Square-Free Words over Partially Commutative Alphabets

Łukasz Mikulski[1,2](✉), Marcin Piątkowski[1,3], and Wojciech Rytter[1,4]

[1] Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland
{lukasz.mikulski,marcin.piatkowski}@mat.umk.pl
[2] School of Computing Science, Newcastle University, Newcastle upon Tyne, UK
[3] Department of Computer Science, University of Helsinki, Helsinki, Finland
[4] Faculty of Mathematics, Informatics and Mechanics, Warsaw University, Warsaw, Poland
rytter@mimuw.edu.pl

Abstract. There exist many constructions of infinite words over three-letter alphabet avoiding squares. However, the characterization of the lexicographically minimal square-free word is an open problem. Efficient construction of this word is not known. We show that the situation changes when some letters commute with each other. We give two characterizations (morphic and recursive) of the lexicographically minimal square-free word $\widetilde{\mathbf{v}}$ in the case of a partially commutative alphabet Θ of size three. We consider the only non-trivial relation of partial commutativity, for which $\widetilde{\mathbf{v}}$ exists: there are two commuting letters, while the third one is blocking (does not commute at all). We also show that the n-th letter of $\widetilde{\mathbf{v}}$ can be computed in time logarithmic with respect to n.

1 Introduction

Problems related to repetitions are crucial in the combinatorics on words due to many practical application, for instance in data compression, pattern matching, text indexing and so on (see [16]). On the other hand, in some cases it is important to consider words avoiding regularities and repetitions. Example applications can be found in such research areas as cryptography and bioinformatics. Languages of words over partially commutative alphabets are fundamental tools for concurrent systems investigation, see [9]. Therefore, the study of repetitions and their avoidability in such languages is significant.

The simplest form of repetition is a square – the factor of the form $x \cdot x$, where x is not empty. Therefore, to show that a word w contains no repetitions, it is sufficient to show that w does not contain squares. Another interesting type of repetition is the abelian square – a factor of the form $x \cdot y$, where x can be obtained from y by permutation of the letters. For example, $baca \cdot caab$ is an abelian square, whereas $bcca \cdot cbba$ is not. A word that contains no abelian squares is called abelian square-free.

A.-H. Dediu et al. (Eds.): LATA 2015, LNCS 8977, pp. 424–435, 2015.
DOI: 10.1007/978-3-319-15579-1_33

Square-free and abelian square-free words have been extensively studied. In 1906 Thue showed that squares are avoidable over three-letter alphabet (see [19]), i.e. there exist infinitely many ternary words without a square. In 1961 Erdös raised the question whether abelian squares are avoidable (see [13]). First attempt to answer this question was made by Evdokimov in 1968 (see [14]), who showed that abelian squares are avoidable over alphabets consisting of at least 25 letters. Then, in 1970, the required size of the alphabet was decreased to 5 by Pleasants (see [18]), and finally, in 1992, to 4 by Keränen (see [15]). Moreover, it can be easily shown that abelian squares cannot be avoided over three-letter alphabet.

A one step further is to study repetitions and their avoidability in words over partially commutative alphabets, see for instance [6–8,10,11]. In contrast to the abelian case, only some fixed pairs of letters from the alphabet are allowed to commute. It complicates considerably the analysis of repetitions in such classes of words.

Our Results. In this paper we deal with the avoidability of repetitions in words over three-letter alphabet Θ with one pair of commuting letters. Then we describe an infinite language of length-increasing square-free words and investigate their combinatorial properties. We use this language, utilising the results of [10], to define the infinite language of partially abelian square-free words over Θ.

As a final result, we give two characterizations of the infinite lexicographically minimal Θ-square-free word $\widetilde{\mathbf{v}}$ and give an efficient construction of this word. The n-th letter of $\widetilde{\mathbf{v}}$ can be computed in logarithmic time with respect to n. The first 176 letters of $\widetilde{\mathbf{v}}$ are:

$$\begin{aligned}\widetilde{\mathbf{v}} = {} & abacabcbacabacbcabacabcbacbcabcbacabacbcabacabcbacabacbcabcb \\ & acbcabacabcbacabacbcabacabcbacbcabcbacabacbcabcbacbcabacabcb \\ & acbcabcbacabacbcabacabcbacabacbcabcbacbcabacabcbacabacbc \ldots\end{aligned}$$

Due to the page limitation, the proofs of some facts were omitted. The full version of this paper, including all proofs, is available as [17].

