Avoiding square-free words on free groups

Golnaz Badkobeh* Tero Harju† Pascal Ochem‡
Matthieu Rosenfeld§

February 8, 2022

Abstract

We consider sets of factors that can be avoided in square-free words on two-generator free groups. The elements of the group are presented in terms of $\{0, 1, 2, 3\}$ such that 0 and 2 (resp.,1 and 3) are inverses of each other so that 02, 20, 13 and 31 do not occur in a reduced word. A Dean word is a reduced word that does not contain occurrences of uu for any nonempty u. Dean showed in 1965 that there exist infinite square-free reduced words. We show that if w is a Dean word of length at least 59 then there are at most six reduced words of length 3 avoided by w. We construct an infinite Dean word avoiding six reduced words of length 3. We also construct infinite Dean words with low critical exponent and avoiding fewer reduced words of length 3. Finally, we show that the minimal frequency of a letter in a Dean word is $\frac{8}{59}$ and the growth rate is close to 1.45818.

1 Introduction

Axel Thue [24] showed in 1912 that there exists an infinite square-free word $\mathbf{m}$ over the alphabet $\{0, 1, 2\}$ that avoids occurrences of the words 010 and 212 of length 3. He also obtained two other infinite ternary square-free words avoiding, respectively, $\{010, 101\}$ and $\{010, 020\}$. We shall consider a related question on the free group generated by two elements.

For the basic notions in combinatorics on words, we refer to Lothaire [10, 11]. Let

$$\Sigma_k = \{0, 1, \ldots, k - 1\}$$

*Department of Computing, Goldsmiths University of London, United Kingdom. G.Badkobeh@gold.ac.uk
†Department of Mathematics and Statistics, University of Turku, Finland. harju@utu.fi
‡LIRMM, CNRS, Université de Montpellier, France. ochem@lirmm.fr
§LIRMM, CNRS, Université de Montpellier, France. rosenfeld@lirmm.fr

be an alphabet of k letters. We denote by Σ^* the set of all finite words over an alphabet Σ. We are interested solely in the words of Σ_3^* and Σ_4^*. The set of infinite words $w\colon \mathbb{N} \to \Sigma$ over an alphabet Σ, represented here as infinite strings $w(1)w(2)\cdots$, is denoted by Σ^ω.

The length, i.e., the number of occurrences of letters, of a word $w \in \Sigma^*$ is denoted by $|w|$.

A factor v of w is *right-special* in w if there exist at least two distinct letters a and b such that va and vb both are factors of w.

A word $w \in \Sigma^* \cup \Sigma^\omega$ is said to *avoid* another word v, if v is not a *factor* (i.e., a finite contiguous subword) of w. Furthermore, w is *square-free* if it avoids all nonempty words of the form vv. A morphism, i.e., a substitution of letters to words, $h\colon \Sigma^* \to \Delta^*$ is said to be *square-free*, if it preserves square-freeness, i.e., if $h(w)$ is square-free for all square-free words w.

Example 1. The word $w = 01210120210121021202101202120$ of length 29 is, up to permutations of the letters, the longest ternary square-free word $w \in \Sigma_3^*$ avoiding three square-free words of length 3. It avoids the words $010, 020, 201$. There are infinite square-free ternary words that avoid two words of length 3; the Hall-Thue word $\mathbf{m}$ as defined below is an example of these. □

The following result is due to Crochemore [4].

Theorem 1. *A morphism $\alpha\colon \Sigma_3^* \to \Delta^*$ is square-free if and only if it preserves all square-free words of length 5.*

The simplest square-free morphism is due to Thue [23] ; see Lothaire [10],

$$h_T(0) = 01201\,,\ h_T(1) = 020121\,,\ h_T(2) = 0212021\,.$$

The *Hall-Thue* word $\mathbf{m} = \tau^\omega(0)$, also known as a *variation of Thue-Morse word* [3], is obtained by iterating the (Hall) morphism [8]:

$$\tau(0) = 012\,,\ \tau(1) = 02\,,\ \tau(2) = 1\,.$$

Thus $\mathbf{m} = 0120210121\cdots$ The morphism τ is not square-free since $\tau(010)$ contains a square 2020. However, the Hall-Thue word $\mathbf{m}$ is square-free as was shown by Thue [23] in 1906. The word $\mathbf{m}$ avoids the words 010 and 212 as is immediate from the form of τ.

2 Dean words

While considering *reduced words* of a free group on two generators, we work on the alphabet Σ_4 by considering 0 and 2 (resp. 1 and 3) as inverses of each other.

Thus a reduced word $w \in \Sigma_4^*$ does not have factors from the set $\{02, 20, 13, 31\}$. In other words, the even and odd letters alternate in the reduced words.

A reduced word $w \in \Sigma_4^*$ is said to be a *Dean word* if it is square-free. Dean words are moderately numerous as suggested in Table 1.

n	#	n	#	n	#	n	#	n	#
1	4	2	8	3	16	4	24	5	40
6	64	7	104	8	144	9	216	10	328
11	496	12	720	13	1072	14	1584	15	2344
16	3384	17	4952	18	7264	19	10632	20	15504

Table 1: The number of Dean words, i.e., square-free reduced words over Σ_4.

Dean [7] proved in 1965 the existence of an infinite square-free reduced word.

Theorem 2. *There exists an infinite Dean word.*

We give here three simple proofs.

