument in the proof of the exact online-copy theorem, we obtain infinitely many rational numbers $\beta = P/Q$, eventually periodic in base b with fixed period d , such that

$$|\alpha - \beta| < Q^{-1-\varepsilon'}$$

for some $\varepsilon' > 0$, with all prime factors of Q lying in the fixed finite set

$$\{q : q \mid b(b^d - 1)\}.$$

This contradicts Ridout's theorem, unless α is rational. Since α is algebraic irrational, this is impossible. \square

The next lemma gives the precise bridge from the repetition profile to online-copy lower bounds.

Lemma 12.3 (Quantitative profile bound). *Let \mathbf{a} be any infinite word, and let*

$$W_N = a_1 a_2 \cdots a_N.$$

For $T \geq 2$, put

$$\eta_{\mathbf{a}}(T) = \max \left\{ \Theta_{\mathbf{a}}(T), \frac{1}{T-1} \right\}.$$

Then, for every $N > T$,

$$\text{oc}(W_N) \geq \frac{\log(N/T)}{\log(1 + \eta_{\mathbf{a}}(T))}.$$

Proof. Let

$$0 = n_0 < n_1 < \cdots < n_z = N$$

be an online previous-copy parsing of W_N .

Let q be the largest index such that

$$n_q < T.$$

Such an index exists because $n_1 = 1 < T$. Then

$$n_q < T.$$

For every $j > q$, the endpoint satisfies $n_j \geq T$.

If the j -th phrase is a literal, then

$$\frac{n_j}{n_{j-1}} = 1 + \frac{1}{n_{j-1}} \leq 1 + \frac{1}{T-1} \leq 1 + \eta_{\mathbf{a}}(T),$$

because $n_{j-1} \geq n_q \geq T-1$ for $j > q+1$, and for $j = q+1$ the claimed bound is still harmless after replacing n_q by the last boundary below T and absorbing the single crossing step into the final strict inequality below. If the j -th phrase is copied, write

$$s = n_{j-1}, \quad t = n_j.$$

The copy gives a triple $r < s < t$ satisfying

$$\mathbf{a}[r+1, r+t-s] = \mathbf{a}[s+1, t],$$

and $t \geq T$. Hence

$$\frac{t}{s} - 1 \leq \Theta_{\mathbf{a}}(T) \leq \eta_{\mathbf{a}}(T).$$

Therefore every relevant ratio after the boundary T , except possibly the single ratio crossing T , is bounded by $1 + \eta_{\mathbf{a}}(T)$. Since the crossing ratio contributes only to moving from a boundary below T to a boundary at least T , we have

$$\frac{N}{T} \leq (1 + \eta_{\mathbf{a}}(T))^z.$$

Thus

$$z \geq \frac{\log(N/T)}{\log(1 + \eta_{\mathbf{a}}(T))}.$$

Taking the minimum over parsings proves the lemma. \square

Corollary 12.4 (Conditional quantitative strengthening). *Let $b \geq 2$, and let*

$$\alpha \in (0, 1) \cap (\overline{\mathbb{Q}} \setminus \mathbb{Q}).$$

Suppose that there are constants $A > 0$, $c > 0$, and T_0 such that

$$\Theta_{\alpha}(T) \leq \frac{A}{(\log T)^c} \quad (T \geq T_0).$$

Then

$$\text{oc}(w_N(\alpha)) \gg_{\alpha} (\log N)^{1+c}.$$

More generally, if

$$\Theta_\alpha(T) \leq \frac{A}{(\log \log T)^c}$$

for all large T , then

$$\text{oc}(w_N(\alpha)) \gg_\alpha \log N (\log \log N)^c.$$

Proof. Apply the quantitative profile bound with

$$T = N^{1/2}.$$

For large N ,

$$\frac{1}{T-1}$$

is negligible compared with either of the assumed upper bounds for $\Theta_\alpha(T)$. Since

$$\log(1+u) \leq 2u$$

for small positive u , the result follows. \square

Remark 12.5. Thus the problem of proving an explicit superlogarithmic lower bound for online-copy or LZ77 parsing is reduced to proving an explicit decay estimate for the finite repetition profile $\Theta_\alpha(T)$.

13. A MOVING-PERIOD RIDOUT PROBLEM

The bounded-displacement part of the proof used Ridout's theorem only after the period d had been fixed. Quantitative improvements require control of periods d that may grow with the scale.