2 Basic Notions

Throughout the paper we use the standard notions of the formal language theory (see [16] for a more detailed introduction). By Σ we denote a finite set, called the *alphabet*. Elements of the alphabet are called *letters*. A finite word over Σ is a finite sequence of letters. The length of a word w is defined as the number of its letters and denoted $|w|$. The set of all finite words over Σ is denoted by Σ^* and is equipped with a binary associative concatenation operation $\cdot$, where $a_1 \ldots a_n \cdot b_1 \ldots b_m$ is simply $a_1 \ldots a_n b_1 \ldots b_m$. An empty sequence of letters, called the *empty word* and denoted by ε, is the neutral element of the concatenation operation. Thus for any word w we have $\varepsilon \cdot w = w \cdot \varepsilon = w$. An infinite word over

Σ is a sequence of letters indexed by non-negative integers. On the other hand, it can be also defined as a limit of infinite sequence of finite words.

A word u is called a *factor* of a word w if there exist words x and y such that $w = xuy$. If $y = \varepsilon$ then u is called a *prefix* of w and if $x = \varepsilon$ then u is called a *suffix* of w. For a word $w = a_1 a_2 \ldots a_n$ and $1 \leq i, j \leq n$ by $w[i..j]$ we denote its factor of the form $a_i a_{i+1} \ldots a_j$.

We assume that the alphabet Σ is given together with a strict total order $<$, called the *lexicographical order*. This notion is extended in a natural way to the level of words. For any two words x and y we have $x < y$ if x is a proper prefix of y or we have $x = uav_1$ and $y = ubv_2$, where a, b are letters and $a < b$.

A mapping $\phi : \Sigma_1^* \to \Sigma_2^*$ is called a *morphism* if we have $\phi(u \cdot v) = \phi(u) \cdot \phi(v)$ for every $u, v \in \Sigma_1^*$. A morphism ϕ is uniquely determined by its values on the alphabet. Moreover, ϕ maps the neutral element of Σ_1^* into the neutral element of Σ_2^*.

A *partially commutative alphabet* is a pair $\Theta = (\Sigma, ind)$, where Σ is an ordered alphabet and $ind \subseteq \Sigma \times \Sigma$ is a symmetric *commutation* relation. Such an alphabet defines an equivalence relation $\equiv_\Theta$ identifying words, which differ only by the ordering of commuting letters. Two words $w, v \in \Sigma^*$ satisfy $w \equiv_\Theta v$ if there exists a finite sequence of commutations of adjacent commuting letters transforming w into v. For example let us consider the partially commutative alphabet $\Theta = (\{a, b, c\}, \{(b, c), (c, b)\})$. Then the word $w = acbcacb$ is equivalent to $v = accbabc$, but is not equivalent to $u = baccacb$. Words over a partially commutative alphabet $\Theta = (\Sigma, ind)$ are called *partially commutative words*. Note that it is usually assumed that for each $a \in \Sigma$ we have $(a, a) \notin ind$, but in the case of this paper such an assumption is not essential and it does not affect the presented results.

A square in a word w is a factor of the form $x \cdot x$, where x is not empty. A word w is called *square-free* if none of its factors is a square. If we consider a partially commutative alphabet $\Theta = (\Sigma, ind)$ a square is called a partially commutative square or a Θ-square in short.

Definition 1 ([10]). *Let $\Theta = (\Sigma, ind)$ be a partially commutative alphabet. A Θ-square is a factor of the form $u \cdot v$ such that $u \equiv_\Theta v$. A word w is Θ-square-free if it does not contain a nonempty Θ-square.*

There are possible other (nonequivalent) definitions of a partially commutative square-free words, see [8]. Moreover, in the case of the full commutation relation (i.e. any pair of letters can commute) Θ-squares are called the *abelian squares*, and words avoiding them – the *abelian square-free* words.

