

# Crucial words and the complexity of some extremal problems for sets of prohibited words

A. Evdokimov<sup>\*†</sup> and S. Kitaev<sup>‡</sup>

27th August 2007

## Abstract

We introduce the notion of a set of prohibitions and give definitions of a complete set and a crucial word with respect to a given set of prohibitions. We consider three special sets which appear in different areas of mathematics and for each of them examine the length of a crucial word. One of these sets is proved to be incomplete. The problem of determining lengths of words that are free from a set of prohibitions is shown to be NP-complete, although the related problem of whether or not a given set of prohibitions is complete is known to be effectively solvable.

## 1 Introduction and Background

In defining or characterising sets of objects in discrete mathematics, “languages of prohibitions” are often used to define a class of objects by listing those prohibited subobjects that are not contained in the objects of the class. To this end the notion of a subobject is defined in different ways. The notion depends on the set under consideration. These sets are subwords for fragmentarily restricted languages, subgraphs for families of graphs and so on.

The research in the direction of words and languages of prohibited subwords was begun by different authors [2, 9, 10, 12, 16, 21] in the 1970s. The interest in the general question in this area arose from considerations of different types of special problems, in particular, in coding theory, combinatorics of symbolic sequences, number theory and problems of Ramsey type (for instance the arithmetic progressions in partitions of the natural row). For algebraic problems it is more typical to study avoidance of infinite sets that are defined by prohibitions

---

<sup>\*</sup>Sobolev Institute of Mathematics, Acad. Koptyug prospect 4, Novosibirsk 630090, Russia, evdok@math.nsc.ru

<sup>†</sup>The work of the first author was partially supported by the Russian Foundation for Fundamental Research (Grant 02-01-00939) and the Ministry for Higher Education of the Russian Federation (Grant E-02-6.0-250)

<sup>‡</sup>Department of Mathematics, University of Kentucky, Lexington, KY 40506-0027, USA, kitaev@ms.uky.edu

of words (called *terms* or *patterns*) in an alphabet of variables that can themselves be words [2, 21]. Different problems on sequences without repetitions, under variation the concept of “strong” or “weak” repetition of subwords, are typical examples of problems of this class (e.g. see [1, 3, 5, 6, 14, 18]). To give a more precise example, one can consider *abelian square-free words*, in which no two adjacent subwords are permutations of each other. In [5] Erdős asked how many letters do we need to construct an infinite abelian square-free word. That is easy to check that such word does not exist in three letter alphabet. The first author [6] gave an example of infinite abelian square-free word over 25 letter alphabet, and improved this result by reducing the number of letters in the alphabet to seven [7]. Pleasants’s improvement [18] contains only five letters in the alphabet. Finally, Keränen [14] constructed an infinite abelian square-free word over a four-letter alphabet and Carpi [3] derived that, on a four-letter alphabet, the number of abelian square-free words of each length grows exponentially, it is uncountable and the monoid of abelian square-free endomorphisms is not finitely generated.

In this paper we consider three types of “prohibitions” connected with a generalisation of the notion of nonrepetitive symbolic sequences, and for each of these sets we consider the structure of crucial words and find their lengths. In Section 5 we investigate the problem of determining lengths of words that are free from any given set of prohibitions. We show that this problem is NP-complete although the related problem whether or not a given set of prohibitions is complete is known to be effectively solvable.

Let  $\mathbf{A} = \{a_1, \dots, a_n\}$  be an alphabet of  $n$  letters. A *word* over the alphabet  $\mathbf{A}$  is a finite sequence of letters of the alphabet. Any  $i$  consecutive letters of a word  $X$  generate a *subword* of length  $i$ . If  $X$  is a subword of a word  $Y$ , we write  $X/Y$ .

The set  $\mathbf{A}^*$  is the set of all the words over the alphabet  $\mathbf{A}$ . The set of nonempty words over  $\mathbf{A}$  is denoted  $\mathbf{A}^+$ . Let  $\mathbf{S} \subseteq \mathbf{A}^*$ . Then  $\mathbf{S}$  is called a *set of prohibited words* or a *set of prohibitions*. A word that does not contain any words from  $\mathbf{S}$  as its subwords is said to be *free from  $\mathbf{S}$* . The set of all words that are free from  $\mathbf{S}$  is denoted by  $\widehat{\mathbf{S}}$ .

**Example 1.** Let  $\mathbf{A} = \{a, b\}$ . The set of prohibitions is  $\mathbf{S} = \{aa, ba\}$ . The word  $abbb$  is in  $\widehat{\mathbf{S}}$ .

If there exists  $k \in \mathbf{N}$  such that the length of any word in  $\widehat{\mathbf{S}}$  is less than  $k$ , then  $\mathbf{S}$  is called a *complete*<sup>1</sup> set.

**Example 2.**  $\mathbf{A} = \{1, 2, 3, 4\}$ . The set of prohibitions is

$$\mathbf{S} = \{123, 13, 14, 11, 22, 33, 44\}.$$

Then  $\mathbf{S}$  is incomplete, since the word  $(124)^k$  is in  $\widehat{\mathbf{S}}$  for any  $k$ .

---

<sup>1</sup>“Complete” is equivalent to “unavoidable” in contemporary literature (e.g. see [17, 19, 20]). However, we use the original terminology from [8, 9, 10, 12].

**Example 3.**  $\mathbf{A} = \{1, 2, 3\}$ . The set of prohibitions is

$$\mathbf{S} = \{12, 23, 31, 32, 11, 22, 33\}.$$

It is easy to check that  $\mathbf{S}$  is complete.

A word  $X \in \widehat{\mathbf{S}}$  is called a *crucial* word (with respect to  $\mathbf{S}$ ), if the word  $Xa_i$  contains a prohibited subword for each letter  $a_i \in \mathbf{A}$ . This means that  $Xa_i$  has the structure  $BB_i a_i$ , where  $B$  is some word and  $B_i a_i \in \mathbf{S}$ . The subword  $B_i$  is called the *i-suffix* of crucial word  $X$ . If for each letter of the alphabet we consider the minimal *i-suffix* (with respect to the length) we obtain a system of *i-suffixes*, which we will use to investigate crucial words.

**Example 4.**  $\mathbf{A} = \{a, b, c\}$ . The set of prohibitions is  $\mathbf{S} = \{aa, cab, acac\}$ . The word  $aca$  is crucial with respect to  $\mathbf{S}$ .

A crucial word of minimal (maximal) length, if it exists, is called a *minimal* (*maximal*) crucial word. Let  $L_{min}(\mathbf{S})$  ( $L_{max}(\mathbf{S})$ ) denote the length of a minimal (maximal) crucial word with respect to  $\mathbf{S}$ .

