Cyclically Repetition-free Words on Small Alphabets

Tero Harju[1] and Dirk Nowotka[2]

[1] Turku Centre for Computer Science (TUCS), Department of Mathematics, University of Turku, Finland
harju@utu.fi

[2] Institute for Formal Methods in Computer Science (FMI), Universität Stuttgart, Germany
nowotka@fmi.uni-stuttgart.de

Abstract. All sufficiently long binary words contain a square but there are infinite binary words having only the short squares 00, 11 and 0101. Recently it was shown by J. Currie that there exist cyclically square-free words in ternary alphabet except for lengths 5, 7, 9, 10, 14, and 17. We consider binary words all conjugates of which contain only short squares. We show that the number $c(n)$ of these binary words of length n grows unboundedly. In order for this, we show that there are morphisms that preserve cyclically square-free words in the ternary alphabet.

1 Introduction

We shall consider binary ($w \in \{0,1\}^*$) and ternary ($w \in \{0,1,2\}^*$) words. A word u is a *factor* of a word w if there are words w_1 and w_2 such that $w = w_1uw_2$. In this case, u *occurs* in w. Two words u and v are *conjugates* if $u = xy$ and $v = yx$ for some words x and y. The *conjugate class* of a word w consists of the words that are conjugates of w. For a given lexicographic order on the alphabet, each conjugate class has a minimal element that is called a *Lyndon word*. A nonempty factor u^2 $(= uu)$ of a word w is a *square* in w. The word w is *square-free* if it has no squares in it. Moreover, w is *cyclically square-free* if its conjugates are square-free.

While each binary word $w \in \{0,1\}^*$ of length at least four contains a square, Entringer, Jackson, and Schatz [3] showed that there exists an infinite word with only 5 different squares. Afterwards Fraenkel and Simpson [4] showed that there exists an infinite binary word having only the three squares 00, 11, and 0101. We say that a binary word w is *short-squared* if its squares belong to the set $\{00, 11, 0101\}$ – but we do not allow the square 1010.

Theorem 1 (Fraenkel–Simpson). *For each $n \geq 1$, there exists a short-squared binary word of length n.*

A simplified proof of Theorem 1 was given by Rampersad, Shallit, and M.-w. Wang [7] which was still shortened by the present authors in [5]. In this paper we consider cyclic words with short squares. The problem was motivated by the following result due to J. Currie [2].

Theorem 2 (Currie). *There exists a cyclically square-free ternary word w of length n if and only if $n \notin \{5, 7, 9, 10, 14, 17\}$.*

A word w is *cyclically short-squared* if its conjugates are all short-squared. We shall show in Theorem 5 that there are arbitrarily long cyclically short-squared binary words.

The exception list of lengths for cyclically short-squared binary words is much more extensive than the list for cyclically square-free ternary words given by Currie. Indeed, it is an open problem to characterize the set L_{cyc} of lengths n for which there exists a cyclically short-squared binary word of length n. Also, even for each length $n \in L_{\text{cyc}}$ there seems to be only a small number of solutions as seen from the next table.

Let $c(n)$ denote the number of conjugate classes of cyclically short-squared binary words of length n, i.e., $c(n)$ is the number of cyclically short-squared binary Lyndon words having length n.

n	4	5	6	7	8	9	10	11	12	13	14	15	16	17
$c(n)$	3	2	2	2	1	0	0	0	3	0	1	0	0	0

n	18	19	20	21	22	23	24	25	26	27	28	29
$c(n)$	0	2	1	0	0	0	3	0	0	0	1	0

n	30	31	32	33	34	35	36
$c(n)$	1	0	0	0	0	0	2

Table 1. Curious sequence of numbers of cyclic short-squared binary words.