Example 1. Let $\Theta = (\{a, b, c\}, \{(b, c), (c, b)\})$ be a partially commutative alphabet. The word $w_1 = abc \cdot acb$ is a Θ-square, but it is not an ordinary square. On the other hand, $w_2 = abc \cdot bac$ is an abelian square, which is neither a Θ-square nor an ordinary square. Therefore, w_1 is a square-free word, which it is neither Θ-square-free nor abelian square-free, while w_2 is square-free and Θ-square-free, but not abelian square-free.

3 Partially Abelian Square-Free Words over Three-Letter Alphabets

It is easy to see that any binary word consisting of at least four letters must contain a square. In 1906 Thue shown that three letters are sufficient to construct an infinite square-free word (see [19]). Moreover, in 1992 Keränen proved that to avoid abelian squares (i.e. factors of the form $x \cdot y$, where x and y differ only by permutation of their letters) four letters are sufficient (see [15]). It follows immediately that any four-letter alphabet with more restricted commutation relation also allows to avoid partially commutative squares. Therefore, the alphabets of size three are the most interesting boundary case.

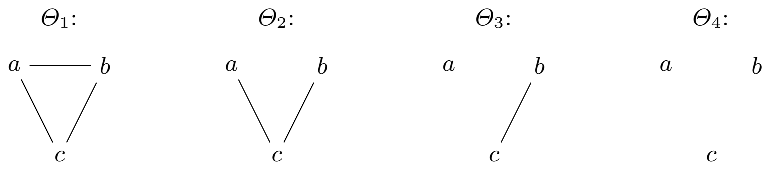


Fig. 1. The possible shapes of the commutation relation over three-letter alphabet. The pairs of letters connected by an edge can commute.

In partially commutative alphabets consisting of three letters one can consider four distinct commutation relations as depicted on Fig. 1. We start with the most restricted case of Θ_4. Observe that the concepts of Θ_4-square freeness and ordinary square-freeness are equivalent. Due to the results of Thue (see [19]), the number of Θ_4-square-free words is infinite. Moreover, the number of finite square-free words of a given length is exponential with respect to this length (see [5] for more details).

On the other hand, the concepts of abelian square-freeness and Θ_1-square-freeness are equivalent. We have only 117 words without Θ_1-square and the longest of them consists of 7 letters. Similarly, the number of Θ_2-square-free words is finite. In this case we have 289 such words with the longest having 15 letters (see [10] for more details).

The most interesting case is the remaining alphabet Θ_3. Similarly as in the case of Θ_4, the number of finite Θ_3-square-free words is infinite, however it is polynomially proportional to the length of the word (see [11]). In what follows, we focus on this alphabet and investigate the combinatorial structure of Θ_3-square-free words in more details.

Remark 1. From now on we only consider the alphabet Θ_3 and denote it by Θ. Thus, by Θ-square and Θ-square-freeness we mean the Θ_3-square and Θ_3-square-freeness.

Conditions for *Θ*-square-freeness

We start with giving some necessary and sufficient conditions for a word to be Θ-square-free. The more detailed study of their combinatorial structure is presented in the subsequent sections. We follow here the results of [10] presented below. Initially we present a statement, which is used further to formulate the conditions for Θ-square-freeness.

Definition 2 (Condition (F), see [10]). *The word $v \in \Sigma^*$ satisfies the condition (F) if neither $abca$ nor $acba$ is a factor of v, where $a, b, c \in \Theta = (\Sigma, ind)$.*

The possible structure of a finite word containing a Θ-square is established by the following fact (see Proposition 3.2 in [10]).

Proposition 1 (see [10]). *Let w be a finite square-free word satisfying the condition (F) and containing a Θ-square as a factor. Then w admits one of the following decompositions:*

(i) $w = w_1bcw_2bcbw_2bw_3$ (ii) $w = w_1cbw_2cbcw_2cw_3$

(iii) $w = w_1bw_2bcbw_2cbw_3$ (iv) $w = w_1cw_2cbcw_2bcw_3$

where w_1, w_2, $w_3 \in \Sigma^*$. Moreover in such a decomposition one of the factors w_1 or w_3 is of length at most 1.

As a corollary to Proposition 1 we can formulate the following fact characterizing the possible building blocks of Θ-square-free words (see the proof of Proposition 2.1 and Proposition 3.2 in [10]). It will be utilized further in construction an infinite Θ-square-free word.

