Square-free words obtained from prefixes by permutations

Tero Harju
Department of Mathematics
University of Turku, Finland
harju@utu.fi

Abstract

An infinite square-free word w over a three letter alphabet T is said to have a k-stem σ if $w = \sigma w_1 w_2 \cdots$ where for each i, there exists a permutation π_i of T which extended to a morphism gives $w_i = \pi_i(\sigma)$. We show that there exists an infinite k-stem word for $k = 1, 2, 3, 9$ and $13 \leq k \leq 19$, but not for $4 \leq k \leq 8$ and $10 \leq k \leq 12$. The problem whether a k-stem words exist for each $k \geq 20$ remains open.

Keywords: Square-free words, Thue-Morse word, morphisms, stem
2000 MSC: 68R15

1. Introduction

We consider square-freeness of ternary words, i.e., words over a three letter alphabet. Without restriction we can choose $T = \{0, 1, 2\}$. Let T^* be the set of all strings, called *words*, over T. The empty word is denoted by ε. The length of a word w is denoted by $|w|$. If $w = u_1 v u_2$, then v is a *factor* of w. If here $u_1 = \varepsilon$ then v is a *prefix* of w and if $u_2 = \varepsilon$ then v is a *suffix* of w. An infinite sequence of letters $w = a_1 a_2 \cdots$ with $a_i \in T$ is an *infinite word*. The above terminology generalizes to infinite words in a natural way. Let T^ω denote the set of all infinite words over T.

A finite or infinite word w is *square-free* if it does not contain any factors $u^2 = uu$ for nonempty words u. Axel Thue showed a hundred years ago that there are infinite square-free words over T. One such example, see Lothaire [11], is obtained by iterating the morphism

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

starting from the word 0. The iteration gives the following square-free word,

$$t = 012021012102012021020121012 \cdots \tag{1}$$

that, as seen from the definition of τ, does not have factors 010 and 212. We call t the *Thue word* although it is due to M. Hall Jr. [8]; see also Istrail [9].

A word $u \in T^*$ is said to be a *permutation* of a word $v \in T^*$, if there exists a permutation π of T which extended to a morphism $\pi\colon T^* \to T^*$ gives $u = \pi(v)$, i.e., if $v = a_1 a_2 \cdots a_k$, with $a_j \in T$, then $u = \pi(a_1)\pi(a_2)\cdots\pi(a_k)$. Note that there are six permutations of each square-free ternary word of length at least two.

We say that a finite prefix σ of a (finite or infinite) word w is a *stem of length* k of w if $w = \sigma w_1 w_2 \cdots$ where for each i there exists a permutation π_i of T with $w_i = \pi_i(\sigma)$. In this case the word w has a *k-stem factorization* of w, or w is a *k-stem word*.

Our first result is immediate.

Theorem 1. *There exist infinite square-free* 1*-stem and* 2*-stem words.*

Proof. Indeed, all infinite square-free ternary words have trivially 1-stem and 2-stem factorizations. For the 2-stem factorizations it is sufficient to observe that each word ab with $a \neq b$ of length two, is a permutation of 01. □

Our main result shows that there exist infinite square-free words with a 3-stem factorization.

Theorem 2. *There exist infinite square-free words with k-stem factorizations for $k = 3, 9$ and $13 \leq k \leq 19$ but not for other k less than 20.*

A quick analysis shows that none of the infinite suffixes of the Thue word t in (1) has a 3-stem factorization. Indeed, 0 is recurrent in t and the word

$$\tau^4(0) = 012021012102012021020121$$

violates the factorization condition of the suffixes of t, i.e., $\tau^4(0)$ has the factor 020 starting from position 19 (congruent to 1 modulo 3) and 020 is not a permutation of the prefix 012. The same nonsynchronization happens for the shorter suffixes of $\tau^4(0)$: (120)210(121)02012021020121 and (202)101(210)2012021020121.

Permutations in constructing square-free words are common when counting the number of ternary square-free words. This counting techniques is due to Brinkhuis [3]; see also Berstel [1]. The *Brinkhuis-type* morphism consists of square-free images that are obtained by taking permutations on T.

2. Preliminaries

Let A be an alphabet. A morphism $h\colon A^* \to A^*$ is said to be *uniform* if, for all $a, b \in A$, $|h(a)| = |h(b)|$. If here the length of the images is n then h is called n-uniform. Also, a morphism h is *square-free*, if it preserves square-freeness of words, i.e., if $v \in A^*$ is square-free, then so is the image $h(v) \in A^*$.