Remark 1. Note that any (not necessarily cyclic) short squared word w that does not have both factors 000 and 111 is not longer than 21. The longest such words are of length 21:

$$110111001101001110010 \quad \text{and} \quad 110111001101001110100$$

and their duals, where 0 and 1 are interchanged. Hence a Lyndon representative of a cyclic short-squared binary word w of length at least 22 starts with 11100 when the order is given as $1 \prec 0$. Indeed, it cannot start with 11101 since it then has a conjugate starting with 0111011 which gives a contradiction at the next bit.

Example 1. Let us consider some examples of cyclically short-squared binary words. We choose the ordering $1 \prec 0$ for the alphabet for our own convenience.

The Lyndon representative of length $n = 12$ are the following three words:

$$111001011000\,,$$
$$111000101100\,,$$
$$111000110010\,.$$

The Lyndon representative of length $n = 24$ are the following words:

$$111001011001110001011000\,,$$
$$111001011100011001011000\,,$$
$$111000110010111000101100\,.$$

There are, however, only two Lyndon representatives of length $n = 36$:

$$111001011001110001100101110001011000\,,$$
$$111001011100010110011100011001011000\,.$$

Despite of Table 1 suggesting a shrinking number of cyclic short-squared binary words when the length grows, we will show

Theorem 3. *The function $c(n)$ is unbounded:*

$$\limsup_{n\to\infty} c(n) = \infty\ .$$

Consider now a uniform morphism $\xi\colon \{0,1,2\}^* \to \{0,1\}^*$ that takes cyclic ternary words to cyclic short-squared binary words. Such a morphism can be found by composing β from Section 3 with α from Section 2 below. Let u and v be two different cyclic square-free ternary words of the same length. Then $\xi(u)$ and $\xi(v)$ are two different cyclic short-squared binary words of the same length. Hence, Theorem 3 follows from the next result. Let $c_3(n)$ denote the number of cyclically square-free ternary Lyndon words of length n w.r.t. some fixed order.

Theorem 4. *The function $c_3(n)$ is unbounded:*

$$\limsup_{n\to\infty} c_3(n) = \infty\ .$$

This result will be proved in Section 3. We also state the following conjecture that is stronger than Theorem 3.

Conjecture 1. There exists an integer N such that $c(n) > 0$ for all $n \geq N$.

2 On Cyclic Binary Words with Short Squares

The following theorem is proven in this section.

Theorem 5. *There are arbitrarily long cyclically short-squared binary words.*

Before we prove Theorem 5 let us recall a morphism that maps square-free ternary words to short-squared binary words.

Let $\alpha\colon \{0,1,2\}^* \to \{0,1\}^*$ be the morphism defined by

$$\alpha(0) = A := 1^3 0^3 1^2 0^2 1 0 1^2 0^3 1^3 0^2 1 0\,,$$
$$\alpha(1) = B := 1^3 0^3 1 0 1^2 0^3 1^3 0^2 1 0 1^2 0^3 1 0\,,$$
$$\alpha(2) = C := 1^3 0^3 1^2 0^2 1 0 1^2 0^3 1 0 1^3 0^2 1 0 1^2 0^2\,.$$

We notice in passing that these words are short-squared, and the words A and C are cyclically short-squared, but B is not. Indeed, B has a conjugate 1000101110001011000111001 01 which has the long square $(10001011)^2$ as its prefix.

The following result was shown in [5].

Theorem 6. *Let $w \in \{0,1,2\}^*$. Then w is a square-free ternary word if and only if $\alpha(w)$ is a short-squared binary word.*

We now turn to the proof of the announced result.

Proof (of Theorem 5). Let then w be a cyclically square-free ternary word provided by Theorem 2, and consider the binary word $\alpha(w)$. By Theorem 6, $\alpha(w)$ is short-squared. The claim follows when $\alpha(w)$ is shown to be cyclically short-squared. Assume, on the contrary, that $\alpha(w)$ has a conjugate v that is not short-squared. Without loss of generality, we can assume that v has a square as a suffix, say

$$v = su^2 \,,$$

where u^2 is a shortest possible square in the conjugates of $\alpha(w)$. One easily checks from the α images of words of length at most two that $|u| \geq 3$ (see also the comment above Theorem 6). Since w is cyclically square-free, it follows that $v \neq \alpha(u)$ for all conjugates u of w.