Corollary 1. *Any infinite Θ-square-free word w starting with a consists of the factors belonging to the following: $\mathcal{B} = \{aba,\ aca,\ abcba,\ acbca,\ abca,\ acba\}$. Moreover, the factors $acba$ and $abca$ can appear only as a prefix of w.*

Remark 2. Note that no two different words created by concatenating factors (without ending a) from the set $\mathcal{B}$ defined in Corollary 1 are equivalent under the relation $\equiv_\Theta$.

Finally, the following theorem gives a sufficient condition for the Θ-square-freeness of an infinite word, see Corollary 3.3 in [10] for the proof.

Theorem 1 (see [10]). *Any infinite square-free word over Σ starting with a and satisfying the (F) condition is Θ-square-free.*

4 The Structure of *Θ*-square-free Words

In the preceding section we presented the necessary and sufficient conditions for a word to be Θ-square-free. Below we investigate the combinatorial structure of such words in more detail.

Recall that due to Corollary 1 any infinite Θ-square-free word w consists only of the factors aba, aca, $abcba$, $acbca$, $acba$ and $abca$, where the last two can appear only as a prefix w. It can be easily proven that neither abc nor acb could be a prefix of the lexicographically minimal infinite Θ-square-free word. Therefore we have to consider only the factors aba, aca, $abcba$, $acbca$.

The above observations are the basis of the idea of encoding the possible building blocks of Θ-square-free words as the symbols of a four-letter meta-alphabet $\Delta = \{A, B, C, D\}$.

Definition 3. *Let Σ=$\{a, b, c\}$ (alphabet) and Δ=$\{A, B, C, D\}$ (meta-alphabet). We define a morphism $M : \Delta^* \longrightarrow \Sigma^*$ as follows:*

$$M \;=\; \begin{cases} A \longrightarrow ab & B \longrightarrow ac \\ C \longrightarrow abcb & D \longrightarrow acbc \end{cases} .$$

It is worth to note that the morphism defined above is a code with finite deciphering delay. This fact is utilized in operations described further in this paper.

In what follows, if a word w over Σ is an image of a word u over Δ we call u an M-reduction of w and w is called M-reducible[1].

The alphabet Δ consists of four letters, hence it allows us to construct words without repetitions. However, not all such words over Δ lead to words with no repetitions over Θ. Since we are interested in Θ-square-free words, the considered words over Δ must satisfy additional conditions presented further.

Lemma 1. *Let w be an infinite, M-reducible and Θ-square-free word starting with abacabcbaca. Then M-reduction of w does not contain any of the factors: AC, CA, BD, DB, ABA, BAB, CBC, DAD, $ADCB$, $BCDA$.*

Remark 3. Let $w \in \Sigma^*$ be a Θ-square-free word satisfying the condition (F) stated in Definition 2. Then w consists of blocks, which are images of letters from the alphabet Δ by the morphism M defined above, hence it is always M-reducible and we can apply the inverse mapping M^{-1} (M-reduction) to w. Moreover, the obtained result is a square-free word. On the other hand, the image by M of a square-free word over Δ does not have to be Θ-square-free word. For instance AC is a square-free word over Δ, but $M(AC) = ababcb$ is not Θ-square-free.

As a corollary to Lemma 1 we can describe the structure of Θ-square-free words in the terms of meta-alphabet.

Corollary 2. *Each M-reduction of a Θ-square-free word starting with aba is an element of the set defined by a following regular expression $\Upsilon = \big((A|C)(B|D)\big)^*$.*

[1] In the approach presented in this paper the morphism M is in fact used as a translation of an infinite word over four-letter alphabet into an infinite word over three-letter alphabet.

Remark 4. The inverse of Corollary 2 is not true, i.e. there exist words whose M-reductions are in the set Υ, but they are not Θ-square-free, for instance *abacabacbc* which is an image of $ABAD$.

