ON STRONGLY CUBE-FREE ω-WORDS GENERATED BY BINARY MORPHISMS

by

Juhani Karhumäki

Department of Mathematics
University of Turku
Turku, Finland

Research Report CS-81-18

April 1981

Faculty

of

Mathematics

University of Waterloo
Waterloo, Ontario, Canada
N2L 3G1

ON STRONGLY CUBE-FREE ω-WORDS GENERATED BY BINARY MORPHISMS

Juhani Karhumäki
Department of Mathematics
University of Turku
Turku, Finland

Abstract

An ω-word is called strongly cube-free if it does not contain a subword of the form vvfirst(v), with $v \neq \lambda$. We show that it is decidable whether a given morphism over a binary alphabet defines, when applied iteratively, a strongly cube-free ω-word. Moreover, an explicit and reasonably small upper bound for the number of iterations needed to be checked is given.

1. Introduction

Repetitions in words, i.e., the existence of occurrences of v^i, with $v \neq \lambda$ and $i \geq 2$, as subwords, was first studied by Thue in [T1] and [T2]. He proved, among other things, that there exists an infinite strongly cube-free word (cf. section 2) over a binary alphabet. Such a word is obtained by iterating the morphism $h(a) = ab$, $h(b) = ba$ when started at a. Few first words obtained are as follows: a → ab → abba → abbabaab → abbabaabbaababba.

Later on this sequence and its interesting properties have been rediscovered several times in different connections. This kinds of nonrepetative sequences have applications in many areas of discrete mathematics, for example in connection with unending games and in group theory to mention only few, cf. [MH].

We call a morphism $h: \Delta^* \to \Delta^*$ prefix-preserving if $h(a) = az$ for some a in Δ and $z \neq \lambda$. As is easily seen such a morphism defines, when applied iteratively starting at a, a unique ω-word. In this paper we are interested in under which conditions such a morphism over a binary alphabet generates a strongly cube-free ω-word and, in particular, whether this can be effectively decided. We shall show that such a morphism must not only be biprefix but also such that h(a) and h(b) must both start and end with a different letter. We call such morphisms strong biprefixes.

The answer to our decision problem is shown to be positive. Furthermore, we are able to give a relatively small upper bound for the number of iterations needed to guarantee an occurrence of a subword of the form vvfirst(v), with $v \neq \lambda$, if such will ever occur in the sequence.

The similar problems are considered in [B] and [K]. In [B] it is shown that the above decision problem for square-free ω-words in a three-letter alphabet is decidable, and in [K] that the same holds true for cube-free ω-words over a binary alphabet.

2. Preliminaries

We use only very basic notions of the formal language theory, see e.g. [H]. For clarity we want to specify the following.

The length of a word v is denoted by $|v|$. For two words u and v, $u^{-1}v$ (resp. vu^{-1}) means the left (resp. right) difference of v by u. Further we write u pref v (resp. u p-pref v) if u is a prefix (resp. a proper prefix) of v, i.e. $v = uw$ holds true for some word w (resp. for some word $w \neq \lambda, v$). For a word u the notation pref(u) denotes the set of all prefixes of u, while the notation $pref_n(u)$ is used to specify the prefix of u of the length n. By definition, $pref_n(u) = u$ if $|u| < n$. The corresponding notions for suffixes are obtained by replacing pref by sub. A word u is a subword of a word v if there exist words u' and u'' such that $v = u'uu''$. By saying that u is a subword in a language L we, of course, mean that u is a subword of some word in L.

By an ω-word we mean an infinite word (from left to right). A word or an ω-word is called strongly cube-free (resp. square-free, cube-free, or fourth power-free) if it does not contain as a subword any word of the form $vv\,first(v)$ (resp. v^2, v^3 or v^4) with $v \neq \lambda$. Here first(v) denotes the first symbol of v, while last(v) is used to denote the last symbol of v. Of course, by saying that a language L is e.g. strongly cube-free we mean that all of its words are such.