Denote $\Delta = \{A, B, C\}$. We have the following *marking property* of $1^3 0^3$:

$$1^3 0^3 \text{ occurs only as a prefix in } A,\ B \text{ and } C.$$

Let z be a shortest prefix of v, say $v = zt$, such that the conjugate tz is in Δ^*. In particular, there exists an $X \in \Delta$ such that $X = yz$ for some y.

Since u^2 is not a factor of the conjugate tz, we must have $|s| < |z|$, say $z = sz'$. Therefore, $u^2 = z't = z'x'y$ for some word x'. However, the marking property and $|u| \geq 3$ implies that $|u| > |y|$ and, hence,

$$u = z'xy \quad \text{and} \quad X = ysz'$$

for some prefix x of a word in Δ^*. Now $tz = xyz'xyz \in \Delta^*$ which ends with the word $X = yz$. It follows that $xyz'x \in \Delta^*$, i.e., x occurs as a suffix and a prefix in Δ^*. This implies that $x \in \Delta^*$ by the marking property. Hence also for the middle part $yz' \in \Delta^*$. Since yz' is shorter than X, it follows that $yz' \in \Delta$. Now both yz' and ysz' are in Δ. This would imply that $|s| = 3$ or 6; however there is no solution for these parameters in Δ. (The length of the longest common prefix, rep. suffix, of two different words of Δ is 18, resp. 4.) □

3 On the Number of Cyclic Square-Free Words

A morphism is called (cyclic) square-free whenever the image of any (cyclic) square-free word is itself (cyclic) square-free. In this section we will construct a set of uniform cyclic square-free morphisms on $\{0,1,2\}^*$ such that an arbitrary number of cyclic square-free words of the same length can be generated.

We start from certain square-free factors taken from an infinite square-free word in order to find substitutions that preserve square-freeness. Then we introduce several markers that allow us to both ensure cyclic square-freeness and the construction of arbitrarily many different substitutions without sacrificing the preservation of square-freeness.

Thue gave in [8] the following morphism ϑ on $\{0,1,2\}^*$ which generates the infinite *Thue word* $\mathbf{t}$ when iterated starting in 0. Consider

$$\vartheta(0) = 012\,, \qquad \vartheta(1) = 02\,, \qquad \vartheta(2) = 1$$

which gives

$$\mathbf{t} = \lim_{k\to\infty} \vartheta^k(0) = \underline{0120210}12102012021\underline{0201210}12021\underline{01210201}2\cdots \tag{1}$$

where we point out three underlined factors of $\mathbf{t}$ which will be used further below. It is well-known that $\mathbf{t}$ is square-free. We will take factors of $\mathbf{t}$ as building blocks for the morphisms $(\gamma_n)_{n\in\mathbb{N}}$. The following morphism $\eta\colon \{0,1,2\}^* \to \{0,1\}^*$ maps $\mathbf{t}$ to an overlap-free binary word [6], the so called *Thue-Morse word*,

$$\eta(0) = 011\,, \quad \eta(1) = 01\,, \quad \eta(2) = 0\,.$$

A word is called *overlap-free* if it has no overlapping factors, i.e., if no factor of the form $awawa$ occurs where a is a letter and w is a (possibly empty) word. In particular the words in the following set do not occur in $\mathbf{t}$:

$$T_{\text{no}} = \{010\,, 212\,, 1021\,, 1201\}\,. \tag{2}$$

Indeed, $\eta(010) = 01101011$ which contains the overlap 10101. Assume that contrary to the claim 212 occurs in $\mathbf{t}$. Then it must be preceded and succeeded by 0 otherwise $\mathbf{t}$ is not square-free. But, $\eta(02120) = 0110010011$ contains the overlap 1001001; a contradiction. If 1021 occurs in $\mathbf{t}$, then it must be preceded by 2 and succeeded by 0 by the previous arguments. But, then $\mathbf{t}$ contains the square 210210; a contradiction. A similar argument holds for the word 1201.

