TUCS

Tero Harju | Dirk Nowotka

Binary Words with Few Squares

Turku Centre *for* Computer Science

TUCS Technical Report
No 737, January 2006

Binary Words with Few Squares

Tero Harju
Turku Centre for Computer Science and Department of Mathematics,
University of Turku, FIN-20014 Turku, Finland
harju@utu.fi
Dirk Nowotka
Institute of Formal Methods in Computer Science,
University of Stuttgart, D-70569 Stuttgart, Germany
dirk.nowotka@informatik.uni-stuttgart.de

TUCS Technical Report
No 737, January 2006

Abstract

A short proof is given for a result of Fraenkel and Simpson [Electronic J. Combinatorics 2 (1995), 159–164] stating that there exists an infinite binary word which has only three different squares u^2.

Keywords: combinatorics on words, Weinbaum factorization, critical points, bordered word, primitive words

TUCS Laboratory
Discrete Mathematics for Information Technology

1 Introduction

Consider the set $\Sigma_k = \{0, 1, \dots, k-1\}$ as an alphabet of symbols, called letters. The set of all *words* over Σ_k, denoted by Σ_k^*, consists of the finite sequences of elements from Σ_k. We denote by $|w|$ the length of the word w, i.e., the number of occurrences of the letters in w. An *infinite word* over Σ_k is a mapping from $w \colon \mathbb{N} \to \Sigma_k$ which is usually presented as the ordered sequence $w(1)w(2)\cdots$ of the images.

A word u is a *factor* of a word $w \in \Sigma_k^*$, if $w = w_1 u w_2$ for some words w_1 and w_2. A nonempty factor u^2 $(= uu)$ is called a *square* in w. The word w is *square-free* if it has no squares.

It is easy to see that every binary word $w \in \Sigma_2^*$ with $|w| \geq 4$ has a square. Indeed, the only square-free binary words are $0, 1, 01, 10, 010, 101$. On the other hand, Entringer, Jackson, and Schatz [1] showed in 1974 that there exists an infinite word with 5 different squares. Later Fraenkel and Simpson [2] showed that there exists an infinite binary word over the alphabet $\{0, 1\}$ that has only three squares 00, 11, and 0101. We shall give a short proof of this result based on square-free words over a three letter alphabet.

Example 2. We can show that the word $w = 101100111000111100111011$ of length 24 has the squares u^2 only for $u \in \{0, 1, 11\}$. Also, it can be checked that the word w has the maximum length 24 among the words having only the squares 0^2, 1^2, and $(1^2)^2$. Indeed, it can be checked that all words of length 25 or more do have three squares 0^2, 1^2, and u^2 for some word $u \notin \{00, 11\}$.

The following result is due to Axel Thue; see, e.g. Lothaire [3] for background and a proof of this theorem.

Theorem 1. *There exists an infinite square-free word over the three letter alphabet* Σ_3.

3 Three squares

In the rest of this paper, we prove

Theorem 2. *There exists an infinite binary word* W *that has only the squares* u^2 *for* $u \in \{0, 1, 01\}$.

Proof. Let w be an infinite square-free word over the alphabet Σ_3 provided by Theorem 1. Also, let W be the word where each 0 (1, 2, resp.) in w is substituted by A (B, C, resp.) where

$$\begin{aligned} A &= 1^3 0^3 1^2 0^2 1 0 1^2 0^3 1^3 0^2 1 0 \,, \\ B &= 1^3 0^3 1 0 1^2 0^3 1^3 0^2 1 0 1^2 0^3 1 0 \,, \\ C &= 1^3 0^3 1^2 0^2 1 0 1^2 0^3 1 0 1^3 0^2 1 0 1^2 0^2 \,. \end{aligned}$$

It is easy to check that these words have only the squares 0^2, 1^2 and $(01)^2$. Denote $\Delta = \{A, B, C\}$. We have the following marking property of $1^3 0^3$:

$$\text{the factor } 1^3 0^3 \text{ occurs only as a prefix of each } A,\ B,\ C \tag{a}$$

