

**FINITE STAMMERING PROFILES AND TUBE BARRIERS  
FOR ALGEBRAIC DIGIT EXPANSIONS**

LUCA BLANCHI

---

*Date:* June 2026.

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. Write

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

for the prefix of length  $N$  of the base- $b$  expansion of  $\alpha$ .

This note isolates the finite repetition obstruction behind online previous-copy compression of algebraic digit expansions. For an infinite word  $\mathbf{a} = a_1 a_2 \cdots$ , define its finite stammering profile by

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

For the base- $b$  digit sequence of an algebraic irrational  $\alpha$ , we prove

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

The proof uses the Adamczewski–Bugeaud Diophantine exponent criterion in the unbounded-displacement case and Ridout’s theorem in the bounded-displacement case.

We then prove a quantitative bridge between  $\Theta_{\mathbf{a}}$  and online previous-copy parsing:

$$\text{oc}(a_1 \cdots a_N) \geq \frac{\log(N/T)}{\log(1 + \max\{\Theta_{\mathbf{a}}(T), 1/(T-1)\})} \quad (N > T \geq 2).$$

As a consequence one recovers

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

The second part studies the Diophantine shape of strong repetitions. A repetition with displacement  $d$  and repeated length  $L$  gives an approximation governed by

$$|\alpha b^r (b^d - 1)|_{\mathbb{Z}} \ll b^{-L},$$

where

$$|x|_{\mathbb{Z}} := \min_{m \in \mathbb{Z}} |x - m|.$$

We prove a subcritical barrier: for fixed  $\varepsilon > 0$  and  $0 < \lambda < 1$ , there are only finitely many repetitions satisfying

$$L \geq \varepsilon s, \quad d \leq (1 - \lambda)L.$$

The remaining critical regime reduces to a linear-height exponential tube problem:

$$|\alpha b^r (b^d - 1)|_{\mathbb{Z}} < \theta^d, \quad r \leq Cd.$$

For fixed  $C$  and  $\theta$ , this tube problem has only finitely many solutions, by the Subspace Theorem. We formulate the corresponding effective threshold problem and show that an exponential, respectively double-exponential, effective bound would imply

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

respectively

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

No effective bound of this form is proved here.

## 1. INTRODUCTION

The base- $b$  expansion of an algebraic irrational number is expected to have little repetitive structure. A theorem of Adamczewski and Bugeaud shows that its factor complexity cannot be bounded linearly: if  $p(n)$  denotes the number of distinct length- $n$  blocks in the expansion, then

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

Their method is based on a Diophantine principle: sufficiently strong repetitions in a base- $b$  expansion yield rational approximations that are too good for an algebraic irrational number.

In a previous note, this principle was applied to online previous-copy compression. In that model, a finite word  $W$  is parsed from left to right into phrases. Each phrase is either a literal symbol or an exact copy whose source starts earlier in the word; overlap between source and target is allowed. Let  $\text{oc}(W)$  be the minimum number of phrases in such a parsing. The main result was

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

for every algebraic irrational  $\alpha$ . Consequently every standard LZ77-type previous-factor parsing of  $w_N(\alpha)$  has  $\omega(\log N)$  phrases.

The present note packages the repetition obstruction in a finite profile  $\Theta_\alpha(T)$ , proves the exact bridge from this profile to previous-copy complexity, and identifies the Diophantine obstruction behind possible quantitative improvements.

The central quantitative issue is the following. A repetition

$$\mathbf{a}[r+1, r+L] = \mathbf{a}[r+d+1, r+d+L]$$

gives a rational approximation with denominator dividing

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

or equivalently a small distance to an integer,

$$\left| \alpha b^r (b^d - 1) \right|_{\mathbb{Z}} \ll b^{-L}.$$

The qualitative tube finiteness proved below is a direct consequence of the Subspace Theorem. The point of formulating it separately is that an effective threshold for this tube problem would yield explicit quantitative previous-copy lower bounds.