So far, we have identified in T_{no} square-free words that do not occur in $\mathbf{t}$. They will serve as markers in the proof of Theorem 4 below. Let us now turn to factors of $\mathbf{t}$ that we can use as building blocks for the morphisms $(\gamma_n)_{n\in\mathbb{N}}$.

Iterating ϑ gives

$$\begin{aligned}
\vartheta(0) &= 012\\
\vartheta^2(0) &= 012021\\
\vartheta^3(0) &= 012021012102\\
\vartheta^4(0) &= 012021012102012021020121\\
&\ \ \vdots
\end{aligned}$$

and

$$\begin{aligned} \vartheta(1) &= 02 \\ \vartheta^2(1) &= 0121 \\ \vartheta^3(1) &= 01202102 \\ \vartheta^4(1) &= 0120210121020121 \\ &\vdots \end{aligned} \qquad \text{and} \qquad \begin{aligned} \vartheta(2) &= 1 \\ \vartheta^2(2) &= 02 \\ \vartheta^3(2) &= 0121 \\ \vartheta^4(2) &= 01202102 \\ &\vdots \end{aligned}$$

Consider the words $\vartheta^4(0)$ and $\vartheta^4(1)$ and $\vartheta^4(2)$ that start with 012021 and that all have an occurrence in $\mathbf{t}$ followed by 0120. Indeed, $\vartheta^6(0)$ is a prefix of $\mathbf{t}$ implying $\vartheta^6(0) = \vartheta^4(012021) = \vartheta^4(0)\vartheta^4(1)\vartheta^4(2)\vartheta^4(0)\vartheta^4(2)\vartheta^4(1)$.

Let δ be a morphism on $\{0,1,2\}^*$ defined by

$$\begin{aligned} \delta(0) &= (012)^{-1}\vartheta^4(0)012 = 021012102012021020121012 \ , \\ \delta(1) &= (012)^{-1}\vartheta^4(1)012 = 0210121020121012 \ , \\ \delta(2) &= (012)^{-1}\vartheta^4(2)012 = 02102012 \ . \end{aligned}$$

We have

Claim 1. *The δ-image of each factor of* $\mathbf{t}$ *occurs itself in* $\mathbf{t}$ *followed by* 021*.*

Indeed, let w be a factor of $\mathbf{t}$, then $\vartheta(w)$, and hence, $\vartheta^4(w)$ is a factor of $\mathbf{t}$. Therefore, $(012)^{-1}\vartheta^4(w)$ is a factor of $\mathbf{t}$ which proves the claim since $(012)^{-1}\vartheta^4(wa)$ occurs in $\mathbf{t}$, for some letter a such that wa occurs in $\mathbf{t}$, and $012 \leq_{\mathrm{p}} \vartheta^4(a)$.

Consider the factors 0201210 and 0120210 and 0121020 of $\mathbf{t}$ as marked in (1). Note that these factors are of the same length and have the same number of occurrences of 0, 1, and 2, respectively.

Let us define the following uniform morphism β on $\{0,1,2\}^*$ where the length of the images of letters is $|\beta(0)| = 122$:

$$\begin{aligned} \beta(0) &= \delta(0201210)01 \\ \beta(1) &= \delta(0120210)01 \\ \beta(2) &= \delta(0121020)01 \end{aligned}$$

Claim 2. *The images $\beta(i)$ are cyclic square-free for all $0 \leq i \leq 2$.*

Proof. The claim can be easily proven by inspection or a computer test. However, we give an alternative proof here for illustrating some arguments also used later below.