First Proof. As noticed by Baker, McNulty and Taylor [2], the construction of Dean [7] corresponds to the fixed point

$$f^{\omega}(0) = 0121032101230321012103230123032 1\cdots$$

of the simple morphism $f\colon \Sigma_4^* \to \Sigma_4^*$ defined by

$$f(0) = 01\,,\ f(1) = 21\,,\ f(2) = 03\,,\ f(3) = 23\,. \tag{1}$$

Clearly, the iterated word $f^{\omega}(0)$ is reduced. Its square-freeness can also be checked with the Walnut software [12], as communicated to us by Jeffrey Shallit [18]:

```
eval dean1 "Ei,n (n>=1) & At (t<n) => DE[i+t]=DE[i+n+t]"
```

Interestingly, $f^{\omega}(0)$ is a complete shuffle of $(02)^{\omega}$ and the paper folding sequence $1131133111331331\cdots$ (after renaming 0 to 3); see Davis and Knuth [6] or Allouche and Shallit [1]. The latter claim was also verified by Shallit using Walnut, or by considering the substitution rules for the paper folding sequence:

$$11 \mapsto 1101\,,\ 01 \mapsto 1001\,,\ 10 \mapsto 1100\,,\ 00 \mapsto 1000\,.$$

We can modify (1) by combining the values $f(1)$ and $f(2)$ to obtain $h\colon \Sigma_3^* \to \Sigma_4^*$ from the three letter alphabet:

$$h(0) = 01\,,\ h(1) = 23\,,\ h(2) = 2103\,.$$

The morphism h is not square-free, since $h(212) = 210(3232)103$ contains a square. However, we can check by Theorem 10 that $h(\mathbf{m})$ is square-free. $\square$

Second Proof. Consider the morphism $h\colon \Sigma_3^* \to \Sigma_4^*$ given by the images of the letters,

$$h(0) = 10\,,\ h(1) = 32\,,\ h(2) = 1230103012\,.$$

It is not difficult to show directly that h is square-free, but applying Theorem 1 also leads to a straighforward proof. Since the words $h(ab)$, for different $a, b \in \Sigma_3$, are reduced, the infinite word $h(\mathbf{m})$ is a Dean word. $\square$

Third Proof. Let $w \in \Sigma_3^\omega$ be an infinite (ternary) square-free word. We construct a word $\overline{w} \in \Sigma_4^\omega$ by adding the letter 3 in the middle of every occurrence of 02 and 20. Then also $\overline{w}$ is square-free. Indeed, if there were a square $\overline{uu}$ in $\overline{w}$, the occurrences of 3 would be aligned (in the same positions) in the two instances of $\overline{u}$, and deleting the occurrences of 3 would evoke a square uu into w; a contradiction. Clearly, $\overline{w}$ is reduced by its construction. $\square$

A morphism $h\colon \Sigma_3^* \to \Sigma_4^*$ that preserves Dean words, is called a *Dean morphism*. For h to be a Dean morphism, it suffices to check that $h(w)$ is square-free for all square-free w of length 5 (Theorem 1) and checking that the words $h(ab)$, for different $a, b \in \Sigma_3$, are reduced.

One can apply the third proof of Theorem 2 to finite words. Then applying this operation to the Thue morphism h_T, the images of which start with the letter 0, and end with 1, yields the following Dean morphism

$$\overline{h}_T(0) = 012301\,,\ \overline{h}_T(1) = 03230121\,,\ \overline{h}_T(2) = 0321230321\,.$$

The reverse operation does not work; deleting the occurrences of 3 from a Dean word does not necessarily give a square-free ternary word.

3 Sets of avoided words

An avoided word is also called an *absent* word. Moreover, v is a *minimal* absent word of w if it is absent in w and all its proper prefixes and suffixes occur in w.

For a Dean word $w \in \Sigma_4^* \cup \Sigma_4^\omega$, let $D_l(w)$ denote the set of minimal absent words of length l of w and $d_l(w) = |D_l(w)| \quad$ (the size of $D_l(w)$).

We are interested in the avoided (absent) words of length 3.

With the aid of a computer program, we find that the longest Dean word that avoids the word $w = 10$ of length 2 has length 58:

3230123032301232123032301230321230123212303230123212301232

By considering the permutations of Σ_4 that preserve reduced words, we have

Lemma 3. *Let w be a Dean word with $|w| \geqslant 59$. Then $d_2(w) = 0$.*

We have already seen, by the third proof of Theorem 2, that every ternary square-free word $w \in \Sigma_3^*$ gives rise to a Dean word $\overline{w} \in \Sigma_4^*$. There are 15504 Dean words of length 20. They are divided into 709 different sets, according to their set of factors of length 3. Though, up to the eight permutations of the letters that preserve reduced words (see the Appendix), there are 'just' 1938 Dean words of length 20.

The word $w = 01030121012321230323$ of length 20 is an example of a Dean word that accommodates all reduced words of length 3, i.e., $d_3(w) = 0$. The length 20 is optimal: a Dean word of length 19 always avoids some word of length 3. Note that w is not 'de Bruijn -type' since both 012 and 123 occur twice in w.

Example 2. The Hall-Thue word $\mathbf{m}$ avoids 010 and 212. The corresponding Dean word $\overline{\mathbf{m}}$ has $D_3(\overline{\mathbf{m}}) = \{010, 030, 212, 232\}$ simply since the letter 3 is never added between two 0's or two 2's. □

The following two simple lemmas deal with *repetitions* of the form $abvab$ for a pair ab. They are verified by the aid of a computer. Up to permutations, the words 010321230 and 012303210 are the longest Dean words without repetitions of pairs.

Lemma 4. *A Dean word that has no repetitions of pairs, has length at most 9.*

A reduced word w is *extendable* if there is a letter $a \in \Sigma_4$ such that wa is a reduced word. In this case, wa is called an extension of w. Clearly, any reduced word has two extensions.