If one knows the crucial words with respect to a given set of prohibitions  $\mathbf{S}$ , one can solve the problem on whether or not the set  $\mathbf{S}$  is complete. Indeed, if there does not exist a crucial word with respect to  $\mathbf{S}$  (as in Example 1) then  $\mathbf{S}$  is not complete, since any word from  $\widehat{\mathbf{S}}$  can be continued to the right by concatenating some letter of the alphabet. On the other hand, if  $X$  is a crucial word with respect to  $\mathbf{S}$  then we consider the words  $B_i a_i \in \mathbf{S}$ , take the longest one (in case we have more than one word with maximum length, we choose any one) and delete the letter  $a_i$  from the chosen word in  $\mathbf{S}$ . By doing this we get the set  $\mathbf{S}'$  – a *reduction* of the set  $\mathbf{S}$ . The completeness of the set  $\mathbf{S}$  is invariant under the operation of reduction described above, that is  $\mathbf{S}$  is complete if and only if  $\mathbf{S}'$  is complete. The notion of a crucial word, as well as the algorithm determining whether or not a given set of prohibitions is complete, which is based on consecutive reductions, was presented by the first author in [8].

We remark that in the case of finite sets of words with fixed lengths, it is helpful to use a geometrical interpretation of the completeness problem on the *de Bruijn graphs*. Determining whether or not a given set of prohibitions is complete reduces to looking for cycles on the subgraph obtained by deleting the set of vertices  $\mathbf{S}$  from a de Bruijn graph (see [9, 10]).

In this paper we consider three sets of prohibitions denoted  $\mathbf{S}_1^n$ ,  $\mathbf{S}_2^n$ ,  $\mathbf{S}_3^{n,k}$ . Here we use  $n$  for indicating the number of letters of the alphabet under consideration and  $k$  is a natural number.

We now give the definitions of these sets:

$\mathbf{S}_1^n = \{XX \mid X \in \mathbf{A}^+\}$ , that is, we prohibit the repetition of two equal consecutive subwords. The elements of  $\mathbf{S}_1^n$  are usually called *squares*.

$\mathbf{S}_2^n = \{XY \mid \bar{\nu}(X) = \bar{\nu}(Y)\}$ , where  $\bar{\nu}(X) = (\nu_1(X), \dots, \nu_n(X))$  is the *content vector*<sup>2</sup> of  $X$ , in which  $\nu_i(X)$  is the number of occurrences of the letter  $a_i$  in

<sup>2</sup>The content vector is also sometimes called the *commutative image*.

$X$ . That is, we prohibit the repetition of two consecutive subwords of the same content. The elements of  $\mathbf{S}_2^n$  are usually called *abelian squares*.

$\mathbf{S}_3^{n,k} = \{XY \mid d(X, Y) \leq k, |X| = |Y| \geq k + 1, k \in \mathbf{N}\}$ , where  $d(X, Y)$  is the number of letters in which the words  $X$  and  $Y$  differ (Hamming metric) and  $|X|$  is the length of the word  $X$ . That is we prohibit any two consecutive subwords of length greater than  $k$  such that the number of positions in which these words differ is less then or equal to  $k$ .

The proofs of the theorems in this paper consist of the constructions of extremal crucial words and of the proofs of their optimality, i.e., the lower bound for  $L_{min}(\mathbf{S})$  and the upper bound for  $L_{max}(\mathbf{S})$ .

## 2 The Set of Prohibitions $\mathbf{S}_1^n$

The result from the following theorem was mentioned by Arshon [1], but no proof was provided. The words appearing in the proof of Theorem 1 are frequently called *Zimin words*, but should perhaps be called *Arshon words*.

**Theorem 1.** *We have*

$$L_{min}(\mathbf{S}_1^n) = 2^n - 1.$$

*Proof.* We define a crucial word  $X$  by induction:

$$X_1 = a_1, X_i = X_{i-1}a_iX_{i-1}, X = X_n.$$

From this construction it follows that  $|X| = 2^n - 1$ . We will prove that  $X$  is a minimal crucial word with respect to  $\mathbf{S}_1^n$ .

Let  $U$  be an arbitrary minimal crucial word. We show that  $U$  coincides with the word  $X$  up to a permutation of letters in  $\mathbf{A}$ .

From the definition of a crucial word it follows that in the word  $Ua_i$  there is a prohibited word of the form  $B_ia_iB_ia_i$ , where  $B_i$  is a certain word and  $B_ia_iB_ia_i$  is a suffix of the word  $Ua_i$  (the suffix may coincide with  $Ua_i$ ). In this case the  $i$ -suffix is the subword  $B_ia_iB_i$ . Let  $\ell_i = B_ia_iB_i$ .

We assume that  $\ell_1/\ell_2/\dots/\ell_n$ , that is  $\ell_i$  is a subword of  $\ell_{i+1}$  since we can make such ordering by permuting the letters of the alphabet, which obviously does not affect the cruciality and minimality of a word.

Note that the minimal crucial word  $U$  has the form

$$U = B_n a_n B_n = B_n a_n Y_n a_1,$$

where  $Y_n$  is a certain word. Actually, if on the right of  $B_n a_n B_n$  there is a certain nonempty word, then it contradicts the minimality of a crucial word, and if instead of  $a_1$  there stands  $a_k$  ( $k > 1$ ) then it contradicts  $\ell_1/\ell_k$ .

We show that  $\ell_{n-1}$  coincides with  $B_n$ . We have  $\ell_{n-1} = B_{n-1}a_{n-1}B_{n-1}$  and assume  $a_n B_n$  be a subword of  $\ell_{n-1}$ . Now  $\ell_{n-1}$  has the form  $Ka_n Pa_{n-1}Ka_n P$ , where  $Ka_n P = B_{n-1}$ , but then

$$\ell_n = Pa_{n-1}Ka_n Pa_n Pa_{n-1}Ka_n P, \text{ where } Pa_{n-1}Ka_n P = B_n,$$

and the word  $U$  contains the prohibited subword  $a_n P a_n P$ . This cannot be the case. It means that  $\ell_{n-1}$  is a subword of the word  $B_n$ , and the word  $U$  has the form:

$$U = \ell_n = Z_n \ell_{n-1} a_n Z_n \ell_{n-1},$$

where  $Z_n$  is a certain word. Since we explore a minimal crucial word, we have  $Z_n = \varepsilon$ , and then  $B_n = \ell_{n-1}$ . In the same way we can show that  $B_i = \ell_{i-1}$  for each  $i = 2, \dots, n-1$  and  $B_1 = \varepsilon$ .

Hence the structure of a minimal crucial word  $U$  coincides with that of the word  $X$  as required.  $\square$

**Remark 2.** From the proof of Theorem 1 it follows that the word  $X$  is the unique minimal crucial word up to a permutation of the letters of the alphabet  $\mathbf{A}$ .

### 3 The Set of Prohibitions $\mathbf{S}_2^n$

**Proposition 3.** *A minimal crucial (with respect to  $\mathbf{S}_2^n$ ) word cannot have three letters, each of which appears exactly twice in the word.*

*Proof.* Since the proposition is obviously true for  $|\mathbf{A}| = 1, 2, 3$ , we will consider the case  $|\mathbf{A}| \geq 4$ .