By Claim 1 the prefix $\beta(i)1^{-1}$ of $\beta(i)$ is a factor of $\mathbf{t}$ for all $0 \leq i \leq 2$. The words $\beta(i)$ end with 1201 which is in the set T_{no} of forbidden factors of $\mathbf{t}$. It follows that the words $\beta(i)$ are square-free. It is also straightforward to verify that $\beta(i)$ are cyclic square-free. Indeed, any cyclic square x^2 must contain the last letter 1 of $\beta(i)$. The case where $|x| < 6$ is easily checked by hand. Note that $1\beta(i)1^{-1}$ begins with 1021 and $\beta(i)$ ends with 1201. Hence, if $|x| \geq 6$ then x

contains 1021 or 1201. But $1021, 1201 \in T_{\text{no}}$ and therefore they occur at most once in any conjugate of $\beta(i)$ which contradicts that x^2 occurs in a conjugate of $\beta(i)$. This concludes the proof of Claim 2.

Let π be any permutation on $\{0, 1, 2\}$. Then we define the following morphisms

$$\beta_\pi(i) = \beta(\pi(i))$$

for all $0 \leq i \leq 2$. Before we show that every β_π is cyclic square-free, we recall the following theorem by Thue [8]; see [1] for a slightly improved version.

Theorem 7. *A morphism α is square-free if the following two conditions are satisfied:*

(1) $\alpha(w)$ is square-free whenever u is square-free and $|u| \leq 3$ and
(2) $\alpha(a)$ is not a proper factor of $\alpha(b)$ for any letters a and b.

In order to show that the constructed morphisms are cyclic square-free we state the following result.

Proposition 1. *A morphism α is cyclic square-free if the following two conditions are satisfied:*

(1) α is square-free and
(2) $\alpha(a)$ is cyclic square-free for all letters a.

Proof. Let $w_{(i)}$ denote ith letter of the word w. Consider a cyclic square-free word w of length n and suppose, contrary to the claim, that $\alpha(w)$ is not cyclic square-free. Let x^2 be a shortest square in $\alpha(w)$. Then x^2 occurs either in $w_{(i)}w_{(i+1)} \cdots w_{(n)}w_{(1)} \cdots w_{(i-1)}$ or in $w_{(i)}w_{(i+1)} \cdots w_{(n)}w_{(1)} \cdots w_{(i-1)}w_{(i)}$ for some i. Both of these words are square-free if w is cyclic square-free, except if $n = 1$; a contradiction in any case. □

It is now straightforward to establish the cyclic square-freeness of any β_π which implies Theorem 4.

Lemma 1. *Let π be any permutation on $\{0, 1, 2\}$. Then β_π is a cyclic square-free morphism.*

Proof. Let $w_{(i)}$ denote ith letter of the word w.

We begin by showing that β_π is square-free. By Theorem 7 the square-freeness of β_π can be checked by hand. However, this is cumbersome and therefore we give an alternative proof without the use of Theorem 7. Suppose contrary to the claim that $\beta_\pi(w)$ contains a square x^2 where w is square-free. Surely, x^2 does not occur in $\beta_\pi(a)$ for any letter a by Claim 2. Note that 1201021 occurs in $\beta_\pi(w)$ only at a point where two β_π images of letters are concatenated. Assume that $|x| \geq 6$; the smaller cases can be easily checked. Then, again as in Claim 2, x contains 1201 or 1021. Both 1021 and 1201 mark the beginnings and ends of the β_π images of letters, and hence, β_π is injective. Let $u \in \{1021, 1201\}$ be such that u occurs in

x. Suppose $u = 1201$, the other case follows analogous reasons. Then either u occurs in the beginning or end of x and the injectivity of β_π gives a contradiction on the square-freeness of w, or $x = yu\beta_\pi(w_{(j)})\beta_\pi(w_{(j+1)})\cdots\beta_\pi(w_{(j+r)})z$ where $1 < j < |w| - r$ and $-1 \leq r < |w|/2$ and $|y| = |z| = 59$ and $zyu = \beta_\pi(w_{(j+r+1)})$. Note that for any two different letters a and b we have that the suffixes of length 61 of $\beta_\pi(a)$ and $\beta_\pi(b)$ differ. Therefore, yu determines the image $\beta_\pi(w_{(j-1)})$ to equal to $\beta_\pi(w_{(j+r+1)})$. But, now we get a contradiction since $w_{(j-1)}w_{(j)}\cdots w_{(j+r)}$ forms a square in w. Therefore, β_π is square-free.

