SEMIRETRACTS - ALGORITHMIC PROBLEMS

TOMASZ KRAWCZYK

Jagiellonian University, Institute of Computer Science
Nawojki 11, 30-072 Krakow, Poland
krawczyk@ii.uj.edu.pl
KBN grant no 3 T11C 010 27

1. Introduction

Semiretracts of free monoids were investigated first by Jim Anderson [1] and then were the subject of the papers - see references [1-6, 10-12, 14-15]. In the paper [1] J.A.Anderson presented a theorem that characterizes any semiretract S by means of two retracts R_α, R_ω. Namely, he showed that for any semiretract S there exist retracts R_α and R_ω such that $S = R_\alpha \cap R_\omega$. In the paper [2] the counterexample to this characteristic was given. In the sequel, in this paper we introduce the notion of dimension of S (written $dim(S)$); namely, $dim(S) = k$ iff k is the minimal number such that $S = \bigcap_{i=1}^{k} R_i$ for some retracts $R_1, ..., R_k$. We present a polynomial time algorithm that test if $dim(S) = k$. On the other hand, we show that a little modification of this problem is $NP-$complete.

2. Basic Notions And Definitions

We assume the reader is familiar with the basic notions and concepts from the theories of semigroups and the the theories of computation.

Let A be any finite set and let A^* denote a free monoid generated by A. The length of a word $w \in A^*$, in symbols $|w|$, is defined to be the number of letters occuring in w (the length of the empty word 1 equals 0).

A retraction $r : A^* \longrightarrow A^*$ is a morphism for which $r \circ r = r$. A retract R of A^* is the image of A^* by a retraction. A semiretract S of A^* is the intersection of a family of retracts of A^*. A dimension of semiretract S - written $dim(S)$ - is equal k iff k is the minimal number such that $S = \bigcap_{i=1}^{k} R_i$ for some retracts $R_1, ..., R_m$. The following theorem is due to J.A.Anderson - see [3].

Theorem 2.1. *$Dim(S)$ is finite for any semiretract S.*

A word $w \in A^*$ is called a key-word if there is at least one letter in A that occurs exactly once in w and the letter is called a key of w. A set $C \subset A^*$ of key-words is called a key-code if there exists an injection $key : C \longrightarrow A$ such that

(1) for any $w \in C$, $key(w)$ is a key of w,
(2) the letter $key(w)$ occurs in no word of C other than w itself.

Note that any key-code is in fact a code and that for a key-code C there is possible to exist more then one injection $key : C \longrightarrow A$. Given a key-code C and a fixed mapping key the set of all keys of words in C is denoted by $key(C)$.

The following characterization of retracts is due to T. Head [**?**].

Theorem 2.2. *$R \subset A^*$ is a retract of A^* if and only if $R = C^*$ where C is a key-code.*

Because we shall be dealing with the complexity problems let us define the set of all inputs (instances) $\mathcal{I}$; namely a sequence $(C_1, ..., C_k, l)$ is in $\mathcal{I}$ iff $C_1, ..., C_n$ are key codes and l is a positive integer. Hence, with any $(C_1, ..., C_n, l) \in \mathcal{I}$ we can associate a semiretract $S = \bigcap_{i=1}^n C_i^*$. The first decision problem (given as a languge) $DIM - SEM \subset \mathcal{I}$ related to the dimension of semiretract can be defined as follows: $(C_1, ..., C_n, l)$ is in $DIM - SEM$ iff there exist l key codes $D_1, ..., D_l$ such that $\bigcap_{i=1}^n C_i^* = \bigcap_{i=1}^l D^i$. We also will consider the decision problem $MIN - SEM \subset \mathcal{I}$; an instance $(C_1, ..., C_n, l)$ is in $MIN - SEM$ iff there exists key codes $C_{i_1}, ..., C_{i_l} \in \{C_1, ..., C_n\}$ for some $i_1, ..., i_l \in \{1, ..., n\}$ such that $\bigcap_{i=1}^n C_i^* = \bigcap_{j=1}^l C_{i_j}^*$.

The main thesis of this paper is as follows: $DIM - SEM$ is in P while $MIN - SEM$ is $NP-$complete.

3. Preliminary results

Let $(C_1, ..., C_n, k) \in \mathcal{I}$. In [2] W. Forys and T. Krawczyk proved the theorem that allows us to narrow down the research on semiretracts to the case when all considered retracts have the same, common key-set K.

Theorem 3.1. *Let $S = \cap_{i=1}^n C_i^*$ be a semiretract given by retracts C_i^* with key-codes $C_i \subset A^*$ for $i = 1, ..., n$. There exist key-codes $D_i \subset A^*$ for $i = 1, ..., n$ such that*

1. *$S \subset D_i^* \subset C_i^*$ for all $i = 1, ..., n$ (it means $S = \bigcap_{i=1}^n C_i^*$)*
2. *$key(D_1) = key(D_2) = ... = key(D_n)$.*

Hence any semiretract S is an intersection of a family of retracts generated by key codes having the common set of keys.

Let $S = \bigcap_{i=1}^n D_i^*$ and let $D_1, ..., D_n$ be key codes with the same set K. In the rest of the paper we assume that any $k \in K$ occurs in some word from the base of semiretract S.

Let us fix the order of retracts - $D_1^*, ..., D_n^*$. For any $k \in K$ there exist words $w_1 \in D_1, ..., w_n \in D_n$ all with the key k. We write this fact in a matrix form (abbreviated $n-$lines):

$$A(k) = \begin{bmatrix} u_1 & k & v_1 \\ \vdots & \vdots & \vdots \\ u_i & k & v_i \\ \vdots & \vdots & \vdots \\ u_n & k & v_n \end{bmatrix}.$$

Hence, in the first column of $A(k)$ there are prefixes u_i of w_i and in the third column there are sufixes v_i of w_i such that $w_i = u_i k v_i$ for all $i = 1, ..., n$. The matrix $A(k)$ is associated with the key $k \in K$. We denote in the sequel by $col_L(k)$ and by $col_R(k)$ the first (left) and the third column of A_k. Since k occurs in some word from the base of semiretract S, then u_i is a suffix of u_j or u_j is a suffix of u_i for all $i, j = 1, ..., n$. For the same reason w_i is a prefix of w_j or w_j is a prefix of w_i for all $i, j = 1, ..., n$. If it is necessary we underline that $A(k)$, $col_L(k)$, $col_R(k)$ were defined relatively to the order $D_1, ..., D_n$.

