The Lexicographically Least Binary Rich Word Achieving the Repetition Threshold

James D. Currie   Narad Rampersad

Submitted: Oct 10, 2023; Accepted: Oct 2, 2024; Published: Dec 27, 2024

Abstract

A word is *rich* if each of its length n factors contains n distinct non-empty palindromes. For a language $\mathcal{L}$, the *repetition threshold* of $\mathcal{L}$ is defined by

$$\mathrm{RT}(\mathcal{L}) = \sup\{k : \text{ every infinite word of } \mathcal{L} \text{ contains a } k\text{-power}\}.$$

Currie et al. (2020) proved that the repetition threshold for binary rich words is $2 + \sqrt{2}/2$. We exhibit the lexicographically least infinite binary rich word attaining this threshold.

Mathematics Subject Classifications: 68R15

1 Introduction

A major branch of combinatorics on words studies words avoiding various powers or patterns. A typical question is whether there exists an infinite word over a certain alphabet avoiding a certain pattern. The earliest known result of this type is by Thue [17], who proved that there is an infinite word over a three-letter alphabet containing no factor of the form $h(xx)$ with h a non-erasing morphism.

We use Σ_n to denote the n-letter alphabet $\Sigma_n = \{0, 1, 2, \ldots, n-1\}$. Let p be an arbitrary finite string. Several generalizations of Thue's result have been explored:

1. Does there exist some n such that there is an infinite word over Σ_n containing no factor of the form $h(p)$ with h a non-erasing morphism?

2. For a fixed n, is there an infinite word over Σ_n containing no factor of the form $h(p)$ with h a non-erasing morphism?

Department of Mathematics and Statistics, University of Winnipeg, 515 Portage Avenue, Winnipeg, Manitoba, Canada (j.currie@uwinnipeg.ca, n.rampersad@uwinnipeg.ca)

The first of these problems was shown to be decidable by Bean et al. [3] and independently by Zimin [18]. It is unknown whether the second problem is decidable.

A word of length ℓ and period p is called a k-power, where $k = \ell/p$. A reformulation of Thue's result is that there is an infinite word over Σ_3 not containing a 2-power. For integer $n \geqslant 2$, the *repetition threshold* function is defined by

$$\mathrm{RT}(n) = \sup\{k : \text{ every infinite word over } \Sigma_n \text{ contains a } k\text{-power}\}.$$

Thus Thue showed that $\mathrm{RT}(3) = 2$. Dejean [9] showed that in fact $\mathrm{RT}(3) = 7/4$, and conjectured that

$$\mathrm{RT}(n) = \begin{cases} 7/4, & \text{if } n = 3; \\ 7/5, & \text{if } n = 4; \\ n/(n-1), & \text{if } n \neq 3, 4. \end{cases}$$

Dejean's conjecture was finally proved by Rao, and independently by Currie and Rampersad [15, 8]. Words over an alphabet which realize the repetition threshold of the alphabet are called **threshold words** and are extremal objects. In the case $n = 2$, the threshold words are the binary overlap-free words, which have a large literature. (A good reference is the thesis of Rampersad [14].) With the solution of Dejean's conjecture, an indexed family of similar languages present themselves for study. As an example of such study, for threshold words on Σ_n with $n \geqslant 27$, Currie et al. [6] have shown that the number of words grows exponentially with length.

Also branching off from the solution of Dejean's conjecture is the study of repetition thresholds for various classes of words. For example, various authors have found the repetition thresholds for binary rich words, for balanced sequences, and for circular words [7, 10, 11, 13]. Other types of repetition thresholds have also been studied, such as undirected repetition thresholds and Abelian repetition thresholds [5, 16].

When investigating the existence of an infinite word over Σ_n with some property, a natural approach is to generate and study long finite words with the property. Such words are typically generated by backtracking, and are therefore the lexicographically least words of a given length. Practically speaking then, solving avoidance problems often involves generating and parsing prefixes of the lexicographically least infinite word with a given property. Allouche et al. [2] characterized the lexicographically least infinite overlap-free binary word starting with any specified prefix. Currie [4] characterized the lexicographically least infinite *good* word, where the good words are closely related to the period-doubling morphism. However, the general study of lexicographically least infinite words with avoidance properties is in its infancy, and more examples are needed.

