Disposability in Square-Free Words

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

November 25, 2019

Abstract

We consider words $w$ over the alphabet $\Sigma = \{0, 1, 2\}$. It is shown that there are irreducibly square-free words of all lengths $n$ except 2, 4, 5, 7 and 12. Such a word is square-free (i.e., it has no repetitions $uu$ as factors), but by removing any one internal letter creates a square in the word.

1 Introduction

Grytczuk et al. [2] showed that there are infinitely many ‘extremal’ square-free ternary words where one cannot augment a single new letter anywhere without creating a square. In this article we consider the dual problem of this and show that there are square-free ternary words of all lengths, except 2, 4, 5, 7 and 12, where removing any single interior letter creates a square. Although the problems resemble each other, the results and the proof techniques are quite different.

Let $\Sigma = \{0, 1, 2\}$ be a fixed ternary alphabet and denote by $\Sigma^*$ and $\Sigma^\omega$ the sets of all finite and infinite length words over $\Sigma$, respectively. A finite word $u$ is called a factor of a word $w \in \Sigma^* \cup \Sigma^\omega$ if $w = w_1u w_2$ for some, possibly empty, words $w_1$ and $w_2$. Moreover, $w$ is square-free if it does not have a nonempty factor of the form $uu$.

Let $w \in \Sigma^*$ be a square-free word with a factorization $w = w_1aw_2$ where $a \in \Sigma$. We say that the occurrence of the letter $a$ is disposable if $w_1w_2$ is square-free. The definition extends naturally to infinite words. An occurrence of a letter $a$ is interior, if $w_1$ and $w_2$ are both nonempty.

If a square-free word $w \in \Sigma^* \cup \Sigma^\omega$ does not have disposable occurrences of interior letters then $w$ is said to be irreducibly square-free, i.e., by deleting any interior occurrence of a letter results in a square in the remaining word.

The nonemptiness condition on the prefixes and suffixes is required since all prefixes and suffixes of square-free words are disposable.
**Example 1.** Let $\tau : \Sigma^* \to \Sigma^*$ be the morphism determined by

$$
\begin{align*}
\tau(0) &= 012, \\
\tau(1) &= 02, \\
\tau(2) &= 1.
\end{align*}
$$

The Thue word $t$ is the fixed point $t = \tau^\omega(0)$ of $\tau$ obtained by iterating $\tau$ on the start word 0. Then $t$ is an infinite square-free word; see, e.g., Lothaire [3]:

$$
t = 012021012012101202102012021020121012021012 \cdots
$$

We show that the Thue word is not irreducibly square-free. For this, we first notice that $t$ avoids 010 and 212 as factors. Also, it avoids 1021, since this word would have to be a factor of $\tau(212)$. By deleting the letter 2 at the third position results in a square-free word $012012102012 \cdots$. Indeed, a potential square would have to start either from the beginning, but the prefix 010 does not occur in $t$, or from the second position, but 1021 does not occur in $t$. In fact, every single position of 2 in a factor $\tau(012) = 012021$ of $t$ is disposable.

Later checking of irreducibility of (infinite) words is based on the following procedure that depends on a morphism $\alpha : \Sigma^* \to \Sigma^*$.

**Procedure I.**

1. Check that the morphism $\alpha$ generates an infinite square-free word; say, $\alpha^\omega(0)$ or $\alpha(w)$, where $w$ is a given infinite square-free word.

2. For any pair $(a, b)$ of different letters, check that $\alpha(ab)$ is irreducibly square-free. This takes care that the last letter of $\alpha(a)$ and the first letter of $\alpha(b)$ are not disposable in $\alpha(ab)$. This guarantees that these occurrences are not disposable in any $\alpha(w)$ where $w = w_1 abw_2$ is square-free.

The first item of Procedure I is often taken care by Crochemore’s criterion [1]:

**Theorem 1.** A morphism $\alpha : \Sigma^* \to \Sigma^*$ preserves square-free words if and only if it preserves square-freeness words of length five.

