Theoretical Computer Science 18 (1982) 221–226
North-Holland Publishing Company

NOTE

SHARP CHARACTERIZATIONS OF SQUAREFREE MORPHISMS

Max CROCHEMORE
Laboratoire d'Informatique, Université de Haute-Normandie, 76130 Mont-Saint-Aignan, France

Communicated by M. Nivat
Received September 1981

1. Introduction

Squarefree morphisms, i.e. morphisms that preserve squarefreeness of words, have been a subject of interest since the works of Thue [8]. They are a powerful tool to prove properties on squarefree words and are used in particular to prove the existence of infinite sequences without repetition. The study of squarefree words is connected with questions in algebra and game theory, and recently several results have been published on that topic [2–7].

In this note we prove the following theorem: "A semigroup morphism defined on a three letter alphabet is squarefree iff the images of the squarefree words of length 5 are squarefree".

The first effective characterization of squarefree morphisms has been given by Berstel [2] who found a bound (on the length of words to check) depending on the morphism. We improve his result in the general case (not only for three letters) and obtain the above theorem as a consequence.

On a related subject, i.e. morphisms defined on a three letter alphabet and which generate infinite squarefree sequences by iteration, Karhumäki [6] gave a bound independent from the morphism.

We also show that the bound 5 in our theorem is optimal and that on a greater alphabet such a result does not exist.

2. Definitions

Given an alphabet A, the free semigroup generated by A is denoted by A^+. By adjunction of the empty word 1 we get A^*.

The *length* of a word w in A^* is denoted by $|w|$. If a word w is equal to $\alpha u\beta$ for words α, β, u in A^*, u is said to be a *factor* of w; it is a *prefix* of w if $\alpha = 1$, a *suffix* of w if $\beta = 1$, and an *internal factor* of w if neither α nor β are empty words.

A word w *contains a square* if it can be written $\alpha u^2\beta$ with a non empty word u; otherwise it is *squarefree.*

Let A and B be two alphabets; a semigroup morphism $h: A^+ \to B^+$ is called *k-squarefree* (for an integer $k \geqslant 1$) if for any squarefree word w in A^+ with $|w| \leqslant k$, $h(w)$ is squarefree. A morphism is *squarefree* if it is ∞-squarefree.

We also introduce the notion of a *pre-square* with respect to a morphism h: let w be a squarefree word in A^+ and u a factor of $h(w)$; an occurrence of u in $h(w)$ is given by words α, β in A^* such that $h(w) = \alpha u\beta$; that occurrence of u is called a pre-square if $u \neq 1$ and if there exists a word w' in A^* satisfying:

ww' is squarefree and u is a prefix of $\beta h(w')$ or $w'w$ is squarefree and u is a suffix of $h(w')\alpha$.

In that case we also say that $h(w)$ contains a pre-square, and that w' duplicates the pre-square u of $h(w)$.

It may be noted that if $h(w)$ includes a square it contains a pre-square.

3. Characterizations of squarefree morphisms

The result mentioned in the introduction is a consequence of Theorem 2 below which gives a precise characterization of squarefree morphisms. Theorem 2 is an improvement on Theorem 1 which is proved first. Both theorems use the notion of pre-square and lead to efficient algorithms for testing squarefreeness of morphisms (on finite alphabets).

Theorem 1. ($|A| \geqslant 2$), *A morphism* $h: A^+ \to B^+$ *is squarefree iff for all distinct letters* a, b *in* A, $h(ab)$ *contains no pre-square.*

Proof. We will prove that if h is not squarefree then there is a pre-square in the image of a word ab of length two, with $a \neq b$.

Let be $w = a_1a_2 \ldots a_k$ a squarefree word in A^+ such that $h(w) = \alpha u^2\bar{\gamma}$, for some u in B^+ and α, $\bar{\gamma}$ in B^*. It can be assumed that α is a prefix of $h(a_1)$ with $\alpha \neq h(a_1)$, and that $\bar{\gamma}$ is a suffix of $h(a_k)$ with $\bar{\gamma} \neq h(a_k)$.

Let us define j as the least integer for which αu is a prefix of $h(a_1 \ldots a_j)$; and let $\bar{\beta}$ be the suffix of $h(a_j)$ which satisfies $h(a_1 \ldots a_j) = \alpha u\bar{\beta}$.