Claim 2 and Proposition 1 conclude the proof.

Now, Theorem 4 follows.

Theorem 4. *The function $c_3(n)$ is unbounded:*

$$\limsup_{n\to\infty} c_3(n) = \infty \ .$$

Indeed, the image of the cyclic square-free word 021 under β_π gives a different cyclic square-free word for any permutation π by Lemma 1. Each of these cyclic square-free words starts with 021, and hence, gives six new cyclic-square-free words (one for each β_π). This process can be arbitrarily often iterated. The uniformness of β_π ensures that the images of a word are of the same length for each π. The number of different cyclic square-free words after k iterations equals 6^k and they are of length $3 \cdot 122^k$.

Remark 2. We mention here shortly another way to prove Theorem 4. Let T be an infinite set $\{t_0, t_1, \ldots t_n, \ldots\}$ of triples of different square-free words of the same length such that the length of those words does not decrease as the index i increases.

It shall be noted that the arguments of Claim 2 and Lemma 1 also imply that for any triple $t = (u_0, u_1, u_2)$ of T of different square-free words of some length m and for any permutation π we have that

$$\gamma_t(i) = 0212\beta_\pi(u_i)$$

is a uniform cyclic square-free morphism. Indeed, that $\beta_\pi(u_i)$ is square-free follows from Lemma 1 and the square-freeness of u_i, and the prefix 0212 serves as a marker that makes γ_t injective.

Since T exists, we get a sequence $(\gamma_{t_i})_{n\in\mathbb{N}}$ of uniform cyclic square-free morphisms which also imply Theorem 4. Indeed, in order to construct k-many cyclic square-free words of the same length one may consider the set $\{\gamma_{t_1}, \gamma_{t_2}, \ldots, \gamma_{t_k}\}$ and the least common multiple m of the length m_j of the words in t_j for all $1 \leq j \leq k$. Then $\{\gamma_{t_j}^{m_j/m}(0) \mid 1 \leq j \leq k\}$ gives a set of cyclic square-free words of length m of the required size k.

References

1. D. R. Bean, A. Ehrenfeucht, and G. F. McNulty. Avoidable patterns in strings of symbols. *Pacific J. Math.*, 85(2):261–294, 1979.

2. J. D. Currie. There are ternary circular square-free words of length n for $n \geq 18$. *Electron. J. Combin.*, 9(1):Note 10, 7 pp. (electronic), 2002.
3. R. C. Entringer, D. E. Jackson, and J. A. Schatz. On nonrepetitive sequences. *J. Combin. Theory, Ser. A*, 16:159–164, 1974.
4. A. S. Fraenkel and J. Simpson. How many squares must a binary sequence contain? *Electronic J. Combin.*, 2(#R2 (electronic)), 1995.
5. T. Harju and D. Nowotka. Binary words with few squares. *Bull. EATCS*, 89:164–166, 2006.
6. M. Lothaire. *Combinatorics on Words*, volume 17 of *Encyclopedia of Mathematics*. Addison-Wesley, Reading, MA, 1983. Reprinted in the Cambridge Mathematical Library, Cambridge Univ. Press, 1997.
7. N. Rampersad, J. Shallit, and Wang M.-w. Avoiding large squares in infinite binary words. *Theoret. Comput. Sci.*, 339, 2005.
8. A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Vid. Skrifter I. Mat.-Nat. Kl., Christiania*, 1:1–67, 1912.