Definition 3.2. We say that $k \in K$ is initial key if $col_L(k) = \begin{bmatrix} u \\ \vdots \\ u \end{bmatrix}$ for some $u \in A^*$. We denote the word u by $left(k)$ as it occurs on the left site of the letter k. We say that $k \in K$ is final if $col_R(k) = \begin{bmatrix} w \\ \vdots \\ w \end{bmatrix}$ for some $w \in A^*$. We denote the word w by $right(k)$ as it occurs on the right site of k.

The set of all initial keys we denote by L_{init}. The set of all final keys we denote by R_{final}.

Definition 3.3. It is said that columns $U = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$ and $V = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ form an $n-$factorization of the word $w \in A^+$ and it is written $U \leftrightarrow_n V$ iff $u_i v_i = w$ for $i = 1, ..., n$ and there exist i, j such that $u_i \neq u_j$. Let $u \in A^*$ be the longest common prefix of $u_1, ..., u_n$ and let v be the longest common suffix of $v_1, ..., v_n$. Then there exist $u_1^{'}, v_1^{'}, ..., u_n^{'}, v_n^{'} \in A^*$ such that $u_i = uu_i^{'}$ and $v_i = v_i^{'}v$ for all $i = 1, ..., n$. Then the columns $U^{'} = \begin{bmatrix} u_1^{'} \\ \vdots \\ u_n^{'} \end{bmatrix}$ and $V^{'} = \begin{bmatrix} v_1^{'} \\ \vdots \\ v_n^{'} \end{bmatrix}$ form an $n-$factorization of some word $w^{'} \in A^+$. The $n-$factorization $U^{'} \leftrightarrow_n V^{'}$ is called the base and the word $w^{'}$ is called the source of the $n-$factorization $U \leftrightarrow_n V$.

Definition 3.4. Let $k_1, k_2 \in K$. We say that k_2 follows k_1 iff $col_R(k_1) \leftrightarrow col_L(k_2)$ constitutes $n-$factorization of some word $w \in A^+$. The word w is denoted by $bk(k_1, k_2)$ as it occurs between keys k_1 and k_2.

The above introduced notations allows us to give a simple lemma that presents a method for obtaining any word in the base of semiretract $S = \bigcap_{i=1}^{n} D_i^*$.

Lemma 3.5. *Let $k_1, ..., k_p \in K$ be a sequence of keys of the semiretract S such that (1) k_1 is initial key, (2) k_p is final key and k_{i+1} follows k_i for $i = 1, ..., p-1$. Then the word*

$$w = left(k_1)k_1bk(k_1, k_2)k_2.......k_{p-1}bk(k_{p-1}, k_p)k_pright(k_p)$$

is in the base (code) C of semiretract S. Moreover, for any word w in C there exist keys $k_1, ..., k_p \in K$ such that the above is true.

Any sequence of keys $k_1, ..., k_p \in K$ fulfilling assumptions (1)-(3) is called a generating key sequence.

Remark 3.6. Finding a word from the base of the semiretract is equivalent to finding a sequence of keys which fulfils the conditions from the above theorem.

Example 3.7. Assume that E_1, E_2 and E_3 are key codes with the same key set $K = \{k_1, k_2, k_3, k_4, k_5\}$.
$E_1 = \{abk_1aba, k_2aa, bk_3b, bk_4baba, k_5aa\}$,
$E_2 = \{abk_1ab, ak_2a, abk_3b, abk_4bab, ak_5a\}$

$E_3 = \{abk_1a, bak_2, aabk_3b, babk_4ba\}$.
Hence $A(k_1)$,$A(k_2)$,$A(k_3)$,$A(k_4)$ and $A(k_5)$ are equal respectively

$$\begin{bmatrix} a & b & k_1 & a & b & a \\ a & b & k_1 & a & b & \\ a & b & k_1 & a & & \end{bmatrix}, \begin{bmatrix} & & k_2 & a & a \\ & a & k_2 & a & \\ b & a & k_2 & & \end{bmatrix}, \begin{bmatrix} & & k_3 & b \\ & a & k_3 & b \\ a & a & k_3 & b \end{bmatrix},$$

$$\begin{bmatrix} & & b & k_4 & b & a & b & a \\ & a & b & k_4 & b & a & b & \\ b & a & b & k_4 & b & a & & \end{bmatrix} \; and \; \begin{bmatrix} & & k_5 & a & a \\ & a & k_5 & a & \\ a & a & k_5 & & \end{bmatrix}.$$

For example:

$$col_L(k_1) = \begin{bmatrix} a & b \\ a & b \\ a & b \end{bmatrix}, \; col_R(k_1) = \begin{bmatrix} a & b & a \\ a & b & \\ a & & \end{bmatrix}, \; col_L(k_2) = \begin{bmatrix} & \\ & a \\ b & a \end{bmatrix}.$$

Hence k_1 is initial key and k_3 is final key. The key k_2 follows k_1, since $col_R(k_1) \leftrightarrow_3 col_L(k_2)$ form $3-$factorization of the word aba. The $3-$factorization $\begin{bmatrix} b & a \\ b & \\ & \end{bmatrix} \leftrightarrow_3 \begin{bmatrix} & \\ & a \\ b & a \end{bmatrix}$ is the base and the word ba is the source of $3-$factorization $col_R(k_1) \leftrightarrow col_L(k_2)$.
Since k_1 is initial key, k_2 follows k_1, k_3 follows k_2 and k_3 is final, then the sequence k_1, k_2, k_3 is the generating key sequence. Hence the word

$$left(k_1)k_1bk(k_1,k_2)k_2bk(k_2,k_3)k_3right(k_3) = abk_1abak_2aak_3b$$

is in the base of semiretract $E_1^* \cap E_2^* \cap E_3^*$.