Let  $X$  be a minimal crucial word, and suppose the system of  $i$ -suffixes for it is  $\ell_1/\ell_2/\dots/\ell_n = X$ . Suppose the letters  $a_{i_1}, a_{i_2}, a_{i_3}$  occur twice in  $X$  and that  $i_1 < i_2 < i_3 < n$  (the fact that  $i_1, i_2, i_3$  do not equal  $n$  follows from the fact that  $a_n$  must occur an odd number of times).

When we pass from  $\ell_{i_3-1}$  to  $\ell_{i_3}$  ( $\ell_{i_3-1}$  is determined, since there are  $i_1, i_2 < i_3$ ) there must appear a letter  $a_{i_3}$ , and when we pass from  $\ell_{i_3}$  to  $\ell_{i_3+1}$  ( $\ell_{i_3+1}$  is determined, since  $i_3 < n$ ) there must appear one more letter  $a_{i_3}$ ; hence, since there are exactly two letters  $a_{i_3}$  in  $X$ , there are no letters  $a_{i_3}$  in  $\ell_j$ , for  $2 < j < i_3$ . Hence, in particular, there are no letters  $a_{i_3}$  in the subword  $\ell_{i_2}$  (both letters  $a_{i_3}$  lie to the left of  $\ell_{i_2}$  in  $X$ ).

Obviously, the letter  $a_{i_1}$  must be in  $\ell_{i_1}$ . The second letter  $a_{i_1}$  appears when we pass from  $\ell_{i_1}$  to  $\ell_{i_2}$ . Since there are only two letters  $a_{i_1}$ , there are no letters  $a_{i_1}$  in the word  $X$  to the left of  $\ell_{i_2}$  (in particular, the letters  $a_{i_3}$  precede the letters  $a_{i_1}$  when reading the word  $X$  from left to right).

If we write the letter  $a_{i_3+1}$  to the right of the word  $X$  we obtain a prohibited word (a word from  $\mathbf{S}_2^n$ ). The words from  $\mathbf{S}_2^n$  are divided into two parts which have the same content vectors. Obviously, the letters  $a_{i_3}$  must be in different parts of the prohibited word, and the letters  $a_{i_1}$  must be in different parts of the same word which is impossible, since the letters  $a_{i_3}$  lie strictly to the left of  $a_{i_1}$ , and this contradicts the assumption.  $\square$

**Remark 4.** From the proof of Proposition 1 we have that if letters  $a_i$  and  $a_j$  occur twice in a word  $X$  (in which  $\ell_1/\ell_2/\dots/\ell_n = X$ ), then either  $i = j + 1$  or  $j = i + 1$ .

Note that a natural approach to a construction of a crucial word that is free from  $\mathbf{S}_2^n$  is possible. It consists of an algorithm of step-by-step optimisation: we ascribe to a crucial word on an  $n$ -letter alphabet a minimal number of letters to obtain a crucial word on an  $(n + 1)$ -letter alphabet.

The algorithm can be written recursively in the following way:

$$\begin{aligned} X_n &= B_{n-1}a_nX_{n-1} \\ B_{n-1} &= B_{n-3}a_{n-1}B_{n-3} \\ B_1 &= a_1, B_2 = a_2, B_{-1} = B_0 = X_0 = \varepsilon. \end{aligned}$$

Some initial values when implementing the algorithm are:

$$\begin{aligned} X_1 &= a_1, \\ X_2 &= a_1a_2a_1, \\ X_3 &= a_2a_3a_1a_2a_1, \\ X_4 &= a_1a_3a_1a_4a_2a_3a_1a_2a_1, \\ X_5 &= a_2a_4a_2a_5a_1a_3a_1a_4a_2a_3a_1a_2a_1. \end{aligned}$$

This is an algorithm by which the minimal crucial word  $X_n$  for the set of prohibitions  $\mathbf{S}_1^n$  can be built. For  $\mathbf{S}_2^n$  such a construction gives an upper bound of the form  $\exp(n/2)$ , or, to be more exact,

$$(3 - (n \bmod 2))2^{\lfloor \frac{n+1}{2} \rfloor} - 3.$$

However, one can give an upper bound that is a linear function (see Theorem 5).

We introduce, as before, the system of  $i$ -suffixes:  $\ell_1/\ell_2/\dots/\ell_n$  (we permute the letters of the alphabet if necessary). We show that the passage from  $\ell_{i-1}$  to  $\ell_i$  is possible by ascribing only two symbols (letters of alphabet  $\mathbf{A}$ ).

When we passed from  $\ell_{i-1}$  to  $\ell_i$  let there appear symbols  $y$  and  $z$ .  $\ell_{i-1}$  may be denoted by  $AB$ , where  $A$  is a certain word,  $B$  consists of the letters of the word  $A$  (which are somehow mixed) and  $B$  contains one letter  $a_{i-1}$  less than  $A$  does. Let  $x$  be the last letter of the word  $A$  on the right. Then  $\ell_i$  may be denoted by  $yzKxB$ , where  $A = Kx$ . From the definition of  $\ell_i$  we have the following equation (recall that  $\bar{v}$  denotes the content vector)

$$\bar{v}(y) + \bar{v}(z) + \bar{v}(K) = \bar{v}(B) + \bar{v}(x) + \bar{v}(a_i),$$

which from the definition of  $K$  and  $B$  is equivalent to

$$2\bar{v}(x) + \bar{v}(a_i) = \bar{v}(y) + \bar{v}(z) + \bar{v}(a_i).$$

It follows necessarily that  $x = a_{i-1}$  and either  $y = a_{i-1}$ ,  $z = a_i$  or  $y = a_i$ ,  $z = a_{i-1}$ . Suppose  $y = a_{i-1}$ ,  $z = a_i$ .

For example, we have the following crucial word for a 6-letter alphabet:

$$a_4a_5a_3a_4a_2a_3a_1a_2 \updownarrow a_6a_4a_3a_2a_1a_2a_3a_4a_6,$$

(the symbol was drawn for a more convenient visual perception of the word).

This word is crucial and its length is equal to 17.

**Theorem 5.** ([11, Theorem 5]<sup>3</sup>) For any  $n > 2$  we have

$$L_{min}(\mathbf{S}_2^n) = 4n - 7.$$

*Proof.* We consider the case of an arbitrary  $n \geq 3$  defining the word  $W$  as

$$W = a_{n-2}a_{n-1}a_{n-3}a_{n-2} \dots a_1a_2 \downarrow a_n a_{n-2} a_{n-3} \dots a_2 a_1 a_2 \dots a_{n-3} a_{n-2} a_n.$$

Then  $|W| = 2(n-2) + n-1 + n-2 = 4n-7$ .

Let us verify that the word  $W$  is crucial.