Our proofs of existence of k-stem words rely on the following result due to M. Crochemore [5].

Theorem 3. *(a) A morphism $h\colon T^* \to T^*$ on $T = \{0, 1, 2\}$ is square-free if and only if h preserves square-freeness of words of length 5.*

(b) For any alphabet A, a uniform morphism $h\colon A^* \to A^*$ is square-free if and only if h preserves square-freeness of words of length 3.

The ternary alphabet is quite special for square-freeness. Indeed, there are only finitely many square-free words over a binary alphabet, and often for larger alphabets special kinds of square-free morphisms are very simple.

Example 1. According to Carpi [4], if a morphism $h\colon T^* \to T^*$ over the ternary alphabet is square-free, such that $h(T) \neq T$, then $\sum_{a \in T} |h(a)| \geq 18$. This bound is the best possible as shown by the following morphism due to Thue [12, 13],

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

Also, there are square-free morphisms over T that fix a letter. The following example can be verified using Theorem 3,

$$\begin{aligned} h(0) &= 0\,, \\ h(1) &= 10212021012\,, \\ h(2) &= 102012021201021012\,. \end{aligned}$$

In general, there exists a rather noninteresting square-free morphism h on the alphabet $A_{n+1} = \{0, 1, \ldots, n\}$ of $n + 1$ letters of total size $n + 5$ for $n \geq 3$:

$$\begin{aligned} &h(i) = i \quad \text{for } i = 0, 1, \ldots, n-2\,, \\ &h(n-1) = (n-1)0n \text{ and } h(n) = (n-1)1n\,. \end{aligned}$$

It was shown by Brandenburg [2] that the smallest square-free uniform morphism has uniform length 11. An example of such a morphism is given by

$$\begin{aligned} h(0) &= 01021012102\,, \\ h(1) &= 01021202102\,, \\ h(2) &= 01210120212\,. \end{aligned}$$

On the other hand, for larger alphabets A_{n+1} with $n \geq 3$, we have square-free uniform morphisms of uniform length 3, e.g.,

$$\begin{aligned} &h(i) = 01(i+2) \quad \text{for } i = 0, 1, \ldots, n-2\,, \\ &h(n-1) = 021 \text{ and } h(n) = 031\,. \end{aligned}$$

We now turn to study stems for square-free words. The following lemma restricts the structure of k-stems.

Lemma 4. *Let $w = \sigma w_1 w_2 \cdots$ be a k-stem factorization of a square-free infinite ternary word $w \in T^\omega$ with a stem σ, where $k \geq 3$, then σ starts with a permutation of* 012 *and ends with a permutation of* 012*.*

Proof. Suppose to the contrary that σ does not start with a permutation of 012, i.e., $\sigma = pqpr\rho$ for $T = \{p, q, r\}$ and some suffix $\rho \in T^*$.

Consider first the case where $|\sigma| = 3$. After each $w_i = aba$ with $T = \{a, b, c\}$, the next permutation w_{i+1} is uniquely determined to be cbc to avoid the square $acac$. Therefore w repeats itself periodically, and it cannot be square-free.

Assume then that $|\sigma| \geq 4$. In order for w not to be ultimately periodic, there must be indices $i < j$ such that $w_i = w_j$ but $w_{i+1} \neq w_{j+1}$, say without loss of generality that $w_{i+1} = 0102\pi_{i+1}(\rho)$ and $w_{j+1} = dedf\pi_{j+1}(\rho)$ for $T = \{d, e, f\}$. Now w_i does not end with 0 or 1 nor d or e to avoid a square in $w_i w_{i+1}$ and $w_j w_{j+1}$, respectively, and hence w_i ends with the letter $f = 2$. Since $\pi_{i+1} \neq \pi_{j+1}$, we have $w_{j+1} = 1012\pi_{j+1}(\rho)$. However, one can check that there does not exist any common predecessor of length four for w_{i+1} and w_{j+1}. Indeed, let x be a square-free word of length $|x| = 4$ with a suffix 2, then we have

$$\begin{aligned}
x = 0102: &\quad x0102 = (0102)^2\,,\\
x = 0212: &\quad x1012 = 0(21)^2012\,,\\
x = 1012: &\quad x1012 = (1012)^2\,,\\
x = 1202: &\quad x0102 = 1(20)^2102\,,\\
x = 2012: &\quad x0102 = (201)^202\,,\\
x = 2102: &\quad x1012 = (210)^212\,.
\end{aligned}$$

This contradiction proves the claim. □

3. Existence of 3-stems

We fix first the order of the permutations on $T = \{0, 1, 2\}$: let

$$\begin{aligned}
&\pi_0 = (0)(1)(2)\,, &&\pi_1 = (0)(1\ 2)\,, &&\pi_2 = (0\ 1)(2)\,,\\
&\pi_3 = (1)(0\ 2) &&\pi_4 = (0\ 1\ 2)\,, &&\pi_5 = (0\ 2\ 1)\,.
\end{aligned}$$

Hence π_0 is the identity permutation.