The current note combines the theme of repetition threshold with that of lexicographically least words. The 2020 paper of Currie et al. [7] established the repetition threshold for binary rich words. Studying such words by backtracking leads naturally to the question: What is the lexicographically least infinite binary rich word attaining the threshold? We answer this question in this note.

2 Preliminaries

A *word* over *alphabet* Σ_n is a finite or infinite sequence over Σ_n. We use lower case letters for finite words, and write, e. g., word $w = w_1 w_2 \cdots w_m$, where each $w_i \in \Sigma_n$. The *length* of w is denoted by $|w| = m$. The word of length 0 is called the *empty word*, and is denoted by ϵ. The *concatenation* of two words $u = u_1 u_2 \cdots u_s$ and $v = v_1 v_2 \cdots v_t$ is given by $uv = u_1 u_2 \cdots u_s v_1 v_2 \cdots v_t$. If u, v, w, z are words and $w = uzv$, we call word z a *factor* of w, word u a *prefix* of w, and word v a *suffix* of w. If $w = uv$, we define $u^{-1}w = v$.

A *morphism* from Σ_n^* to Σ_m^* is a function f respecting concatenation; i.e., $f(xy) = f(x)f(y)$ for all $x, y \in \Sigma_n^*$. If $f^{-1}(\epsilon) = \{\epsilon\}$, we call f *non-erasing*.

We use bold-face letters for infinite words, writing $\boldsymbol{w} = w_1 w_2 w_3 \cdots$, where each $w_i \in \Sigma_n$. The set of finite words over Σ_n is denoted by Σ_n^*, and the set of infinite words is denoted by Σ_n^ω.

Iteration of a morphism f is written as exponentiation:

$$f^i(x) = \begin{cases} x, & \text{if } i = 0; \\ f(f^{i-1}(x)), & \text{if } i > 0. \end{cases}$$

If $f : \Sigma_n^* \to \Sigma_n^*$ is a non-erasing morphism such that for some $a \in \Sigma_n$, $f(a) = au$, $u \neq \epsilon$, then $f^{n-1}(a)$ is a proper prefix of $f^n(a)$ for every positive integer n. We can then define $\boldsymbol{w} = \lim_{n\to\infty} f^n(a)$ to be the infinite word such that, for each n, word $f^n(a)$ is a prefix of $\boldsymbol{w}$.

Let w be a finite word over Σ_n. Write $w = w_1 w_2 \cdots w_m$ where each $w_i \in \Sigma_n$. The *reversal* of w is the word $w^R = w_m w_{m-1} \cdots w_1$. We call word w a *palindrome* if $w = w^R$. Any word w contains at most $|w|$ distinct non-empty palindromic factors. If w in fact contains $|w|$ distinct non-empty palindromic factors, we say that w is *rich*. A good reference on rich words is the paper of Glen et al. [12]. One of their results which we will use is

Theorem 1. *[12, Theorem 2.14] For any finite or infinite word w, the following properties are equivalent:*

(i) w is rich;

(ii) for any factor u of w, if u contains exactly two occurrences of a palindrome p as a prefix and as a suffix only, then u is itself a palindrome.

A factor u of w containing exactly two occurrences of a factor p as a prefix and as a suffix is called a *return word* of p. An infinite word is defined to be rich if each of its finite factors is rich.

Let $\mathbf{w}$ be an infinite word. The *critical exponent* of $\mathbf{w}$ is defined to be

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

(We allow $\mathrm{ce}(\boldsymbol{w}) = \infty$. This happens, for example, in the case where $\boldsymbol{w}$ is periodic.)

Let L be a set of infinite words. The repetition threshold of L is defined to be

$$\mathrm{RT}(L) = \sup\{k : \text{ every word of } L \text{ contains a } k\text{-power}\} = \inf\{\mathrm{ce}(\boldsymbol{w}) : \boldsymbol{w} \in L\}.$$

Thus $\mathrm{RT}(n) = \mathrm{RT}(\Sigma_n^\omega)$.

