The repetition threshold for ternary rich words

Jarkko Peltomäki

Information Technology
Åbo Akademi University

6.5.2025

Joint work with J. D. Currie and L. Mol

Contents

- Repetition threshold for rich words
- Known results
- Outline of the proof in the ternary case
- Recent progress

Critical Exponent and Repetition Threshold

Definition

The *critical exponent* $\mathrm{ce}(\mathbf{w})$ of an infinite word $\mathbf{w}$ is

$$\mathrm{ce}(\mathbf{w}) = \sup\{k \in \mathbb{Q} : \mathbf{w} \text{ contains a } k\text{-power}\}.$$

Let $\mathcal{L}$ be a nonempty set of infinite words.

Definition

The *repetition threshold* $\mathrm{RT}(\mathcal{L})$ of $\mathcal{L}$ is

$$\mathrm{RT}(\mathcal{L}) = \inf\{\mathrm{ce}(\mathbf{w}) : \mathbf{w} \in \mathcal{L}\}.$$

For example, if $\mathcal{S}$ is the set of Sturmian words, then $\mathrm{RT}(\mathcal{S}) = 2 + \varphi \approx 3.6180$, where φ is the Golden ratio. The repetition threshold is achieved by the Fibonacci word.

Question

In this talk, we are interested in the following question.

Question

What is $\mathrm{RT}(R_k)$ *for the set* R_k *of rich infinite words over a k-letter alphabet?*

Rich Words

Definition

A finite word u is *rich* if it contains $|u| + 1$ distinct palindromes (including the empty word).

For example, 0010110 is rich as it contains the palindromes ε, 0, 00, 1, 010, 101, 11, 0110. The word 00101100 is not rich as it contains the same palindromes as the preceding word.

Every word of length n contains at most $n + 1$ palindromes as appending a new letter to a word can introduce at most one new palindromic factor.

Definition

An infinite word $\mathbf{w}$ is *rich* if its factors are rich.

Examples of Infinite Rich Words

- Sturmian words
- Episturmian words
- Codings of symmetric interval exchange transformations

The Thue-Morse word is not rich.

Some First Observations

Observation (Vesti (2017))

The k-bonacci words are rich, so their critical exponent gives an upper bound. More precisely, $\mathrm{RT}(R_k) \leq 2 + 1/(\varphi_k - 1)$, *where* φ_k *is the generalized Golden ratio. The upper bound tends to* 3 *as* $k \to \infty$.

Observation (Pelantová, Starosta (2013))

Every infinite rich word contains infinitely many overlaps (factors of the form axaxa for a word x and a letter a), so $\mathrm{RT}(R_k) > 2$ *for all* k.

First Results

Theorem (Baranwal, Shallit (2019))

There is a rich binary word with critical exponent $2 + \sqrt{2}/2 \approx 2.7071$, *so* $\mathrm{RT}(R_2) \leq 2 + \sqrt{2}/2$.

Baranwal and Shallit conjectured that the upper bound is optimal.

Theorem (Currie, Mol, Rampersad (2020))

$\mathrm{RT}(R_2) = 2 + \sqrt{2}/2$

Answer in the Ternary Case

Theorem (Currie, Mol, P. (2024))

$\mathrm{RT}(R_3) = 1 + 1/(3 - \mu) \approx 2.2588$, *where* μ *is the unique real root of the polynomial* $x^3 - 2x^2 - 1$.

The result follows from a structure theorem for $16/7$-free ternary rich words.