If we write the letters  $a_1, a_2, a_n$  to the right of the word  $W$  we will obviously have prohibited subwords. Let  $2 < i < n$ . Then if we write the letters  $a_i$  we will have the prohibition

$$a_{i-1}a_i \dots a_1 a_2 a_n a_{n-2} \dots a_i \uparrow a_{i-1} \dots a_2 a_1 a_2 \dots a_{n-2} a_n a_i,$$

since the composition vectors of the left and right subwords with respect to the vertical line are equal.

Before proving that  $W \in \widehat{\mathbf{S}}_2^n$  we make the following remark.

In the word  $W$  we have  $\ell_n/\ell_1/\dots/\ell_{n-2}/\ell_{n-1}$ . Substituting  $a_1$  for  $a_n$ ,  $a_2$  for  $a_1, \dots, a_n$  for  $a_{n-1}$  we obtain another word

$$U = a_{n-1}a_n \dots a_2 a_3 \downarrow a_1 a_{n-1} \dots a_3 a_2 a_3 \dots a_{n-1} a_1,$$

for which  $\ell_1/\ell_2/\dots/\ell_n$ .

In both cases (before and after substitution of letters of the alphabet) we have the construction of a crucial word (which will be proved below) hence the same upper bound for the length of a minimal crucial word.

For  $W$  it is more convenient to show further that  $W \in \widehat{\mathbf{S}}_2^n$ .

We rewrite  $W$  making in it the marks  $(1),(2), \dots, (2n-4)$ , which number the gaps between letters of a word like this:

$$(2n-4)a_{n-2}(2n-5)a_{n-1} \dots (2)a_1(1)a_2 \uparrow a_n a_{n-2} \dots a_2 a_1 a_2 \dots a_{n-2} a_n.$$

In a possible prohibition we mark the left and right bounds. Note that the length of a prohibition is an even number, and each letter must occur an even number of times in a prohibition. The left bound of the prohibition must lie to the right of the mark  $(2n-5)$ , since the letter  $a_{n-1}$  occurs in  $W$  once;

It must lie to the left of the mark  $(1)$ , since to the right of the mark  $(1)$  there is one letter  $a_1$ .

Note that if  $m$  is even then  $(m)$  is not the left bound of the possible prohibition. Actually in this case two variants are possible:

- 1) the prohibition does not cover the left letter  $a_n$ .
- 2) the prohibition covers the left letter  $a_n$ .

In the second case we do not have a prohibition, since if the prohibition begins from the even mark, then it cannot cover the second  $a_n$ .

---

<sup>3</sup>An anonymous referee remarked that the lower bound for the result obtained in the theorem appears in [15].

In the first case the right bound of the prohibition lies to the left of  $a_n$ , hence the letter  $a_{\lfloor \frac{m}{2} \rfloor + 1}$  enters the prohibition only once.

Suppose the prohibition begins from the mark  $(m)$  and  $m$  is odd.

There are two possible cases.

1) The prohibition does not cover the left letter  $a_n$  (this case is impossible since the letter  $a_{\lfloor \frac{m}{2} \rfloor}$  occurs in the prohibition once).

2) The prohibition covers the left  $a_n$ . Then it covers the right  $a_n$  too, and the letter  $a_{\lfloor \frac{m}{2} \rfloor}$  occurs an odd number of times in the prohibition. So  $W \in \widehat{\mathbf{S}}_2^n$  and hence  $L_{min}(\mathbf{S}_2^n) \leq 4n - 7$  for  $n > 2$ .

We give now a lower bound.

Since the length of a minimal crucial word must be odd, and the passage from  $\ell_i$  to  $\ell_{i+1}$  requires at least two letters, we have that a trivial lower bound of length of a minimal crucial word is  $2n - 1$ .

Let us now improve the lower bound. Obviously a minimal crucial word in which  $\ell_1/\ell_2/\dots/\ell_n$  has an even number of occurrences of the letter  $a_i$  for  $i = 1, \dots, n - 1$  and an odd number of occurrences of the letter  $a_n$ . The word  $U$  has two letters  $a_1$ , two letters  $a_2$ , one letter  $a_n$  and four of any other letter. From Proposition 3 we know that there does not exist a crucial word that has the fewer number of letters, hence the word  $U$  gives us the lower bound of length of a minimal crucial word.  $\square$

## 4 The Set of Prohibitions $\mathbf{S}_3^{n,k}$

**Theorem 6.** *We have*

$$L_{min}(\mathbf{S}_3^{n,k}) = 2k + 1.$$

*Proof.* For the set of prohibitions  $\mathbf{S}_3^{n,k}$  we must have  $|A| = |B| \geq k + 1$ , where  $AB$  is an arbitrary prohibition. So we have

$$L_{min}(\mathbf{S}_3^{n,k}) \geq 2k + 1.$$

An upper bound is given by the construction  $p_1 p_2 \dots p_k x p_1 p_2 \dots p_k$ , where  $x, p_i \in \mathbf{A}$ ,  $i = 1, \dots, k$ .  $\square$

**Remark 7.** The crucial word with respect to  $\mathbf{S}_3^{1,k}$  is unique and its length is  $2k + 1$ .

**Theorem 8.** *We have*

$$L_{max}(\mathbf{S}_3^{2,k}) = 3k + 3.$$

*Proof.* Let

$$\bar{a} = \begin{cases} 1, & \text{if } a = 2, \\ 2, & \text{if } a = 1. \end{cases}$$

Moreover, let us consider an arbitrary crucial word  $A$ , with respect to  $\mathbf{S}_3^{2,k}$ , of length greater than  $3k + 3$ . It is easy to see that if  $a_1 a_2 \dots a_{k+1}$  are the first



$k + 1$  letters of  $A$  then the next  $k + 1$  letters of  $A$  must be  $\bar{a}_1\bar{a}_2 \dots \bar{a}_{k+1}$ , because otherwise the first  $2k + 2$  letters of  $A$  will form a prohibited subword. By the same argument, we can show that

$$A = a_1a_2 \dots a_{k+1}\bar{a}_1\bar{a}_2 \dots \bar{a}_{k+1}a_1a_2 \dots a_{k+1}\bar{a}_1 \dots$$

Let us consider the subwords  $A_i$  of  $A$  of the length  $2k + 4$  which start from the  $i$ th letter, where  $1 \leq i \leq k$ :

$$A_i = \underbrace{a_i a_{i+1} \dots a_{k+1} \bar{a}_1 \dots \bar{a}_i}_{k+2} \underbrace{\bar{a}_{i+1} \dots \bar{a}_{k+1} a_1 \dots a_{i+1}}_{k+2}$$

If  $a_i = \bar{a}_{i+1}$  then the underbraced subwords of  $A_i$  are the same in the first and in the last positions, so they differ in at most  $k$  positions, hence  $A_i$  is prohibited. So we must have  $a_i = a_{i+1}$  for  $i = 1, \dots, k$ .