Let $S = \{0, 1, 2, 3, 4, 5\}$ be a fixed alphabet of six letters.

The following construction gives an infinite square-free word having a k-stem factorization simultaneously for $k = 3$ and $k = 9$. The proof is instructive also for the other positive cases considered in a later section. The square-free word w with a k-stem factorization is obtained as $h(u)$, where u is a square-free ternary word and h is a uniform morphism constructed using square-free words over the six element alphabet S. Most of the checking and constructions are done by a computer program aided by a human mind (for selecting good candidate words for the images $h(a)$).

Theorem 5. *There exists an infinite word w over T that has 3-stem and 9-stem factorizations.*

Proof. Consider the morphism $\gamma\colon S^* \to T^*$ defined by $\gamma(i) = \pi_i(\sigma)$ for $\sigma = 012021201$. Here $|\sigma| = 9$ and σ is a concatenation of permutations of 012. Denote

$$s_0 = 014103\,, \quad s_1 = 014241\,, \quad s_2 = 014253\,.$$

Finally let $h_9\colon T^* \to T^*$ be defined by $h_9(i) = \gamma(s_i)$, i.e.,

$$\begin{aligned} h_9(0) &= 012021201021012102120102012021012102012021201210201021\,,\\ h_9(1) &= 012021201021012102120102012102120210120102012021012102\,,\\ h_9(2) &= 012021201021012102120102012102120210201210120210201021\,, \end{aligned}$$

where the images have uniform length 54. A computer check shows that the words $h_9(v)$, with $v \in T^*$ square-free and of length 3, are all square-free. (Indeed, the common prefix 0120212010 of length 10 does not occur in any $h_9(v)$ except as a prefix of the images $h_9(i)$, $i \in T$.) Therefore by Theorem 3, the morphism h_9 is square-free. Hence, e.g., the infinite word $h_9(t)$ is square-free for the Thue word t, and it clearly has 3-stem and 9-stem factorizations by the definition of h_9. □

Remark. For $k = 3$ there is a simpler uniform morphism that also does the job. The stem word is naturally $\sigma = 012$, and the new morphism h_3 is defined as follows:

$$\begin{aligned} h_3(0) &= 012021012102120102\,,\\ h_3(1) &= 012021201021012102\,,\\ h_3(2) &= 012021201210120102\,. \end{aligned}$$

The 18-uniform morphism h_3 is square-free by Theorem 3, and hence $h_3(t)$ is an infinite word having a 3-stem factorization. We note that although $h_3(0) = 012021012\pi_2(012021012)$ the word 012021012 is not a 9-stem for any infinite word as can be checked using a computer program.

4. Negative cases

Before going to other positive cases, we consider the small cases where no k-stem factorizations exist for infinite ternary words. These cases are checked by a computer program that systematically checks square-free words over the six element alphabet S and determines that after some point a square is always obtained no matter what k-stem candidate is chosen.

Proposition 6. *There are no infinite square-free words over T having a k-stem factorization for $4 \le k \le 8$ and for $10 \le k \le 12$.*

It turns out that the cases in Proposition 6 are determined quite easily, i.e., the squares are found already in all words of short lengths. Table 1 shows for which lengths $|v| = kn$, the square-free word v cannot have k-stem factorization. For instance, if $k = 4$, then $n = 6$, i.e., no word of length 24 has a 4-stem

n	no k-stems of length
3	12
4	10
6	$4, 6, 8, 11$
10	$5, 7$

Table 1: No k-stem factorizations for $k = 4, 5, 6, 7, 8, 10, 11, 12$.

factorization. In this case we have the word $\gamma(02030) = 01201021012021020120$ of length 20, where γ is defined in the proof of Theorem 5 for $\sigma = 0120$.

In the case for $k = 5$ the following square-free word of length 45

$$012101202120102012101202101210201021202101210$$

has a 5-stem factorization with $\sigma = 01210$, but it cannot be extended to length 50.