4. The problem $DIM - SEM$ is in P.

Suppose now that $(C_1, ..., C_n, l) \in \mathcal{I}$. By the previous paragraph there exists a sequence of key codes $D_1, ..., D_n$ with the same set of keys K such that $S = \bigcap_{i=1}^{n} D_i^*$.

Let $k_1, k_2 \in K$ be any keys such that k_2 follows k_1. Assume that $n-$factorization $U \leftrightarrow_n V$ is the base of $col_R(k_1) \leftrightarrow_n col_L(k_2)$. If k_3 and k_4 are such that k_4 follows k_3 and the base of $n-$factorization $col_R(k_3) \leftrightarrow_n col_L(k_4)$ is equal $U \leftrightarrow_n V$, then k_4 follows k_1 and k_2 follows k_3 as well and the bases of $n-$factorizations $col_R(k_1) \leftrightarrow_n col_R(k_4)$ and $col_L(k_3) \leftrightarrow_n col_R(k_2)$ are equal $U \leftrightarrow_n V$. Hence, with the pair $U \leftrightarrow_n V$ we can associate two sets $R, L \subset K$ such that for all $k \in R, \overline{k} \in L$ the key $\overline{k}$ follows k and the base of $n-$factorization $col_R(k) \leftrightarrow_n col_L(\overline{k})$ is equal $U \leftrightarrow_n V$.

Let us denote by $\mathcal{B}(D_1, ..., D_n)$ the set of all $n-$factorizations that occur as the base of $n-$factorization $col_R(k) \leftrightarrow col_L(\overline{k})$ for some $k, \overline{k} \in K$ such that $\overline{k}$ follows k. It may happen that the set R or L associated with an element $U \leftrightarrow_n V \in \mathcal{B}(D_1, ..., D_n)$ consists of exactly one element. Suppose that $L = \{l\}$ and $R = \{r_1, ..., r_m\}$ for some $l, r_1, ..., r_m \in K$. Note that in any generating key sequence the key l has to occur after any r_i whenever r_i occurs in a generating key sequence. Let us define for $i = 1, ..., n$

$$D_i^{'} = (D_i \setminus \{v_i(l), v_i(r_1), ..., v_i(r_m)\}) \cup \{v_i(r_1)v_i(l), ..., v_i(r_m)v_i(l)\},$$

where $v_i(k)$ for any $k \in K$ denotes key word in D_i with k as the key letter. Of course, for $i = 1, ..., n$ the set $D_i^{'}$ is a key code (fix the letter r_j as the key of word $v_i(l)v_i(r_j)$ for $j = 1, ..., m$). By the previous considerations $S = \bigcap_{i=1}^{n} D_i^{'}$. Note that the number of elements in $\mathcal{B}(D_1^{'}, ..., D_n^{'})$ relatively to $\mathcal{B}(D_1, ..., D_n)$ diminish to 1. We could repeat the following procedure in the case R consists of exactly one element. Hence, we can state:

Lemma 4.1. *Let $S = \bigcap_{i=1}^{n} D_i^*$ and let $D_1, ..., D_n$ be key codes with the same key set K. Then there exist key codes $E_1, ..., E_n$ such that*

1. *$S \subset E_i^* \subset D_i^*$ for $i = 1, ..., n$ (it means $S = \bigcap_{i=1}^{n} E_i^*$)*
2. *$key(E_1) = key(E_2) = ... = key(E_n)$*
3. *if $U \leftrightarrow_n V \in \mathcal{B}(E_1, ..., E_n)$ then the sets R, L associated with $U \leftrightarrow_n V$ have at least two members.*

Suppose now that $S = \bigcap_{i=1}^{n} E_i^*$ and the sequence $E_1, ..., E_n$ fulfills the properties listed in the previous lemma.

Definition 4.2. Let $U \leftrightarrow_n V \in \mathcal{B}(C_1, ..., C_n)$ be an $n-$factorization of the word $w_1 \in A^+$. Let $L, R \subset K$ be associated with $U \leftrightarrow_n V$. We say that $w_2 \in A^+$ separates R and L iff w_2 is the word of the maximal length containing w_1 and the equality

$$\{kbk(k, \overline{k})\overline{k} \mid k \in R, \overline{k} \in L\} = \{kright(k)w_2left(\overline{k})\overline{k} \mid k \in R, \overline{k} \in L\}$$

is true for some words $right(k), left(\overline{k}) \in A^*$. For any $k \in K$ the word $left(k)kright(k)$ is now defined and we denote this word by $root(k)$. Note that the word w_2 is properly defined. It may happened that $w_1 = w_2$ of course.

Let us fix the order of all members of the set $\mathcal{B}(E_1, ..., E_n)$ - $U_1 \leftrightarrow_n V_1, ..., U_m \leftrightarrow_n V_n$. Assume that sets $R_j, L_j \subset K$ are associated with the base $U_j \leftrightarrow_n V_j$ and denote the separating word for the pair R_j, L_j by sep_j. Note that the families $\{L_{init}, L_1, ..., L_m\}$ and $\{R_{final}, R_1, ..., R_m\}$ constitute the partitions of the set K. Note that by the previous lemma every set of those families except L_{init} or R_{final} has to contain at least 2 members.

