On factor and permutation complexity of infinite words

Alexandr Valyuzhenich

Sobolev Institute of Mathematics

May 9, 2018

Basic definitions

Let Σ be a finite alphabet.
An infinite word over Σ is a sequence of the form $w = w_1 w_2 w_3 \dots$ where $w_i \in \Sigma$.

Definition

A finite word u is called a factor or subword of length n of an infinite word w if $u = w_{i+1} \dots w_{i+n}$ for some $i \geq 0$.

Let $F_w(n)$ be the set of all distinct factors of w of length n.

Definition

The factor complexity (or subword complexity) of w is $f_w(n) = |F_w(n)|$.

Basic definitions

Definition

An infinite word w is called periodic if $w = uvvv\ldots$ for some finite words u and v.

Definition

An infinite word w is called aperiodic if w is not a periodic word.

We note that if w is a periodic word, then $f_w(n) \leq C$ for some constant C.

Theorem (Morse, Hedlund, 1940)

Let w be an infinite aperiodic word. Then $f_w(n) \geq n + 1$.

Definition

An infinite word w is called a Sturmian word if $f_w(n) = n + 1$ for arbitrary n.

So, Sturmian words have the minimum factor complexity in the class of all infinite aperiodic words.

Morphisms

Definition

A map $\varphi : \Sigma^* \longrightarrow \Sigma^*$ is called a morphism if $\varphi(xy) = \varphi(x)\varphi(y)$ for any words $x, y \in \Sigma^*$.

Let $u = u_1 u_2 \dots u_n$ be a word.
Then $\varphi(u) = \varphi(u_1)\varphi(u_2)\dots\varphi(u_n)$.

So, every morphism is uniquely determined by the images of letters.

Consider a morphism $\varphi : \Sigma^* \longrightarrow \Sigma^*$.

Let $a \in \Sigma$ and $\varphi(a) = ax$ for some nonempty word x. Then $\varphi^2(a) = \varphi(\varphi(a)) = \varphi(ax) = \varphi(a)\varphi(x) = ax\varphi(x)$ and $\varphi^3(a) = \varphi(\varphi^2(a)) = ax\varphi(x)\varphi^2(x)$.

We see that $\varphi^n(a) = ax\varphi(x)\ldots\varphi^{n-1}(x)$. We have that $\varphi^{n-1}(a)$ is a prefix $\varphi^n(a)$ for any n. We define the infinite word w as follows: let the prefix $w_1 \ldots w_{|\varphi^n(a)|}$ of w is $\varphi^n(a)$. The word w is denoted by $\lim\limits_{n\to\infty} \varphi^n(a)$. We note that $w = \varphi(w)$.

Fixed point of morphisms

Definition

An infinite word w is a fixed point of a morphism φ if $w = \varphi(w)$.

Proposition

There exists only one fixed point w of a morphism φ that starts with symbol a. Moreover, $w = \lim_{n\to\infty} \varphi^n(a)$.

Fixed point of morphisms

Definition

An infinite word w is a fixed point of a morphism φ if $w = \varphi(w)$.

Proposition

There exists only one fixed point w of a morphism φ that starts with symbol a. Moreover, $w = \lim_{n\to\infty} \varphi^n(a)$.

Important morphisms

Definition

The morphism $\varphi(0) = 01$, $\varphi(1) = 0$ is the Fibonacci morphism. A fixed point of the Fibonacci morphism is called the Fibonacci word. The Fibonacci word is a Sturmian word.

Definition

The morphism $\varphi(0) = 01$, $\varphi(1) = 10$ is the Thue-Morse morphism. A fixed point of the Thue-Morse morphism is called the Thue-Morse word.

Morphisms

Definition

A finite word u is called a square (cube) if $u = vv$ ($u = vvv$) for some word v.

Definition

An infinite word w is called square-free (cube-free) if all factors of w are not squares (cubes).

Problem

Does there exist an infinite square-free (cube-free) word?