5. k-Stems for $k < 20$

We divide the cases for $13 \leq k \leq 19$ to two subcases.

Theorem 7. *There exists an infinite square-free word with a k-stem factorization for $k \in \{13, 17, 18, 19\}$.*

Proof. In each of the following cases of the morphism h_k is of Brinkhuis-type, i.e., the images $h_k(i)$ of the letters are square-free, and permutations of each other. Also, the infinite words $h_k(t)$ will be square-free with a k-stem factorization. Here again t is the Thue word.

Case $k = 13$. As shown by J. Leech [10], see also U. Grimm [7], the morphism

$$\begin{aligned} h_{13}(0) &= 0121021201210\,, \\ h_{13}(1) &= 1202102012021\,, \\ h_{13}(2) &= 2010210120102 \end{aligned} \tag{2}$$

consisting of palindromes, is square-free, and hence the word $\sigma = 0121021201210$ is a stem for the infinite word $h_{13}(t)$. The length 13 is the smallest where this kind of case applies.

Case $k = 17$. In this case, we choose also palindromes as the images of the letters,

$$\begin{aligned} h_{17}(0) &= 01202120102120210\,, \\ h_{17}(1) &= 12010201210201021\,, \\ h_{17}(2) &= 20121012021012102\,. \end{aligned}$$

By applying Theorem 3, the morphism h_{17} can be shown to be square-free. Hence the palindromic word $\sigma = 01202120102120210$ is a stem for the infinite word $h_{17}(t)$.

Case $k = 18$. As shown by Ekhad and Zeilberger [6] the morphism defined by

$$\begin{aligned} h_{18}(0) &= 012021020102120210\,, \\ h_{18}(1) &= 120102101210201021\,, \\ h_{18}(2) &= 201210212021012102\,, \end{aligned}$$

is square-free. Hence the word $\sigma = 012021020102120210$ is a stem for the infinite word $h_{18}(t)$.

Case $k = 19$. We can choose palindromic images,

$$\begin{aligned} h_{19}(0) &= 0120212012102120210\,, \\ h_{19}(1) &= 1201020120210201021\,, \\ h_{19}(2) &= 2012101201021012102\,. \end{aligned}$$

Again, by applying Theorem 3, the morphism h_{19} can be shown to be square-free. Hence $\sigma = 012021201020120210$ is a stem for the infinite word $h_{19}(t)$. $\square$

The following cases require a different approach since for these cases no Brinkhuis-type morphisms as in Theorem 7 exists as can be seen by a systematic computer search.

Theorem 8. *There exists an infinite square-free word with a k-stem factorization for $k \in \{14, 15, 16\}$.*

Proof. Case $k = 14$. We show that the stem $\sigma = 01202120102012$ will do. Let

$$s_0 = 015342\,, \quad s_1 = 015351\,, \quad s_2 = 015102$$

be words over the alphabet $\{0, 1, 2, 3, 4, 5\}$ of six letters, and $h_{14} \colon T^* \to T^*$ be defined by $h_{14}(i) = \gamma(s_i)$, where $\gamma(i) = \pi_i(\sigma)$ as in the proof of Theorem 5. Therefore

$$\begin{aligned} h_{14}(0) =& 012021201020120210121020102120121012021201\cdot \\ & \cdot 210201021202101201020121012010212021012102\,, \\ h_{14}(1) =& 012021201020120210121020102120121012021201\cdot \\ & \cdot 210201021202102012101202120102101210201021\,, \\ h_{14}(2) =& 012021201020120210121020102120121012021201\cdot \\ & \cdot 021012102010210120212010201210212021012102\,, \end{aligned}$$

where the images have length 84. Again a computer check on words of length three reveals that h_{14} is square-free. Hence, for the Thue word t, the infinite word $h_{14}(t)$ is square-free and, due to the permutations π_i, one detects that it has a 14-stem factorization.

