Sets of prohibitions and some problems related to them

Sergey Kitaev

Reykjavík University

Let $\mathbf{A} = \{a_1, \ldots, a_n\}$ be an alphabet of n letters.

A word in the alphabet $\mathbf{A}$ is a finite sequence of letters of the alphabet.

Any i consecutive letters of a word X generate a factor of length i.

The set $\mathbf{A}^*$ is the set of all the words on the alphabet $\mathbf{A}$.

Let Σ be an alphabet.

A map $\varphi : \Sigma^* \to \Sigma^*$ is called a morphism, if we have

$$\varphi(uv) = \varphi(u)\varphi(v)$$

for any $u, v \in \Sigma^*$.

A morphism φ can be defined by defining $\varphi(i)$ for each $i \in \Sigma$.

Fix a set $\mathbf{S} \subseteq \mathbf{A}^*$ and call it a set of prohibited words.

A word that does not contain any word from $\mathbf{S}$ as its factor is called free from $\mathbf{S}$.

The set of all words that are free from $\mathbf{S}$ is denoted by $\widehat{\mathbf{S}}$.

If $\widehat{\mathbf{S}}$ has finitely many words in $\mathbf{A}^*$, then the set of prohibitions $\mathbf{S}$ is called an unavoidable set.

$\mathbf{A} = \{1, 2\}$;

$\mathbf{S} = \{111, 21, 22\}$ is an unavoidable set since

$\widehat{\mathbf{S}} = \{1, 11, 12, 112, 2\}$.

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 factor for any letter $a_i \in \mathbf{A}$.

A crucial word of minimal (maximal) length, if it exists, is called a minimal (maximal) crucial word.

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 factor for any letter $a_i \in \mathbf{A}$.

A crucial word of minimal (maximal) length, if it exists, is called a minimal (maximal) crucial word.

$\mathbf{S} = \{111, 21, 22\}$ is an unavoidable set of prohibitions;

$\widehat{\mathbf{S}} = \{1, 11, 12, 112, 2\}$.

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 factor for any letter $a_i \in \mathbf{A}$.

A crucial word of minimal (maximal) length, if it exists, is called a minimal (maximal) crucial word.

$\mathbf{S} = \{111, 21, 22\}$ is an unavoidable set since

$\widehat{\mathbf{S}} = \{1, 11, 12, 112, 2\}$.

Let $L_{\text{min}}(\mathbf{S})$ ($L_{\text{max}}(\mathbf{S})$) denote the length of a minimal (maximal) crucial word with respect to $\mathbf{S}$.

$\mathbf{S}_1^n = \{XX \mid X \in \{a_1, ..., a_n\}^*\}$, that is, we prohibit the repetition of two equal consecutive factors. We prohibit squares.

$a_2 a_2 a_1 a_2 a_1 a_1 a_2 \notin \widehat{\mathbf{S}}_1^2$;

$a_1 a_2 a_4 a_1 a_2 a_3 \in \widehat{\mathbf{S}}_1^4$.

$\mathbf{S}_1^n = \{XX \mid X \in \{a_1, ..., a_n\}^*\}$, that is, we prohibit the repetition of two equal consecutive factors. We prohibit squares.

$a_2 a_2 a_1 a_2 a_1 a_1 a_2 \notin \widehat{\mathbf{S}}_1^2$;

$a_1 a_2 a_4 a_1 a_2 a_3 \in \widehat{\mathbf{S}}_1^4$.

Theorem. [Fraenkel, Simpson (1998); Ilie (2005)] A word of length n has at most $2n$ distinct squares.

Question. [Fraenkel, Simpson (1998)] Is the number of distinct squares in a word of length n at most n?

A semigroup is a set S of elements $a, b, c, ...$ in which an associative operation $\cdot$ is defined.

The element z is a zero element if $z \cdot a = a \cdot z = z$ for all a in S.

Let S be a semigroup generated by three elements, such that the square of every element in S is zero (thus $a \cdot a = z$ for all a in S).

Does S have an infinite number of elements?