Example 4.3.

$$\mathcal{B}(E_1, E_2, E_3) = \left\{ \begin{bmatrix} b & a \\ b & \\ & \end{bmatrix} \leftrightarrow_3 \begin{bmatrix} & \\ & a \\ b & a \end{bmatrix}, \begin{bmatrix} a & a \\ a & \\ & \end{bmatrix} \leftrightarrow_3 \begin{bmatrix} & \\ & a \\ a & a \end{bmatrix} \right\}.$$

$L_{init} = \{k_1\}$, $L_1 = \{k_2, k_4\}$, $L_2 = \{k_3, k_5\}$.
$R_{final} = \{k_3\}$, $R_1 = \{k_1, k_4\}$, $R_2 = \{k_2, k_5\}$.
The families $\{L_{init}, L_1, L_2\}$ and $\{R_{final}, R_1, R_2\}$, where R_1, L_1 and R_2, L_2 are associated respectively with the first and the second element of $\mathcal{B}(E_1, ..., E_n)$, form the partitions of the set K.
The word $aba \in A^+$ separates R_1 and L_1. The word aa separates R_2 and L_2.
The roots of k_1, k_2, k_3, k_4 and k_5 are equal respectively bak_1, k_2, k_3b, bk_4b, k_5.

Now we are ready to give the basic for our considerations lemma.

Lemma 4.4. *Let $S = \bigcap_{i=1}^{n} E_i^*$ be a semiretract such that the sequence of key codes $E_1, ..., E_n$ with a common key set K fulfills the conditions given in Lemma 4.1. Then, for any key code F with key set $\overline{K}$ such that $S \subset F^*$ there exists a key code G with K as the key set such that*

1. *$S \subset G^* \subset F^*$*

(2) *Let $k \in K$. Assume that if k is not final, then $k \in R_s$ for some $s \in \{1, ..., m\}$ and if k is not initial, then $k \in L_t$ for some $t \in \{1, ..., m\}$. If $v(k) \in G$ is the key word with $k \in K$ as the key letter, then $root(k)$ is a subword of $v(k)$. Moreover, if*
  (a) *k is initial and final key, then $v(k) = root(k)$,*
  (b) *k is initial and not final key, then $v(k)$ is a subword of $root(k)sep_t$,*
  (c) *k is initial and not final key, then $v(k)$ is a subword of $sep_s root(k)$,*
  (d) *k is not final and not initial key, then v is a subword of $sep_s root(k) sep_t$.*

Proof. Let us denote by $w(\overline{k})$ the key word in F with $\overline{k} \in \overline{K}$ as the key letter. For any $k \in K$ let $\overline{k_1}, ..., \overline{k_p} \in \overline{K}$ be the sequence of all keys that occur in $root(k)$. We denote the word $w(\overline{k_1})...w(\overline{k_p}) \in F^*$ by $root^F(k)$. Note that $root^F(k)$ is uniquely determined.

For any separating word sep_j let $\overline{k_1}, ..., \overline{k_p}$ be the sequence of all keys in $\overline{K}$ that occur in sep_j for $j = 1, ..., m$. We denote the word $w(\overline{k_1})...w(\overline{k_p}) \in F^*$ by sep_j^F. Note that sep_j^F is uniquely determined.

Let w be a word in the base of semiretract S and let $k_1, ..., k_p \in K$ be the generating key sequence for w. Let us consider the double factorization of the word w. Assume that for any $i = 1, ..., n$ the number $j_i \in \{1, ..., m\}$ is such that $U_{j_i} \leftrightarrow_n V_{j_i}$ is the base of $n-$factorization $col_R(k_i) \leftrightarrow_n col_L(k_{i+1})$. By Lemma 3.5 and by Definition 4.2.

$$w = root(k_1) sep_{j_1} root(k_2) sep_{j_2} .....sep_{j_{p-1}} root(k_p).$$

On the other hand, by $S \subset F^*$

$$w = root^F(k_1) sep_{j_1}^F root^F(k_2) sep_{j_2}^F .....sep_{j_{p-1}}^F root^F(k_p).$$

Since any set $R_1, L_1, ..., R_m, L_m$ has at least 2 elements, then the word $sep_{j_i}^F$ has to be a subword of sep_{j_i}. Hence the word $root^F(k_i)$ contains $root(k_i)$ as a subword. Since any letter $k \in K$ occurs in some word from the base of S, then the word $root(k)$ is a subword of $root^F(k)$ and for any $j \in \{1, ..., m\}$ the word sep_j contains sep_j^F as a subword.

Let $k \in K$. If k is not final, then assume that $k \in R_s$ for some $s \in \{1, ..., m\}$. If k is not initial, then assume that $k \in L_t$ for some $t \in \{1, ..., m\}$. For any $k \in K$ let $v(k)$ (with k as the key letter) denote the word

- $root^F(k)$ if k is initial and final,
- $root^F(k) sep_t^F$ if k is initial and not final,
- $sep_s^F root^D(k)$ if k is final and not initial,
- $sep_s^F root^F(k) sep_t^F$ if k is not initial and not final.

Then the key code

$$G = \{v(k) \mid k \in K\}$$

makes our theorem true. □

Definition 4.5. Let $w_1, ..., w_m \in A^+$ be a sequence of words and let $U(w_j) \leftrightarrow V(w_j)$ be an $l-$factorization of w_j for $j = 1, ..., m$. We say that the sequence $U(w_1) \leftrightarrow_l V(w_1), ..., U(w_m) \leftrightarrow_l V(w_m)$ constitute $l-$factorization of the sequence $w_1, ..., w_m$ if and only if the columns $U(w_i), V(w_j)$ for $i, j = 1, ..., m$ constitute $l-$factorization only if $i = j$.

Hence, the sequence $U_1 \leftrightarrow_n V_1, ..., U_m \leftrightarrow_n V_m$ forms $n-$factorization of the sequence $w_1, ..., w_m \in A^+$, where w_i is a subword of sep_i for $i = 1, ..., m$. As a consequence, there exists $n-$factorization of the sequence $sep_1, ..., sep_m$ (it is obtained by modifying a little bit the columns $U_1, V_1, ..., U_m, V_m$).