Without loss of generality we can assume that  $a_1 = 1$ , so

$$A = \underbrace{11 \dots 1}_{k+1} \underbrace{22 \dots 2}_{k+1} \underbrace{11 \dots 1}_{k+1} 2 \dots$$

It is easy to see that if the length of  $A$  is greater than  $3k + 3$  then  $A$  has a prohibited subword of length  $2k + 4$ :

$$A = \underbrace{11 \dots 1}_k \underbrace{122 \dots 2}_{k+1} \underbrace{11 \dots 1}_{k+1} 22 \dots$$

(here and then two braces above a word show us a disposition of a prohibited subword and, in particular, a disposition of parts of this subword that correspond to  $X$  and  $Y$  from the definition of the set of prohibitions  $\mathbf{S}_3^{n,k}$ ).

So  $L_{max}(\mathbf{S}_3^{2,k}) \leq 3k + 3$ .

To prove the theorem it is sufficient to check that there are no prohibited subwords in the word  $A = \underbrace{11 \dots 1}_{k+1} \underbrace{22 \dots 2}_{k+1} \underbrace{11 \dots 1}_{k+1}$ .

Obviously the left end of a possible prohibition can be only in the left block  $\underbrace{1 \dots 1}_{k+1}$ :

$$\underbrace{1 \dots 1}_j \underbrace{2 \dots 2}_i \underbrace{2 \dots 2}_{k-i+1} \underbrace{1 \dots 1}_{2i+j-k-1}$$

with

$$j + i \geq k + 1 \tag{1}$$

Two cases are possible:

1.  $j \geq k - i + 1$
2.  $j < k - i + 1$

In the first case there is non-coincidence between the left and right parts of the prohibition in the first  $k - i + 1$  letters and in the last  $i$  letters that is non-coincidence in  $k + 1$  letters. So this case is impossible.

In the second case we have non-coincidence in the first  $j$  letters and in the last  $2i + j - k - 1$  letters. Hence we have non-coincidence in  $2(i + j) - k - 1$  letters, that according to (1) is greater than or equal to  $k + 1$ .

It follows that the word  $\underbrace{1\dots 1}_{k+1}\underbrace{2\dots 2}_{k+1}\underbrace{1\dots 1}_{k+1}$  does not contain a prohibition and thus the theorem is proved.  $\square$

**Theorem 9** (Incompleteness). *The set of prohibitions  $\mathbf{S}_3^{n,k}$  for  $n \geq 3$  is incomplete.*

*Proof.* Since the alphabet  $\mathbf{A}$  is finite, there is no trivial solution of the problem (such as taking all letters of  $\mathbf{A}$  and obtaining an infinite sequence with the properties needed). So to prove the incompleteness of the set  $\mathbf{S}_3^{n,k}$  we have to show the existence of an infinite word which is free from the set of prohibitions  $\mathbf{S}_3^{n,k}$ .

We consider the case  $n = 3$  and the alphabet  $\mathbf{A} = \{1, 2, 3\}$ , since the incompleteness of the set of prohibitions  $\mathbf{S}_3^{n,k}$  for the case  $n > 3$  will follow from the incompleteness of the set of prohibitions for the case  $n = 3$ .

Let  $\mathbf{B} = \{a, b, c\}$  be an alphabet. We define the mapping  $h$  by  $h(a) = 1^{k+1}$ ,  $h(b) = 2^{k+1}$ ,  $h(c) = 3^{k+1}$ . Thus,  $h$  is a mapping from  $\mathbf{B}^*$  to the set of all words over the alphabet  $\mathbf{C} = \{1^{k+1}, 2^{k+1}, 3^{k+1}\}$ . Clearly, the inverse  $f = h^{-1}$  is well-defined. The domain of the mapping  $f$  is the set of all words over the alphabet  $\mathbf{C}$ , and the image of  $f$  is the set  $\mathbf{B}^*$ .

Let the set of prohibitions  $\mathbf{S}' = \{XX|X \in \mathbf{B}^+\}$ . Obviously, the set  $\mathbf{S}'$  coincides with the set  $\mathbf{S}_2^n$  whenever  $\mathbf{A} = \mathbf{B}$ .

It is known [2] that for the alphabet  $\mathbf{B}$  there exists an infinite sequence  $L'$  which is free from the set of prohibitions  $\mathbf{S}'$ .  $L'$  is built by iteration of the morphisms:

$$\begin{aligned} a &\rightarrow abc \\ b &\rightarrow ac \\ c &\rightarrow b \end{aligned}$$

The morphism iteration procedure is as follows.

We start from the letter  $a$ . Then we substitute this letter with  $abc$ . Then we substitute each letter in  $abc$  by the rule above. We obtain after this step  $abcacb$ . And so on. Executing this procedure an infinite number of times gives us the sequence  $L'$ .

Let us prove that the sequence  $L = h(L')$  does not contain words prohibited by  $\mathbf{S}_3^{3,k}$ .

We are going to prove the statement by considering  $L$  and all possible dispositions of words prohibited by  $\mathbf{S}_3^{3,k}$ .

The sequence  $L$  is built up from the letters of the alphabet  $\mathbf{C}$  or in other words from the *blocks*  $x^k = \underbrace{x \dots x}_{k+1}$ , where  $x \in \{1, 2, 3\}$ . It means that there are only three different cases for a disposition of a possible prohibition in  $L$ .

$$\text{Case 1. } \overbrace{x^{k+1} \dots y^{k+1}} \overbrace{z^{k+1} \dots t^{k+1}};$$

$$\text{Case 2. } x^i \overbrace{x^{k-i+1} \dots y^{k+1}} \overbrace{z^{k+1} \dots t^{k-i+1}} t^i, \text{ where } 0 < i < k + 1;$$

$$\text{Case 3. } x^i \overbrace{x^{k-i+1} \dots y^\ell} \overbrace{y^{k-\ell+1} \dots t^{k-j+1}} t^j, \text{ where } 0 \leq i, j, \ell \leq k + 1.$$

Now we will consider these cases and show that each of them is impossible.

**Case 1.** Let  $\mathbf{P}$  denote the prohibited subword (prohibition) under consideration,  $\mathbf{R}$  and  $\mathbf{L}$  denote the right and the left parts of  $\mathbf{P}$  respectively.

It is obvious that  $\mathbf{L}$  and  $\mathbf{R}$  have the same number of blocks. Moreover, the  $i$ th block of  $\mathbf{L}$  (from the left to the right) is equal to the  $i$ th block of  $\mathbf{R}$ , because otherwise we have non-coincidence of  $\mathbf{L}$  and  $\mathbf{R}$  in at least  $k + 1$  letters which contradicts the fact that  $\mathbf{P} \in \mathbf{S}_3^{3,k}$ . So we have that  $\mathbf{P} = WW$  for some  $W \in \mathbf{C}^*$ .