Our basic notion is that of a morphism of a finitely generated free monoid Δ^*. Let $h: \Delta^* \to \Delta^*$ be a λ-free morphism, i.e., nonerasing. We say that h is prefix-preserving if

$$(*) \qquad h(a) = az \qquad \text{for some } a \text{ in } \Delta \text{ and } z \neq \lambda$$

If this is the case, then $h^2(a) = h(az) = h(a)\,h(z)$ and, in general,

$$h^i(a) = h^{i-1}(a)\,h^{i-1}(z) \qquad \text{for } i \geq 1.$$

Consequently, $h^{i-1}(a)$ is a proper prefix of $h^i(a)$ for each i, which means that the iterative application of h starting at a defines as a limit an ω-word. Infinite words thus obtained are called ω-words generated by morphisms. Morphisms satisfying (*) are called prefix-preserving morphisms or pp-morphisms for short.

As usual we call a morphism $h: \{a,b\}^* \to \{a,b\}^*$ a biprefix if each of $h(a)$ and $h(b)$ is neither a prefix nor a suffix of the other. By a strong biprefix we mean a morphism h over $\{a,b\}$ such that $first(h(a)) \neq first(h(b))$ and $last(h(a)) \neq last(h(b))$.

3. Simple Properties

In this section we present a necessary condition for a morphism to generate a strongly cube-free ω-word. We show that such a morphism must be a strong biprefix. The proof goes along the lines presented in [K], where it is shown that only biprefixes may generate cube-free ω-words.

The following lemma is established in [K] simply by generating, step by step, all words which are cube-free and does not contain aa as a subword.

Lemma 1 Every cube-free (and hence also strongly cube-free) word over a binary alphabet $\{a,b\}$ and of the length at least 18 contains aa and bb as subwords.

Corollary 1 Every cube-free (and hence also strongly cube-free) ω-word over a binary alphabet $\{a,b\}$ contains aa and bb as subwords.

From the corollary it is easy to conclude

Theorem 1 Every prefix-preserving morphism generating a strongly cube-free ω-word is a strong biprefix.

Combining the above with results in [K] we conclude this section with the following remarks concerning ω-words generated by morphisms over a binary alphabet. As is well-known square-free ω-words over $\{a,b\}$ do not exist at all, and hence such cannot be generated by morphisms, either. As shown by Thue, strongly cube-free ω-words over $\{a,b\}$ can be generated by morphisms, but, because of Theorem 1, only by strong biprefixes. If, in turn, we want to generate cube-free ω-words over $\{a,b\}$, then, as shown in [K], only biprefixes are suitable candidates. Finally, fourth power-free ω-words over $\{a,b\}$ can be generated by nonbiprefixes, as shown in [K] by using the Fibonacci morphism: $h(a) = ab$, $h(b) = a$.

4. Main Result

In this section we prove our main result.

Theorem 2 It is decidable whether a given prefix-preserving morphism over a binary alphabet generates a strongly cube-free ω-word.

Proof Let $h: \{a,b\}^* \to \{a,b\}^*$ be a pp-morphism, say $h(a) = az$ with $z \neq \lambda$. We denote $h(a) = \alpha$, $h(b) = \beta$ and $L = \{h^n(a) \mid n \geq 0\}$. By Theorem 1, we may assume that h is a strong biprefix, i.e., $\text{first}(\alpha) \neq \text{first}(\beta)$ and $\text{last}(\alpha) \neq \text{last}(\beta)$.

The basic idea behind the proof is to show that if L contains long enough subwords of the form

$$(1) \qquad vv\,\mathrm{first}(v) \qquad \text{with } v \neq \lambda,$$

then it contains shorter, too.