We denote by $\bar{\alpha}$, β, γ the words defined by the relations: $h(a_1) = \alpha\bar{\alpha}$, $h(a_j) = \beta\bar{\beta}$, $h(a_k) = \gamma\bar{\gamma}$. The words $\bar{\alpha}$, β and γ are not empty.

If $1 \leqslant j \leqslant 2$ or $k-1 \leqslant j \leqslant k$, then the conclusion holds. Thus, in the following j is assumed to be strictly between 2 and $k-1$.

Two cases are first examined in which $\bar{\alpha}$, $\bar{\beta}$ on one hand and β, γ on the other are assumed to be of the same length; this implies $\bar{\alpha} = \bar{\beta}$ since these words are both prefixes of u; we also have $\beta = \gamma$ as suffixes of u.

Case 1: If a_j is different from a_1 and a_k, then $h(a_1a_ja_k)$ includes the square $\bar{\gamma}\beta = \bar{\beta}\gamma$. The conclusion holds.

Case 2: If for instance $a_1 = a_j$, let i be the least integer with $a_i \neq a_{j+i-1}$. This integer exists and is strictly between 1 and the minimum of j and $k-j+1$, otherwise w itself would be a square. In that situation, $h(a_i)$ and $h(a_{j+i-1})$ are prefixes of a same suffix of u. Then, one of them is a prefix of the other, which means that $h(a_ia_{j+i-1})$ or $h(a_{j+i-1}a_i)$ contain a square.

There are several cases still to examine where $|\bar{\alpha}| \neq |\bar{\beta}|$. The cases where $|\beta| \neq |\gamma|$ can be handled symmetrically.

First, $\bar{\alpha}$ is assumed to be of smaller length than $\bar{\beta}$. Since both are prefixes of u, we can find a word δ in B^+ satisfying $\bar{\alpha}\delta = \bar{\beta}$.

Case 3: $a_2 \neq a_j$. δ is a prefix of $h(a_2 \ldots a_i)$. Let i be the least integer such that δ is a prefix of $h(a_2 \ldots a_i)$. The word δ being a factor of $h(a_j)$ if the letter a_j has an occurrence in $a_2 \ldots a_i$, then it appears only once and $a_i = a_j$ necessarily. Adding to that fact the squarefreeness of $a_2 \ldots a_i$ we deduce that $a_ja_2 \ldots a_i$ is squarefree. Then $h(a_ja_2)$ contains a pre-square which is an occurrence of δ.

Case 4: $a_2 = a_j$. From $j > 2$ we deduce that $h(a_2)$ is a factor of u. So the non-empty word $(\beta\bar{\alpha})^2$ is a prefix of $k(a_j \ldots a_k)$. Thus, $h(a_ja_{j+1})$ has a pre-square.

Finally let us assume $|\bar{\alpha}| > |\bar{\beta}|$, and define δ by $\bar{\alpha} = \bar{\beta}\delta$.

Case 5: $a_1 \neq a_{j+1}$. The word δ is a prefix of $h(a_{j+1} \ldots a_k)$. If i is the least integer greater than $j+1$ such that δ is a prefix of $h(a_{j+1} \ldots a_i)$, then $a_1a_{j+1}a_{j+2} \ldots a_i$ is squarefree for the same reason as in Case 3. therefore, $h(a_1a_{j+1})$ contains a pre-square which is an occurrence of δ.

Case 6: $a_1 = a_{j+1}$ and $\alpha\bar{\beta} \neq 1$. The word $h(a_{j+1})$ is a factor of u because $j < k$. The non-empty word $(\alpha\bar{\beta})^2$ is then a prefix of $h(a_1 \ldots a_j)$ which means that $h(a_1a_2)$ includes a pre-square.

Case 7: $\alpha\bar{\beta} = 1$. This implies $u = h(a_1 \ldots a_j)$. As in Case 2, there must exists an integer i between 1 and the minimum of j and $k-j$ such that $a_i \neq a_{j+i}$ and one of the two words, $h(a_i)$, $h(a_{j+i})$, is a prefix of the other. If there were no such i, w itself would be a square. Therefore, $h(a_ia_{j+i})$ or $h(a_{j+i}a_i)$ contains a square.

Theorem 2. *A morphism $h: A^+ \to B^+$ is squarefree iff:*

(i) *h is 3-squarefree;*

(ii) *for any a in A, $h(a)$ does not have any internal pre-square.*

Proof. We have to show that a non-squarefree morphism catisfies (non i) or (non ii). By Theorem 1, we consider two different letters a and b and a factor u of $h(ab)$ with a pre-square occurrence. There exists a word $w = c_1 \ldots c_k$ of minimal length such that abw is squarefree and u^2 is a factor of $h(abw)$. It can be supposed that $k > 1$, otherwise (non i) holds.