Now,  $f(\mathbf{P}) = f(W)f(W)$  is a subword of  $L'$ . But  $f(W)f(W) \in \mathbf{S}'$  which is impossible by the properties of  $L'$ . So Case 1 is impossible.

We note that an important consequence of Case 1 is the following. If  $x^{k+1}y^{k+1}$  is a subword of  $L$  then  $x \neq y$ .

**Case 2.** If there are no letters between  $x^{k+1}$  and  $y^{k+1}$ , that is

$$\mathbf{P} = \overbrace{x^{k-i+1}y^{k+1}} \overbrace{z^{k+1}t^{k-i+1}},$$

then we must have  $x = z$ , because otherwise we have  $x \neq z$  and  $y \neq z$  which gives us that  $\mathbf{L}$  and  $\mathbf{R}$  differ in the first  $k + 1$  positions, but this contradicts  $\mathbf{P} \in \mathbf{S}_3^{3,k}$ .

By the same argument we have  $y = t$ , so

$$\mathbf{P} = \overbrace{x^{k-i+1}y^{k+1}} \overbrace{x^{k+1}y^{k-i+1}}.$$

But if we consider now  $f(L) = L'$  then it has

$$\mathbf{P}' = \overbrace{f(x^{k+1})f(y^{k+1})} \overbrace{f(x^{k+1})f(y^{k+1})}.$$

as a subword, which is impossible since  $\mathbf{P}' \in \mathbf{S}'$ .

So there is some nonempty subword in  $\mathbf{L}$  between  $x^{k+1}$  and  $y^{k+1}$ , and  $\mathbf{P}$  can be written as

$$\mathbf{P} = \overbrace{x^{k-i+1}x_1^{k+1} \dots x_p^{k+1}y^{k+1}} \overbrace{z^{k+1}z_1^{k+1} \dots z_p^{k+1}t^{k-i+1}}.$$

There are two possible subcases here.

1.  $x = z$ .

Since  $x \neq x_1$  we have  $x_1 \neq z$ . If  $x_1 \neq z_1$  then  $\mathbf{L}$  and  $\mathbf{R}$  differ in  $k+1$  position starting from the  $(k-i+2)$ th position, which is impossible since  $\mathbf{P} \in \mathbf{S}_3^{3,k}$ . So  $x_1 = z_1$ .

In the same way, for each of  $x_2, x_3, \dots, x_p, y$ , we can obtain that

$$\mathbf{P} = \overbrace{z^{k-i+1} z_1^{k+1} \dots z_p^{k+1} t^{k+1}} \overbrace{z^{k+1} z_1^{k+1} \dots z_p^{k+1} t^{k-i+1}}$$

which leads us to the fact that  $\mathbf{L}$  has a subword  $WW$  for some  $W \in \mathbf{C}^*$ , hence  $\mathbf{L}'$  has a subword  $f(W)f(W)$  which is impossible.

So the subcase 1 is impossible.

**2.**  $x \neq z$ .

If  $x_1 \neq z$  then  $\mathbf{L}$  and  $\mathbf{R}$  differ in  $k+1$  position starting from the first position, which is impossible since  $\mathbf{P} \in \mathbf{S}_3^{3,k}$ . So  $x_1 = z$ .

If  $x_2 \neq z_1$  then  $\mathbf{L}$  and  $\mathbf{R}$  differ in  $k+1$  position starting from the  $(k+2)$ th position, which is impossible by the same arguments as above. So  $x_2 = z_1$ . And so on.

We have

$$\mathbf{P} = \overbrace{x^{k-i+1} z^{k+1} z_1^{k+1} \dots z_p^{k+1}} \overbrace{z^{k+1} z_1^{k+1} \dots z_p^{k+1} t^{k-i+1}}.$$

Applying  $f$  to  $L$  gives us a subword  $\mathbf{P}'$  of  $L'$ ,

$$\mathbf{P}' = \overbrace{f(z^{k+1})f(z_1^{k+1}) \dots f(z_p^{k+1})} \overbrace{f(z^{k+1})f(z_1^{k+1}) \dots f(z_p^{k+1})},$$

which is prohibited in  $L'$  by  $\mathbf{S}'$ .

We have got that subcase 2 is impossible and hence Case 2 is impossible.

**Case 3.** We can assume that  $\ell \neq 0$  and  $\ell \neq k+1$ , because otherwise we deal with either Case 1 or Case 2 which are impossible.

We suppose that  $i \geq \ell$  (the case  $i < \ell$  can be considered in the same way).

If there are no letters between  $y^{k-\ell+1}$  and  $t^{k-j+1}$ , then we have either

$$\mathbf{P} = \overbrace{x^{k-i+1} y^\ell} \overbrace{y^{k-\ell+1} t^{k-j+1}}$$

or

$$\mathbf{P} = \overbrace{x^{k-i+1} z^{k+1} y^\ell} \overbrace{y^{k-\ell+1} t^{k-j+1}}.$$

In the first of these cases we have that  $x \neq y$  and  $y \neq t$  which gives us that  $\mathbf{L}$  and  $\mathbf{R}$  have noncoincidence in at least  $k+1$  letters, but this contradicts  $\mathbf{P} \in \mathbf{S}_3^{3,k}$ .

In the second case we must have  $z = t$ , because otherwise since  $z \neq y$  and  $t \neq y$ ,  $\mathbf{L}$  and  $\mathbf{R}$  have noncoincidence in the last  $k+1$  letters which is impossible. So in the second case we have

$$\mathbf{P} = \overbrace{x^{k-i+1} t^{k+1} y^\ell} \overbrace{y^{k-\ell+1} t^{k-j+1}}.$$

If  $x \neq y$  then  $\mathbf{L}$  and  $\mathbf{R}$  have noncoincidence in the first  $k - \ell + 1$  positions and in the last  $\ell$  positions, that is they have noncoincidence in at least  $k + 1$  positions which is impossible. So  $x = y$ .

Now applying  $f$  to  $L$  gives us that  $L'$  contains the subword

$$\mathbf{P}' = \overbrace{f(x^{k+1})f(t^{k+1})} \overbrace{f(x^{k+1})f(t^{k+1})},$$

which is impossible.

So there is some nonempty subword in  $\mathbf{R}$  between  $y^{k-\ell+1}$  and  $t^{k-j+1}$ , and  $\mathbf{P}$  can be written in the form

$$\mathbf{P} = \overbrace{x^{k-i+1}L_1 \dots L_p y^\ell} \overbrace{y^{k-\ell+1}R_1 \dots R_{p'} t^{k-j+1}},$$

where  $L_s, R_m \in \mathbf{C}$ , for  $1 \leq s \leq p$ ,  $1 \leq m \leq p'$ , and either  $p = p'$  or  $p = p' + 1$ .

We define  $\Delta(L_s) = x_s$  if  $L_s = x_s^{k+1}$ . In the same way we define  $\Delta(R_m)$ .