## 2. DIOPHANTINE INPUTS

Let

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

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

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

We use the following form of the Adamczewski–Bugeaud repetition criterion.

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

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

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

$$\sum_{n \geq 1} a_n b^{-n}$$

is rational or transcendental.

We also use Ridout’s theorem in the following form.

**Theorem 2.2** (Ridout). *Let  $S$  be a finite set of rational primes, let  $\xi$  be algebraic, and let  $\varepsilon > 0$ . Then there are only finitely many rational numbers  $P/Q$ , written in lowest terms, such that all prime divisors of  $Q$  belong to  $S$  and*

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

Finally, we use the qualitative Subspace Theorem over a number field. We shall apply it with  $K = \mathbb{Q}(\alpha)$ , with one distinguished real place corresponding to the given embedding of  $\alpha$ , and with the finite places of  $K$  above the rational primes dividing  $b$ .

**Theorem 2.3** (Subspace Theorem, qualitative form). *Let  $K$  be a number field, let  $S$  be a finite set of places of  $K$ , and for each  $v \in S$  let*

$$L_{v,1}, \dots, L_{v,n}$$

be linearly independent linear forms in  $n$  variables with coefficients in  $K$ . For every  $\delta > 0$ , the solutions  $x \in K^n \setminus \{0\}$  of

$$\prod_{v \in S} \prod_{i=1}^n |L_{v,i}(x)|_v < H(x)^{-\delta}$$

lie in finitely many proper  $K$ -linear subspaces of  $K^n$ .

In our applications, the points  $x$  are rational integer vectors. Hence, after passing to an infinite subsequence inside one of the exceptional  $K$ -subspaces, the rational integer points lie in a proper rational subspace. Equivalently, one may take a non-trivial rational linear relation among them.

We use normalized absolute values. If  $q \in \mathbb{Q}^\times$ , then for a rational prime  $p$ ,

$$\prod_{v|p} |q|_v = |q|_p.$$

In particular, for the finite places above primes dividing  $b$ , the contribution of  $b^r$  is  $b^{-r}$ , and the contribution of  $b^{r+d}$  is  $b^{-(r+d)}$ .

### 3. THE FINITE STAMMERING PROFILE

**Definition 3.1.** Let  $\mathbf{a} = a_1a_2\cdots$  be an infinite word. For  $T \geq 1$ , define

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

If the set is empty, we put  $\Theta_{\mathbf{a}}(T) = 0$ .

If  $\mathbf{a}$  is the base- $b$  digit sequence of  $\alpha$ , write  $\Theta_{\alpha}(T)$ .

The quantity  $\Theta_{\mathbf{a}}(T)$  measures the strongest exact repetition, in the Adamczewski–Bugeaud sense, whose endpoint is at least  $T$ .

**Theorem 3.2** (Qualitative decay). 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

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

*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

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

for every  $m$ . Put

$$d_m = s_m - r_m.$$

If  $d_m \rightarrow \infty$  along an infinite subsequence, the exact repetition criterion implies that  $\alpha$  is rational or transcendental, contradicting the hypothesis that it is algebraic irrational.

Thus  $d_m$  is bounded along an infinite subsequence. Passing to a further subsequence, assume

$$d_m = d$$

is fixed. The equality above gives

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

Hence the block

$$a_{r_m+1} \cdots a_{t_m}$$

is periodic with period  $d$ .