Lemma 5. *Let w be a Dean word that has only one right-special factor of length 2. Then $|w| \leqslant 23$.*

Example 3. The word $w = 01032301032123010323010$ is a longest Dean word with only one right-special factor of length 2, $v = 32$. We have $d_2(w) = 0$ and $D_3(w) = \{012, 030, 101, 121, 210, 232, 303\}$ of seven elements. □

A prefix v of w is *proper* if $|v| < |w|$.

Lemma 6. *Let w be a Dean word of length $|w| \geqslant 24$ such that every reduced pair occurs in a proper prefix of w. Then $d_3(w) \leqslant 6$.*

Proof. By the hypothesis, each of the eight reduced pairs ab has at least one extension in w. Hence $d_3(w) \leqslant 8$.

Suppose that $d_3(w) = 8$. Then no reduced pair ab is a right-special factor in w. However, by Lemma 4, each Dean word of length 10 has a repetition of some pair $abzab$. By the uniqueness of the extensions, this would eventually result in a square $abz \cdot abz$ in the prefix of w of length 16.

The case $d_3(w) = 7$ allows that exactly one pair is a right-special factor in w, and the other seven pairs are not right-special. By Lemma 5, $|w| \leqslant 23$, and the claim follows. $\square$

The bound 24 in Lemma 6 is optimal as witnessed by Example 3.

Corollary 7. *If w is an infinite Dean word then $d_3(w) \leqslant 6$.*

By applying a computer search on all Dean words of length 59, we see that there are only four possible sets of avoided Dean words, as specified in Theorem 8. The claim of Theorem 8 for words $|w| > 59$ immediately follows from this result. We shall show in Theorem 12 that these sets are actual for infinite Dean words.

Theorem 8. *Let w be a Dean word of length $|w| \geqslant 59$ with $d_3(w) = 6$. Then $D_3(w)$ is one of the following sets*

$$\begin{aligned}
S_1 &= \{101, 123, 212, 232, 303, 321\} \\
S_2 &= \{012, 030, 101, 121, 210, 232\} \\
S_3 &= \{010, 032, 212, 230, 303, 323\} \\
S_4 &= \{010, 030, 103, 121, 301, 323\}
\end{aligned}$$

As expected, the family of the sets S_i in Theorem 8 is closed under permutations that preserve reduced words. E.g., the third set S_3 can be obtained from S_1 by the permutation $(0\,1)(2\,3)$, and $(1\,3)$ fixes S_1. Moreover, each S_i is closed under reversals of their elements.

Remark 9. *For the length 58, there are still 12 different sets of cardinality six of avoided words. The drop in the number between the lengths 58 to 59 is dramatic due to the fact that there are no longer any avoided words of length 2.*

The next proof relies on a strong result due to Currie [5].

Theorem 10. *Let $h\colon \Sigma_3^* \to \Delta^*$ be a morphism, and let $w \in \Sigma_3^\omega$ be an infinite word with $\{010, 212\} \subseteq D_3(w)$. Then $h(w)$ is square-free if and only if h preserves square-freeness of the factors of w of length 7.*

We apply Theorem 10 to the Hall-Thue word $\mathbf{m}$ which does avoid the words 010 and 212.

Corollary 11. *Let $h\colon \Sigma_3^* \to \Sigma_4^*$ be a non-erasing morphism. Then $h(\mathbf{m})$ is square-free if and only if h is square-free on the factors of $\mathbf{m}$ of length 7.*

For the sake of completeness, the factors of length 7 of $\mathbf{m}$ are listed in the Appendix.

Theorem 12. *There exists an infinite Dean word $w \in \Sigma_4^\omega$ such that $D_3(w) = S_i$, for each $i = 1, 2, 3, 4$.*

Proof. We obtain a solution $h(\mathbf{m})$ for the first set

$$S_1 = \{101, 123, 212, 232, 303, 321\};$$

solutions for the other three sets are obtained by suitable permutations applied to the following morphism.

Let $g\colon \Sigma_3^* \to \Sigma_4^*$ be defined by

$$g(0) = 103230\,,\ g(1) = 1030\,,\ g(2) = 12\,,$$

where the images have lengths $6, 4, 2$, respectively. The morphism g is *not* square-free since the image $g(010)$ contains a square. However, a computer check can verify that g preserves square-freeness of the factors of $\mathbf{m}$ of length 7. Therefore, by Corollary 11, the word $g(\mathbf{m})$ is square-free. Clearly, the images $g(w)$ are reduced for all $w \in \Sigma_3^*$. By inspection, the words of S_1 are avoided by all words $g(ab)$ with $a \neq b$. Since $d_3(g(\mathbf{m})) \leqslant 6$ by Corollary 7, it follows that $D_3(g(\mathbf{m})) = S_1$. □

By the next theorem and the remark that follows, g is, in effect, the only morphism $\Sigma_3^* \to \Sigma_4^*$ that produces Dean words avoiding the words of S_1.

Theorem 13. *Let w be an infinite Dean word such that $S_1 \subseteq D_3(w)$. Then $w = ug(v)$, where $|u| \leqslant 5$, and v is a square-free ternary word.*

Proof. Let $w = x1z$, where x has no occurrences of 1. (The word x can be empty.) By analysing the words without factors in S_1, we conclude that 1 is preceded in w only by a suffix of $g(0) = 103230$, $g(1) = 1030$, or $g(2) = 12$. This gives the bound $|x| \leqslant |g(0)| - 1 = 5$ for the prefix x. We conclude that the suffix $1z$ has a decomposition in terms of $g(a)$, $a \in \Sigma_3$. This means that $w = ug(v)$, where v is square-free since w is. □

Remark 14. *The set S_1 is closed under the permutation $(1\,3)$, and thus the morphism g', where 1 and 3 are interchanged, also satisfies Theorem 13. Also, besides permuting with $(1\,3)$, we can conjugate the images with respect to a common prefix: if $g(a) = uv_a$, for all $a \in \Sigma_3$ then also g'' with $g''(a) = v_a u$ satisfies Theorem 13 with only slight changes. We notice that the morphism g^R where the images $g(a)$ are reversed in order, is obtained from g by conjugation and the permutation $(1\,3)$.*

4 Critical exponent VS number of factors of length 3

Theorem 12 exhibits a Dean word with critical exponent 2 that avoids 6 factors of length 3. In this section, we complement this result by considering the trade-off between the critical exponent of an infinite Dean word and the number of avoided factors of length 3. Negative results are obtained by standard backtracking. Positive results are proved by uniform morphisms that generate infinite Dean words with suitable properties using the method described in [13].