Suppose now that $dim(S) \leq l$. By definition $S = \bigcap_{i=1}^{l} F_i^*$ for some key codes $F_1, ..., F_l$. Since $S \subset F_i^*$, then by the previous lemma there exists key code G_i with the key set K such that $S \subset G_i^* \subset F_i^*$ for $i = 1, ..., l$. The form of any key word in G_i and the equality $S = \bigcap_{i=1}^{l} G_i^*$ imply, that there exist $l-$factorization of the sequence $sep_1, ..., sep_m$.

Suppose now that a sequence $X^1 \leftrightarrow_l Y^1, ..., X^m \leftrightarrow_l Y^m$ forms an $l-$factorization of the sequence $sep_1, ..., sep_m$. Assume that $k \in K$ is not initial and not final key. Then $k \in R_s$ and $k \in L_t$ for some $s, t \in \{1, ..., m\}$. Let us define l-key words with k as the key letters as follows (we use the matrix form):

$$A(k) = \begin{bmatrix} X_1^t left(k) & k & right(k) Y_1^s \\ \vdots & \vdots & \vdots \\ X_i^t left(k) & k & right(k) Y_i^s \\ \vdots & \vdots & \vdots \\ X_l^t left(k) & k & right(k) Y_l^s \end{bmatrix},$$

where X_i^t and Y_i^s for $i = 1, ..., l$ denote the entries in the $i-$th rows of columns X^t and Y^s respectively. In the case k is initial the left column of $A(k)$ consist entirely of $left(k)$ and in the case k is final the right column of $A(k)$ consist entirely of $right(k)$. It is not hard to verify that the intersection of l retracts with l key codes defined above is equal with S. As a consequence we have the following statement true.

Theorem 4.6. *Let $S = \bigcap_{i=1}^{n} E_i$, where the sequence of key codes $E_1, ..., E_n$ fulfills the conditions given in Lemma 4.5. Then, $dim(S) \leq l$ iff there exist $l-$factorization of the sequence $sep_1, ..., sep_m$.*

To verify if there exist an $l-$factorization of the sequence $sep_1, ..., sep_m$ let us consider a network $D = (V, A)$ with a capacity function $c : A \to \mathbb{N}$. Let $V = \{s, t\} \cup V_1 \cup V_2$ be the set of all vertices in a digraph $D = (V, A)$, where $s, t \in V$ are respectively the source and the sink of the network,

$$V_1 = \{sep_j |\ j \in \{1, ..., m\}\}$$

and

$$V_2 = \{w \mid w \text{ is a subword of some } sep_j,\ j \in \{1, ..., m\}\}.$$

Let

$$A = \{s, V_1\} \cup E \cup V_2 \times \{t\},$$

where $E \subset V_1 \times V_2$ is the set of edges defined as follows: $(v_1, v_2) \subset V_1 \times V_2$ is in E iff v_2 is a subword of v_1. Finally, we define the capacity function by the following rules:

- $c(s, v_1) = x$ for $(s, v_1) \in \{s\} \times V_1$ if the word v_1 occurs exactly x times in the sequence $sep_1, ..., sep_m$,
- $c(v_1, v_2) = \infty$ for $(v_1, v_2) \in E$,

- $c(v_2,t) = max(m, l(v_n))$ for $(v_2,t) \in \{v_2\} \times V_2$, where $l(v_2)$ is the number of all different $l-$factorization of the word v_2 with v_2 as the source. Since such an $l-$factorization of v_2 is fully determined by the left column of $l-$factorization, then

$$l(v_2) = \sum_{k_1,k_2 \geq 1, k_1+k_2 \leq l} \binom{l}{k_1}\binom{l-k_1}{k_2}(|v_2|-1)^{l-(k_1+k_2)},$$

where the term $\binom{l}{k_1}\binom{l-k_1}{k_2}(|v_2|-1)^{l-(k_1+k_2)}$ denotes the number of columns with exactly:
  - k_1 rows filled up with 1,
  - k_2 rows filled up with v_2,
  - $l-(k_1+k_2)$ rows filled up with nonempty, proper prefix of v_2.

Lemma 4.7. *There exist an $l-$factorization of the sequence $sep_1, ..., sep_m$ iff the maximal flow of the network $D = (V, A)$ with the capacity function $c : E \to \mathcal{N}$ is equal m.*

Proof. Let $U_1 \leftrightarrow_l V_1, ..., U_m \leftrightarrow_l V_m$ be an $l-$factorization of the sequence $sep_1, ..., sep_m$ with the sources respectively $w_1, ..., w_n$. Let us consider the function $f : A \to \mathbb{N}$ defined as follows:

- $f(s,v_1) = c(s,v_1)$ for $(s,v_1) \in \{s\} \times V_1$,
- $f(v_1,v_2) = x$ for $(v_1,v_2) \in E$ if the pair (v_1,v_2) occurs x time in the sequence $(sep_1,w_1), ..., (sep_m,w_m)$,
- $f(v_2,t) = y$ for $(v_2,t) \in V_2 \times \{t\}$ if the word v_2 occurs in the sequence $w_1, ..., w_m$ exactly y times.

We can easily check that f satisfy the conservation and feasibility rules and hence f is a flow function with the flow value m. By the max-flow min-cut theorem for the cut $(\{s\}, V \setminus \{s\})$ with the capacity m we conclude that f is the maximal flow in the network.

Suppose now that $f : A \to \mathbb{N}$ is a maximal flow function in the network and the flow value is m. Let $v_1 \in V_1$. Since the cut $(\{s\}, V \setminus \{s\})$ has the capacity m, then $f(s,v_1) = c(s,v_1) = x$ for some $x \in \mathbb{N}$. Thus, the word v_1 occurs on the list $sep_1, ..., sep_m$ exactly x times. Assume, that $j_1, ..., j_x \in \{1, ..., m\}$ are such that $sep_{j_i} = v_1$ for $i = 1, ..., x$. Hence, by the conservation rule for the vertex v_1 there exists a list $L(v_1) = w_{j_1}, ..., w_{j_k}$ such that w_{j_i} is the subword of $v_1 = sep_{j_i}$ and any word $v_2 \in L(v_1)$ occurs on the list $L(v_1)$ exactly $f(v_1,v_2)$ times. Hence, with any separating word sep_{j_i} we can associate a subword w_{j_i} for all $i = 1, ..., x$. Repeating this step for any vertex $v_1 \in V_1$ we obtain a sequence $w_1, ..., w_m$ such that w_i is associated with sep_i for $i = 1, ..., m$.