In the two first cases the pre-square in $h(ab)$ is assumed not to be in $h(a)$ nor in $h(b)$. There are words α, $\bar{\beta}$ in B^* and $\bar{\alpha}$, β in B^+ such that: $h(a) = \alpha\bar{\alpha}$, $h(b) = \beta\bar{\beta}$, $u = \bar{\alpha}\beta$.

Case 1: $|\bar{\beta}|<|\bar{\alpha}|$. As $\bar{\alpha}$ and $\bar{\beta}$ are prefixes of u we may define δ by the relation $\bar{\alpha}=\bar{\beta}\delta$. This word δ is a prefix of $h(c_1\dots c_i)$ for a least integer i. If $c_i\neq a$, (non i) holds because $h(ac_i)$ contains a square. If $c_i=a$ and $i>1$, then $ac_{i-1}c_i$ is squarefree and $h(ac_{i-1}c_i)$ contains a square; so (non i) holds again. If δ is a prefix of $h(c_1)$ and $c_1=a$, by the minimality of $|w|$ and the assumption on k, $h(c_1)$ $(=h(a))$ is a factor of u. Two occurrences of $h(a)$ overlap in $h(ab)$, which means that $h(ab)$ includes a square.

Case 2: $|\bar{\beta}|\geq|\bar{\alpha}|$. Let δ be the word such that $\bar{\beta}=\bar{\alpha}\delta$. The non-empty word $\delta h(c_1)$ is a factor of u which is both a suffix of $h(bc_1)$ and a prefix of $h(b)$. Then $h(bc_1b)$ has a square and (non i) holds.

Let us suppose now that u is a factor of $h(b)$, for instance.

Case 3: u is a prefix of $h(b)$. We have $h(b)=u\delta$ for a word δ in B^*. The word $\delta h(c_1)$ is non empty and is a prefix of u by the assumption on k. The image of bc_1b is not squarefree; (non i) holds.

Case 4: u is a suffix of $h(b)$. If $c_k\neq b$, then $h(bc_k)$ is not squarefree (minimality of k) and (non i) holds; if not, then certainly $k>1$, $bc_{k-1}c_k$ is squarefree and its image is not squarefree; (non i) holds.

Case 5: u is an internal factor of $h(b)$; (non ii) holds.

4. Bounds for squarefreeness

Two corollaries may be immediately deduced from Theorem 2. For a morphism h let us set

$$M(h)=\max\{|h(a)|\mid a\in A\},\qquad m(h)=\min\{|h(a)|\mid a\in A\}.$$

Let us also denote by $\lceil x\rceil$ the least integer greater than or equal to a real number x. With these notations, we have:

Corollary 3. *A morphism h is squarefree iff it is k-squarefree with* $k=\max(3,\lceil(M(h)-3)/m(h)\rceil)$.

The following corollary applies to a particular class of morphisms: uniform morphisms. For these morphisms we have $|h(a)|=|h(b)|$, for any a,b in A.

Corollary 4. *A uniform morphism is squarefree iff it is* 3*-squarefree.*

The above corollary is also a consequence of a theorem of Thue [8] reformulated in [1] which gives a sufficient condition for a morphism to be squarefree.

Another consequence of Theorem 2 is the result mentioned in the introduction.

Corollary 5. *A morphism defined on a three letter alphabet is squarefree iff it is* 5*-squarefree.*

Proof. Only the second assertion of Theorem 2 is of interest here. Let us consider a letter a for which $h(a)$ contains a pre-square u. Let $w = c_1 \ldots c_k$ be a word that duplicates the pre-square occurrence of u in $h(a)$. The word w may be chosen of minimal length; so if a appears in $c_1 \ldots c_k$, then it can occur only once at the end ($a = c_k$) because $|u| \leq |h(a)|$. The word $c_1 \ldots c_{k-1}$ being squarefree on two letters is thus of length 3 at most. The conclusion follows.

Our last consequence is:

Corollary 6. ($k \geq 3$). *A morphism h is k-squarefree iff it is 3-squarefree and* $\min\{|w| \mid w$ *duplicates an internal pre-square of* $h(a)$, $a \in A\} \geq k$.