Thue (1906) Arshon (1937) Morse (1938)

Theorem. [SK, 1996] We have $L_{\mathsf{min}}(\mathbf{S}_1^n) = 2^n - 1$.

We define a crucial word X by induction:

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

Theorem. [SK, 1996] We have $L_{\text{min}}(\mathbf{S}_1^n) = 2^n - 1$.

We define a crucial word X by induction:

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

There is a generalization of the theorem:

Theorem. [R. Travkin, 2005] If $\mathbf{S} = \{X^k \mid X \in \{a_1, ..., a_n\}^*\}$ then $L_{\text{min}}(\mathbf{S}) = k^n - 1$.

We define a crucial word X by induction:

$$X_1 = a_1^{k-1},\ X_i = (X_{i-1} a_i)^{k-1} X_{i-1},\ X = X_n.$$

The $\overline{\nu}$(X) = $(\nu_1(X), \ldots, \nu_n(X))$ is the content vector of X, if $\nu_i(X)$ is the number of occurrences of the letter a_i in X.

$\overline{\nu}(a_1a_3a_1a_1a_2a_3a_1) = (4, 1, 2)$;

The $\overline{\nu}(X) = (\nu_1(X), \ldots, \nu_n(X))$ is the content vector of X, if $\nu_i(X)$ is the number of occurrences of the letter a_i in X.

$\overline{\nu}(a_1a_3a_1a_1a_2a_3a_1) = (4, 1, 2)$;

$\mathbf{S}_2^n = \{XY \mid \overline{\nu}(X) = \overline{\nu}(Y)\}$. That is, we prohibit the repetition of two consecutive factors of the same content. We prohibit abelian squares.

In each word from $\widehat{\mathbf{S}}_2^n$, no two adjacent factors are permutations of each other.

$131232123 \notin \widehat{\mathbf{S}}_2^3$;

$1232413 \in \widehat{\mathbf{S}}_2^4$.

In 1961, Erdös asked whether $\mathbf{S}_2^4$ is unavoidable.

$\mathbf{S}_2^{25}$ – Evdokimov (1968)

$\mathbf{S}_2^{7}$ – Evdokimov (1971)

$\mathbf{S}_2^{5}$ – Pleasants (1970)

In 1961, Erdös asked whether $\mathbf{S}_2^4$ is unavoidable.

$\mathbf{S}_2^{25}$ – Evdokimov (1968)

$\mathbf{S}_2^7$ – Evdokimov (1971)

$\mathbf{S}_2^5$ – Pleasants (1970)

This was answered in the affirmative in 1992 by Keränen.

$|f(a)| = |f(b)| = |f(c)| = |f(d)| = 85$

$f(a) = abcacdcbcdcadcdbdabacabadbabcbdbcbacbcdcacbabdabacadcb$ -
$cdcacdbcbacbcdcacdcbdcdadbdcbca$

$f(b) = bcdbdadcdadbadacabcbdbcbacbcdcacdcbdcdadbdcbcabcbdbadc$ -
$dadbdacdcbdcdadbdadcadabacadcdb$

$f(c) = cdacabadabacbabdbcdcacdcbdcdadbdadcadabacadcdbcdcacbad$ -
$abacabdadcadabacabadbabcbdbadac$

$f(d) = dabdbcbabcbdcbcacdadbdadcadabacabadbabcbdbadacdadbdcba$ -
$bcbdbcabadbabcbdbcbacbcdcacbabd$

A natural approach to a construction of a crucial word is:

We add to a crucial word of an n-letter alphabet a minimum number of letters to obtain a crucial word of an $(n+1)$-letter alphabet.

This gives a crucial word of length $(3-(n \bmod 2))2^{\lfloor \frac{n+1}{2} \rfloor}-3$.

A natural approach to a construction of a crucial word is:

We add to a crucial word of an n-letter alphabet a minimum number of letters to obtain a crucial word of an $(n+1)$-letter alphabet.

This gives a crucial word of length $(3-(n \bmod 2))2^{\lfloor\frac{n+1}{2}\rfloor}-3$.