Case $k = 15$. In this case we show that the stem $\sigma = 012021201210120$ can be chosen. Let

$$s_0 = 025143\,, \quad s_1 = 025152\,, \quad s_2 = 025203\,.$$

and $h_{15} \colon T^* \to T^*$ be defined by $h_{15}(i) = \gamma(s_i)$, where $\gamma(i) = \pi_i(\sigma)$ as in the above. Then

$$\begin{aligned}
h_{15}(0) =& 012021201210120102120210201021201210120102012\cdot \\
&\cdot 021012102120210120102012021201210201021012102\,, \\
h_{15}(1) =& 012021201210120102120210201021201210120102012\cdot \\
&\cdot 021012102120210201210120102012102120210201021\,, \\
h_{15}(2) =& 012021201210120102120210201021201210120102012\cdot \\
&\cdot 102120210201021012021201210120210201021012102\,,
\end{aligned}$$

where the images are of length 90. Again a computer check on words of length three reveals that h_{15} is square-free. Hence, for the Thue word t, the infinite word $h_{15}(t)$ is square-free and it has a 15-stem factorization.

Case $k = 16$. We show that the stem $\sigma = 0121021201021012$ will do. Let

$$s_0 = 024342\,, \quad s_1 = 024351\,, \quad s_2 = 024201\,.$$

and $h_{16} \colon T^* \to T^*$ be defined by $h_{16}(i) = \gamma(s_i)$, where $\gamma(i) = \pi_i(\sigma)$ as in the above. Now

$$\begin{aligned}
h_{16}(0) =& 012102120102101210201202101201021202102012102120\cdot \\
&\cdot 210120102120121012021020121021201020120210120102\,, \\
h_{16}(1) =& 012102120102101210201202101201021202102012102120\cdot \\
&\cdot 210120102120121020102101202102010212012102012021\,, \\
h_{16}(2) =& 012102120102101210201202101201021202102012102120\cdot \\
&\cdot 102012021012010201210212010210120212012102012021\,,
\end{aligned}$$

where the images are of length 96. Again a computer check on words of length three reveals that h_{16} is square-free. Hence, for the Thue word t, the infinite word $h_{16}(t)$ is square-free and it has a 16-stem factorization. □

6. Open questions

There remains many open questions on the k-stem factorizations.

Problems

1. Are there infinite square-free ternary words with a k-stem factorization for each $k \geq 13$?

   Note that there are infinitely many values k for which a k-stem factorization exists of an infinite square-free word. Indeed, consider the square-free morphism h_{13} from (2). Then the power h^i is also square-free and it gives a 13^{i+1}-stem factorization of the infinite word $h^i(t)$.

2. Does the Thue word t have a k-stem factorization for some $k > 3$?

[1] Berstel, J., 2005. Growth of repetition-free words—a review. Theoret. Comput. Sci. 340 (2), 280–290.
URL `http://dx.doi.org/10.1016/j.tcs.2005.03.039`

[2] Brandenburg, F.-J., 1983. Uniformly growing kth power-free homomorphisms. Theoret. Comput. Sci. 23 (1), 69–82.
URL `http://dx.doi.org/10.1016/0304-3975(88)90009-6`

[3] Brinkhuis, J., 1983. Nonrepetitive sequences on three symbols. Quart. J. Math. Oxford Ser. (2) 34 (134), 145–149.

[4] Carpi, A., 1983. On the size of a square-free morphism on a three-letter alphabet. Inform. Process. Lett. 16 (5), 231–235.
URL `http://dx.doi.org/10.1016/0020-0190(83)90094-7`

[5] Crochemore, M., 1982. Sharp characterizations of squarefree morphisms. Theoret. Comput. Sci. 18 (2), 221–226.
URL `http://dx.doi.org/10.1016/0304-3975(82)90023-8`

[6] Ekhad, S. B., Zeilberger, D., 1998. There are more than $2^{n/17}$ n-letter ternary square-free words. J. Integer Seq. 1, Article 98.1.9 (electronic).

[7] Grimm, U., 2001. Improved bounds on the number of ternary square-free words. J. Integer Seq. 4 (2), Article 01.2.7, 14 pp. (electronic).

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

[9] Istrail, S., 1977. On irreducible languages and nonrational numbers. Bull. Math. Soc. Sci. Math. R. S. Roumanie (N.S.) 21(69) (3-4), 301–308.

[10] Leech, J., 1957. A problem on strings of beads. Math. Gazette 41 (338), 277–278.

[11] Lothaire, M., 1997. Combinatorics on words. Cambridge Mathematical Library. Cambridge University Press, Cambridge.
URL `http://dx.doi.org/10.1017/CBO9780511566097`

[12] Thue, A., 1906. Über unendliche Zeichenreihen. Norske Vid. Skrifter I. Mat.-Nat. Kl., Christiania 7, 1–22.

[13] Thue, A., 1912. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. Norske Vid. Skrifter I. Mat.-Nat. Kl., Christiania 1, 1–67.