Lemma 15. *Let w be a $\frac{5}{3}$-free Dean word, then $|w| < 62$.*

Lemma 16. *There exists an infinite $\frac{5}{3}^+$-free Dean word avoiding 323.*

The image of any $\frac{7}{5}^+$-free 4-ary word by the following 136-uniform morphism is a $\frac{5}{3}^+$-free Dean word avoiding 323.

```
0 →  01230321230121030123210123032103012103212301030
     21012321030123032123010321012303210301232101230
     10321230321012321030121032123010321012321030
1 →  01230321230121030123210123032103012103212301030
     21012303210301232101230103212303210123210301230
     32123010321012303210301210321230321012321030
2 →  01230321230121030123210123010321230121030123030
     21230103210123210301210321230103210123032103010
     23210123010321230321012321030121032123010321
3 →  01230321230121030123210123010321230121030123030
     21230103210123032103012321012301032123032101230
     21030123032123010321012321030121032123010321
```

Lemma 17. *Let w be a $\frac{17}{10}$-free Dean word and $d_3(w) \geqslant 2$, then $|w| < 289$.*

Lemma 18. *There exists an infinite $\frac{17}{10}^+$-free Dean word avoiding 030 and 232.*

The image of any $\frac{7}{5}^+$-free 4-ary word by the following 358-uniform morphism is a $\frac{17}{10}^+$-free Dean word avoiding 030 and 232.

```
0 →  0103230121032123010321012303212301210323010321230
     3210123010321230121032301032101230321230121032123
     0321012301032301210321230103210123032123010323012
     1032123032101230103212301210323010321230321012301
     0323012103212301032101230321230121032123032101230
     1032123012103230103210123010323012103212301032101
     2303212301032301210321230321012301032123012103230
     103210123032123
1 →  0103230121032123010321012303212301210323010321230
     3210123010321230121032301032101230321230121032123
     0321012301032301210321230103210123032123010323012
     1032123032101230103212301210323010321012301032301
     2103212301032101230321230121032123032101230103212
     3012103230103212303210123010323012103212301032101
     2303212301032301210321230321012301032123012103230
     103210123032123
2 →  0103230121032123010321012303212301210323010321230
     3210123010321230121032301032101230103230121032123
     0103210123032123010323012103212303210123010321230
     1210323010321230321012301032301210321230103210123
     0321230121032301032101230103230121032123032101230
     1032123012103230103210123032123010323012103212301
     0321012303212301210321230321012301032123012103230
     103212303210123
3 →  0103230121032123010321012303212301210323010321230
     3210123010321230121032301032101230103230121032123
     0103210123032123010323012103212303210123010321230
     1210323010321230321012301032301210321230103210123
     0321230121032123032101230103212301210323010321012
     3032123010323012103212301032101230321230121032301
     0321012301032301210321230321012301032123012103230
     103212303210123
```

Lemma 19. *Let w be a $\frac{7}{4}$-free Dean word and $d_3(w) \geqslant 3$, then $|w| < 68$.*

Lemma 20. *There exists an infinite $\frac{7}{4}^+$-free Dean word avoiding 232, 212, 303, and 323.*

The image of any $\frac{7}{5}^+$-free 4-ary word by the following 46-uniform morphism is a $\frac{7}{4}^+$-free Dean word avoiding 232, 212, 303, and 323.

$0 \to$ `0103210123012103010321030121032101230121030123`
$1 \to$ `0103012101230121032101230103012103210121030123`
$2 \to$ `0103012101230121030103210123012103210121030123`
$3 \to$ `0103012101230103210301210321012103010321030123`

Lemma 21. *Let w be a $\frac{15}{8}$-free Dean word and $d_3(w) \geqslant 5$, then $|w| < 136$.*

Lemma 22. *There exists an infinite $\frac{15}{8}^+$-free Dean word avoiding 010, 032, 212, 303, and 323.*

The image of any $\frac{7}{4}^+$-free ternary word by the following 100-uniform morphism is a $\frac{15}{8}^+$-free Dean word avoiding 010, 032, 212, 303, and 323.

$0 \to$ `01232101230121030123210121030121012321030123210123`
`01210123210301210123012103012321030121012321012103`
$1 \to$ `01210123012103012321012301210123210121030121012321`
`03012321012103012101230121030123210123012101232103`
$2 \to$ `01210123012103012321012103012101232103012321012301`
`21030123210301210123210121030123210123012101232103`

5 Critical exponent VS directedness

A word u is *d-directed* if for every factor f of u of length d, the reverse of f, denoted by f^R, is not a factor of u. In this section, we consider the trade-off between the critical exponent of an infinite Dean word and the smallest d such that it is d-directed. We use the same techniques as in the previous section for positive and negative results. In order to verify that a word is d-directed, we only have to check for occurrences of factors of length d and their reverse.