Let  $\beta_m$  be the rational number whose base- $b$  expansion agrees with  $\alpha$  through position  $r_m$ , and then continues periodically with period

$$a_{r_m+1} \cdots a_{r_m+d}.$$

Then  $\alpha$  and  $\beta_m$  agree through at least the first  $t_m$  digits, so

$$|\alpha - \beta_m| \ll_b b^{-t_m}.$$

Write

$$\beta_m = P_m/Q_m$$

in lowest terms. Since  $\beta_m$  has preperiod  $r_m$  and period  $d$ , its denominator divides

$$b^{r_m}(b^d - 1).$$

Thus every prime divisor of  $Q_m$  belongs to the fixed finite set

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

and

$$Q_m \ll_{b,d} b^{r_m}.$$

If  $r_m$  is bounded along an infinite subsequence, only finitely many  $\beta_m$  occur. Since  $t_m \rightarrow \infty$ , one of them agrees with  $\alpha$  to arbitrarily many digits, hence equals  $\alpha$ , contradicting irrationality.

Thus  $r_m \rightarrow \infty$ . Since

$$s_m = r_m + d$$

and

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

there is  $\varepsilon' > 0$  such that, for all sufficiently large  $m$ ,

$$t_m \geq (1 + \varepsilon')r_m.$$

Consequently

$$b^{-t_m} \ll Q_m^{-1-\varepsilon''}$$

for some  $\varepsilon'' > 0$ . Hence

$$|\alpha - \beta_m| < Q_m^{-1-\varepsilon''}$$

for infinitely many rationals  $P_m/Q_m$  whose denominators have all prime factors in  $S$ . This contradicts Ridout's theorem.  $\square$

#### 4. ONLINE PREVIOUS-COPY PARSING AND THE PROFILE BRIDGE

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

**Definition 4.1.** *An online previous-copy parsing of  $W$  is a factorization*

$$W = F_1 \cdots F_z$$

*with boundaries*

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

*such that each phrase*

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

*is either:*

*i a literal, in which case  $|F_j| = 1$ , or*

ii a copy: writing  $s = n_{j-1}$ ,  $t = n_j$ ,  $L = t - s$ , there exists  $p \leq s$  such that

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

The source may overlap the target. Let  $\text{oc}(W)$  be the minimum possible number of phrases.

**Theorem 4.2** (Profile-to-compression bridge). *Let  $\mathbf{a} = a_1 a_2 \cdots$  be an infinite word and let*

$$W_N = a_1 \cdots a_N.$$

For  $T \geq 2$ , define

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

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

$$n_q < T.$$

Since  $T \geq 2$  and  $n_1 = 1$ , such  $q$  exists. We have  $n_q < T$ .

Consider any phrase with index  $j > q$ .

If it is copied, put  $s = n_{j-1}$  and  $t = n_j$ . The copy condition gives a triple  $0 \leq r < s < t$  such that

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

Since  $t = n_j \geq T$ , the definition of  $\Theta_{\mathbf{a}}(T)$  gives

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

If the phrase is a literal, then

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

If  $j = q + 1$ , then  $n_j \geq T$  and  $n_{j-1} = n_q < T$ . Since the phrase is literal,  $n_j = n_q + 1$ , hence  $n_q = T - 1$ . Thus

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

If  $j > q + 1$ , then  $n_{j-1} \geq T$ , and again

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

Therefore, for every  $j > q$ ,

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

Multiplying,

$$\frac{N}{n_q} = \prod_{j=q+1}^z \frac{n_j}{n_{j-1}} \leq (1 + \eta_{\mathbf{a}}(T))^{z-q} \leq (1 + \eta_{\mathbf{a}}(T))^z.$$

Since  $n_q < T$ ,

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

Taking logarithms gives the result.  $\square$

**Corollary 4.3.** *Let  $b \geq 2$  and let*

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

*Then*

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

*Proof.* Let  $M > 0$ . Choose  $\varepsilon > 0$  such that

$$\frac{1}{\log(1 + \varepsilon)} > 2M.$$

By the decay theorem, choose  $T$  so large that

$$\Theta_{\alpha}(T) \leq \varepsilon \quad \text{and} \quad \frac{1}{T-1} \leq \varepsilon.$$