## 2 Irreducibly square-free words of almost all lengths

By a systematic search we find that there are no irreducibly square-free words of lengths 2,4,5,7 and 12. In the following table we have counted the irreducibly square-free words of lengths 3,…,30 up to isomorphism (produced by permutations of the letters) and reversal (mirror image) of the words. For instance, 010212010 is the only irreducibly square-free word of length nine up to isomorphism and reversal. It is a palindrome. The table suggests that the irreducibly square-free words are quite rare among the square-free words, e.g., there are (up to isomorphism and reversal) 202 square-free words of length 20, but only 12 of those are irreducibly square-free. Counting the numbers of (irreducibly) square-free words must take into considerations those words that are palindromes or isomorphic to their reversals.
Table 1: The number of irreducibly square-free words of lengths from 3 to 30 up to isomorphism and reversal.

| length | card |
|--------|------|
| 3      | 1    |
| 10     | 1    |
| 17     | 9    |
| 24     | 24   |
| 4      | 0    |
| 11     | 3    |
| 18     | 7    |
| 25     | 34   |
| 5      | 0    |
| 12     | 0    |
| 19     | 12   |
| 26     | 36   |
| 6      | 1    |
| 13     | 3    |
| 20     | 12   |
| 27     | 48   |
| 7      | 0    |
| 14     | 4    |
| 21     | 16   |
| 28     | 55   |
| 8      | 1    |
| 15     | 4    |
| 22     | 18   |
| 29     | 69   |
| 9      | 1    |
| 16     | 7    |
| 23     | 23   |
| 30     | 78   |

Theorem 2. There exists an infinite irreducibly square-free word.

Proof. Let \( \varphi \) be the following uniform palindromic morphism of length 17 (with the permutation \((0 \ 1 \ 2)\) of the letters giving the images):

\[
\varphi(0) = 01202120102120210 \\
\varphi(1) = 12010212010201021 \\
\varphi(2) = 20121012021012102
\]

By Theorem 1, \( \varphi \) preserves square-freeness. It is easy to check that \( \varphi(0) \), and so also the isomorphic copies \( \varphi(1) \) and \( \varphi(2) \), are irreducibly square-free. Finally, Procedure I entails that deleting the ‘middle’ 17th letters 0 of \( \varphi(0) \) and of \( \varphi(2) \) gives squares: 11 and 02120212, respectively. Similarly, deleting the 18th letter of \( \varphi(0) \) and of \( \varphi(2) \) gives squares: 10201020 and 00, respectively. These observations suffice for the proof of the theorem, since now \( \varphi(w) \) is irreducibly square-free for all square-free, finite or infinite, words \( w \).

Remark 3. The morphism \( \varphi \) has an alignment property, i.e., \( \varphi(a) \) is not a factor of \( \varphi(bc) \) for any \( b \) and \( c \) letters different from \( a \).

The morphism \( \varphi \) has an infinite fixed point

\[
\Phi = \varphi^\omega(0)
\]

that is the limit of the sequence \( \varphi(0), \varphi^2(0), \ldots \).

Note that the finite prefixes of \( \Phi \) are not always irreducibly square-free. For instance, none of the prefixes of \( \Phi \) of length \( n \) with 19 \( \leq n \leq 29 \) are irreducibly square-free. However, we do have the following result with the help of \( \varphi \).

Theorem 4. There are irreducibly square-free words of all lengths \( n \) except 2, 4, 5, 7 and 12.

Proof. Table 2 gives an example for the cases \( n \leq 17 \).

For \( n \geq 18 \), we rely on the morphism \( \varphi \) in order to have solutions for the lengths \( n \equiv p \) (mod 17) for \( p = 0, 1, \ldots, 16 \).
Claim A. Let \( w \) be a nonempty suffix of \( \phi(1) \) or \( \phi(2) \) of length \( |w| < 17 \). Then the word \( w\Phi \) is square-free (but not necessarily irreducibly square-free).