Let us consider any w_i for $i = 1, ..., m$ and assume that w_i occurs exactly y ($y \in \mathbb{N}$) times on the list $w_1, ..., w_m$. Suppose that $w_i = w_{k_1} = ... = w_{k_y}$ for some $k_1, ..., k_y \in \{1, ..., m\}$. The conservation rule for the vertex $w_i \in V_2$ and the feasibility rule for the edge (w_i, t) asserts that we can find y different $l-$factorizations of the word w_i; let us denote them by $U_{k_1} \leftrightarrow_l V_{k_1}, ..., U_{k_y} \leftrightarrow_l V_{k_y}$. Repeating this step for any $w_i \in \{w_1, ..., w_m\}$ we obtain a sequence of $l-$factorizations $U_1 \leftrightarrow_l V_1, ..., U_m \leftrightarrow_l V_m$, where $U_j \leftrightarrow V_j$ is an $l-$factorization of w_j for $j = 1, ..., m$. Note that if $U^1 \leftrightarrow_l V^1$ and $U^2 \leftrightarrow_l V^2$ form $l-$factorizations with different source words, then U^1, V^2 and U^2, V^1 as well does not form $l-$factorization. It follows that the sequence $U_1 \leftrightarrow_l V_1, ..., U_m \leftrightarrow_l V_m$ forms the $l-$factorization of the sequence

$w_1, ..., w_m$. Thus, sine w_i is a subword of sep_i for $i = 1, ..., m$, then there exists an $l-$factorization of the sequence $sep_1, ..., sep_m$.

□

Assume that $(C_1, ..., C_m, l) \in \mathcal{I}$. Then $S = \bigcap_{i=1}^{n} C_i^*$. Then we compute the sequence of key codes $E_1, ..., E_n$ that satisfy the properties listed in the Lemma 4.1. Next, we produce the sequence $sep_1, ..., sep_m$ of all separating word. We refer to [2] to show that the list $sep_1, ..., sep_m$ can be computed in polynomial time. After all, for the sequence $sep_1, ..., sep_m$ we construct the network as presented above. The instance $(C_1, ..., C_m, l) \in DIM - SEM$ iff the maximal flow in the network is equal m. Since $MAX - FLOW$ is in P, then $DIM - SEM$ is also in P.

5. Problem $MIN - SEM$ is NP-complete.

The problem $MIN - SEM$ is in NP. For any $(C_1, ..., C_n, l) \in \mathcal{I}$ a nondeterministic Turing machine indicates l key codes $C_{i_1}, ..., C_{i_l} \in \{C_1, ..., C_n\}$ for some $i_1, ..., i_l \in \{1, ..., m\}$. Next, it constructs minimal, deterministic automatons A_1, A_2 that recognize the base of semiretracts $\bigcap_{i=1}^{n} C_i^*$ and $\bigcap_{j=1}^{l} C_{i_j}^*$ respectively. Finally, it tests if $A_1 = A_2$. In [2] the polynomial time algorithm for constructing minimal, deterministic automatons that recognizes the base of semiretract is presented. Finally, we can test if $A_1 = A_2$ in polynomial time.

We prove that $3 - SAT \leq_P MIN - RET$. Let $\{x_1, ..., x_p\}$ be the set of all variables that occur in the formula $\alpha = \bigwedge_{j=1}^{m} \alpha_j$, where $\alpha_j \equiv \alpha_j^1 \vee \alpha_j^2 \vee \alpha_j^3$, $j = 1, ..., m$. The transformation $\mathcal{T}$, for given formula α, produces $2p$ key codes $C_{x_1}, C_{\neg x_1}, ..., C_{x_p}, C_{\neg x_p}$ and the special key code denoted by C_s. We will prove that α is satisfiable iff $(C_{x_1}, C_{\neg x_1}, ..., C_{x_p}, C_{\neg x_p}, C_s, p+1)$ is in $MIN - SEM$. Let us describe the transformation $\mathcal{T}(\alpha)$.

All key codes $C_{x_1}, C_{\neg x_1}, ..., C_{x_p}, C_{\neg x_p}$ have the same key set

$$K = \{f, h, x_1, ..., x_p, \alpha_1, ..., \alpha_m\}$$

and are defined over the alphabet

$$A = K \cup \{h', x_1', ..., x_p', \alpha_1', ..., \alpha_m'\}.$$

Let us fix the order $C_{x_1}, C_{\neg x_1}, ..., C_{x_p}, C_{\neg x_p}, C_s$ of all key codes. We define any key code by giving all columns $col_L(k), col_R(k)$ for any $k \in K$ with respect to the order $C_{x_1}, C_{\neg x_1}, ..., C_{x_p}, C_{\neg x_p}, C_s$.