Lemma 2. *Let $w < v \in \Sigma^*$ be two M-reducible words starting with aba, which M-reductions have equal length and are contained in Υ, and let $u \in \Sigma^*$ be the longest M-reducible word such that $M^{-1}(w) = M^{-1}(u)M^{-1}(w')$ and $M^{-1}(v) = M^{-1}(u)M^{-1}(v')$. Then $M^{-1}(v')$ starts with C and $M^{-1}(w')$ starts with A or $M^{-1}(v')$ starts with D and $M^{-1}(w')$ starts with B.*

Theorem 2. *An infinite word w starting with the letter a and not starting with abca or acba is Θ-square-free if and only if w is square-free and M-reducible.*

5 Two Equivalent Characterizations of the Infinite Word $\widetilde{\mathbf{v}}$

In this section we present two alternative definitions of the language of square-free words over the meta-alphabet $\Delta = \{A, B, C, D\}$ introduced in the previous section. We start with a definition using a morphism, then we show a recurrent procedure, which generates the same class of words.

5.1 Morphic Characterization

Let us define the sequence of words $\{X_i\}_{i \geq 0}$ over Δ together with the languages L_m and L_{dep} similar to $M(L_m)$. Both of them are based on the following morphism.

Definition 4. *We define a morphism $m : \{A, B, C, D\}^* \longrightarrow \{A, B, C, D\}^*$ as:*

$$m = \begin{cases} A \longrightarrow BCB \\ B \longrightarrow ADA \\ C \longrightarrow BCDCB \\ D \longrightarrow ADCDA \end{cases}.$$

Definition 5. *Let $\{X_i\}_{i>0}$ be defined as:*

$$X_i = \begin{cases} AB & \text{for} \quad i = 0 \\ A \cdot m(X_{i-1}) \cdot B & \text{for } i > 0 \end{cases}.$$

We define the languages

$$L_m = \Big\{ m^i(AB) : \ i \geq 0 \Big\}; \qquad \text{and} \qquad L_{dep} = \Big\{ M(X_i) : \ i \geq 0 \Big\}.$$

The subsequent fact describes the combinatorial structure of words contained in the language L_{dep}.

Lemma 3. *For every two words $u, v \in L_{dep}$ either u is the prefix of v or v is the prefix of u.*

Note that L_{dep} is an infinite set of words with strictly growing lengths. Therefore, for any $k > 0$ there exists a word $w \in L_{dep}$ such that $|w| > k$. This observation, together with Lemma 3, constitutes the correctness of the definition of the infinite word $\widetilde{\mathbf{v}}$.

Id Z_i is a sequence of length increasing words, such that Z_i is a prefix of Z_{i+1} for each i then $lim_{i\to\infty} Z_i$ denotes the infinite word containing all Z_i as its prefixes.

Definition 6. *Define $\widetilde{\mathbf{v}} = lim_{i\to\infty} M(X_i)$, or equivalently as $\widetilde{\mathbf{v}} = \sup(L_{dep})$.*

We show that $\widetilde{\mathbf{v}}$ is the lexicographically least word over our partially commutative alphabet.

5.2 Recurrent Characterization of $\widetilde{\mathbf{v}}$

In this subsection we define two sequences of words using recurrence. Furthermore, at the end of this section, we show that one of them is equivalent to the sequence $\{X_i\}_{i\geq 0}$ defined previously. We start with defining the operation of so-called complement for letters.

Definition 7. *We define the operation $\widehat{\ }$: $\{A, B, C, D\} \to \{A, B, C, D\}$ as follows:*

$$\widehat{\ } = \begin{cases} A \to B, & B \to A \\ C \to D, & D \to C \end{cases}.$$

The mapping defined above is in a natural way extended to the level of words. It allows us to define two recurrent sequences of words.

Definition 8. *We define the sequences of words $\{Y_i\}_{i\geq -1}$ and $\{S_i\}_{i\geq 0}$ over the alphabet $\Delta = \{A, B, C, D\}$ as follows:*

$$Y_{-1} = \varepsilon, \qquad Y_0 = AB, \qquad S_0 = C,$$
$$Y_{n+1} = Y_n S_n \widehat{Y}_n \widehat{S}_n Y_n, \qquad S_{n+1} = S_n \widehat{Y}_{n-1} \widehat{S}_n Y_{n-1} S_n.$$

Example 2. The first few elements of sequences defined above are as follows:

$$Y_0 = AB, \quad S_0 = C, \qquad Y_1 = ABCBADAB, \quad S_1 = CDC,$$
$$Y_2 = ABCBADABCDCBADABCBADCDABCBADAB,$$
$$S_2 = CDCBADCDABCDC.$$

The following facts describe some of the combinatorial properties of the sequences defined above.