Morphisms

Theorem (Thue, 1906)

There exists an infinite cube-free word over an alphabet of size two.

For example, the Thue-Morse word is a cube-free word.

Theorem (Thue, 1906)

There exists an infinite square-free word over an alphabet of size three.

For example, the fixed point of the morphism $\varphi(a) = abc$, $\varphi(b) = ac$ and $\varphi(c) = b$ is a square-free word.

Factor complexity

For the problem of finding the factor complexity of fixed point morphisms was developed a good general approach:

- In 1997 Cassaigne developed an algorithm for the calculating the factor complexity of a fixed point of biprefix morphisms.
- In 1998 Avgustinovich and Frid found the exact formula for the factor complexity of a fixed point of biprefix morphisms.

Linear order on the shifts

Let $w = w_1 w_2 w_3 \dots$ be an infinite aperiodic word and $<_L$ is a lexicographic order on Σ.
The word $w_i w_{i+1} \dots$ is denoted by $w[i]$.

We will write $w[i] < w[j]$ if $w[i] = xa \dots$, $w[j] = xb \dots$ and $a <_L b$.

Definition

A permutation $\pi = \pi_1 \dots \pi_n$ of numbers $\{1, \dots, n\}$ is a subpermutation of an infinite aperiodic word w if there exists $i \geq 0$ such that $\pi_k < \pi_m$ iff $w[i+k] < w[i+m]$.

Example of subpermutation

Let T be the Thue-Morse word:

$$T = 0110100110010110\ldots$$

We have that $T[4] = 010\ldots$, $T[5] = 100\ldots$, $T[6] = 001\ldots$ and $T[7] = 011\ldots$.

So $T[6] < T[4] < T[7] < T[5]$. Therefore $\pi = 2413$ is a subpermutation of the Thue-Morse word.

Let $P_w(n)$ be the set of all distinct subpermutations of w of length n.

Definition

The permutation complexity of w is $p_w(n) = |P_w(n)|$.

Permutation complexity of aperiodic words is a relatively new notion in combinatorics on words. This complexity was introduced by Makarov:

M. A. Makarov. On permutations generated by infinite binary words. Sib. Elektron. Mat. Izv., 3 (2006), 304–311.

Valid permutations

Definition

A finite permutation π is called a valid if π is a subpermutation of some infinite binary word w.

Example. $\pi = 132$ is a valid permutation.
Indeed, we consider the word $w = 0101110101^5 \ldots 01^{2n+1} \ldots$.

We have $w[3] = 011\ldots$, $w[4] = 111\ldots$ and $w[5] = 110\ldots$.

So $w[3] < w[5] < w[4]$. Therefore $\pi = 132$ is valid.

Valid permutations

We note that not all permutations are valid.

For example, consider permutation $\pi = 2134$.
Suppose that there exists a binary word w such that 2134 is a subpermutation of w.

Then for some i we have $w[i+2] < w[i+1] < w[i+3] < w[i+4]$.

Since $w[i+1] > w[i+2]$, $w[i+2] < w[i+3]$ and $w[i+3] < w[i+4]$, we see that $w_{i+1} = 1$, $w_{i+2} = 0$ and $w_{i+3} = 0$.

Hence $w[i+1] = 1\ldots > w[i+3] = 0\ldots$ and we obtain a contradiction.

Valid permutations

Let $p(n)$ be the number of all distinct valid permutations of length n.

Theorem (Makarov, 2006)

$$p(n+1) = \sum_{t=1}^{n} \Psi(t) \cdot 2^{n-t}$$

for $n \geq 1$.

Corollary (Makarov, 2006)

$$p(n+1) = 2^n(n - c + O(n2^{-n/2})).$$

So the maximum permutation complexity of an infinite binary aperiodic word is

$$2^n(n - c + O(n2^{-n/2})).$$

Permutation complexity

Theorem (Makarov, 2009)

Let w be a Sturmian word. Then $p_w(n) = n$.