For any key x_i, $i = 1, ..., p$ associated with the variable x_i, we define $col_R(x_i)$ putting x_i' at the positions that correspond to key codes C_{x_i} and $C_{\neg x_i}$ and putting 1 at the other positions. For any key α_j, $j = 1, ..., m$ associated with the clause α_j we define $col_R(\alpha_j)$ putting α_j' at the positions that correspond to the key codes $C_{\alpha_j^1}$, $C_{\alpha_j^2}$ and $C_{\alpha_j^3}$ and putting 1 at the other positions. For the key $h \in K$ we define $col_R(h)$ putting h' at the position that correspond to the key code C_s and putting 1 at the other positions. To make x_1 the one, initial key and f the one, final key we define $col_L(x_1)$ and $col_R(f)$ putting 1 on any positions. The columns $col_L(x_2), ..., col_L(x_p)$, $col_L(\alpha_1)$,....,$col_R(\alpha_m)$, $col_L(h)$ and $col_L(f)$ are defined such that the sequence of keys

$$(x_1, x_2, ..., x_p, \alpha_1, \alpha_2, ..., \alpha_m, h, f)$$

is the only one possible generating key sequence. By Lemma 3.5, the base of semiretracts consists of exactly one word, namely

$$x_1x_1'x_2x_2'....x_px_p'\alpha_1\alpha_1'\alpha_2\alpha_2'....\alpha_m\alpha_m'hh'f.$$

Example 5.1. Let

$$\phi \equiv (x_1 \vee \neg x_2 \vee x_3) \wedge (\neg x_1 \vee x_2 \vee \neg x_3) \wedge (\neg x_1 \vee \neg x_2 \vee \neg x_3).$$

The set of all variables is equal to $\{x_1, x_2, x_3\}$. Hence we define key codes C_{x_1}, $C_{\neg x_1}$, C_{x_2}, $C_{\neg x_2}$, C_{x_3}, $C_{\neg x_3}$, C_s with the same set of keys $K = \{f, h, x_1, x_2, x_3, \alpha_1, \alpha_2, \alpha_3\}$ over the alphabet $K \cup \{h', y_1, y_2, y_3, a_1, a_2, a_3\}$. Key codes C_{x_1}, $C_{\neg x_1}$, C_{x_2}, $C_{\neg x_2}$, C_{x_3}, $C_{\neg x_3}, C_s$ are presented in the matrix form:

$$\begin{array}{cc} x_1 & - \\ \neg x_1 & - \\ x_2 & - \\ \neg x_2 & - \\ x_3 & - \\ \neg x_3 & - \\ s & - \end{array}
\begin{bmatrix} 1 & x_1 & x_1' \\ 1 & x_1 & x_1' \\ 1 & x_1 & 1 \\ 1 & x_1 & 1 \\ 1 & x_1 & 1 \\ 1 & x_1 & 1 \\ 1 & x_1 & 1 \end{bmatrix},
\begin{bmatrix} 1 & x_2 & 1 \\ 1 & x_2 & 1 \\ x_1' & x_2 & x_2' \\ x_1' & x_2 & x_2' \\ x_1' & x_2 & 1 \\ x_1' & x_2 & 1 \\ x_1' & x_2 & 1 \end{bmatrix},
\begin{bmatrix} x_2' & x_3 & 1 \\ x_2' & x_3 & 1 \\ 1 & x_3 & 1 \\ 1 & x_3 & 1 \\ x_2' & x_3 & x_3' \\ x_2' & x_3 & x_3' \\ x_2' & x_3 & 1 \end{bmatrix},
\begin{bmatrix} x_3' & \alpha_1 & \alpha_1' \\ x_3' & \alpha_1 & 1 \\ x_3' & \alpha_1 & 1 \\ x_3' & \alpha_1 & \alpha_1' \\ 1 & \alpha_1 & \alpha_1' \\ 1 & \alpha_1 & 1 \\ x_3' & \alpha_1 & 1 \end{bmatrix},$$

$$\begin{array}{cc} x_1 & - \\ \neg x_1 & - \\ x_2 & - \\ \neg x_2 & - \\ x_3 & - \\ \neg x_3 & - \\ s & - \end{array}
\begin{bmatrix} 1 & \alpha_2 & 1 \\ \alpha_1' & \alpha_2 & \alpha_2' \\ \alpha_1' & \alpha_2 & \alpha_2' \\ 1 & \alpha_2 & 1 \\ 1 & \alpha_2 & 1 \\ \alpha_1' & \alpha_2 & \alpha_2' \\ \alpha_1' & \alpha_2 & 1 \end{bmatrix},
\begin{bmatrix} \alpha_2' & \alpha_3 & 1 \\ 1 & \alpha_3 & \alpha_3' \\ 1 & \alpha_3 & 1 \\ \alpha_2' & \alpha_3 & \alpha_3' \\ \alpha_2' & \alpha_3 & 1 \\ 1 & \alpha_3 & \alpha_3' \\ \alpha_2' & \alpha_3 & 1 \end{bmatrix},
\begin{bmatrix} \alpha_3' & h & 1 \\ 1 & h & 1 \\ \alpha_3' & h & 1 \\ 1 & h & 1 \\ \alpha_3' & h & 1 \\ 1 & h & 1 \\ \alpha_3' & h & h' \end{bmatrix},
\begin{bmatrix} h' & f & 1 \\ h' & f & 1 \\ h' & f & 1 \\ h' & f & 1 \\ h' & f & 1 \\ h' & f & 1 \\ 1 & f & 1 \end{bmatrix}.$$

The sequence $(x_1, x_2, x_3, \alpha_1, \alpha_2, \alpha_3, h, f)$ is the only one possible generating key sequence. It follows that in the base of semiretract S there is exactly one word, namely

$$x_1x_1'x_2,x_2'x_3x_3'\alpha_1\alpha_1'\alpha_2\alpha_2'\alpha_3\alpha_3'hh'f.$$

Assume that the formula α is satisfiable by an assignment $l_1 = TRUE, ..., l_p = TRUE$, where l_j for all $j = 1, ..., m$ is a literal from the set $\{x_j, \neg x_j\}$. Let us fix the order of key codes $C_{l_1}, ..., C_{l_p}, C_s$. Note that $col_L(x_i)$ for $i = 1, ..., p$ relatively to the order $C_{l_1}, ..., C_{l_p}, C_s$ contains elements x_i and 1 at the positions that corresponds to the key codes C_{l_i} and C_s respectively. Quite similar, $col_L(\alpha_j)$, $j = 1, ..., m$ relatively to the order $C_{l_1}, ..., C_{l_p}, C_s$ contains elements α_j' at the position that corresponds to the key code indexed by the literal that makes clause α_j true and contains 1 at position that corresponds to the key code C_s. Since elements $x_1', ..., x_p', \alpha_1', ..., \alpha_m', h'$ are pairwise different then the only possible key sequence in semiretract generated by $C_{l_1}, ..., C_{l_p}, C_s$ is still $(x_1, ..., x_p, \alpha_1, ..., \alpha_m, h, f)$. It follows that

$$(\bigcap_{i=1}^{p} C^*_{x_i} \cap C^*_{\neg x_i}) \cap C^*_s = (\bigcap_{i=1}^{p} C^*_{l_i}) \cap C^*_s$$

and hence $(C_{x_1}, C_{\neg x_1}, ..., C_{x_p}, C_{\neg x_p}, C_s, p+1)$ is in $MIN - SEM$.