Lemma 4. *For each i, the word S_i is a palindrome and the word Y_i is a pseudo-palindrome, i.e. for each $1 \leq k \leq l$ we have $Y[l-k+1] = \widehat{Y}[k]$, where $l = |Y_i|$.*

Lemma 5. *Let X_i be as in Definition 5. Then*
(1) *for each $i \geq 0$ we have $X_i = Y_i$;*
(2) $\widetilde{\mathbf{v}} = \lim_{i\to\infty} M(Y_i)$.

Taking into account the images of the words Y_i and S_i by morphism M we can formulate the following fact, which will be very useful further.

Proposition 2 (Block lengths). *Let Y_i and S_i be as defined above. Then for each $i \geq 1$ we have:*

$$\left|M(Y_i)\right| = \frac{4(4^{i+1}-1)}{3} \quad \text{and} \quad \left|M(S_i)\right| = \frac{4(2\cdot 4^i+1)}{3}.$$

6 Combinatorial Properties of the Word $\widetilde{\mathbf{v}}$

In this section we formulate and prove the main results of the paper. Namely, we show the Θ-square-freeness (Theorem 3) and lexicographical minimality (Theorem 4) of the word $\widetilde{\mathbf{v}}$ and the time complexity of the computing the n-th letter of $\widetilde{\mathbf{v}}$ (Theorem 5). The proof of the latter yields in fact a very efficient procedure.

We start with a series of facts which lead to the proof of Θ-square-freeness of $\widetilde{\mathbf{v}}$. Let us recall languages L_m and $L_{dep} = M(L_m)$ from Definition 5.

Proposition 3. *The language L_m consists of square-free words only.*

Lemma 6. $L_m \subseteq \Upsilon = \Big((A|C)(B|D)\Big)^*$.

Proposition 4. *Let $v \in L_m$. Then v does not contain a factor of the form $wxwy$, where $w \in \Delta^*$, and $(x = A \ \wedge \ y = C)$ or $(x = B \ \wedge \ y = D)$.*

Theorem 3. *The languages $M(L_m)$ and L_{dep} consists of square-free words only. The word $\widetilde{\mathbf{v}}$ is an infinite Θ-square-free word.*

Theorem 4. *The word $\widetilde{\mathbf{v}}$ is the lexicographically minimal infinite Θ-square-free word.*

Proof. The Θ-square-freeness of $\widetilde{\mathbf{v}}$ follows from Theorem 3.

Suppose that there exists an infinite Θ-square-free word $\widetilde{\mathbf{w}}$ that is lexicographically smaller than $\widetilde{\mathbf{v}}$. Then, by the analysis of short Θ-square-free words, $\widetilde{\mathbf{w}}$ has to start with *abacabcbaca*. Moreover, due to Theorem 2 the word $\widetilde{\mathbf{w}}$ is M-reducible. Let us consider $u \in \Sigma^*$ – the longest common M-reducible prefix of $\widetilde{\mathbf{v}}$ and $\widetilde{\mathbf{w}}$. Moreover, let $X, Y \in \Delta$ be such that $v = uM(X)$ and $w = uM(Y)$ are prefixes of $\widetilde{\mathbf{v}}$ and $\widetilde{\mathbf{w}}$, respectively. We have that $M(X) > M(Y)$. Potentially there are 6 cases for X, Y, however from our previous results it follows that the only cases to consider are:

$$(X, Y) = (C, A), \quad \text{or } (X, Y) = (D, B).$$

Precisely, w, v and u satisfy all assumptions of Lemma 2, hence indeed, one of the following conditions holds:

1. $M^{-1}(v) = M^{-1}(u)C$ and $M^{-1}(w) = M^{-1}(u)A$,
2. $M^{-1}(v) = M^{-1}(u)D$ and $M^{-1}(w) = M^{-1}(u)B$.