Then, for all  $N > T$ ,

$$\text{oc}(w_N(\alpha)) \geq \frac{\log(N/T)}{\log(1 + \varepsilon)}.$$

For  $N$  sufficiently large this is at least  $M \log N$ . Since  $M$  is arbitrary, the result follows.  $\square$

## 5. STRONG REPETITIONS AND RATIONAL APPROXIMANTS

Let

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

be an exact repetition, with

$$0 \leq r < s, \quad d = s - r, \quad L \geq 1.$$

Thus

$$s = r + d.$$

The repetition says

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

Let  $\beta$  be the rational number whose base- $b$  expansion agrees with  $\alpha$  through position  $r$ , and thereafter is periodic with period

$$a_{r+1} \cdots a_{r+d}.$$

Then  $\beta$  has denominator dividing

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

and  $\alpha$  and  $\beta$  agree through at least position

$$r + d + L = s + L.$$

Hence, for some integer  $P$ ,

$$\beta = \frac{P}{b^r(b^d - 1)}$$

and

$$\left| \alpha - \frac{P}{b^r(b^d - 1)} \right| \ll_b b^{-(r+d+L)}.$$

Equivalently,

$$\left| \alpha b^r(b^d - 1) - P \right| \ll_b b^{-L}.$$

Thus

$$\left| \alpha b^r(b^d - 1) \right|_{\mathbb{Z}} \ll_b b^{-L}.$$

We call the repetition  $\varepsilon$ -strong if

$$L \geq \varepsilon s = \varepsilon(r + d).$$

## 6. A SUBCRITICAL BARRIER

The following result eliminates repetitions in which the two copies overlap too much relative to the repeated length.

**Theorem 6.1** (Subcritical barrier). *Let  $b \geq 2$ , and let*

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

*Fix  $\varepsilon > 0$  and  $0 < \lambda < 1$ . Then there are only finitely many exact repetitions*

$$\mathbf{a}[r + 1, r + L] = \mathbf{a}[r + d + 1, r + d + L]$$

*such that*

$$L \geq \varepsilon(r + d)$$

*and*

$$d \leq (1 - \lambda)L.$$

*Proof.* Suppose infinitely many such repetitions exist. For each one, choose  $P \in \mathbb{Z}$  such that

$$|\alpha Q - P| \ll_b b^{-L}, \quad Q = b^r(b^d - 1).$$

Let

$$K = \mathbb{Q}(\alpha).$$

Let  $v_0$  be the real place of  $K$  corresponding to the given embedding of  $\alpha$ , and let  $S$  consist of  $v_0$  together with all finite places of  $K$  lying above rational primes dividing  $b$ .

We apply the Subspace Theorem over  $K$  to the rational integer vector

$$X = (X_1, X_2) = (P, Q).$$

At the distinguished real place  $v_0$ , use

$$L_{v_0,1}(X) = \alpha X_2 - X_1, \quad L_{v_0,2}(X) = X_2.$$

At every finite place  $v \in S$ , use the coordinate forms

$$L_{v,1}(X) = X_1, \quad L_{v,2}(X) = X_2.$$

At  $v_0$ ,

$$|L_{v_0,1}(X)|_{v_0} \ll_b b^{-L}, \quad |L_{v_0,2}(X)|_{v_0} = Q \asymp_b b^{r+d}.$$

For the finite places above primes dividing  $b$ ,

$$\prod_{v|b} |X_2|_v = b^{-r},$$

because

$$Q = b^r(b^d - 1) \quad \text{and} \quad \gcd(b, b^d - 1) = 1.$$

Also

$$\prod_{v|b} |X_1|_v \leq 1.$$

Therefore

$$\prod_{v \in S} \prod_{i=1}^2 |L_{v,i}(X)|_v \ll_{K,b} b^{-L} b^{r+d} b^{-r} = b^{d-L}.$$