Baranwal and Shallit [1] showed that there is an infinite binary rich word with critical exponent $2 + \sqrt{2}/2$, and Currie et al. [7] proved that this word achieves the repetition threshold for infinite binary rich words. Let $\boldsymbol{L}$ be the set of infinite binary rich words. The set $\boldsymbol{T}$ of *threshold words* is the set of infinite binary rich words whose critical exponent is the repetition threshold. Thus

$$\boldsymbol{T} = \{\boldsymbol{w} \in \boldsymbol{L} : \mathrm{ce}(\boldsymbol{w}) = 2 + \sqrt{2}/2\}.$$

Define morphisms $f : \Sigma_3^* \to \Sigma_2^*$ and $g, h : \Sigma_3^* \to \Sigma_3^*$ by

$$f(\mathtt{0}) = \mathtt{0}$$
$$f(\mathtt{1}) = \mathtt{01}$$
$$f(\mathtt{2}) = \mathtt{011}$$

$$g(\mathtt{0}) = \mathtt{011}$$
$$g(\mathtt{1}) = \mathtt{0121}$$
$$g(\mathtt{2}) = \mathtt{012121}$$

$$h(\mathtt{0}) = \mathtt{01}$$
$$h(\mathtt{1}) = \mathtt{02}$$
$$h(\mathtt{2}) = \mathtt{022}$$

Word $f(h^\omega(\mathtt{0}))$ is the word constructed by Baranwal and Shallit [1]. The word $f(g(h^\omega(\mathtt{0})))$ was shown to be a binary rich word with the same critical exponent by Currie et al. [7].

Remark 2. It is shown by Currie et al. [7] that f, g and h preserve non-richness of words.

The lexicographic order on Σ_n^* and Σ_n^ω is defined as follows:

- We order letters in the natural way: $0 < 1 < 2 < \cdots < n-1$. We also insist that $\epsilon < 0$.
- Let the longest common prefix of u and v be p. We say that $u < v$ if and only if the first letter of $p^{-1}u$ is less than the first letter of $p^{-1}v$, where the first letter of ϵ is taken to be ϵ.

One checks that morphisms f, g, and h are order-preserving: Let $\phi \in \{f, g, h\}$. If $u \leqslant v$ then $\phi(u) \leqslant \phi(v)$.

Theorem 3 (Main Theorem)**.** *Word* $\boldsymbol{\ell} = f(01g(h^\omega(\mathtt{0})))$ *is the lexicographically least word in* $\boldsymbol{T}$.

3 Proof of Main Theorem

We say that a word $w \in \Sigma_2^*$ is **eligible** if it is both rich and 14/5-free. We say that an infinite word over Σ_2 is eligible if its factors are eligible. Since $2 + \sqrt{2}/2 < 14/5$, each word of $\boldsymbol{T}$ is eligible.

Observation 4. *Since an infinite eligible word $\boldsymbol{w}$ must be 3-free, it can be written as $pf(\boldsymbol{u})$, where $p \in \{\epsilon, 1, 11\}$, and $\boldsymbol{u} \in \Sigma_3^\omega$.*

We use slight amplifications of the results of Currie et al. [7]:

Lemma 5. *Suppose $f(\boldsymbol{u})$ is eligible, where $\boldsymbol{u} \in 011\Sigma_3^\omega$. Then*

1. *Word $\boldsymbol{u} = g(\boldsymbol{W})$ for some word $\boldsymbol{W} \in \Sigma_3^\omega$.*

2. *Word $\boldsymbol{W}$ has the form $h(\boldsymbol{U})$ for some word $\boldsymbol{U} \in \Sigma_3^\omega$.*

Proof. To show that 1 holds, replacing $\boldsymbol{u}$ by a suffix if necessary, write $\boldsymbol{u} = u_1u_2u_3u_4\cdots$ where each u_i starts with 0 and contains no other 0. It suffices to show that each u_i is among 011, 0121, and 012121. If not, $\boldsymbol{u}$ contains a factor u from among 00, 010, 0111, 0112, 0120, 01211, 012120, 0121211, 0121212, 012122, 0122, and 02. A backtracking search shows that for each such u, $f(u)$ cannot be extended to an eligible word of length 50.