Without the loss of generality, we can assume the first case. According to Lemma 1, the last meta-letter of $M^{-1}(u)$ is either B or D. We deal with each of those cases separately.

1° ($M^{-1}(u)$ ends with B): Due to the morphic definition of $\widetilde{\mathbf{v}}$, the last but one letter of $M^{-1}(u)$ is A. Hence, the word $M^{-1}(w)$ contains a forbidden factor ABA. Therefore, by Lemma 1, the infinite word $\widetilde{\mathbf{w}}$ cannot be Θ-square-free.

2° ($M^{-1}(u)$ ends with D): Following similar reasoning as above, we obtain that $M^{-1}(u)$ ends with BCD. Hence, the word $M^{-1}(w)$ contains a forbidden factor $BCDA$. Therefore, due to Lemma 1, the infinite word $\widetilde{\mathbf{w}}$ cannot be Θ-square-free.

The contradictions obtained above prove that the initial assumption concerning the existence of $\widetilde{\mathbf{w}}$ was wrong. Therefore, $\widetilde{\mathbf{v}}$ is indeed the lexicographically minimal infinite Θ-square-free word.

Theorem 5. *For each $n > 0$ the n-th letter of the word $\widetilde{\mathbf{v}}$ can be determined in time $O(\log n)$.*

Proof. To prove the above lemma we present a simple recurrent procedure.

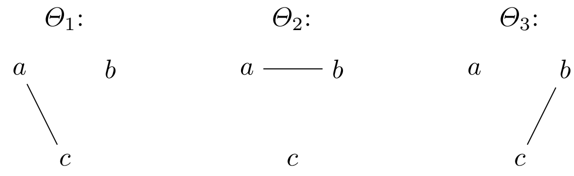
For given $n \geq 0$ we find the shortest word Y_i of length $l \geq n$, where Y_i's are as in Definition 8. Due to Proposition 2, both the index i and length l are given by simple arithmetic formulas. By Definition 8, the word Y_i consists of five factors (either Y_{i-1} or S_{i-1} or their complements) with lengths given by Proposition 2. Thus, we can determine the factor $F \in \{Y_{i-1}, \widehat{Y}_{i-1}, S_{i-1}, \widehat{S}_{i-1}\}$ containing the considered position k in a constant time. It remains to determine (using a recurrent call) the letter in F on a position n', which is obtained by subtracting from k the starting position of F. The recurrence stops when $i = 1$ and we have F equal to one of the words $M(Y_1)$, $M(S_1)$, $M(\widehat{Y_1})$ or $M(\widehat{S_1})$.

Note that at each call of the recurrence it is necessary to memorize whether we are looking for a letter in one of words ($M(Y_1)$ or $M(S_1)$) or their complements ($M(\widehat{Y_1})$ or $M(\widehat{S_1})$). It could be done by using a single boolean variable.

It is easy to see that the number of iterations performed by the procedure described above is logarithmic with respect to n. Moreover, the required computations on each level of recurrence can be performed in a constant time.

7 Final Remarks

In partially commutative alphabets of size three with one pair of commuting letters one can consider three commutation relations:



From the point of view of the partially abelian square-freeness all above alphabets are equivalent. We considered in this paper square-free words over the partially commutative alphabet Θ_3, but any Θ_3-square-free word could be transformed by a morphism (precisely an isomorphism) to a Θ_1-square-free or Θ_2-square-free word and almost all the results follow.

However, if we are interested in construction of the lexicographically minimal partially abelian square-free word, the choice of the alphabet is very important. Such a choice determines the blocking letter and the structure of the lexicographically minimal word. In the case of alphabets Θ_1 and Θ_2 it requires further investigation.

In [2] Allouche and Shallit presented an open problem of characterizing the lexicographically minimal square-free word over three-letter alphabet without any commutation allowed. The construction of Thue (see [19]) leads to a word which is not lexicographically minimal.

On the other hand, there is a procedure, proposed by Currie [12], which allows to determine if a given finite word is a prefix of an infinite word avoiding some repetitions. It immediately gives an algorithm computing arbitrary long prefix of the lexicographically least infinite word. However, generating n-th letter is definitely not logarithmic with respect to n. Moreover, in the case of square-freeness, it seems to be directly applicable for alphabets consisting more than four letters.