The word \( w \) is a suffix of exactly one of the words \( \phi(a) \), \( a \in \Sigma \). Suppose there is a square in \( w\Phi \) and assume that \( w \) is of minimal length with this property. Then \( w\Phi \) has a prefix \( uu \) for \( u = w\phi(x)z \) for some words \( x \) and nonempty \( z \) with \( |z| < 17 \) (when \( |x| \) is chosen to be maximal). Hence \( uu = w\phi(x)z\phi(x)z \), and so \( zw = \phi(a) \) for \( a \in \Sigma \). By the alignment property, \( uu \) must continue in \( w\Phi \) by the word \( w \). This delivers a square in \( \Phi \), namely \( \phi(x)zw\phi(x)zw = \phi(xaxa) \); a contradiction since \( \Phi \) is square-free. This proves Claim A.

Clearly, there are irreducibly square-free words of lengths \( n \equiv 0 \pmod{17} \), since we can take a prefix of \( \Phi \) of length \( n/17 \) and apply \( \phi \) to it. Next we extend \( \Phi \) to the left by considering words of the form \( uw \), where \( w \) is a prefix of \( \Phi \).

Claim B. The words \( 121\Phi \) and \( 0102\Phi \) are square-free.

First, the words \( 121\phi(0) \) and \( 0102\phi(0) \) are not factors of \( \Phi \), since \( \phi \) has the alignment property and the given words are not suffixes of any \( \phi(a) \), \( a \in \Sigma \). Therefore, if \( 121\Phi \) contains a square, then the square must be a prefix \( 21v \) of \( 21\Phi \) for some \( v \) (and \( 1\Phi \) is square-free by Claim A).

Assume that \( 21v = 21u21u \) (where \( v = u21u \) is a prefix of \( \Phi \)). By the alignment property, \( u21 = \phi(z)\phi(1) \) for some \( z \), since only \( \phi(1) \) ends in 21. Now, \( \phi(1) = y21 \) and \( u = \phi(x)y \). This means that the square \( 21v = 21\phi(z1z)y \) is necessarily continued by the rest of \( \phi(1) \), i.e., by 21, giving a prefix \( v21 = u21u21 \) of \( \Phi \); a contradiction, since \( \Phi \) is square-free.

In the case of \( 0102\Phi \), Claim A guarantees that \( 102\Phi \) is square-free. For the full prefix \( 0102 \), the claim follows since the prefix \( 01020120 \) of \( 0102\Phi \) does not occur in \( \Phi \). This proves Claim B.

The special words \( w_i \) of Table 3 are chosen such that

(iii) \( |w_i| = i \),

(iv) \( w_i\Phi \) is square-free. By Claim A, this follows for \( k = 1, 2, 3, 4, 5 \) and 10. By Claim B, the claim follows for the other cases.

(v) \( w_i\phi(0) \) is irreducibly square-free (by a simple computer check).
The words $w_i$, themselves, are not (and, indeed, cannot be) all irreducibly square-free, but they are square-free.

| $w_i$  | $w_i$  |
|--------|--------|
| $w_1$  | 1      |
| $w_2$  | 02     |
| $w_3$  | 121    |
| $w_4$  | 2102   |
| $w_5$  | 12102  |
| $w_6$  | 020121 |
| $w_7$  | 2120102|
| $w_8$  | 01020121|
| $w_9$  | 121020121 |
| $w_{10}$ | 2021012102 |
| $w_{11}$ | 10121020121 |
| $w_{12}$ | 101202120121 |
| $w_{13}$ | 0210121020121 |
| $w_{14}$ | 01021201020121 |
| $w_{15}$ | 010201202120121 |
| $w_{16}$ | 0201021201020121 |

Table 3: The special words with $w_i \equiv i \pmod{17}$. The words $w_i$ with $i > 1$ end in 121 or 0102 as called for by Claim A.