Also, we notice that the longest common prefix (suffix, resp.) of two words from Δ is $1^3 0^3 1^2 0^2 1 0 1^2 0^3 1$ of length 18 ($0^2 1 0$ of length 4, resp.). Since the words A, B and C are longer than 22, we obtain

no word $Z \in \Delta$ can be factored as $Z = yx$ where x is a suffix of a word $X \in \Delta \setminus \{Z\}$ and y is a prefix of a word $Y \in \Delta \setminus \{Z\}$. (b)

Suppose that W has a square U^2, where $U \notin \{0, 1, 01\}$. It is straightforward using (a) to verify that the short words XY for different $X, Y \in \Delta$, do not contain other squares than 0^2, 1^2, and $(01)^2$. In particular, we may assume that U^2 has a factor Z from Δ.

Suppose first that U does not have a factor from Δ, and thus that U has at most one occurrence of the marker $1^3 0^3$. Now $U^2 = xZy$ for some $Z \in \Delta$, where x is a suffix of a word $X \in \Delta$ and y is a prefix of a word $Y \in \Delta$ for $X, Y \neq Z$, since w is square-free. We divide our considerations into two cases.

(1) Assume that U has exactly one occurrence of $1^3 0^3$. In this case, $Z = yx$, since y begins with the unique occurrence of $1^3 0^3$ inside the second U. This, however, contradicts the property (b).

(2) Assume that U has no occurrences of $1^3 0^3$. In this case, $U = xu = vy$ so that $Z = uv$, where u and y are proper prefixes of $1^3 0^3$, and as they are both suffixes of U, one of them is a suffix of the other. Necessarily $u = y$, and hence also $x = v$. But now $Z = yx$ contradicts the property (b).

Suppose then that U has a factor from Δ. If $U = uXv$ with $X \in \Delta$, then necessarily $U^2 = uXvuXv$ is a factor in W such that $vu \in \Delta^*$, because of the marking property (a). Therefore, $U = xX_1 X_2 \cdots X_n y$ such that $X_0 X_1 \cdots X_n Z X_1 \cdots X_n X_{n+1}$ is a factor of W where $X_i \in \Delta$ for each i, x is a suffix of X_0, $Z = yx \in \Delta$, and y is a prefix of X_{n+1}. Now $Z \neq X_0$, since otherwise $(X_0 X_1 \cdots X_n)^2$ would be a factor of W and this would contradict the square-freeness of w. Similarly, $Z \neq X_{n+1}$. Again $Z = yx$ contradicts (b). $\square$

We observe that the word

$$w = 10110011100011001000111000100$$

of length 29 has only the squares u^2 for $u \in \{0, 1, 100\}$. It can be shown, with the aid of a computer, that each word w of length $|w| \geq 30$ with exactly three different squares, has the squares 0^2, 1^2 and $(01)^2$ or 0^2, 1^2 and $(10)^2$.

References

[1] R. C. Entringer, D. E. Jackson, and J. A. Schatz. On nonrepetitive sequences. 16, 1974.

[2] A. S. Fraenkel and J. Simpson. How many squares must a binary sequence contain? *Electronic J. Combinatorics*, 2(#R2), 1995.

[3] M. Lothaire. *Combinatorics on Words*. Addison-Wesley, Reading, MA, 1983. Reprinted in the Cambridge Mathematical Library, Cambridge Univ. Press, 1997.

Turku
Centre *for*
Computer
Science

Lemminkäisenkatu 14 A, 20520 Turku, Finland | www.tucs.fi

University of Turku

- Department of Information Technology
- Department of Mathematics

Åbo Akademi University

- Department of Computer Science
- Institute for Advanced Management Systems Research

Turku School of Economics and Business Administration

- Institute of Information Systems Sciences

ISBN 952-12-1662-X
ISSN 1239-1891