Proof. Let us first suppose that h is k-squarefree. Then h is also 3-squarefree because $k \geq 3$. If w duplicates a pre-square in $h(a)$, then aw is squarefree and $h(aw)$ contains a square, for instance. This implies $|aw| > k$ or $|w| \geq k$. If h is not k-squarefree, there exists a squarefree word w, $|w| \leq k$, which image contains a square. Going through the proofs of Theorem 1 and 2, we can find a squarefree word aw' such that: $|aw'| \leq 3$ or w' duplicates an internal pre-square of $h(a)$. The main point to be noted is that in the second case $|aw'| \leq |w|$. Therefore if h is 3-squarefree the inequality cannot be true.

5. Optimality of bounds

Let us show that the bounds in Corollaries 3 and 5 are optimal; a property which is trivially true for Corollary 4.

Lemma 7. *Let h be a morphism from* $\{a_1, \ldots, a_n\}^+$ *into* $\{b, a_1, \ldots, a_n\}^+$ *defined by*:

$$h(a_1) = ba_1a_2 \ldots a_na_1, \qquad h(a_i) = a_i \quad \textit{for } 2 \leq i \leq n.$$

Then h is k-squarefree with $k = \max(3, \lceil (M(h) - 3)/m(h) \rceil - 1$ *and h is not squarefree.*

Proof. We have $k = n - 1$, and the image of $a_1 \ldots a_n$ is

$$ba_1 \ldots a_na_1a_2 \ldots a_n$$

which contains the square $(a_1 \ldots a_n)^2$. The $(n-1)$-squarefreeness comes from Corollary 6: h is 3-squarefree and $a_1 \ldots a_n$ is the only one pre-square of the $h(a)$'s.

Lemma 8. *The morphism* $h : \{a, b, c\}^+ \to \{a, b, c, d, e\}^+$ *defined by* $h(a) = deabcbda$, $h(b) = b$, $h(c) = c$ *is 4-squarefree but not squarefree.*

Proof. We have $h(abcba) = deabcbdabcbdeabcbda$ which contains the square

$(abcbd)^2$. Then, h is 3-squarefree and $abcbd$ is only one internal pre-square of $h(a)$. The result follows from Corollary 6.

6. Bound on more than three letters

We now show that on an alphabet with more than three letters the bound on the length of words that must be examinated to decide the squarefreeness of a morphism depends on the length of the images of the letters by the morphism.

Theorem 9. *Let* $|A|>3$. *For any integer* n, *there exists a morphism which is* n-*squarefree and not squarefree.*

Proof. It suffices to prove the theorem for a four-letter alphabet, $\{a, b, c, d\}$.

Let x be a squarefree word on $\{b, c, d\}$ of length n. We know, from Thue [8], that there exists such a word. We define a morphism $h : \{a, b, c, d\} \to \{a, b, c, d, e\}$ by:

$$h(a) = eaxa, \qquad h(b) = b, \qquad h(c) = c, \qquad h(d) = d.$$

The word ax is squarefree of length $n+1$ and its image by h is $eaxax$. So h is not squarefree. The n-squarefreeness of h is a consequence of Corollary 6.

References

[1] D. Bean, A. Ehrenfeucht and G. McNulty, Avoidable patterns in strings of symbols, *Pacific J. Math.* **85** (1979) 261–294.

[2] J. Berstel, Sur les mots sans carré définis par un morphisme, *Proc. 6th ICALP Symposium*, Lecture Notes in Computer Science **71** (Springer, Berlin, 1979) 16–25.

[3] J. Berstel and C. Reutenauer, Squarefree words and idempotent semigroups, in: Lothaire, Ed., *Combinatorics on Words* (Addison-Wesley, Reading, MA) Chapter 2, to appear.

[4] F.J. Brandenburg, On non-repetitive strings, prepublication (1981).

[5] A. Ehrenfeucht and G. Rosenberg, On the separating power of EOL systems, *RAIRO Informat. Théor.*, to appear.

[6] J. Karhumaki, On cubicfree ω-words generated by binary morphisms, prepublication (1981).

[7] R. Ross and R. Winklmann, Repetitive strings are not context-free, CS-81-070, Washington State University, Pullman, Washington (1981).

[8] A. Thue, Uber die gegenseitige Lage gleicher teile gewisser Zeichenreihen, *Norske Vid. Selsk. Skr., I. Mat. Nat. Kl., Christiana*, **1** (1912) 1–67.