Finally, let $n = 17k + i$. By Table 2, we can assume that $n \geq 18$. We then choose $w_i$ from Table 3 and pick a prefix $\varphi(v)$ of $\Phi$ of length $17k$. This creates an irreducibly square-free word $w_i \varphi(v)$ of length $n$.

\[ \square \]

3 Problems on Longer Words to Dispose

Irreducibly square-freeness can be generalized to longer factors than just letters. Let $w \in \Sigma^*$ be a square-free word with a factorization $w = w_1vw_2$ such that both $w_1$ and $w_2$ are nonempty. We say that the (occurrence of the) factor $v$ is disposable if also $w_1w_2$ is square-free. If a finite or infinite square-free word $w$ does not have disposable factors of length $k$ then $w$ is called $k$-irreducibly square-free.

**Example 2.** The powers of the Thue morphism $\tau$ produce different solutions for the factor lengths $k$ that giving $k$-irreducibly square-free. Indeed, $\tau^{2n}(0)$ is not 2-irreducibly square-free, but the limit $t = \tau^\omega(0)$ is 2-irreducibly square-free. This is because a disposable pair $ab$ in $\tau^{2n}(0)$ occurs only close to the end of the word, and it becomes nondisposable after $\tau^{2n}(0)$ is continued to $\tau^{2n+1}(0)$.

Indeed, $\tau^{2n}(0)$ has a suffix 121, and a 2-irreducible square-free word cannot be of the form $w121$, since by removing the pair 12 we obtain a (square-free) prefix of $w1$.

**Example 3.** Each square-free word $w \in \Sigma^* \cup \Sigma^\omega$ of length at least 5 is not $k$-irreducible square-free for some $k$ with $2 \leq k \leq 4$. This is because for each prefix of $w$ of the form $ava$ with $|v| \geq 1$, by deleting $va$ one has a (square-free) suffix of $w$. Now, the first letter of every square-free word reoccurs in a prefix of length 5.

These considerations raise many problems.
Problem 1. Does there exist an infinite ternary $k$-irreducibly square-free word for every $k \geq 1$?

Problem 1 has a positive solution for $k = 1, 2$ since $\Phi$ is $k$-irreducibly square-free for these values. For small values of $k$ can be found using square-free morphisms. E.g., the morphism

$$
\alpha_3(0) = 0121012
\alpha_3(1) = 01020120212
\alpha_3(2) = 0102101210212
$$

generates a $3$-irreducibly square-free word $\alpha^a(0)$. The Thue word $t$ is $4$-irreducibly square-free.

Problem 2. Does there exist for every $k$, a $k$-irreducibly square-free word of sufficient length $n$?

The following table gives the known cases for which there are no $k$-irreducibly square-free words of length $k + 2$.

| $k$  | no k-irred. |
|------|-------------|
| 1    | 4, 5, 7, 12 |
| 2    | 6           |
| 3    | -           |
| 4    | 8, 9        |
| 5, ..., 10 | -        |

Table 4: Not $k$-irreducibly square-free lengths.

Finally, we state a problem of the opposite nature:

Problem 3. Does there exist an infinite square-free word that is not $k$-irreducibly square-free for all $k \geq 1$?

Every infinite square-free word $w$ does have an infinite number of integers $k$ such that $w$ is not $k$-irreducibly square-free. Indeed, one needs only to consider ‘near prefixes’ of $w = auaw_0$, where removing the factor $ua$ gives $aw_0$, a (square-free) suffix of $w$.

References

[1] Max Crochemore. Sharp characterizations of squarefree morphisms. Theoret. Comput. Sci., 18(2):221–226, 1982.

[2] Jarosław Grytczuk, Hubert Kordulewski, and Artur Niewiadomski. Extremal square-free words. ArXiv: 1920.06226v1, 2019.
[3] M. Lothaire. *Combinatorics on words*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1997.