We start with

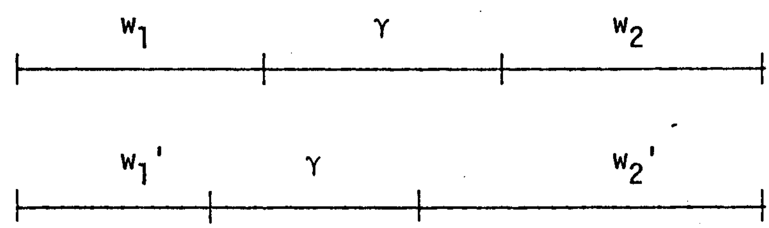
<u>Claim I</u> Assume that $ww\,\mathrm{first}(w)$ is a subword in L satisfying

$$(2) \qquad w = w_1 \gamma w_2 = w_1' \gamma w_2'$$

for some w_1, w_2, w_1' and w_2' with $0 < ||w_1| - |w_1'|| < |\gamma|$ and $\gamma \in \{\alpha,\beta\}$. Then $ww\,\mathrm{first}(w)$ is not the shortest subword of L of the form (1).

The claim is proved as follows. Without loss of generality let $|w_1| > |w_1'|$. Then (2) can be illustrated in the following way:

w:



Consequently, the word w, and hence also L, has a subword $\gamma'\gamma'\,\mathrm{first}(\gamma')$, where $\gamma' = \mathrm{pref}_{|w_1|-|w_1'|}(\gamma)$.

<u>Claim II</u> The shortest subword of L of the form (1) (if there are any) is of the length at most $4|\alpha\beta|$.

To prove claim II we proceed as follows. Let $xuu\,\mathrm{first}(u)y$ be a word in L such that $uu\,\mathrm{first}(u)$ is a minimal subword (with respect to the length) of L of the form (1). We derive from the assumption

$$(3) \qquad |uu\,\mathrm{first}(u)| > 4|\alpha\beta|$$

a contradiction.

Let u_1 be a word satisfying

$$(4) \qquad \begin{array}{l} x\,u_1 \in h(\Delta^*) \quad \text{with} \quad u_1 \notin \Delta^* \{\alpha,\beta\}, \\ u_1\,\delta \text{ pref } u \quad \text{for some } \delta \text{ in } \{\alpha,\beta\}. \end{array}$$

Without loss of generality we set $\delta = \alpha$.

Since xuufirst(u)y is in $h(\Delta^*)$ and h is a biprefix there exists a word u_2 such that

$$u_2 \in h(\Delta^*),$$
(5)
$$u_1\ \alpha\ u_2 \text{ pref } u\ u_1 \text{ p-pref } u_1\ \alpha\ u_2\ \psi \text{ pref } uu$$

for some ψ in $\{\alpha,\beta\}$. We show that $|\alpha u_2| = |u|$, which implies, since h is a biprefix, that $\psi = \alpha$, and therefore writing $u_2 = u_2'u_1$ we obtain

(6) $$xuu = xu_1\alpha u_2'u_1\alpha u_2' \quad \text{with} \quad xu_1,\ xu_1\alpha u_2'u_1 \in h(\Delta^*).$$

To prove (6) assume that $|\alpha u_2| \neq |u|$ which means that $|\alpha u_2| < |u|$. If now $\psi = \alpha$, then, by (4) and (5), the assumptions of Claim I are satisfied with $w_1 = u_1$, $w_1' = u^{-1}u_1\alpha u_2$, $\gamma = \alpha$, and so we derive a contradiction with the minimality of uufirst(u). Consequently, it remains the case $\psi = \beta$. In this case we conclude from (4) and (5) that

(7) $$uu_1\alpha,\ u_1\alpha u_2\beta \in \text{pref}(uu).$$

Moreover, since xu_1, $xu_1\alpha u_2$ and xuufirst(u)y $\in h(\Delta^*)$ and h is a biprefix, there exist words A and B in $\{\alpha,\beta\}^*$ such that uu pref $uu_1\alpha A$ and uu pref $u_1\alpha u_2\beta B$. We show that any choice of A and B leads to a contradiction.