Theorem. [SK, 1996] For any $n>2$ we have $L_{\mathrm{min}}(\mathbf{S}_2^n)=4n-7$.

Construction of a minimal crucial word:

$a_{n-2}a_{n-1}a_{n-3}a_{n-2}\cdots a_1a_2$

A natural approach to a construction of a crucial word is:

We add to a crucial word of an n-letter alphabet a minimum number of letters to obtain a crucial word of an $(n+1)$-letter alphabet.

This gives a crucial word of length $(3-(n \bmod 2))2^{\lfloor \frac{n+1}{2} \rfloor}-3$.

Theorem. [SK, 1996] For any $n>2$ we have $L_{\text{min}}(\mathbf{S}_2^n)=4n-7$.

Construction of a minimal crucial word:

$a_{n-2}a_{n-1}a_{n-3}a_{n-2}\cdots a_1a_2a_na_{n-2}a_{n-3}\cdots a_2a_1a_2\cdots a_{n-3}a_{n-2}a_n$

The construction might be helpful in the following problems:

- Find the length of a minimal crucial word in the case of abelian k-th power-free words.

The construction might be helpful in the following problems:

- Find the length of a minimal crucial word in the case of abelian k-th power-free words.

- Suppose $\ell_{\min}(n)$ is the length of the minimal two-sided crucial word over an n-letter alphabet. Improve the following bounds:

  [M. Korn (2003)] $4n - 7 \leqslant \ell_{\min}(n) \leqslant 6n - 10$ for $n \geqslant 6$

  [E. Bullock (2004)] $6n - 29 \leqslant \ell_{\min}(n) \leqslant 6n - 12$ for $n \geqslant 8$

  Of course, the second bounds are better than the first ones.

$\mathbf{S}_3^{n,k_1,\dots,k_n} = \{X \mid \nu_i(X) \equiv 0 \pmod{k_i},\ k_i \in \mathbf{N},\ i = 1,\dots,n\}$, that is, we prohibit words in which the number of letters a_i is congruent to zero modulo k_i for each $i = 1,\dots,n$.

$2123312331 \notin \widehat{\mathbf{S}}_3^{3,2,3,2}$;

$12122111 \in \widehat{\mathbf{S}}_3^{2,3,4}$.

Theorem. [SK, 1998] We have $L_{min}(\mathbf{S}_3^{n,k_1,\ldots,k_n}) = \sum_{i=1}^{n} k_i - 1.$

An optimal construction:

$$a_{n-1}\underbrace{a_n\ldots a_n}_{k_n-1}a_{n-2}\underbrace{a_{n-1}\ldots a_{n-1}}_{k_{n-1}-1}\ldots a_1\underbrace{a_2\ldots a_2}_{k_2-1}\underbrace{a_1\ldots a_1}_{k_1-1}$$

Theorem. [SK, 1998] We have $L_{min}(\mathbf{S}_3^{n,k_1,\ldots,k_n}) = \sum_{i=1}^{n} k_i - 1.$

An optimal constructions:

$$a_{n-1}\underbrace{a_n\ldots a_n}_{k_n-1}a_{n-2}\underbrace{a_{n-1}\ldots a_{n-1}}_{k_{n-1}-1}\ldots a_1\underbrace{a_2\ldots a_2}_{k_2-1}\underbrace{a_1\ldots a_1}_{k_1-1}$$

Theorem. [SK, 1998] We have $L_{max}(\mathbf{S}_3^{n,k_1,\ldots,k_n}) = \prod_{i=1}^{n} k_i - 1.$

Let $d(X,Y)$ be the number of letters in which the words X and Y differ (Hamming metric).

$d(10212, 21213) = 3$;

Let $d(X,Y)$ be the number of letters in which the words X and Y differ (Hamming metric).

$d(10212, 21213) = 3$;

$\mathbf{S}_4^{n,k} = \{XY \mid d(X,Y) \leqslant k, |X| = |Y| \geqslant k+1, k \in \mathbf{N}\}$, where $|X|$ is the length of the word X. That is we prohibit any two consecutive subwords X, Y of length greater then k such that $d(X,Y) \leqslant k$.