By the subcritical hypothesis,

$$d - L \leq -\lambda L.$$

Thus

$$\prod_{v,i} |L_{v,i}(X)|_v \ll_{K,b} b^{-\lambda L}.$$

Since

$$L \geq \varepsilon(r + d)$$

and

$$H(X) \asymp_{\alpha,b} b^{r+d},$$

there is  $\delta > 0$  such that, for all sufficiently large repetitions,

$$\prod_{v,i} |L_{v,i}(X)|_v < H(X)^{-\delta}.$$

By the Subspace Theorem, all such  $X$  lie in finitely many proper  $K$ -subspaces of  $K^2$ . Since the points  $X$  are rational integer points, after passing to an infinite subsequence they lie in a proper rational line

$$uX_1 + vX_2 = 0, \quad u, v \in \mathbb{Q},$$

not both zero. If  $u = 0$ , then  $Q = 0$ , impossible. Hence

$$\frac{P}{Q} = -\frac{v}{u}$$

is constant on this subsequence. But

$$\left| \alpha - \frac{P}{Q} \right| \ll_b b^{-(r+d+L)} \rightarrow 0.$$

Therefore  $\alpha = -v/u \in \mathbb{Q}$ , contradiction.  $\square$

## 7. THE LINEAR-HEIGHT EXPONENTIAL TUBE PROBLEM

The subcritical theorem leaves the regime

$$d > (1 - \lambda)L.$$

Together with

$$L \geq \varepsilon(r + d),$$

this gives

$$r + d \leq L/\varepsilon < \frac{d}{\varepsilon(1 - \lambda)}.$$

Hence

$$r \leq \left( \frac{1}{\varepsilon(1 - \lambda)} - 1 \right) d.$$

Moreover

$$\left| \alpha b^r (b^d - 1) \right|_{\mathbb{Z}} \ll_b b^{-L}.$$

Since  $L \geq \varepsilon d$ , for all sufficiently large  $d$  this implies

$$\left| \alpha b^r (b^d - 1) \right|_{\mathbb{Z}} < b^{-\varepsilon d/2}.$$

This motivates the following auxiliary problem.

**Definition 7.1.** For  $C > 0$  and  $0 < \theta < 1$ , let  $D_{\alpha, b}(C, \theta)$  be the least integer  $D$ , if it exists, such that there are no solutions

$$r, d \geq 1, \quad r \leq Cd, \quad d \geq D,$$

to

$$\left| \alpha b^r (b^d - 1) \right|_{\mathbb{Z}} < \theta^d.$$

The next theorem proves that  $D_{\alpha, b}(C, \theta)$  is finite. No effective upper bound is obtained.

**Theorem 7.2** (Qualitative tube finiteness). Let  $b \geq 2$ , let

$$\alpha \in \overline{\mathbb{Q}} \setminus \mathbb{Q}$$

be real, let  $C > 0$ , and let  $0 < \theta < 1$ . Then there are only finitely many pairs

$$r, d \geq 1, \quad r \leq Cd,$$

such that

$$\left| \alpha b^r (b^d - 1) \right|_{\mathbb{Z}} < \theta^d.$$

*Proof.* Suppose infinitely many such pairs exist. For each pair choose  $m \in \mathbb{Z}$  such that

$$|\alpha b^r (b^d - 1) - m| < \theta^d.$$

Set

$$X = (X_1, X_2, X_3) = (b^{r+d}, b^r, m).$$

Let

$$K = \mathbb{Q}(\alpha).$$

Let  $v_0$  be the distinguished real place corresponding to the given embedding of  $\alpha$ , and let  $S$  consist of  $v_0$  together with all finite places of  $K$  above rational primes dividing  $b$ .

At  $v_0$ , use the forms

$$\begin{aligned} L_{v_0,1}(X) &= \alpha X_1 - \alpha X_2 - X_3, \\ L_{v_0,2}(X) &= X_1, \quad L_{v_0,3}(X) = X_2. \end{aligned}$$

At every finite place  $v \in S$ , use the coordinate forms

$$L_{v,1}(X) = X_1, \quad L_{v,2}(X) = X_2, \quad L_{v,3}(X) = X_3.$$

The forms are linearly independent at each place.