Now we have that either  $p = p'$  or  $p = p' + 1$ . Each of these cases has two possible subcases: either  $x = y$  or  $x \neq y$ . Let us consider the case  $p = p' + 1$ . The other case can be considered by similar reasoning. Thus we must consider the following subcases a) and b):

**a)**  $x = y$ ; It must be that  $L_1 = R_1$ , because otherwise  $\mathbf{L}$  and  $\mathbf{R}$  differ in  $k + 1$  positions starting from the  $(k - i + 2)$ th position. Then we consider one by one  $L_2, L_3, \dots, L_p$ . One can see that in this subcase

$$\mathbf{P} = \overbrace{y^{k-i+1}R_1 \dots R_{p'} t^{k+1} y^\ell} \overbrace{y^{k-\ell+1}R_1 \dots R_{p'} t^{k-j+1}},$$

and  $L$  has  $WW$  as a subword, where  $W = y^{k+1}R_1 \dots R_{p'} t^{k+1}$ , which is impossible.

**b)**  $x \neq y$ ; There are two special subcases here, namely either  $\Delta(L_1) = y$  or  $L_1 = R_1$ .

If  $\Delta(L_1) = y$  then

$$\mathbf{P} = \overbrace{x^{k-i+1}y^{k+1}R_1 \dots R_{p'} y^\ell} \overbrace{y^{k-\ell+1}R_1 \dots R_{p'} t^{k-j+1}},$$

and  $L$  has  $WW$  as a subword, where  $W = y^{k+1}R_1 \dots R_{p'}$ , which is impossible.

So  $L_1 = R_1$ . In this case we have

$$\mathbf{P} = \overbrace{x^{k-i+1}R_1 \dots R_{p'} t^{k+1} y^\ell} \overbrace{y^{k-\ell+1}R_1 \dots R_{p'} t^{k-j+1}}.$$

Since  $y \neq x$ ,  $y \neq \Delta(R_1)$  and  $y \neq t$ ,  $\mathbf{L}$  and  $\mathbf{R}$  have noncoincidence in the first  $k - l + 1$  positions and in the last  $l$  positions, so they have noncoincidence in  $k + 1$  positions which contradicts  $\mathbf{P} \in \mathbf{S}_3^{3,k}$ .

We have got that Case 3 is impossible.

We have proved that the infinite word  $L$  contains no word from the set  $\mathbf{S}_3^{3,k}$  as a subword, therefore  $\mathbf{S}_3^{n,k}$  is incomplete for  $n \geq 3$ .  $\square$

## 5 The Complexity of Problems on Completeness of Sets of Words

It is known [9, 10] that the complexity of deciding whether or not an arbitrary finite set of prohibited words  $\mathbf{S}$  is complete is  $O(|\mathbf{S}| \cdot n)$ , where  $n$  is the greatest length of a word in  $\mathbf{S}$  (see also [17] and references therein).

It is interesting in its own right to be able to effectively (in polynomial time) recognise whether a set is complete, but also to give a more detailed characterisation of the set of words  $\widehat{\mathbf{S}}$ , in particular to find the greatest length of a word that is free from  $\mathbf{S}$ . The set  $\mathbf{A}^n$  is the set of all the words over the alphabet  $\mathbf{A}$  whose length is equal to  $n$ . If  $\mathbf{S} \subseteq \mathbf{A}^n$  and  $L(n) = \max_{\mathbf{S}} L(\widehat{\mathbf{S}})$ , where  $L(\widehat{\mathbf{S}})$  is the greatest length of a word that is free from  $\mathbf{S}$ , then [9] we have

$$L(n) = |\mathbf{A}|^{n-1} + n - 2 = C(n) + n - 1.$$

Here  $C(n)$  is the greatest length of a single path in the de Bruijn graph of order  $n$  that has no chords and does not go through the vertices with loops corresponding to the constant words  $(x, \dots, x)$  where  $x \in \mathbf{A}$ .

One can find all words that are free from  $\mathbf{S}$ , in particular all crucial words, simply by considering all words of length less than or equal to  $L(\widehat{\mathbf{S}})$  and checking for each word, if it is free from  $\mathbf{S}$ . Such an algorithm is not effective since it can require considering  $\exp(L(n))$  words.

The question of deciding the possible lengths of words that are free from  $\mathbf{S}$ , in particular of crucial words, can be formulated as a problem of recognising properties of “languages of prohibitions” in the terminology of the theory of NP-completeness [13].

**Problem A:**

*Given:* An arbitrary set of words  $\mathbf{S}$  and a natural number  $\ell$ .

*The question:* Does there exist a word of length at least  $\ell$  that is free from  $\mathbf{S}$ ?

In order to compare, we formulate the problem of completeness of a set of words  $\mathbf{S}$  in the same form.

**Problem B:**

*Given:* An arbitrary set of words  $\mathbf{S}$  in an alphabet  $\mathbf{A}$ .

*The question:* Does there exist  $\ell \in \mathbf{N}$  such that  $|X| \leq \ell$  for any word  $X$  that is free from  $\mathbf{S}$ ?

Considering problems A and B as problems of recognising properties of finite sets  $\mathbf{S}$ , we observe that problem B is a question of *existence of a bound* on the length of the words that are free from  $\mathbf{S}$ . This problem, as we have already mentioned, can be solved effectively with complexity of order  $|\mathbf{S}| \cdot n$ . In the same time the problem A is a question of *localization of this bound*. We will show that problem A, as opposed to problem B, is NP-complete. We observe that problems A and B for infinite sets  $\mathbf{S}$  do not make sense if one does not consider particular constructive methods for generating a set  $\mathbf{S}$ .

Let  $\mathbf{A} = \{a_1, \dots, a_n\}$  be an alphabet and  $\mathbf{A}_\ell$  be the set of all those words on the alphabet  $\mathbf{A}$  whose length is less than or equal to  $\ell$ . We assume also

that the empty word belongs to  $\mathbf{A}_\ell$  and that  $\mathbf{S}_1$  is an arbitrary set such that  $\mathbf{S}_1 \subseteq \mathbf{A}_2 \setminus \mathbf{A}_1$ . We define  $\mathbf{S}_2$  by

$$\mathbf{S}_2 = \{xXx \mid x \in \mathbf{A}, X \in \mathbf{A}^{n-1}\}.$$

So the set  $\mathbf{S}_2$  contains all possible words of length less than or equal to  $n+1$  whose first letter coincides with their last letter. Suppose  $\mathbf{S} = \mathbf{S}_1 \cup \mathbf{S}_2$ .

We now consider an “auxiliary” problem  $A'$ .

**Problem  $A'$ :**

*Given:* A set  $\mathbf{S}$  of the type described above and a natural number  $\ell$ ,  $\ell \leq n$ .

*The question:* Does there exist a word of length at least  $\ell$  that is free from  $\mathbf{S}$ ?