Let $(C_{x_1}, C_{\neg x_1}, ..., C_{x_p}, C_{\neg x_p}, C_s, p+1)$ in $MIN - SEM$ and assume that

$$C_{l_1}, ..., C_{l_p}, C_{l_{p+1}} \in \{C_{x_1}, C_{\neg x_1}, ..., C_{x_p}, C_{\neg x_p}, C_s\}$$

for some $l_1, ..., l_{p+1} \in \{x_1, \neg x_1, ..., x_p, \neg x_p, s\}$ are such that the equality

$$(\bigcap_{i=1}^{p} C^*_{x_i} \cap C^*_{\neg x_i}) \cap C^*_s = \bigcap_{i=1}^{p+1} C^*_{l_i}$$

is true. Since $f \in A^*$ is not in the base of semiretract $\bigcap_{i=1}^{p+1} C^*_{l_i}$ (more precisely, since f is not final key), then C_s has to be in $\{C_{l_1}, ..., C_{l_{p+1}}\}$. Assume that $C_s = C_{l_{p+1}}$. Since the column $col_R(x_i)$ for all $i = 1, ..., p$ relatively to the order $C_{l_1}, ..., C_{l_p}, C_s$ has to contain $x_i^{'}$ (x_i is not a final key) at some position, then C_{x_i} or $C_{\neg x_i}$ is in the set $C_{l_1}, ..., C_{l_p}$. It follows that an assignment $l_1 = TRUE, ..., l_p = TRUE$ is well defined. Quite similar, the column $col_R(\alpha_j)$ for all $j = 1, ..., m$ with respect to the order $C_{l_1}, ..., C_{l_p}, C_s$ has to contain $\alpha_j^{'}$ at some position, exactly at positions that corresponds to key codes $C_{\alpha_j^1}$, $C_{\alpha_j^2}$ or $C_{\alpha_1^3}$. Hence, there exist a literal $l \in \{l_1, ..., l_p\}$ that makes the clause $\alpha_j \equiv \alpha_j^1 \vee \alpha_j^1 \vee \alpha_j^3$ true. Hence, α is satisfiable.

Example 5.2. Formula ϕ is satisfiable by the assignment

$$x_1 = TRUE, x_2 = TRUE, \neg x_3 = FALSE.$$

Let us consider blocks A_k for all $k \in K$ relatively to $C_{x_1}, C_{x_2}, C_{\neg x_3}, C_s$:

$$\begin{array}{cc} x_1 & - \\ x_2 & - \\ \neg x_3 & - \\ s & - \end{array} \begin{bmatrix} 1 & x_1 & x_1^{'} \\ 1 & x_1 & 1 \\ 1 & x_1 & 1 \\ 1 & x_1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & x_2 & 1 \\ x_1^{'} & x_2 & x_2^{'} \\ x_1^{'} & x_2 & 1 \\ x_1^{'} & x_2 & 1 \end{bmatrix}, \begin{bmatrix} x_2^{'} & x_3 & 1 \\ 1 & x_3 & 1 \\ x_2^{'} & x_3 & x_3^{'} \\ x_2^{'} & x_3 & 1 \end{bmatrix}, \begin{bmatrix} x_3^{'} & \alpha_1 & \alpha_1^{'} \\ x_3^{'} & \alpha_1 & 1 \\ 1 & \alpha_1 & 1 \\ x_3^{'} & \alpha_1 & 1 \end{bmatrix},$$

$$\begin{array}{cc} x_1 & - \\ x_2 & - \\ \neg x_3 & - \\ s & - \end{array} \begin{bmatrix} 1 & \alpha_2 & 1 \\ \alpha_1^{'} & \alpha_2 & \alpha_2^{'} \\ \alpha_1^{'} & \alpha_2 & \alpha_2^{'} \\ \alpha_1^{'} & \alpha_2 & 1 \end{bmatrix}, \begin{bmatrix} \alpha_2^{'} & \alpha_3 & 1 \\ 1 & \alpha_3 & 1 \\ 1 & \alpha_3 & \alpha_3^{'} \\ \alpha_2^{'} & \alpha_3 & 1 \end{bmatrix}, \begin{bmatrix} \alpha_3^{'} & h & 1 \\ \alpha_3^{'} & h & 1 \\ 1 & h & 1 \\ \alpha_3^{'} & h & h^{'} \end{bmatrix}, \begin{bmatrix} h^{'} & f & 1 \\ h^{'} & f & 1 \\ h^{'} & f & 1 \\ 1 & f & 1 \end{bmatrix}.$$

According to the previous considerations the key x_1 is still the one initial key, f is still the one final key and key sequence $(x_1, x_2, x_3, \alpha_1, \alpha_2, \alpha_3, h, f)$ is the one possible key sequence relatively to the order $C_{x_1}, C_{x_2}, C_{\neg x_1}, C_s$.

References

[1] W.Forys, T.Krawczyk, J.A.Anderson, Semiretracts - a counterexample and some results, Theoretical Computer Science, 307, 2003

[2] W.Forys, T.Krawczyk, The algorithmic approach to ... , 2005

[3] J.A.Anderson, The intersection of retracts of A^*,, Theoretical Computer Science, 237, 2000

[4] J.A.Anderson, Code properties of minimal generating sets of retracts and semiretracts, SEA Bull.Math, 18, 1994