At the distinguished real place,

$$\begin{aligned} |L_{v_0,1}(X)|_{v_0} &< \theta^d, \\ |L_{v_0,2}(X)|_{v_0} &= b^{r+d}, \quad |L_{v_0,3}(X)|_{v_0} = b^r. \end{aligned}$$

At the finite places above primes dividing  $b$ ,

$$\prod_{v|b} |X_1|_v = b^{-(r+d)}, \quad \prod_{v|b} |X_2|_v = b^{-r},$$

and

$$\prod_{v|b} |X_3|_v \leq 1.$$

Therefore

$$\prod_{v \in S} \prod_{i=1}^3 |L_{v,i}(X)|_v \ll_{K,b} \theta^d.$$

Since

$$H(X) \asymp_{\alpha,b} b^{r+d}$$

and

$$r + d \leq (C + 1)d,$$

there exists  $\delta > 0$  such that, for all sufficiently large solutions,

$$\theta^d \ll H(X)^{-\delta}.$$

Hence

$$\prod_{v,i} |L_{v,i}(X)|_v < H(X)^{-\delta/2}$$

for all sufficiently large solutions.

By the Subspace Theorem, all such  $X$  lie in finitely many proper  $K$ -subspaces. Passing to an infinite subsequence, assume the rational integer points  $X$  lie in a fixed proper rational hyperplane

$$uX_1 + vX_2 + wX_3 = 0$$

with  $u, v, w \in \mathbb{Q}$ , not all zero.

If  $w = 0$ , then

$$ub^{r+d} + vb^r = 0.$$

Dividing by  $b^r$ ,

$$ub^d + v = 0.$$

If  $u = 0$ , then  $v = 0$ , contradiction. Thus  $d$  is fixed. Since  $r \leq Cd$ , only finitely many  $r$  occur, contradiction.

If  $w \neq 0$ , then

$$m = -\frac{u}{w}b^{r+d} - \frac{v}{w}b^r.$$

Substituting into

$$|\alpha b^{r+d} - \alpha b^r - m| < \theta^d$$

gives

$$\left| \left( \alpha + \frac{u}{w} \right) b^{r+d} + \left( -\alpha + \frac{v}{w} \right) b^r \right| < \theta^d.$$

Divide by  $b^{r+d}$ :

$$\left| \left( \alpha + \frac{u}{w} \right) + \left( -\alpha + \frac{v}{w} \right) b^{-d} \right| < \theta^d b^{-(r+d)}.$$

If  $d$  is bounded, then  $r \leq Cd$  is bounded, so only finitely many solutions occur. Thus along an infinite subsequence we have  $d \rightarrow \infty$ . Letting  $d \rightarrow \infty$ , the right-hand side tends to 0, and the term involving  $b^{-d}$  also tends to 0. Hence

$$\alpha + \frac{u}{w} = 0.$$

Thus  $\alpha = -u/w \in \mathbb{Q}$ , contradiction.  $\square$

**Corollary 7.3** (Critical repetitions are finite). *Fix  $\varepsilon > 0$  and  $0 < \lambda < 1$ . There are only finitely many exact repetitions satisfying*

$$L \geq \varepsilon(r + d)$$

and

$$d > (1 - \lambda)L.$$

*Proof.* As observed above, all sufficiently large such repetitions give solutions of

$$\left| \alpha b^r (b^d - 1) \right|_{\mathbb{Z}} < b^{-\varepsilon d/2}$$

with

$$r \leq \left( \frac{1}{\varepsilon(1 - \lambda)} - 1 \right) d.$$

The qualitative tube finiteness theorem applies.  $\square$

## 8. CONDITIONAL QUANTITATIVE CONSEQUENCES

The preceding finiteness results are qualitative. We now record exactly what kind of effective input would imply explicit lower bounds for online previous-copy complexity.