To show that 2 holds, we show that $\boldsymbol{W}$ cannot contain a factor v from among 00, 11, 12, or 21. However if $\boldsymbol{W}$ contains such a factor, then $g(\boldsymbol{W})$ contains a factor u from among 11011, 0121012, and 2101210. A backtracking search shows that for each such u, $f(u)$ cannot be extended to an eligible word of length 50. □

Lemma 6. *Let $\boldsymbol{u} \in 0\Sigma_3^\omega$. Suppose that for some positive integer n, one of $f(g(h^n(\boldsymbol{u})))$ and $f(h^n(\boldsymbol{u}))$ is eligible. Then $\boldsymbol{u} = h(\boldsymbol{W})$ for some word $\boldsymbol{W} \in \Sigma_3^\omega$.*

Proof. Since f, g, and h preserve non-richness, $\boldsymbol{u}$ must be rich. Since $3 > 14/5$, one of $f(g(h^n(\boldsymbol{u})))$ and $f(h^n(\boldsymbol{u}))$ (and hence $\boldsymbol{u}$ also) must be cube-free.

Let $F = \{00, 01210, 10101, 11, 1221, 212\}$. We claim that there is a suffix of $\boldsymbol{u}$ that does not contain any factor u from F.

In the case $u \in \{00, 10101, 11, 1221\}$, for any $a, b \in \Sigma_3$, one checks that $f(h(aub))$, $g(h(aub))$, and $h^2(aub)$ all contain cubes, so that $f(g(h^n(aub)))$ and $f(h^n(aub))$ also contain cubes. Thus there is a suffix of $\boldsymbol{u}$ containing none of 00, 10101, 11, and 1221.

One shows by induction on n that $h^n(212)$ and $g(h^n(212))$ contain a cube or a factor of one of the forms $v0v0v2$ or $v1v1v2$. This forces $f(g(h^n(212)))$ and $f(h^n(212))$ to contain cubes. Thus 212 is not a factor of $\boldsymbol{u}$.

Finally, there is no word $v01210$ where $|v| = 8$, such that $v01210$ is rich and cube-free, and contains none of $00, 10101, 11, 1221, 212$. Thus there is a suffix of $\boldsymbol{u}$ not containing 01210. This establishes the claim.

Replacing $\boldsymbol{u}$ by a suffix if necessary, write $\boldsymbol{u} = u_1u_2u_3u_4\cdots$ where each u_i starts with 0 and contains no other 0. To show that $\boldsymbol{u}$ has the form $\boldsymbol{u} = h(\boldsymbol{W})$, it suffices to show that each u_i is among 01, 02, and 022. If not, $\boldsymbol{u}$ contains a factor from among 00, 011, 0120, 01210, 01211, 01212, 01220, 01221, 01222, 0210, 0211, 0212, 02210, 02211, 02212, and 0222. Each of these is either in F, is not rich, or contains a cube. $\square$

Theorem 7. *The infinite binary word* $\boldsymbol{v} = f(g(h^\omega(0)))$ *is eligible.*

Theorem 7 was proved by Currie et al. [7].

Lemma 8. *The lexicographically least infinite eligible word with prefix* 001010 *is*

$$\boldsymbol{v} = f(g(h^\omega(0))).$$

Proof. Suppose that $\boldsymbol{V}$ is an infinite eligible word with prefix 001010, and $\boldsymbol{V} \leqslant \boldsymbol{v}$. By Observation 4, write $\boldsymbol{V} = f(\boldsymbol{u})$, where $\boldsymbol{u} \in \Sigma_3^\omega$. Since $\boldsymbol{V}$ has prefix 001010, word $\boldsymbol{u}$ has prefix 011. It follows from Lemma 5 that $\boldsymbol{u} = g(\boldsymbol{U})$ for some word $\boldsymbol{U} \in \Sigma_3^\omega$, where $\boldsymbol{U}$ has the form $h(\boldsymbol{W}_1)$ for some word $\boldsymbol{W}_1 \in \Sigma_3^\omega$.