Lemma 23. *There exists an infinite $\frac{5}{3}^+$-free 12-directed Dean word.*

The image of any $\frac{7}{5}^+$-free 4-ary word by the following 72-uniform morphism is a $\frac{5}{3}^+$-free Dean word that is 12-directed.

0 $\to$ 010321012303210301210323010321230321
030123210323012103212303210123210323
1 $\to$ 010321012303210301210323010321230321
012321032301210321230321030123210323
2 $\to$ 010321012303210301210323010321012321
032301210321230321012321030121032123
3 $\to$ 010321012303210301210323010321012321
030121032123032101232103230121032123

Lemma 24. *Let w be a $\frac{27}{16}$-free 11-directed Dean word, then $|w| < 129$.*

Lemma 25. *There exists an infinite $\frac{27}{16}^+$-free 10-directed Dean word.*

The image of any $\frac{7}{5}^+$-free 4-ary word by the following 564-uniform morphism is a $\frac{27}{16}^+$-free Dean word that is 10-directed.

0 $\to$ 01032101230321030123210123032123010323012103012321012303210301232103230121032123010321
01230321230103230121030123210123032123010321012303210301232103230121030123210123032123
01032301210321230103210123032103012321012303212301032301210301232103230121032123010321
01230321230103230121030123210123032123010321012303210301232103230121032123010323012103
01232101230321230103230121032123010321012303210301232103230121030123210123032123010323
01210301232103230121032123010321012303212301032301210301232101230321030123210323012103
212301032301210301232101230321230103230121032123
1 $\to$ 01032101230321030123210123032123010323012103012321012303210301232103230121032123010321
01230321230103230121030123210123032123010321012303210301232103230121030123210123032123
01032301210321230103210123032103012321012303212301032301210301232103230121032123010321
01230321230103230121030123210123032103012321032301210321230103230121030123210123032123
01032301210321230103210123032103012321032301210301232101230321230103230121030123210323
01210321230103210123032123010323012103012321012303212301032101230321030123210323012103
212301032301210301232101230321230103230121032123
2 $\to$ 01032101230321030123210123032123010323012103012321012303210301232103230121032123010321
01230321230103230121030123210123032123010321012303210301232103230121030123210123032123
01032301210321230103210123032103012321012303212301032301210301232101230321030123210323
01210321230103230121030123210123032123010323012103212301032101230321030123210323012103
01232101230321230103230121030123210323012103212301032101230321230103230121030123210123
03210301232103230121032123010323012103012321012303212301032101230321030123210323012103
012321012303212301032301210301232103230121032123
3 $\to$ 01032101230321030123210123032123010323012103012321012303210301232103230121032123010321
01230321230103230121030123210123032123010321012303210301232103230121030123210123032123
01032301210321230103210123032103012321012303212301032301210301232101230321030123210323
01210321230103230121030123210123032123010321012303210301232103230121030123210123032123
01032301210301232103230121032123010321012303212301032301210301232101230321030123210323
01210321230103230121030123210123032123010323012103212301032101230321030123210323012103
012321012303212301032301210301232103230121032123

Lemma 26. *Let w be a $\frac{7}{4}$-free 9-directed Dean word, then $|w| < 114$.*

Lemma 27. *There exists an infinite $\frac{7}{4}^+$-free 6-directed Dean word.*

The image of any $\frac{7}{5}^+$-free 4-ary word by the following 40-uniform morphism is a $\frac{7}{4}^+$-free Dean word that is 6-directed.

$$
\begin{aligned}
\mathtt{0} &\to \mathtt{01030121012321230121012301030123212300323}\\
\mathtt{1} &\to \mathtt{010301210123212301030123032301232123030323}\\
\mathtt{2} &\to \mathtt{010301210123032301232123012101232123030323}\\
\mathtt{3} &\to \mathtt{010301210123010301230323012101232123030323}
\end{aligned}
$$

Lemma 28. *Let w be a 5-directed Dean word, then $|w| < 10$.*

6 Letter frequencies

Theorem 29. *The minimum frequency of a letter in an infinite Dean word is $\frac{8}{59}$.*

The image of every Dean word by the morphism f below is a Dean word such that the frequency of the letter $\mathtt{3}$ is $\frac{8}{59}$.

$$
\begin{aligned}
f(\mathtt{0}) = \;& \mathtt{010}\\
f(\mathtt{1}) = \;& \mathtt{30121012321012103012101230121032101210}\\
& \mathtt{30121012321012103210123012101232101210}\\
& \mathtt{301210123012103210121030121012321012103}\\
f(\mathtt{2}) = \;& \mathtt{212}\\
f(\mathtt{3}) = \;& \mathtt{32101210301210123210121032101230121012}\\
& \mathtt{32101210301210123012103210121030121012}\\
& \mathtt{321012103210123012101232101210301210123}
\end{aligned}
$$

It is easy to see that the f-image of a Dean word is reduced. Let us check that it is also square free. A computer check rules out squares of period at most 500. Notice that $|f(\mathtt{0})| = |f(\mathtt{2})| = 3$ and $|f(\mathtt{1})| = |f(\mathtt{3})| = 115$. Moreover, the factor $\mathtt{010}$ (resp. $\mathtt{212}$) only appears as the f-image of $\mathtt{0}$ (resp. $\mathtt{2}$). So the period of a potential square in the f-image of a Dean word must be a multiple of $|f(\mathtt{01})| = 118$. Since the longest common prefix (resp. suffix) of $f(\mathtt{1})$ and $f(\mathtt{3})$ has length 1, our square implies the existence of a square with the same period and that contains the f-image of a letter as a prefix. This forces a square in the pre-image by f, which is a contradiction.

A computer check shows that every Dean word of length 118 and containing only 15 occurrences of the letter $\mathtt{3}$ is not extendable. Thus $\frac{8}{59}$ is an optimal bound.