Assume first that $u_1\alpha u_2\beta$ p-pref $uu_1\alpha$. Please, observe that the equality $uu_1\alpha = u_1\alpha u_2\beta$ is excluded since h is a biprefix. Now B must start with β, otherwise we obtain, by claim I, a contradiction to the minimality of uufirst(u). It also follows that u_2 must end with α, otherwise we would contain three consequtive occurrences of β, again a contradiction to the minimality of uufirst(u). Consequently, by (3), $u_1\alpha u_2\beta\beta$ is a proper prefix of uu. If it is also a proper prefix of $uu_1\alpha$ we are done: there are no ways to continue $u_1\alpha u_2\beta\beta$ without a contradiction, since α-continuation is excluded by claim I and β-continuation since three consequtive β's in a word of L contradicts with (3). So it follows that $uu_1\alpha$ pref $u_1\alpha u_2\beta\beta$. Again the equality is excluded since h is a biprefix. Hence, by claim I, A must start with α. This means that last (u_1) must be different from last (α), otherwise we have a contradiction, and, consequently, (3) implies that $|u_1\alpha\alpha| < |u|$. So there is still a way to continue the proper prefixes $u_1\alpha u_2\beta\beta$ and $uu_1\alpha\alpha$ of uu. However, any continuation leads to a contradiction as in the case of the word $u_1\alpha u_2\beta\beta$ above.

Since the possibility $uu_1\alpha$ p-pref $u_1\alpha u_2\beta$ can be handled with the very same manner we have proved the identity (6). Now we consider the word uufirst(u) in the form

$$u_1\alpha u_2'\ u_1\alpha u_2' \text{ first } (u).$$

Let $h(p) = u_2'u_1$. If $u_1 = \lambda$, then first (u) = first (α) and, consequently, since h is a strong biprefix, $xuu\text{first}(u)y$ can be rewritten as $x\alpha u_2'\alpha u_2'\alpha y'$ for some word y'. So L contains a word $h^{-1}(\alpha u_2'\alpha u_2'\alpha) = apapa$ as a subword. This contradicts with the minimality of $uu\text{first}(u)$. Observe here that also the case when instead of setting $\gamma = \alpha$ we set $\gamma = \beta$ leads to a contradiction. If, in turn, $u_1 \neq \lambda$ then we may rewrite $xuu\text{first}(u)y$ as $x'\rho\alpha u_2'u_1\alpha u_2'u_1y''$ for some words x' and y'' and some ρ in $\{\alpha,\beta\}$ such that ρ is a suffix of $u_2'u_1$. Consequently, L contains a subword last (p) a p a p which is again a contradiction with the minimality of $uu\text{first}(u)$.

So our proof for claim II is complete. Now, the theorem follows from claim II and from the following easily provable lemma (see [K], cf. also section 5).

Lemma 2 Given a morphism $h: \Delta^* \to \Delta^*$ and words w, $\omega \in \Delta^*$. It is decidable whether the language $\{h^n(\omega) \mid n \geq 0\}$ contains w as a subword.

5. An Effective Upper Bound

In this section we strengthen Theorem 2 by establishing an upper bound for the number of applications of h to guarantee the existence of a word of the form $vv\text{first}(v)$ as a subword in $\{h^n(a) \mid n \geq 0\}$ if such will ever occur. More precisely, we define for each $i \geq 1$ a number $\delta(i)$ as follows. Let h be a pp-morphism over $\{a,b\}$ such that $|h(a)| \geq i$, $|h(b)| \geq i$ and $h(a) = az$ for some $z \neq \lambda$. Then $\delta(i)$ is defined to be the smallest integer satisfying for every h of the above form:

$$\{h^n(a) \mid n \geq 0\} \quad \text{is strongly cube-free,}$$

if and only if,

$$\{h^n(a) \mid n \leq \delta(i)\} \quad \text{is strongly cube-free.}$$

We continue with three simple lemmas, the proofs of which can be easily derived from the arguments in section 6 of [K].