Since f is order-preserving, $\boldsymbol{u} \leqslant g(h^\omega(0))$. Since g is order-preserving, $\boldsymbol{U} \leqslant h^\omega(0)$. Since h is order-preserving, $\boldsymbol{W}_1 \leqslant h^\omega(0)$. In particular, since the first letter of $h^\omega(0)$ is 0, the first letter of $\boldsymbol{W}_1$ is 0. Using Lemma 6, write $\boldsymbol{W}_1 = h(\boldsymbol{W}_2)$ for some word $\boldsymbol{W}_2 \in \Sigma_3^\omega$. Again, since h is order-preserving, the first letter of $\boldsymbol{W}_2$ is 0. By induction, we find that for each positive integer n we have $\boldsymbol{W}_1 = h^{n-1}(\boldsymbol{W}_n)$, for some word $\boldsymbol{W}_n \in 0\Sigma_3^\omega$. It follows that $h^n(0)$ is a prefix of $\boldsymbol{W}_1$ for each n, so that $\boldsymbol{W}_1 = h^\omega(0)$.

We conclude that the lexicographically least infinite eligible word with prefix 001010 is

$$\boldsymbol{v} = f(g(h^\omega(0))).$$ $\square$

The Main Theorem follows from the following three lemmas.

Lemma 9. *Let* $\boldsymbol{m}$ *be an infinite eligible word. Let*

$$\boldsymbol{\ell} = f(01g(h^\omega(0))).$$

Then

$$\boldsymbol{\ell} \leqslant \boldsymbol{m}.$$

Proof. The least binary 3-free word of length 8 is 00100100. However, 00100100 cannot be extended on the right to a binary 3-free word. It follows that $001001010 \leqslant \boldsymbol{m}$. If $\boldsymbol{m} \leqslant \boldsymbol{\ell}$, then

$$f(01g(0))0 = 001001010 \leqslant \boldsymbol{m} \leqslant \boldsymbol{\ell} = f(01g(h^\omega(0))),$$

forcing $\boldsymbol{m}$ to have prefix 001001010. Then $(001)^{-1}\boldsymbol{m}$ is an infinite eligible word with prefix 001010. By Lemma 8, this forces

$$f(g(h^\omega(0))) \leqslant (001)^{-1}\boldsymbol{m}$$

forcing

$$\boldsymbol{\ell} = 001f(g(h^\omega(0))) \leqslant \boldsymbol{m}.$$ $\square$

Lemma 10. *Word* $01f(g(h^\omega(0)))$ *is recurrent.*

Proof. Word $g(h^\omega(0))$ is recurrent. However, the only letter preceding a 0 in $g(h^\omega(0))$ is 1, so that if p is a prefix of $g(h^\omega(0))$, word $1p$ must be a (necessarily recurrent) factor of $g(h^\omega(0))$. Any u factor of $01f(g(h^\omega(0)))$ is a factor of $01f(p) = f(1p)$ for some prefix p of $g(h^\omega(0))$. Since $1p$ is recurrent in $g(h^\omega(0))$, $f(1p)$ is recurrent in $f(g(h^\omega(0)))$, and so is u. Then u is recurrent in $01f(g(h^\omega(0)))$. We conclude that $01f(g(h^\omega(0)))$ is recurrent. $\square$

Lemma 11. *The word*

$$\boldsymbol{\ell} = f(01g(h^\omega(0)))$$

contains no factor with exponent greater than $2 + \sqrt{2}/2$.

Proof. Currie et al. [7] proved that $f(g(h^\omega(0)))$ contains no factor with exponent greater than $2 + \sqrt{2}/2$. By Lemma 10, word $01f(g(h^\omega(0)))$ has the same factors as $f(g(h^\omega(0)))$ and therefore also contains no factor with exponent greater than $2 + \sqrt{2}/2$. Therefore, any $(2 + \sqrt{2}/2)^+$ power in $f(01g(h^\omega(0))) = 001f(g(h^\omega(0)))$ must be a prefix. The word 00100 is a prefix of $f(01g(h^\omega(0)))$, but does not occur in $f(g(h^\omega(0)))$. It follows that any $(2 + \sqrt{2}/2)^+$ power which is a prefix of $f(01g(h^\omega(0)))$ has period 4 or less. A very short finite check shows no such $(2 + \sqrt{2}/2)^+$ power is a prefix of $f(01g(h^\omega(0)))$. $\square$

Lemma 12. *The word*

$$\boldsymbol{\ell} = f(01g(h^\omega(0)))$$

is rich. Thus $\boldsymbol{\ell} \in \boldsymbol{T}$.