7 Growth rate

Let T_n be the set of Dean words of length n. We use the same technique as in [16] (which is really close to the technique introduced in [21] that was itself inspired by [9]) to show the following.

Theorem 30. *For all n,*

$$|T_n| \geqslant 1.4581846^n .$$

Let $p = 36$ and let $\mathcal{F}^{\leqslant p}$ be the set of reduced words that contain no squares of period at most p. Let Λ be the set of reduced words that are prefixes of minimal squares of period at most p. For any w, we let $\Lambda(w)$ be the longest word from Λ that is a suffix of w. For any set of words S, and any $w \in \Lambda$, we let $S^{(w)}$ be the set of words from S whose longest suffix that belongs to Λ is w, that is $S^{(w)} = \{u \in S : \Lambda(u) = w\}$.

Lemma 31. *There exist coefficients $(C_w)_{w\in\Lambda}$ such that $C_0 > 0$ and for all $v \in \Lambda$,*

$$\alpha C_v \leqslant \sum_{\substack{a\in\{0,1,2,3\}\\ va\in\mathcal{F}^{\leqslant p}}} C_{\Lambda(va)} \tag{2}$$

where $\alpha = 33075185/22682414 \approx 1.4581862847578$.

The proof of this lemma relies on computer verifications. Let $M \in \mathbb{Z}^{|\Lambda|\times|\Lambda|}$ be the matrix indexed over Λ such that for all $u, v \in \Lambda$,

$$M_{u,v} = |\{a \in \{0,1,2,3\} : u = \Lambda(va)\}| .$$

Notice that $M_{u,v}$ is either 0 or 1. Then we choose for the coefficient α the largest eigenvalue and for $(C_w)_{w\in\Lambda}$ any corresponding eigenvector. To find a vector close enough to this eigenvector, we simply iterate the matrix (starting with the vector containing 1 everywhere) and renormalize the vector. Then we simply verify that the vector obtained after n iterations (say $n = 100$) has the desired property. We implemented this procedure in C++[1] and it verifies Lemma 31 using exact computation over the rational numbers in less than 10 minutes and using 9GB of RAM on a laptop.

For the rest of this section, let us fix coefficients $(C_w)_{w\in\Lambda}$ that respect the conditions given by Lemma 31. For each set S of words, we let

$$\widehat{S} = \sum_{w\in\Lambda} C_w |S^{(w)}| .$$

Whenever we mention the weight of a word w in informal definitions, we mean $C_{\Lambda(w)}$. We are now ready to state our main Lemma.

[1]The program can be dowloaded at https://www.lirmm.fr/~mrosenfeld/codes/Finding_the_coefficient_for_Dean_words.cpp

Lemma 32. *Let $\beta > 1$ be a real number such that*

$$\alpha - \frac{\beta^{3-2\lceil\frac{p+1}{2}\rceil}}{\beta^2 - 1} \geqslant \beta\,.$$

Then for all n,

$$\widehat{T_{n+1}} \geqslant \beta\widehat{T_n}\,.$$

Proof. We proceed by induction on n. Let n be an integer such that the lemma holds for any integer smaller than n and let us show that $\widehat{T_{n+1}} \geqslant \beta\widehat{T_n}$.

By the induction hypothesis, for all i,

$$\widehat{T_n} \geqslant \beta^i \widehat{T_{n-i}} \tag{3}$$

A word of length $n+1$ is *good*, if its prefix of length n is in T_n, if it is a reduced word and if it contains no square of period at most p. The set of good words is G. A word is *wrong*, if it is good, but contains a square of period larger than p. The set of wrong words is F. Then for any w, $T_{n+1} = G \setminus F$ and

$$\widehat{T_{n+1}} \geqslant \widehat{G} - \widehat{F} \tag{4}$$

Let us first lower-bound $\widehat{G} = \sum_{w\in\Lambda} |G^{(w)}|C_w$.

The *extensions* of any word $v \in T_n$ that belongs to G are the words of the form va where $a \in \{0,1,2,3\}$ and such that $va \in \mathcal{F}^{\leqslant p}$. By definition, $\Lambda(v)$ is the longest suffix of v amongst prefixes of squares of period of length at most p. This implies that for any Dean word v and for any letter u, $vu \in \mathcal{F}^{\leqslant p}$ if and only if $\Lambda(v)u \in \mathcal{F}^{\leqslant p}$. For the same reason, for any square-free word v and for any word u, $\Lambda(vu) = \Lambda(\Lambda(v)u)$. We then deduce that the contribution of the extensions of any word $v \in T_n$ to $\widehat{G}$ is

$$\sum_{\substack{a\in\{0,1,2,3\}\\ va\in\mathcal{F}^{\leqslant p}}} C_{\Lambda(va)} = \sum_{\substack{a\in\{0,1,2,3\}\\ \Lambda(v)a\in\mathcal{F}^{\leqslant p}}} C_{\Lambda(\Lambda(v)a)}\,.$$

By Lemma 31, we deduce that the contribution of the extensions of any word $v \in T_n$ to $\widehat{G}$ is at least $\alpha C_{\Lambda(v)}$. We sum the contribution over T_n

$$\widehat{G} \geqslant \sum_{v\in T_n} \alpha C_{\Lambda(v)} = \sum_{u\in\Lambda} \alpha C_u |T_n^{(u)}| = \alpha\widehat{T_n} \tag{5}$$

Let us now bound F. For all i, let F_i be the set of words from F that end with a square of period i. Clearly, $F = \cup_{i\geqslant 1} F_i$ and

$$\widehat{F} \leqslant \sum_{i\geqslant 1} \widehat{F}_i\,. \tag{6}$$

By definition of G and F, for every $i \leqslant p$, $|F_i| = 0$ and $\widehat{F_i} = 0$. Moreover, since reduced words contain no square of odd period, for all i, $|F_{2i+1}| = 0$ and $\widehat{F_{2i+1}} = 0$.