In case of the problem  $A'$ , the restriction on  $\ell$  is natural, because any word free from  $\mathbf{S}$  is free from  $\mathbf{S}_2$  and therefore consists of different letters of the alphabet, whence its length is less than or equal to  $n$ .

Checking whether a given word of length  $\ell$  (a solution of  $A'$  that we “guessed”) is free from  $\mathbf{S}$  can be done in polynomial time. Indeed, the freeness from  $\mathbf{S}_2$  of the word is equivalent to the absence of identical letters in the word (which can be checked in linear time) and the freeness from  $\mathbf{S}_1$  is recognised by considering all subwords of length 2 (there are  $\ell - 1$  such subwords) and by checking for each of them whether it belongs to  $\mathbf{S}_1$  (polynomial checking time).

We now introduce the problem of “the longest path in a graph”, which is known to be NP-complete (see [13]).

**Problem “path”:**

*Given:* A directed graph  $\vec{G}(V, E)$  and a natural number  $\ell$ ,  $\ell \leq |V| = n$ .

*The question:* Does there exist a simple directed path (without self-intersections in vertices) of length at least  $\ell$ ?

One can obtain a correspondence between problem  $A'$  and problem “path” as follows. We compare vertices  $v_1, \dots, v_n$  from  $V(\vec{G})$  to the letters  $a_1, \dots, a_n$  over the alphabet  $\mathbf{A}$ . Also we compare each edge  $v_i \vec{v}_j$  from  $E(\vec{G})$  to the word  $a_i a_j$ . We form the set  $\mathbf{S}_1$  from all such words of  $\mathbf{A}^2$  that correspond to the edges of the graph that is the complement of  $\vec{G}$  with respect to the complete directed graph.

Now to any oriented simple path  $v_{i_1}, \dots, v_{i_\ell}$  of length  $\ell$  in  $\vec{G}$  there corresponds the word  $a_{i_1} \dots a_{i_\ell}$  of length  $\ell$ , consecutive letters of which correspond to vertices in the order in which the path passed through them. This word is free from  $\mathbf{S}_1$  because  $a_{i_j} a_{i_{j+1}} \notin \mathbf{S}_1$  for any  $i = 1, 2, \dots, \ell - 1$ . The word is free from the set  $\mathbf{S}_2$  as well because in the path there is no repetition of vertices (a property of a simple path) and therefore  $a_{i_1} \dots a_{i_\ell}$  does not contain a subword of the form  $a_i X a_i$  for any word  $X$  and any letter  $a_i \in \mathbf{A}$ .

Conversely, to any word over the alphabet  $\mathbf{A}$  that is free from  $\mathbf{S}$  there corresponds a path in  $\vec{G}(V, E)$  that goes through edges from  $E(\vec{G})$  since the word is free from  $\mathbf{S}_1$  and that is not self-intersected since the word is free from  $\mathbf{S}_2$ .

Now NP-completeness of problem  $A'$  and the more general problem A follows from NP-completeness of the problem “path”.

## Acknowledgments

The authors are very grateful to the anonymous referees for many helpful comments and suggestions.

## References

- [1] Arshon S. E., The Proof of the Existence of  $n$  - letter Non-Repeated Asymmetric Sequences, Math. collected papers, New series, Publ. H. AS USSR, Vol. **2**, Issue 3, Moscow (1937), 769–779. (Russian)
- [2] Bean D. R., Ehrenfeucht A., McNulty G. F., Avoidable patterns in strings of symbols, *Pacific J. Math.*, Vol. **85**, no. 2 (1979), 261–294.
- [3] Carpi A., On the number of abelian square-free words on four letters, *Discrete Appl. Math.* **81** (1998), no. 1-3, 155–167.
- [4] Choffrut C., Karhumäki J., *Handbook of formal languages*, Vol. **1**: Word, language, grammar, Berlin, Springer (1997), 329–438.
- [5] Erdős P., Some unsolved problems, *Publ. Math. Inst. Hung. Acad. Sci. Ser. A* **6** (1961), 221–254.
- [6] Evdokimov A., Strongly asymmetric sequences generated by a finite number of symbols, *Dokl. Akad. Nauk. SSSR* **179** (1968), 1268–1271 (English transl. *Soviet. Math. Dokl.* **9** (1968), 536–539)
- [7] Evdokimov A., The existence of a basis that generates 7-valued iteration-free sequences. *Diskret. Analiz* **18** (1971), 25–30. (in Russian). See MR#45:4914.
- [8] Evdokimov A., Completeness of a word set. (Talk at the international conference FCT-79, Wendishritz, Germany (1979)).
- [9] Evdokimov A., Complete sets of words and their numerical characteristics, *Metody Diskret. Analiz., Novosibirsk, IM SB RAS* **39** (1983), 7–32 (in Russian). See MR#86e:68087.
- [10] Evdokimov A., The completeness of sets of words, Proceedings of the All-Union seminar on discrete mathematics and its applications (Russian) (Moscow, 1984), Moskov. Gos. Univ., Mekh.-Mat. Fak., Moscow (1986), 112–116 (in Russian). See MR#89e:68066.
- [11] Evdokimov A., Kitaev S., Crucial words and the complexity of some extremal problems for sets of prohibited words, Chalmers University of Technology and Göteborg University, preprint 2001:64 (2001).
- [12] Evdokimov A., Krainev V., Problems on completeness of sets of words, Proceedings of the 22nd regional scientific conference by the Popov association, Novosibirsk (1979), 105–107 (in Russian).



- [13] Garey M., Johnson D., *Computers and intractability: a guide to the theory of NP-completeness*, W. H. Freeman (1979).
- [14] Keränen V., Abelian squares are avoidable on 4 letters, *Automata, languages and programming* (Vienna, 1992), 41–52, Lecture Notes in Comput. Sci., 623, Springer, Berlin (1992).
- [15] Korn M., Maximal abelian square-free words of short length, *J. Comb. Theory A* **102** (2003), 207–211.
- [16] Lothaire M., *Combinatorics on Words*, Encyclopedia of Mathematics, Vol. **17**, Addison-Wesley (1983).
- [17] Lothaire M., *Algebraic Combinatorics on Words*, Encyclopedia of Mathematics 90, Cambridge University Press (2002).
- [18] Pleasants P. A. B., Non-repetitive sequences, *Proc. Cambridge Philos. Soc.* **68** (1970) 267–274.
- [19] Saker C., Higgins P., Unavoidable sets of words of uniform length, *Information and Computation* **173** (2002).
- [20] Schützenberger M.-P., On the synchronizing properties of certain prefix codes, *Information and Control* **7** (1964), 23–36.
- [21] Zimin A. I., Blocking sets of terms, *Mat. Sbornik*, Vol. **119**, No. 3 (1982), 363–375, 447 (in Russian). See MR#84d:20072. English translation: *Math. USSR Sbornik*, **47** (1984), 353–364.