For  $\varepsilon > 0$ , let  $R_{\alpha,b}(\varepsilon)$  be any threshold such that there are no repetitions

$$\mathbf{a}[r+1, r+L] = \mathbf{a}[r+d+1, r+d+L]$$

with endpoint

$$r+d+L \geq R_{\alpha,b}(\varepsilon)$$

and

$$L \geq \varepsilon(r+d).$$

The qualitative results above imply that  $R_{\alpha,b}(\varepsilon) < \infty$  for each fixed  $\varepsilon > 0$ , but they do not provide an effective bound.

**Lemma 8.1.** *If*

$$R_{\alpha,b}(\varepsilon) \leq F(\varepsilon)$$

*for all sufficiently small  $\varepsilon$ , then*

$$\Theta_{\alpha}(T) \leq \varepsilon \quad (T \geq F(\varepsilon)).$$

*Proof.* A triple counted by  $\Theta_{\alpha}(T)$  with

$$\frac{t}{s} - 1 \geq \varepsilon$$

has

$$L = t - s \geq \varepsilon s.$$

Writing  $s = r + d$ , it is a repetition with endpoint  $t = r + d + L \geq T$ . If  $T \geq F(\varepsilon) \geq R_{\alpha,b}(\varepsilon)$ , no such repetition exists.  $\square$

**Theorem 8.2** (Exponential threshold implies power-logarithmic compression). *Assume that there exist constants  $K, A > 0$  such that, for all sufficiently small  $\varepsilon > 0$ ,*

$$R_{\alpha,b}(\varepsilon) \leq \exp(K\varepsilon^{-A}).$$

*Then*

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

*Proof.* Set

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

Choose

$$\varepsilon = \left( \frac{2K}{\log T} \right)^{1/A}.$$

Then

$$\exp(K\varepsilon^{-A}) = T^{1/2} < T.$$

Thus

$$\Theta_{\alpha}(T) \leq \varepsilon.$$

For  $N$  large, also  $1/(T-1) \leq \varepsilon$ . The profile-to-compression bridge gives

$$\text{oc}(w_N(\alpha)) \geq \frac{\log(N/T)}{\log(1+\varepsilon)} \gg \frac{\log N}{\varepsilon}.$$

Since

$$\varepsilon \asymp (\log N)^{-1/A},$$

we get

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

□

**Theorem 8.3** (Double-exponential threshold). *Assume that, for some  $K, A > 0$ ,*

$$R_{\alpha,b}(\varepsilon) \leq \exp \exp(K\varepsilon^{-A})$$

*for all sufficiently small  $\varepsilon > 0$ . Then*

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

*Proof.* Again take  $T = N^{1/2}$ , and choose

$$\varepsilon = \left( \frac{2K}{\log \log T} \right)^{1/A}.$$

Then

$$\exp \exp(K\varepsilon^{-A}) < T$$

for  $N$  large. The bridge gives

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

□

**Remark 8.4.** *The exponential or double-exponential hypotheses above are not proved in this note. They would require effective forms of the Diophantine finiteness results used in the subcritical and tube arguments. The qualitative Subspace Theorem does not provide such thresholds.*

## 9. THE EFFECTIVE TUBE PROBLEM

The critical part of the threshold  $R_{\alpha,b}(\varepsilon)$  is governed by the linear-height exponential tube problem.