Let us now upper-bound $\widehat{F_i}$ for any even $i > p$.

Let $u \in F_i$ be a word. For the sake of contradiction suppose, $i \leqslant |\Lambda(u)|$ and let v be the square of period i at the end of u and let $k = |\Lambda(u)|$. By hypothesis, $v_1 \cdots v_i = v_{i+1} \cdots v_{2i}$. There exists $j \leqslant p$ such that $\Lambda(u)$ has period j, but since $\Lambda(u)$ is the suffix of length k of v, $v_{2i+1-k} \cdots v_{2i-j} = v_{2i+1+j-k} \cdots v_{2i}$ (using that $j \leqslant p < i \leqslant k$ one easily verifies that the indices are valid). So in particular, using the two previous equations together,

$$v_{j+1} \cdots v_i = v_{i+j+1} \cdots v_{2i} = v_{i+1} \cdots v_{2i-j}$$

(it is easy to check that all the indices are valid). So there is a square inside u which is not a suffix of u which is a contradiction, since by definition of F the only squares inside u are suffixes of u. Hence, $i > |\Lambda(u)|$.

For any $u \in F_i$, u ends with a square of period i so the last i letters are uniquely determined by the prefix v of length $|u| - i$. By definition, $v \in T_{n+1-i}$. Moreover, by the previous paragraph the suffix of size $|\Lambda(u)|+1$ of u and v are the same which implies $\Lambda(v) = \Lambda(u)$ and $v \in T_{n+1-i}^{(\Lambda(u))}$. So for every $w \in \Lambda$, any word of $F_i^{(w)}$ is uniquely determined by a word of $T_{n+1-i}^{(w)}$, which implies $|F_i^{(w)}| \leqslant |T_{n+1-i}^{(w)}|$. By summing over all $w \in \Lambda$,

$$\widehat{F_i} \leqslant \widehat{T_{n+1-i}}\,.$$

Together with (3), it implies that for even $i > p$,

$$\widehat{F_i} \leqslant \widehat{T_n}\beta^{1-i}\,.$$

We can now sum the F_i to upper bound F using equation (6),

$$\widehat{F} \leqslant \sum_{i \geqslant \lceil \frac{p+1}{2} \rceil} \widehat{T_n}\beta^{1-2i} = \frac{\beta^{3-2\lceil \frac{p+1}{2} \rceil}}{\beta^2 - 1}\widehat{T_n}\,.$$

Using this bound and (5) with (4) yields

$$\widehat{T_{n+1}} \geqslant \widehat{T_n}\left(\alpha - \frac{\beta^{3-2\lceil \frac{p+1}{2} \rceil}}{\beta^2 - 1}\right)\,.$$

By theorem hypothesis we deduce

$$\widehat{T_{n+1}} \geqslant \beta\widehat{T_n}$$

which concludes our proof. $\square$

It is easy to check that $\beta = 1.4581846$ satisfies the conditions of Lemma 32 and we deduce the following corollary.

Lemma 33. *For all* n*,*

$$\widehat{T_{n+1}} \geqslant 1.4581846\widehat{T_n}\,.$$

It implies that for all n, $\widehat{T_n} \geqslant 1.4581846^{n-1}\widehat{T_1} \geqslant 1.4581846^{n-1}C_0$. There exists a constant C, such that $|T_n| \geqslant C1.4581846^n$. Using the fact that the set of Dean words is factorial, it is routine to deduce that for all n,

$$|T_n| \geqslant 1.4581846^n\,.$$

This proves Theorem 30.

We can use the technique from [22], to show that the growth rate of Dean words is at most $\frac{101454247}{69575642} \approx 1.458186286$. It can be done by a simple modification of our program. The gap between the lower bound and the upper bound is less than 2×10^{-6} and can be reduced even more by simply running our program with a larger value of p (with $p = 40$ the lower bound becomes 1.458185888, but running the program takes approximately 2 hour and 97GB of memory).

Let us mention that, using the same technique, we were able to show that there exist at least 1.12^n $\frac{5}{3}^+$-free Dean words of length n.

8 Problems

There are Dean words of length $n \geqslant 7$ that do not occur in any infinite Dean words, i.e., they cannot be extended in any infinite Dean word (but a finite amount). One example of such a word is 0103010.

Due to minimality of $D_l(w)$, there is no obvious relation between $d_l(w)$ and $d_{l+1}(w)$.

Example 4. We have $D_3(g(\mathbf{m})) = S_1$ for the morphism g of Theorem 12. However, $d_4(g(\mathbf{m})) = d_5(g(\mathbf{m})) = d_6(g(\mathbf{m})) = 0$. □

Problem 1. Are there infinite Dean words w such that $d_l(w) > 0$ for all $l \geqslant 3$?

If Problem 1 has a solution w then the (nonempty) sets $D_l(w)$ stabilise, i.e., for each l, there is a bound N_l such that for any prefix v of w with $|v| \geqslant N_l$ one has $D_l(v) = D_l(w)$ (since always $D_l(w) \subseteq D_l(v)$). We conjecture that in such a case the integers $d_l(w)$ are very small.

As mentioned, there are Dean words that are non extendable. We formally define the set of extendable Dean words as follows:

$$\Omega = \{w \mid \text{ for all } n\text{, there is a Dean word } u_n w v_n \text{ with } |u_n|, |v_n| \geqslant n\}.$$

Notice that Ω has the same growth rate as the set of all Dean words [20].