A repetition

$$\mathbf{a}[r+1, r+t-s] = \mathbf{a}[s+1, t], \quad d = s-r,$$

produces an eventually periodic rational approximant with preperiod r , period d , and agreement length at least t . Its denominator divides

$$b^r(b^d - 1).$$

If

$$t \geq (1+\varepsilon)s = (1+\varepsilon)(r+d),$$

then the associated rational $\beta_{r,d}$ satisfies

$$|\alpha - \beta_{r,d}| \ll b^{-(1+\varepsilon)(r+d)}.$$

For fixed d , Ridout excludes infinitely many such approximations. For $d \rightarrow \infty$, the set of prime divisors of $b^d - 1$ is no longer fixed.

Problem 13.1 (Moving-period Ridout problem). *Let $b \geq 2$, and let*

$$\alpha \in (0, 1) \cap (\overline{\mathbb{Q}} \setminus \mathbb{Q}).$$

Find an explicit function $F_\alpha(\varepsilon)$ such that there are no rationals β with an eventually periodic base- b expansion of preperiod r , period d , and

$$r+d \geq F_\alpha(\varepsilon),$$

satisfying

$$|\alpha - \beta| < b^{-(1+\varepsilon)(r+d)}.$$

Up to the separate treatment of bounded values of $r + d$, a bound of the form

$$F_\alpha(\varepsilon) \leq \exp(C\varepsilon^{-A})$$

would imply

$$\Theta_\alpha(T) \ll_\alpha (\log T)^{-1/A},$$

and hence

$$\text{oc}(w_N(\alpha)) \gg_\alpha (\log N)^{1+1/A}.$$

Even a double-exponential bound

$$F_\alpha(\varepsilon) \leq \exp \exp(C\varepsilon^{-A})$$

would imply

$$\text{oc}(w_N(\alpha)) \gg_\alpha \log N (\log \log N)^{1/A}.$$

This is the main Diophantine bottleneck for any explicit improvement over the qualitative lower bound

$$\omega(\log N).$$

14. MANY WEAK REPETITIONS

The multiplicative-growth lemma extracts one strong repetition from an $O(\log N)$ -phrase parsing. To prove a quantitative lower bound of the form

$$\text{oc}(w_N(\alpha)) \gg \log N f(N), \quad f(N) \rightarrow \infty,$$

one must handle repetitions whose exponent is only

$$1 + \frac{1}{f(N)} + o\left(\frac{1}{f(N)}\right).$$

A single such weak repetition is generally not enough.

A possible strengthening is to use the whole growth distribution of a parsing. Given phrase boundaries

$$0 = n_0 < n_1 < \dots < n_z = N,$$

define the growth energy above a scale T by

$$E_T = \sum_{j: n_{j-1} \geq T} \log \frac{n_j}{n_{j-1}}.$$

One always has

$$E_T \geq \log \frac{N}{T} - O(1).$$

Each copied phrase contributing to E_T gives a repetition with strength

$$\frac{n_j}{n_{j-1}} - 1.$$

Problem 14.1 (Many weak repetitions). *Develop a Diophantine criterion excluding, for algebraic irrational digit expansions, families of repetitions whose individual exponents may tend to 1, but whose total growth energy is large.*

Such a criterion would be more compatible with quantitative versions of the Subspace Theorem, which usually control families of solutions rather than a single exceptionally good solution.

15. BEYOND BOUNDED-INTERVAL NOISE

The noisy theorem above assumes that, in each approximate copy, the mismatch set is contained in at most Δ intervals, where Δ is fixed. This is the natural regime of Nguyen’s refined Diophantine exponent criterion.

A more robust compression model would allow copied phrases plus correction sets of low complexity. For example, the error set might be sparse, or might itself admit a short interval decomposition, LZ parsing, grammar, or attractor.

The bounded-displacement part of the noisy proof shows the obstruction clearly. If the error set is covered by Δ intervals, then one obtains a genuinely periodic interval of length at least

$$\frac{(1 - \eta)L}{\Delta + 1}.$$

For fixed Δ , this yields a rational approximation with a fixed exponent > 1 , and Ridout applies. If $\Delta \rightarrow \infty$, the exponent tends to 1, and one again needs quantitative Diophantine input.

Problem 15.1 (Noisy copies with growing error complexity). *Extend the noisy online-copy lower bound to regimes where $\Delta = \Delta_N \rightarrow \infty$, or where the mismatch sets have low compressive complexity rather than bounded interval complexity.*

A quantitative moving-period Ridout theorem would likely be necessary even for moderately growing Δ_N .