Morphisms and a Transducer

Define morphisms f, g and a transducer τ by

$$
\begin{aligned}
f(0) &= 01 & \qquad g(0) &= 20 & \qquad \tau(0) &= 001/002 \\
f(1) &= 022 & \qquad g(1) &= 21 & \qquad \tau(1) &= 00101101/00202202 \\
f(2) &= 02 & \qquad g(2) &= 2 & \qquad \tau(2) &= 001011010010101101/ \\
& & & & & \quad 002022020020202202
\end{aligned}
$$

The Structure Theorem

Theorem (Currie, Mol, P. (2024))

Let $\mathbf{z} \in \{0,1,2\}^\omega$ *be a* $16/7$*-free rich word. For all* $n \geq 0$, $\mathbf{z}$ *has a suffix* $\tau(g(f^n(\mathbf{x}_n)))$ *(possibly after a permutation of the letters) for some* $\mathbf{x}_n \in \{0,1,2\}^\omega$.

In particular $\mathbf{z}$ contains all factors obtained from $\tau(g(f^\omega(0)))$ by permuting the letters.

Proposition

The word $\tau(g(f^\omega(0)))$ *is rich and its critical exponent equals* $1 + 1/(3-\mu)$.

Proof Outline

The proof of the structure theorem consits of three steps: analysis of two 'outer' layers (τ and g) and a more intricate 'inner' layer analysis (f). The proof spans 20 pages, and involves frequent use of backtracking searches to exclude forbidden factors.

Here, we sketch the analysis of the first outer layer.

First Outer Layer Analysis

Let **z** be a ternary rich word that avoids 16/7 powers. The goal is to show that a suffix of **z** is of the form $\tau(\mathbf{y})$ for some ternary infinite word **y**.

A backtracking search shows that a longest 16/7-free ternary rich word avoiding the factors 001002, 112110, 220221 has length 388. By permuting letters and removing a prefix, we may assume that **z** has prefix 001002.

First Outer Layer Analysis

Lemma

The word $\mathbf{z}$ *contains neither* 12 *nor* 21*.*

Proof.

Say $\mathbf{z}$ contains 12 or 21. Let p be the shortest prefix of $\mathbf{z}$ that ends in 12 or 21, and let s be the suffix of p of length 2. By the minimality of p, the reversal of s does not appear in p. Thus the only nonempty palindromic suffix of p has length 1. Since p has prefix 001002, every word of length 1 occurs more than once in p. Thus the longest palindromic suffix of p occurs multiple times in p. As the longest palindromic suffix of a rich word is unioccurrent, it follows that p is not rich. A contradiction. □

First Outer Layer Analysis

Let us then find the complete first returns to 00 in **z**. As 000 does not occur, we may focus on returns that begin with 001. Next, we enumerate all possible continuations of 001 and exclude all but few possibilities. Recall that 12 and 21 do not occur.

- 0010: OK
- 0011: Backtracking shows that the longest 16/7-free ternary rich word with prefix 0011 has length 498.
- 0012: Has suffix 12.
- 00100: OK, is a complete first return.
- 00101: OK
- 00102: Backtracking shows that the longest 16/7-free ternary rich word with prefix 102 has length 152.
- 001010: Has 5/2-power 01010 as a suffix.
- 001011: OK
- 001012: Has suffix 12.

First Outer Layer Analysis

- 0010110: OK
- 0010111: Has suffix 111.
- 0010112: Has suffix 12.
- 00101100: Has a permutation of 0011 as a suffix.
- 00101101: OK
- 00101102: Has 102 as suffix.
- 001011010: OK
- 001011011: Has 7/3-power 1011011 as a suffix.
- 001011012: Has suffix 12.
- 0010110100: OK, is a complete first return.
- 0010110101: Has 5/2-power 10101 as a suffix.
- 0010110102: Has 102 as a suffix.

Thus 00 has two complete returns starting with 001: 00100 and 0010110100. There are two complete first returns starting with 002: 00200 and 0020220200.

First Outer Layer Analysis

What remains is to show which complete first returns can appear next to each other to conclude that $\mathbf{z}$ is of the form $\tau(\mathbf{y})$ for a ternary word $\mathbf{y}$. The analysis is similar to what we already saw.