$313221223 \notin \widehat{\mathbf{S}}_4^{3,2}$;

$101121231 \in \widehat{\mathbf{S}}_4^{3,3}$.

Theorem. [SK, 1998] We have $L_{min}(\mathbf{S}_4^{n,k}) = 2k + 1.$

Theorem. [SK, 1998] We have $L_{max}(\mathbf{S}_4^{2,k}) = 3k + 3.$

Theorem. [SK, 1998] The set of prohibitions $\mathbf{S}_4^{n,k}$ for $n \geqslant 3$ is avoidable.

$\mathbf{A} = \{1, 2, 3\}$; $\mathbf{B} = \{a, b, c\}$;

L is built by iteration of morphisms:

$$a \to abc, \quad b \to ac, \quad c \to b;$$

a, abc, $abcacb$, $abcacbabcbac$, ...

We define the mapping f as follows:

$$a \to \underbrace{1 \ldots 1}_{k+1}, \; b \to \underbrace{2 \ldots 2}_{k+1}, \; c \to \underbrace{3 \ldots 3}_{k+1}.$$

Then $f(L)$ is free from $\mathbf{S}_4^{n,k}$.

The Basic Problem:

Given: A set of prohibitions $\mathbf{S}$.

The question: Is $\mathbf{S}$ unavoidable?

Other Questions about $\widehat{\mathbf{S}}$:

If $\mathbf{S}$ is unavoidable then:

Find or estimate $L(\widehat{\mathbf{S}}) = \max_{X \in \widehat{\mathbf{S}}} |X|$;

Construct a word of length $L(\widehat{\mathbf{S}})$ that is free from $\mathbf{S}$;

$|\widehat{\mathbf{S}}| = ?$

If $\mathbf{S}$ is avoidable then:

Find a sequence that is free from $\mathbf{S}$;

Describe all such sequences;

Find the cardinality of the set of these sequences.

$\mathbf{A} = \{0, 1\}$;

$\mathbf{S_1} = \{000, 001, 101\underline{1}, 0101, 1111\}$

$\mathbf{S_2} = \{000, 001, 101\underline{0}, 0101, 1111\}$

$\mathbf{S_1}$ is unavoidable; $L(\widehat{\mathbf{S}_1}) = 8$; 01110100.

$\mathbf{A} = \{0, 1\}$;

$\mathbf{S_1} = \{000, 001, 101\underline{1}, 0101, 1111\}$

$\mathbf{S_2} = \{000, 001, 101\underline{0}, 0101, 1111\}$

$\mathbf{S_1}$ is unavoidable; $L(\widehat{\mathbf{S}_1}) = 8$; 01110100.

$\mathbf{S_2}$ is avoidable since

$\underbrace{011}\underbrace{011}\ldots$ is free from $\mathbf{S_2}$;

$\underbrace{0111}\underbrace{0111}\ldots$ is free from $\mathbf{S_2}$;

$\underbrace{011}\underbrace{0111}$ and $\underbrace{0111}\underbrace{011}$ are free from $\mathbf{S_2}$.

So taking **any** sequence in the alphabet $\mathbf{A}$ and substituting all 0s with 011 and all 1s with 0111 we get a sequence that is free from $\mathbf{S_2}$ whence the cardinality of $\widehat{\mathbf{S}_2}$ is the **continuum**!

In general it is important to know criteria of unavoidability for different classes of sets of prohibited words. An example of such a class can be the following:

$$\{\mathbf{S} : \mathbf{S} \subseteq \mathbf{A}^n\};$$

that is, we are interested in words of length n. It is interesting to know numerical characteristics that describe the bound between unavoidable and avoidable sets of prohibitions. For instance, let

$$M(n) = \min |\mathbf{S}| \quad \text{and} \quad L(n) = \max L(\widehat{\mathbf{S}}),$$

where the extremum is taken with respect to all unavoidable $\mathbf{S} \subseteq \mathbf{A}^n$ and $L(\widehat{\mathbf{S}}) = \max\limits_{X \in \widehat{\mathbf{S}}} |X|$.