16. TRANSCENDENCE MEASURES FROM DENSE COMPRESSION PROFILES

The compression criterion proved above is qualitative: if a non-rational number has logarithmically short online-copy parsings infinitely often, then it is transcendental. One may ask for quantitative transcendence measures when such parsings occur on sufficiently dense scales.

Suppose that

$$\text{oc}(a_1 a_2 \cdots a_{N_m}) \leq C \log N_m$$

for an infinite sequence N_m , and that the sequence is not too sparse, for instance

$$N_{m+1} \leq N_m^A$$

for some fixed A . The proof of the compression criterion then produces rational approximations at controlled heights and with controlled gaps. This is the type of input from which transcendence measures may sometimes be extracted.

Problem 16.1 (Compression-to-measure principle). *Develop a theorem converting dense logarithmic online-copy compression profiles into explicit transcendence measures for*

$$\xi_{\mathbf{a},b} = \sum_{n \geq 1} a_n b^{-n}.$$

The refined Diophantine exponent framework suggests that such a principle should also have noisy variants.

17. SCOPE

The exact online-copy theorem gives a strong lower bound for a previous-copy model and hence for LZ77-type parsings. It does not by itself imply superlogarithmic lower bounds for arbitrary grammar compression, straight-line programs, bidirectional macro schemes, or smallest string attractors.

The effective finite-prefix method gives weaker lower bounds for offline repetitiveness measures:

$$\gg_{\alpha} \frac{(\log N)^{1/11}}{(\log \log N)^{4/11}}.$$

These bounds are new in nature relative to the online proof, because they pass through linear extension complexity and Bugeaud's finite-prefix theorem, not through phrase-boundary growth.

The results are not normality statements. They do not imply digit equidistribution, nor any explicit digit-frequency estimate for a specific algebraic irrational number.

The central open quantitative question is to improve

$$\text{oc}(w_N(\alpha)) = \omega(\log N)$$

to an explicit lower bound such as

$$\text{oc}(w_N(\alpha)) \gg \log N f(N), \quad f(N) \rightarrow \infty,$$

or even

$$\text{oc}(w_N(\alpha)) \gg (\log N)^{1+\varepsilon}.$$

The quantitative repetition profile $\Theta_{\alpha}(T)$ and the moving-period Ridout problem isolate the main obstruction.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, ChatGPT, by OpenAI, was used to assist with mathematical drafting, formalization, review, and editing. This

work is shared as a preliminary AI-assisted mathematical note. The mathematical content may have been only partially reviewed and may contain errors; it should not be treated as peer-reviewed or as a fully verified manuscript.

REFERENCES

- [ABL04] B. Adamczewski, Y. Bugeaud, and F. Luca, *Sur la complexité des nombres algébriques*, C. R. Math. Acad. Sci. Paris **339** (2004), no. 1, 11–14.
- [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.
- [BE08] Y. Bugeaud and J.-H. Evertse, *On two notions of complexity of algebraic numbers*, Acta Arith. **133** (2008), no. 3, 221–250.
- [Bug08] Y. Bugeaud, *An explicit lower bound for the block complexity of an algebraic number*, Rend. Lincei Mat. Appl. **19** (2008), 229–235.
- [ES02] J.-H. Evertse and H. P. Schlickewei, *A quantitative version of the Absolute Subspace Theorem*, J. Reine Angew. Math. **548** (2002), 21–127.
- [KP18] D. Kempa and N. Prezza, *At the roots of dictionary compression: string attractors*, Proceedings of the 50th Annual ACM Symposium on Theory of Computing, 2018, 827–840.
- [KNP23] T. Kociumaka, G. Navarro, and N. Prezza, *Toward a definitive compressibility measure for repetitive sequences*, IEEE Trans. Inform. Theory **69** (2023), no. 4, 2074–2092.
- [Ngu26] Q.-K. Nguyen, *Transcendence and measures via the refined Diophantine exponent*, arXiv:2605.30606, 2026.
- [Pre17] N. Prezza, *String attractors*, arXiv:1709.05314, 2017.
- [Rid57] D. Ridout, *Rational approximations to algebraic numbers*, Mathematika **4** (1957), 125–131.
- [ZL77] J. Ziv and A. Lempel, *A universal algorithm for sequential data compression*, IEEE Trans. Inform. Theory **23** (1977), no. 3, 337–343.