Finding the Critical Exponent

Determining the critical exponent of $\tau(g(f^\omega(0)))$ is straightforward but tedious. Standard methods of determining large repetitions in morphic words do not quite apply due to the transducer τ.

Indeed, due to alternation in τ, periods of odd length are destroyed. E.g., 00 has period 1 but $\tau(00)$ does not have period $|\tau(0)|$: $\tau(00) = 001002$.

We thus need to keep track of large powers in $g(f^\omega(0))$ that have even periods. These powers have a substitutive structure, but some extra work is needed.

4-letter Case

Question

What is the value of $\mathrm{RT}(R_4)$*?*

We do not know and we do not have a candidate for the word achieving $\mathrm{RT}(R_4)$. Backtracking search suggests that $\mathrm{RT}(R_4)$ is approximately 2.117.

Asymptotic Critical Exponent

Definition

The *asymptotic critical exponent* $\mathrm{ce}^*(\mathbf{w})$ of an infinite word $\mathbf{w}$ is

$$\mathrm{ce}^*(\mathbf{w}) = \limsup_{n\to\infty}\{k \in \mathbb{Q} : \mathbf{w} \text{ has a factor of exponent } k \text{ and period } n\}.$$

Note that $\mathrm{ce}(\mathbf{w}) \geq \mathrm{ce}^*(\mathbf{w})$.

Definition

The *asymptotic repetition threshold* $\mathrm{RT}(\mathcal{L})$ of a language $\mathcal{L}$ is

$$\mathrm{RT}^*(\mathcal{L}) = \inf\{\mathrm{ce}^*(\mathbf{w}) : \mathbf{w} \in \mathcal{L}\}.$$

Question

What is the asymptotic repetition threshold $\mathrm{RT}^*(R_k)$ of k-ary rich words?

Asymptotic Critical Exponent

Theorem (Dvořáková, Klouda, Pelantová (2025))

$\mathrm{RT}^*(R_k) = 2$ *for all* k.

Their proof consists of two parts:

1. finding a k-ary rich word $\sigma_k^\omega(0)$ with asymptotic critical exponent E_k^* such that $E_k^* \to 2$ as $k \to \infty$,
2. projecting $\sigma_k^\omega(0)$ to binary alphabet preserving the asymptotic critical exponent and richness.

Asymptotic Critical Exponent

The morphism σ_k and the projection π_k are as follows.

$$\begin{aligned}\sigma\colon 0 &\mapsto 01,\\ 1 &\mapsto 02,\\ &\dots,\\ k-2 &\mapsto 0(k-1)\\ k-1 &\mapsto 0(k-1)(k-1)\end{aligned}$$

$$\pi_k\colon\, i \mapsto 01^i \text{ for every } i \in \{0, 1, \dots, k-1\}$$

Asymptotic Critical Exponent

Interestingly $\mathrm{RT}(R_2)$ is achieved by the word $\pi_2(\sigma_3^\omega(0))$ (Currie, Mol, Rampersad (2020)). Moreover, Dvořáková and Pelantová (2025) show that the word $\tau(g(f^\omega(0)))$ in fact equals $\zeta(\sigma_5^\omega(0))$ for the projection

$$\begin{aligned} \zeta\colon 0 &\mapsto A \\ 1 &\mapsto AD \\ 2 &\mapsto AC \\ 3 &\mapsto ACC \\ 4 &\mapsto ACCBCC, \end{aligned}$$

where $A = 00101101$, $B = 001$, $C = 00202202$, $D = 002$.

Thank You

Thank you for your attention!

J. D. Currie, L. Mol, J. Peltomäki.
The repetition threshold for ternary rich words.
Preprint (2024), arXiv:2409.12068.

Dvořáková, Klouda, Pelantová.
The asymptotic repetition threshold of sequences rich in palindromes.
Eur. J. Combin. **126**:104124 (2025).

Dvořáková, Pelantová.
A note on symmetries of rich sequences with minimum critical exponent
Preprint (2025), arXiv:2501.15135.