Theorem. [Evdokimov, 1983] We have

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

Theorem. [Evdokimov, 1983] We have

$$M(n) = \frac{1}{n} \sum_{d|n} \varphi(n/d) |\mathbf{A}|^d,$$

where $\varphi(n)$ is the number of integers among $1, 2, \ldots n-1$ that are relatively prime to n (Euler's φ-function).

Since any set of prohibitions $\mathbf{S}$ for $|\mathbf{S}| \leqslant M(n)$ is avoidable, it is helpful to have a table for $M(n)$. If $|\mathbf{A}| = 2$ and $1 < n \leqslant 10$ then we have

n	2	3	4	5	6	7	8	9	10
M(n)	3	4	6	8	14	20	36	60	108

In particular, any set of binary words of length 10 that has less than 108 words is avoidable.

$M(n) \sim |\mathbf{A}|^n/n$, when $n \to \infty$.

The definitions of unavoidability and the function $L(n)$ immediately lead to the following.

Proposition. A set of prohibitions $\mathbf{S} \subseteq \mathbf{A}^n$ is complete iff any word of length $L(n) + 1$ is not free from $\mathbf{S}$.

So in order to decide if a given set of prohibitions $\mathbf{S}$ is complete or not, we can consider all words of length $L(n)+1 = |\mathbf{A}|^{n-1}+n-1$ and check if there is a word that is free from $\mathbf{S}$.

The definitions of unavoidability and the function $L(n)$ immediately lead to the following.

Proposition. A set of prohibitions $\mathbf{S} \subseteq \mathbf{A}^n$ is complete iff any word of length $L(n) + 1$ is not free from $\mathbf{S}$.

So in order to decide if a given set of prohibitions $\mathbf{S}$ is complete or not, we can consider all words of length $L(n)+1 = |\mathbf{A}|^{n-1}+n-1$ and check if there is a word that is free from $\mathbf{S}$.

Theorem. [Evdokimov, 1984] The complexity of deciding whether or not an arbitrary set $\mathbf{S} \subseteq \mathbf{A}^n$ is complete is $O(|\mathbf{S}| \cdot n)$.

Problem A:

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

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

Problem B that is equivalent to the Basic Problem:

Given: An arbitrary set of words $\mathbf{S}$.

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

$\mathbf{A} = \{a_1, \ldots, a_n\}$ is an alphabet;

$\mathbf{A}_\ell$ is the set of all the words of the alphabet $\mathbf{A}$ whose length is at most ℓ. We assume that the empty word belongs to $\mathbf{A}_\ell$.

$\mathbf{S}_1 \subseteq \mathbf{A}_2 \setminus \mathbf{A}_1$;
$\mathbf{S}_2 = \{xXx \mid x \in \mathbf{A}, X \in \mathbf{A}_{n-1}\}$.

So $\mathbf{S}_2$ contains all possible words of length at most $n+1$ whose first letter coincides with their last letter.

Suppose $\mathbf{S} = \mathbf{S}_1 \cup \mathbf{S}_2$.

Problem $\mathbf{A}'$:

Given: An arbitrary set $\mathbf{S}$ of the type described above and a natural number $\ell \leqslant n$.

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

Problem "path":

Given: Directed graph $\vec{G}(V, E)$ and a natural number $\ell \leqslant |V| = n$.

The question: Does there exist a simple directed path of length at least ℓ?

Correspondence between problem A′ and problem "path":

We compare vertices $v_1, \ldots, v_n$ from $V(\vec{G})$ to the letters $a_1, \ldots, a_n$ in the alphabet $\mathbf{A}$.

We compare each arc $v_i v_j$ 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 arc of the complement of $\vec{G}$ with respect to the complete directed graph.

$v_{i_1}, \ldots, v_{i_\ell}$ is a simple directed path iff $a_{i_1} \ldots a_{i_\ell}$ is free from $\mathbf{S}$.

Problem "path" is NP-complete $\Rightarrow$

Problem A$'$ is NP-complete $\Rightarrow$

Problem A is NP-complete!

[Evdokimov and SK, 2004]