**Problem 9.1** (Effective linear-height exponential tube problem). *Let*

$$\alpha \in \overline{\mathbb{Q}} \setminus \mathbb{Q}, \quad b \geq 2.$$

*Find an explicit upper bound for*

$$D_{\alpha,b}(C, \theta),$$

*where  $D_{\alpha,b}(C, \theta)$  is the least integer  $D$  such that there are no solutions*

$$r, d \geq 1, \quad r \leq Cd, \quad d \geq D$$

*to*

$$\left| \alpha b^r (b^d - 1) \right|_{\mathbb{Z}} < \theta^d.$$

In the compression application one has

$$C \asymp \varepsilon^{-1}, \quad \theta = b^{-c\varepsilon}.$$

Thus a bound of the form

$$D_{\alpha,b}(C, \theta) \leq \exp\left(K_{\alpha,b} C^A (1 - \theta)^{-B}\right)$$

would imply an exponential threshold for  $R_{\alpha,b}(\varepsilon)$ , up to the corresponding effective subcritical estimate. A double-exponential bound would imply the double-exponential threshold.

## 10. A PARTIAL EFFECTIVE RANGE

Liouville's inequality gives an effective bound only in a range much stronger than the one needed for compression.

**Proposition 10.1** (Liouville range). *Let  $D = \deg(\alpha)$ . If*

$$\theta < b^{-(D-1)(C+1)},$$

*then  $D_{\alpha,b}(C, \theta)$  is effectively bounded in terms of  $\alpha, b, C, \theta$ .*

*Proof.* Suppose

$$\left| \alpha b^r (b^d - 1) \right|_{\mathbb{Z}} < \theta^d$$

with  $r \leq Cd$ . Put

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

Then

$$q \leq b^{(C+1)d}.$$

For some integer  $p$ ,

$$\left| \alpha - \frac{p}{q} \right| < \frac{\theta^d}{q}.$$

Liouville's inequality gives an effective constant  $c_\alpha > 0$  such that

$$\left| \alpha - \frac{p}{q} \right| \geq c_\alpha q^{-D}$$

for all rationals  $p/q$ . Therefore

$$\theta^d > c_\alpha q^{1-D}.$$

Since

$$q^{1-D} \geq b^{-(D-1)(C+1)d},$$

the inequality is impossible for all sufficiently large  $d$ , effectively, provided

$$\theta < b^{-(D-1)(C+1)}.$$

□

**Remark 10.2.** *In the compression regime,*

$$C \asymp \varepsilon^{-1}, \quad \theta = b^{-c\varepsilon}.$$

*For small  $\varepsilon$ , the Liouville condition is far stronger than what is available. Thus Roth–Ridout–Subspace type input is genuinely needed.*

## 11. SCOPE

The results in this note are qualitative except for the elementary Liouville range and the conditional implications in Section 8. In particular, we do not prove

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

or

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

Such bounds would follow from effective estimates for the repetition threshold  $R_{\alpha,b}(\varepsilon)$ , or more specifically from effective bounds for the linear-height exponential tube problem.

The main unconditional conclusions are:

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

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

the subcritical finiteness theorem, and the qualitative finiteness of the tube problem.

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

- [AB07] B. Adamczewski and Y. Bugeaud, *On the complexity of algebraic numbers I. Expansions in integer bases*, Ann. of Math. 2 **165** 2007, 547–565.
- [ABL04] B. Adamczewski, Y. Bugeaud, and F. Luca, *Sur la complexité des nombres algébriques*, C. R. Math. Acad. Sci. Paris **339** 2004, 11–14.
- [Bug08] Y. Bugeaud, *An explicit lower bound for the block complexity of an algebraic number*, Rend. Lincei Mat. Appl. **19** 2008, 229–235.
- [BE08] Y. Bugeaud and J.-H. Evertse, *On two notions of complexity of algebraic numbers*, Acta Arith. **133** 2008, 221–250.
- [EF13] J.-H. Evertse and R. Ferretti, *A further improvement of the Quantitative Subspace Theorem*, Ann. of Math. 2 **177** 2013, 513–590.
- [ES02] J.-H. Evertse and H. P. Schlickewei, *A quantitative version of the Absolute Subspace Theorem*, J. Reine Angew. Math. **548** 2002, 21–127.
- [Rid57] D. Ridout, *Rational approximations to algebraic numbers*, Mathematika **4** 1957, 125–131.