Problem 2. Give a characterisation of the set Ω.

For a Dean word w, let $m_l(w) = |M_l(w)|$ denote the size of the set of *minimal extendable avoided words*:

$$M_l(w) = \Omega \cap D_l(w).$$

Problem 3. Are there infinite Dean words w such that $m_l(w) = 0$ for all l?

A solution word for Problem 3 would contain all words in Ω. Clearly, if $m_j(w) = 0$ then also $m_k(w) = 0$ for all $k < j$.

Example 5. The Dean word $w = 010301210123032123210323010301$ of length 30 satisfies $d_2(w) = d_3(w) = d_4(w) = 0$ but $d_5(w) = 16$. $\square$

As shown by Shallit and Shur [19], this problem is related to the existence of transition words connecting a pair of words in a given language [15]. The existence of such a universal word is conjectured for all power-free languages [19, 21] and has been proved for cube-free words [14] and for some other power-free words [17], but the case of ternary square-free words is still open.

Problem 4. Each Dean word $w \in \Sigma_3^*$ alternates between even and odd letters, and thus w is a letter-to-letter shuffle of even $w_e \in \{0,2\}^*$ and odd words $w_o \in \{1,3\}^*$. What are the *forbidden words* $v \in \{0,2\}^*$ that are not of the form $v = w_e$?

Appendix

(1) Below we list in the cycle form the permutations of Σ_4 that preserve reduced words (apart from the identity):

$$(0\,2),\ (1\,3),\ (0\,2)(1\,3),\ (0\,1)(2\,3),\ (0\,3)(2\,1),\ (0\,1\,2\,3),\ (0\,3\,2\,1).$$

(2) The list of the factors of length 7 of the Hall-Thue word $\mathbf{m}$ needed in Theorem 12:

0120210	1202101	2021012	0210121	2101210
1012102	0121020	1210201	2102012	1020120
0201202	2012021	1202102	2021020	0210201
1020121	0201210	2012101	0121012	1210120
2101202	1012021			

References

[1] Jean-Paul Allouche and Jeffrey Shallit. *Automatic Sequences. Theory, applications, generalizations*. Cambridge: Cambridge University Press, 2003.

[2] K. A. Baker, G. F. McNulty, and W. Taylor. Growth problems for avoidable words. *Theoret. Comput. Sci.*, 69(3):319–345, 1989.

[3] Francine Blanchet-Sadri, James D. Currie, Narad Rampersad, and Nathan Fox. Abelian complexity of fixed point of morphism $0 \mapsto 012, 1 \mapsto 02, 2 \mapsto 1$. *Integers*, 14:paper a11, 17, 2014.

[4] Max Crochemore. Sharp characterizations of squarefree morphisms. *Theoret. Comput. Sci.*, 18(2):221–226, 1982.

[5] James D. Currie. Finite test sets for morphisms that are squarefree on some of Thue's squarefree ternary words. *J. Integer Seq.*, 22(8):article 19.8.2, 9, 2019.

[6] Chandler Davis and Donald. E. Knuth. Number representations and dragon curves - I. *J. Recreational Math.*, 3:61–81, 1970.

[7] Richard A. Dean. A sequence without repeats on x, x^{-1}, y, y^{-1}. *Amer. Math. Monthly*, 72:383–385, 1965.

[8] Marshall Hall, Jr. Generators and relations in groups—The Burnside problem. In *Lectures on Modern Mathematics, Vol. II*, pages 42–92. Wiley, New York, 1964.

[9] Roman M. Kolpakov. On the number of repetition-free words. *J. Appl. Ind. Math*, 1:453–462, 2007.

[10] M. Lothaire. *Combinatorics on Words*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1997.

[11] M. Lothaire. *Algebraic Combinatorics on Words*, volume 90. Cambridge: Cambridge University Press, 2002.

[12] Hamoon Mousavi. Walnut software. https://cs.uwaterloo.ca/ shallit/walnut.html.

[13] Pascal Ochem. A generator of morphisms for infinite words. *RAIRO: Theoret. Informatics Appl.*, 40:427–441, 2006.

[14] A.M. Petrova and Arseny M. Shur. Transition property for cube-free words. In *CSR*, 2020.

[15] A. Restivo and S. Salemi. Some decision results on non-repetitive words. *NATO ASI Ser.*, 1985.

[16] Matthieu Rosenfeld. Avoiding squares over words with lists of size three amongst four symbols. *CoRR*, abs/2104.09965, 2021.

[17] J. Rukavicka. Transition property for α-power free languages with $\alpha \geqslant 2$ and $k \geqslant 3$ letters. In *DLT*, pages 294–303, 2020.

[18] Jeffrey Shallit. Electronic mail correspondence, April 2021.

[19] Jeffrey Shallit and Arseny Shur. Subword complexity and power avoidance. *Theor. Comput. Sci.*, 792:96–116, 2019.

[20] Arseny M. Shur. Comparing complexity functions of a language and its extendable part. *RAIRO: Theoret. Informatics Appl.*, 42(3):647–655, 2008.

[21] Arseny M. Shur. Two-sided bounds for the growth rates of power-free languages. In Volker Diekert and Dirk Nowotka, editors, *Developments in Language Theory*, pages 466–477, 2009.

[22] Arseny M. Shur. Growth rates of complexity of power-free languages. *Theoretical Computer Science*, 411(34):3209–3223, 2010.

[23] Axel Thue. Über unendliche Zeichenreihen. *Norske Vid. Selsk. Skr. I Math-Nat. Kl.*, 7:1–22, 1906.

[24] Axel Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Vid. Selsk. Skr. I Math-Nat. Kl.*, 1:1–67, 1912.