So, for the Sturmian words we have $f_w(n-1) = p_w(n)$.

Period doubling word is the fixed point of the morphism $\varphi(0) = 0100$, $\varphi(1) = 0101$.

Theorem (Makarov, 2010)

Let D be the period doubling word. Then

$$p_D(n) = \begin{cases} n + 6 \cdot 2^t - 1, & \text{if } 5 \cdot 2^t + 1 \leq n \leq 6 \cdot 2^t \text{ and } t \geq 1; \\ 2n + 2 \cdot 2^t - 2, & \text{if } 6 \cdot 2^t + 1 \leq n \leq 10 \cdot 2^t \text{ and } t \geq 0. \end{cases}$$

for $n \geq 7$.

Permutation complexity

Theorem (Widmer, 2011)

Let T be the Thue-Morse word, $n = 2^a + b$ and $0 < b \leq 2^a$. Then $p_T(n) = 2(2^{a+1} + b - 2)$ for $n \geq 6$.

Morphisms

- A morphism is called l–uniform if its blocks are of the same length l.
- A morphism $\varphi : \Sigma^* \longrightarrow \Sigma^*$ is called a marked if its blocks are of the form $\varphi(a_i) = b_i x_i c_i$, where x_i is an arbitrary word, b_i and c_i are symbols of the alphabet Σ, and all b_i (as well as all c_i) are distinct.
- A morphism $\varphi : \Sigma^* \longrightarrow \Sigma^*$ where $\Sigma = \{0, 1\}$ is called binary.

Class Q_l

We say that a l–uniform marked binary morphism φ such that $\varphi(0)$ starts with 0 belongs to the class Q_l if it satisfies the following properties (we assume that $l \geq 2$):

- If $\varphi(0) = 0u0x$ for some word x, then $0u1$ is not a subword of $\varphi(0)$ and $\varphi(1)$ and $0u$ is not a suffix of $\varphi(0)$ and $\varphi(1)$.
- If $\varphi(1) = 1u1x$ for some word x, then $1u0$ is not a subword of $\varphi(0)$ and $\varphi(1)$ and $1u$ is not a suffix of $\varphi(0)$ and $\varphi(1)$.

Examples

- Morphism $\varphi(0) = 01^n$, $\varphi(1) = 10^n$ for $n \geq 1$ belongs to Q_l (for $l = n + 1$).
- Morphism $\varphi(0) = 01^{2n}01^n$, $\varphi(1) = 10^{2n}10^n$ for $n \geq 2$ belongs to Q_l (for $l = 3n + 2$).
- Morphism $\varphi(0) = 01^{5n}01^{3n}$, $\varphi(1) = 10^{6n}10^{2n}$ for $n \geq 1$ belongs to Q_l (for $l = 8n + 2$).
- Morphism $\varphi(0) = 0101^3$, $\varphi(1) = 10^5$ does not belong to Q_6.

Thus the Thue-Morse morphism belongs to Q_2.

Let $n \geq l^2 + 1$. Then for n there exists a unique pair of numbers $k(n)$ and $s(n)$ such that $s(n) > 0$, $k(n) \in \{l, ..., l^2 - 1\}$ and $k(n)l^{s(n)} < n \leq (k(n) + 1)l^{s(n)}$. Let $r(n) = n - k(n)l^{s(n)}$.

Theorem (V., 2014)

Let w be a fixed point of φ, $\varphi \in Q_l$ and $n \geq l^2 + 1$. Then

$$p_w(n) = (r(n)-1)\mu(k(n)+2)+(l^{s(n)}-r(n)+1)\chi(k(n)+1)-\beta(k(n)+1)$$

for $r(n) > 1$ and

$$p_w(n) = l^{s(n)}\chi(k(n) + 1) - \alpha(k(n))$$

for $r(n) = 1$.

Since the Thue-Morse morphism belongs to Q_2, we automatically obtain an alternative way to compute the permutation complexity of the Thue-Morse word.