Lemma 3 If a prefix-preserving morphism over $\{a,b\}$ is not a strong biprefix, then it generates a word of the form $vv\text{first}(v)$, with $v \neq \lambda$, in not more than 7 steps.

Lemma 4 Any prefix-preserving strong biprefix h over $\{a,b\}$ satisfying $|h(a)| = 1$ or $|h(b)| = 1$ generates a word of the form vv first (v), with $v \neq \lambda$, in two steps.

Lemma 5 Any prefix-preserving strong biprefix h over $\{a,b\}$ satisfying $|h(a)| = |h(b)| = 2$ generates a word of the form $vv\text{first}(v)$, with $v \neq \lambda$, in two steps, if at all.

Finally, we state our basic lemma of this section. For its proof we again refer to [K].

Lemma 6 Let h be a strong biprefix over $\{a,b\}$ such that $\min\{|h(a)|, |h(b)|\} \geq 2$, $\max\{|h(a)|, |h(b)|\} \geq 3$ and $h(a) = az$ for some $z \neq \lambda$. If a word x with the length at most $4|h(a)\ h(b)|$ occurs as a subword in the language $\{h^n(a) \mid n \geq 0\}$, it occurs also in $\{h^n(a) \mid n \leq 7\}$.

Now from Lemmas 3-5 and from claim II in the proof of Theorem 2, we conclude

Theorem 3 $\delta(1) \leq 7$ and hence also $\delta(2) \leq 7$.

In special cases Lemmas 3 and 5 can be strengthened to yield

Theorem 4 $\delta(3) \leq 6$ and $\delta(i) \leq 5$ for $i \geq 4$.

We are not claiming that our upper bounds for the values of $\delta(i)$ are the best possible. On the other hand, they are quite small.

6. Discussion

We have shown that it is decidable whether a given prefix-preserving morphism over a binary alphabet generates a strongly cube-free ω-word. Moreover, we proved that the problem can be settled by checking only relatively few iteration steps, and consequently it might be possible to use a computer in searching such morphisms.

The techniques we have used is similar to that used in [K], when we have shown that it is decidable whether a given pp-morphism over a binary alphabet generates a cube-free ω-word. However, our considerations are now shorter and, moreover, upper bounds for the number of iterations needed to be checked are smaller. So the approach seems to be very suitable for the problem solved here, i.e. for the strongly cube-freeness problem over a binary alphabet. We want to finish this section by mentioning that the same techniques can be used to establish the Berstel's result, cf. [B], which states that it is decidable whether a given prefix-preserving morphism over a three-letter alphabet generates a square-free ω-word. Moreover, our approach would give a constant upper bound independent of the given morphism for the number

Acknowledgements

The author is grateful to Finnish Academy for the excellent working conditions under which this research was done. Completion of this paper was supported by the Natural Sciences and Engineering Council Canada, under grant A7403.

References

[B] J. Berstel, Sur les mots sans carré définis par un morphisme. Springer Lecture Notes in Computer Science v. 71, 1979, 16-25.

[H] M. Harrison, Introduction to Formal Language Theory, Addison-Wesley, Reading, 1978.

[K] J. Karhumäki, On cube-free ω-words generated by morphisms, manuscript.

[MH] M. Morse and G.A. Hedlund, Unending chess, symbolic dynamics and a problem in semigroups, Duke Math. J. 11, 1944, 1-7.

[T1] A. Thue, Über unendliche Zeichenreihen, Norsk. Vid. Selsk. Skr. I, Mat. - Nat. Kl. Nr. 7, 1906, 1-22.

[T2] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, Norsk. Vid. Selsk. Skr. I, Mat.-Nat. Kl. Nr. 1 (1912), 1-67.