Another problem related to square-freeness is the overlap-freeness (i.e., avoiding pattern $axaxa$, where a is a letter and x is a word). Berstel proved [3], (see also [4]), that the lexicographically *greatest* infinite overlap-free word on the binary alphabet $\Sigma = \{0, 1\}$ that begins with 0 is the Thue-Morse overlap-free sequence τ.

Moreover, it has been shown in [1] that the lexicographically least infinite overlap-free binary word is $001001\bar{\tau}$, where $\bar{\tau}$ is the negation of overlap-free Thue-Morse word τ. This makes the problem of extremal cases for overlap-freeness closed. However, its solution relies on Thue-Morse word, which is a fix point of a morphism. This supports the claim that there is also an efficient construction of the lexicographically least square-free word over a ternary alphabet without commutation. We believe that the techniques utilized in this paper might be helpful in finding such a construction.

Acknowledgements. We would like to thank the anonymous referees for their comments and useful suggestions.

This research was supported by Polish National Science Center under the grant No.2013/09/D/ST6/03928 and by research fellowship within project "Enhancing Educational Potential of Nicolaus Copernicus University in the Disciplines of Mathematical and Natural Sciences" (project no. POKL.04.01.01-00-081/10).

References

1. Allouche, J., Currie, J.D., Shallit, J.: Extremal infinite overlap-free binary words. Electr. J. Comb. 5 (1998)
2. Allouche, J., Shallit, J.O.: Automatic Sequences - Theory, Applications, Generalizations. Cambridge University Press (2003)
3. Berstel, J.: A rewriting of Fife's theorem about overlap-free words. In: Karhumäki, J., Maurer, H., Rozenberg, G. (eds.) Results and Trends in Theoretical Computer Science, pp. 19–29. Springer, Heidelberg (1994)
4. Berstel, J.: Axel Thue's work on repetitions in words. Séries formelles et combinatoire algébrique **11**, 65–80 (1992)
5. Brandenburg, F.J.: Uniformly growing k-th power-free homomorphisms. Theoretical Computer Science **23**(1), 69–82 (1983)
6. Carpi, A.: On the number of abelian square-free words on four letters. Discrete Applied Mathematics **81**(1–3), 155–167 (1998)
7. Carpi, A.: On abelian squares and substitutions. Theoretical Computer Science **218**(1), 61–81 (1999)
8. Carpi, A., de Luca, A.: Square-free words on partially commutative free monoids. Information Processing Letters **22**(3), 125–131 (1986)
9. Cartier, P., Foata, D.: Problèmes combinatoires de commutation et réarrangements. LNM, vol. 85. Springer, Heidelberg (1969)
10. Cori, R., Formisano, M.R.: ITA. Partially abelian squarefree words **24**, 509–520 (1990)
11. Cori, R., Formisano, M.R.: On the number of partially abelian square-free words on a three-letter alphabet. Theoretical Computer Science **81**(1), 147–153 (1991)
12. Currie, J.D.: On the structure and extendibility of k-power free words. European Journal of Combinatorics **16**(2), 111–124 (1995)
13. Erdös, P.: Some unsolved problems. Magyar Tud. Akad. Mat. Kutató **6**, 221–254 (1961)
14. Evdokimov, A.: Strongly asymmetric sequences generated by a finite number of symbols. Dokl. Akad. Nauk. SSSR **179**, 1268–1271 (1968)
15. Keränen, V.: Abelian squares are avoidable on 4 letters. In: Kuich, W. (ed.) ICALP 1992. LNCS, vol. 623. Springer, Heidelberg (1992)
16. Lothaire : Algebraic Combinatorics on Words. Encyclopedia of mathematics and its application, vol. 90. Cambridge University Press (2002)
17. Mikulski, Ł., Piątkowski, M., Rytter, W.: Square-free words over partially commutative alphabets. CS-TR 1439, Newcstle University (2014)
18. Pleasants, P.A.B.: Non-repetitive sequences. Mathematical Proceedings of the Cambridge Philosophical Society **68**, 267–274 (1970)
19. Thue, A.: Über unendliche Zeichenreihen. Videnskabsselskabets Skrifter, I Mat.-nat. K1, 1–22 (1906)