Results for nonuniform morphisms

Let us consider the morphism $\varphi(0) = 01^k, \varphi(1) = 0$ for $k \geq 2$.

Let $\lambda_1 = \frac{1+\sqrt{1+4k}}{2}$ and $\lambda_2 = \frac{1-\sqrt{1+4k}}{2}$ be an eigenvalues of $A = \begin{pmatrix} 1 & 1 \\ k & 0 \end{pmatrix}$,

$c_1(x, y) = \frac{(k+\frac{1+\sqrt{1+4k}}{2})x+\frac{1+\sqrt{1+4k}}{2}y}{\sqrt{1+4k}}$ and

$c_2(x, y) = \frac{(\frac{\sqrt{1+4k}-1}{2}-k)x+\frac{\sqrt{1+4k}-1}{2}y}{\sqrt{1+4k}}$.

Consider the sequences $a_s = c_1(k+1, k^2)\lambda_1^s + c_2(k+1, k^2)\lambda_2^s + 1$ and

$b_s = c_1(2, k)\lambda_1^{s-1} + c_2(2, k)\lambda_2^{s-1} + 1$.

Results for nonuniform morphisms

Theorem (V., 2015)

Let w be a fixed point of φ, $\varphi(0) = 01^k, \varphi(1) = 0$, $n > k^2 + k + 1$ and $k > 2$. Then the following statements are true:

1. $p_w(n) = 2n + M_1\lambda_1^{s+1} + M_2\lambda_2^{s+1} + W_1$ for $a_s < n < b_{s+3}$, where $M_1 = \frac{1}{\lambda_1 - 1}(c_1(1+k, k^2) + c_1(1,0)\lambda_1 - c_1(1,1)\lambda_1^2)$, $M_2 = \frac{1}{\lambda_2 - 1}(c_2(1+k, k^2) + c_2(1,0)\lambda_2 - c_2(1,1)\lambda_2^2)$ and W_1 is a constant;
2. $p_w(n) = 3n - a_{s+1} + M_1\lambda_1^{s+2} + M_2\lambda_2^{s+2} + W_2$ for $b_{s+3} \leq n \leq a_{s+1}$, where $M_1 = \frac{1}{\lambda_1 - 1}(c_1(1+k, k^2) + c_1(1,0)\lambda_1 - c_1(1,1)\lambda_1^2)$, $M_2 = \frac{1}{\lambda_2 - 1}(c_2(1+k, k^2) + c_2(1,0)\lambda_2 - c_2(1,1)\lambda_2^2)$ and W_2 is a constant.

Recent results

- (Lu, Chen, Guo, Wen, 2016). Lu et. al find the permutation complexity of the fixed point of the morphism $\varphi(0) = 01^k0, \varphi(1) = 1^{k+2}$ (Cantor-like sequence).
- (Borchert, Rampersad, 2017). Borchert and Rampersad show that the permutation complexity of the image of a Sturmian word by a binary marked morphism is $n + k$ for some constant k and all lengths n sufficiently large.

References

1. M. A. Makarov. On permutations generated by infinite binary words. Sib. Elektron. Mat. Izv. 3 (2006) 304–311.

2. M. A. Makarov. On the permutations generated by Sturmian words. Sib. Math. J. 50 (4) (2009) 674–680.

3. M. A. Makarov. On the infinite permutation generated by the period doubling word. European J. Combin. 31 (1) (2010) 368–378.

4. A. Valyuzhenich. On permutation complexity of fixed points of some uniform binary morphisms. Discrete Mathematics and Theoretical Computer Science. 2014. Vol.16, no. 3. p. 95–128.

5. A. Valyuzhenich. On permutation complexity of fixed points of some nonuniform binary morphisms. Siberian Electronic Mathematical Reports. 2015. Vol. 12. p. 64–79.

6. S. Widmer. Permutation complexity of the Thue–Morse word. Adv. in Appl. Math. 47 (2) (2011) 309–329.