Proof. Currie et al. [7] proved that $f(g(h^\omega(0)))$ is rich. By Lemma 10, $01f(g(h^\omega(0)))$ has the same factors as $f(g(h^\omega(0)))$ and is also rich.

Suppose that $001f(g(h^\omega(0)))$ is not rich. It will therefore have a complete return to a palindrome which is not a palindrome. Since $01f(g(h^\omega(0)))$ is rich, some prefix of $001f(g(h^\omega(0)))$ must be a complete return to a palindrome which is not a palindrome. Let this prefix be pqp where p is a palindrome and q is not. The palindrome 00100 is a prefix of $f(01g(h^\omega(0)))$, but does not occur in $f(g(h^\omega(0)))$. It follows that $|p| \leqslant 4$. The only possibility is seen to be $p = 00$. However the complete return to 00 is 00100, which is a palindrome. This is a contradiction. $\square$

Acknowledgements

The work of James Currie is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number DDG-2024-00005]. The work of Narad Rampersad is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number 2019-04111].

References

[1] A. Baranwal and J. Shallit, Repetitions in infinite palindrome-rich words, in R. Mercas and D. Reidenbach (Eds.), *Proc. WORDS 2019, LiNCS* **11682**, Springer, (2019), 93–105.

[2] J.-P. Allouche, J. Currie, J. Shallit, Extremal infinite overlap-free binary words, *Electron. J. Combin.* **5** (1998), #R27.

[3] D. R. Bean, A. Ehrenfeucht, G. F. McNulty, Avoidable Patterns in Strings of Symbols. *Pacific J. Math.* **85** (1979), 261–294.

[4] J. D. Currie, The analog of overlap-freeness for the period-doubling sequence, *J. Int. Seq.* **26** (2023) 23.8.2.

[5] J. D. Currie and L. Mol, The undirected repetition threshold and undirected pattern avoidance, *Theoret. Comput. Sci.* **866** (2021), 56–69.

[6] J. D. Currie, L. Mol, and N. Rampersad, The number of threshold words on n letters grows exponentially for every $n \geqslant 27$, *J. Int. Seq.* **23(3)** (2020), 20.3.1.

[7] J. D. Currie, L. Mol, and N. Rampersad, The repetition threshold for binary rich words, *Discrete Mathematics and Theoretical Computer Science* **22(1)** (2020).

[8] J. D. Currie, N. Rampersad, A proof of Dejean's conjecture. *Math. Comp.* **80** (2011), 1063–1070.

[9] F. Dejean, Sur un théorème de Thue, *J. Combin. Theory Ser. A* **13** (1972), 90–99.

[10] F. Dolce, L. Dvořáková, and E. Pelantová, On balanced sequences and their critical exponent, Theoret Comput. Sci. **939** (2023), 18–47.

[11] L. Dvořáková, D. Opočenská, E. Pelantová, A. M. Shur, On minimal critical exponent of balanced sequences, *Theoret. Comput. Sci.* **922** (2022), 158–169.

[12] A. Glen, J. Justin, S. Widmer, L. Q. Zamboni, Palindromic richness, *Eur. J. Combin.* **30** (2009), 510–531.

[13] I. A. Gorbunova, Repetition threshold for circular words, *Electron. J. Combin.* **19(4)** (2012), #P11.

[14] N. Rampersad, *Overlap-Free Words and Generalizations*, Ph.D. thesis, University of Waterloo, 2007.

[15] M. Rao, Last cases of Dejean's conjecture, *Theoret. Comput. Sci.* **412** (2011), 3010–3018.

[16] A. V. Samsonov and A. M. Shur, On Abelian repetition threshold, *RAIRO-Theor. Inf. Appl.* **46** (2012) 147–163.

[17] A. Thue, Über unendliche Zeichenreihen. *Kra. Vidensk. Selsk. Skrifter. I. Mat. Nat. Kl.* **7** (1906), pp. 1–22. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell et al., editors, Universitetsforlaget, Oslo, 1977, pp. 139–158.

[18] A. I. Zimin, Blocking sets of terms, *Math. USSR Sbornik*, **47** (1984), 353–364.