On Subword Complexity of Morphic Sequences

Rostislav Deviatov

Moscow State University and Independent University of Moscow
deviatov1@rambler.ru

Abstract. We sketch the proof of the following result: the subword complexity of arbitrary morphic sequence is either $\Theta(n^2)$, or $O(n^{3/2})$.

1 Introduction

Morphisms and morphic sequences are well known and well studied in combinatorics on words (e. g., see [1]). We study their subword complexity.

Let Σ be a finite alphabet. A mapping $\varphi\colon \Sigma^* \to \Sigma^*$ is called a *morphism* if $\varphi(uv) = \varphi(u)\varphi(v)$ for all $u, v \in \Sigma^*$. A morphism is determined by its values on single-letter words. A morphism is *non-erasing* if $|\varphi(a)| \geq 1$ for each $a \in \Sigma$, and is called *coding* if $|\varphi(a)| = 1$ for each $a \in \Sigma$. Let $|\varphi|$ denote $\max_{a\in\Sigma} |\varphi(a)|$.

Let $\varphi(s) = su$ for some $s \in \Sigma$, $u \in \Sigma^*$, and suppose $\forall n$ $\varphi^n(u)$ is not empty. Then an infinite sequence $\varphi^\infty(s) = \lim_{n\to\infty} \varphi^n(s)$ is well-defined and is called *pure morphic*. Sequences of the form $\psi(\varphi^\infty(s))$ with coding ψ are called *morphic*.

In this paper we study a natural combinatorial characteristics of sequences, namely subword complexity. *Subword complexity* of a sequence β is a function $p_\beta\colon \mathbb{N} \to \mathbb{N}$ where $p_\beta(n)$ is the number of all different n-length subwords occurring in β. For a survey on subword complexity, see, e. g., [2]. Pansiot showed [4] that the subword complexity of an arbitrary pure morphic sequence adopts one of the five following asymptotic behaviors: $O(1)$, $\Theta(n)$, $\Theta(n \log\log n)$, $\Theta(n \log n)$, or $\Theta(n^2)$. Since codings can only decrease subword complexity, the subword complexity of every morphic sequence is $O(n^2)$. We formulate the following main result.

Theorem 1. *The subword complexity p_β of a morphic sequence β is either $p_\beta(n) = \Theta(n^{1+\frac{1}{k}})$ for some $k \in \mathbb{N}$, or $p_\beta(n) = O(n \log n)$.*

Note that for each k the complexity class $\Theta(n^{1+\frac{1}{k}})$ is non-empty [3].

However in this extended abstract we show the technics used in the proof of this main result considering the following weaker case of Theorem 1.

Theorem 2. *The subword complexity p_β of a morphic sequence β is either $p_\beta(n) = \Theta(n^2)$, or $p_\beta(n) = O(n^{3/2})$.*

We give an example of a morphic sequence β with $p_\beta = \Theta(n^{3/2})$ in Section 6.

Let Σ be a finite alphabet, $\varphi\colon \Sigma^* \to \Sigma^*$ be a morphism, $\psi\colon \Sigma^* \to \Sigma^*$ be a coding, α be a pure morphic sequence generated by φ, and $\beta = \psi(\alpha)$ be a morphic

E.A. Hirsch et al. (Eds.): CSR 2008, LNCS 5010, pp. 146–157, 2008.

sequence. By Theorem 7.7.1 from [1] every morphic sequence can be generated by a non-erasing morphism, so further we assume that φ is non-erasing.

To prove Theorem 2, we prove the following two propositions:

Proposition 1. *If there are evolutions of 2-blocks arising in α that are not continuously periodic, the subword complexity of β is $\Omega(n^2)$.*

Proposition 2. *If all the evolutions of 2-blocks arising in α are continuously periodic, the subword complexity of β is $O(n^{3/2})$.*

Most part of the paper is devoted to formulation of what k-blocks, evolutions and continuously periodic evolutions are. Similarly, one can generalize the notion of a continuously periodic evolution of k-blocks to each $k \in \mathbb{N}$, generalize Propositions 1 and 2 to an arbitrary k, and thus prove Theorem 1. (Actually, the notion of continuously periodic evolution of k-blocks needs more technical details, but Propositions 1 and 2 can be reformulated easily: 2-blocks, n^2 and $n^{3/2}$ should be replaced by k-blocks, $n^{1+1/(k-1)}$ and $n^{1+1/k}$, respectively). However, the full detailed proof of Theorem 1 (and Theorem 2 as well) needs much more space and will be published elsewhere.

We will speak about occurrences in α. Strictly speaking, we call a pair of a word γ and a location i in α *an occurrence* if the subword of α that starts from position i in α and is of length $|\gamma|$ is γ. This occurrence is denoted by $\alpha_{i\ldots j}$ if j is the index of the last letter that belongs to the occurrence. In particular, $\alpha_{i\ldots i}$ denotes a single-letter occurrence, and $\alpha_{i\ldots i-1}$ denotes an occurrence of the empty word between the $(i-1)$-th and the i-th letters. Since $\alpha = \alpha_1\alpha_2\alpha_3\ldots = \varphi(\alpha) = \varphi(\alpha_1)\varphi(\alpha_2)\varphi(\alpha_3)\ldots$, φ might be considered either as a morphism on words (which we call abstract words sometimes), or as a mapping on the set of occurrences in α. Usually we speak of the latter, unless stated otherwise.

We call a finite word γ *p-periodic* with *left* (resp. *right*, *complete*) period δ if $|\delta| = p$ and $\gamma = \delta\delta\ldots\delta\delta_{1\ldots k}$ (resp. $\gamma = \delta_{p-k+1\ldots p}\delta\ldots\delta$, $\gamma = \delta\ldots\delta$). If δ is known, we will shortly call γ a *left* (resp. *right*, *completely*) *δ-periodic* word. δ will be always considered as an abstract word. The subword $\gamma_{|\gamma|-k+1\ldots|\gamma|}$ is called the *incomplete occurrence*, where $0 \le k < |\delta|$. All the same is with sequences of symbols or numbers.

The function $r_a\colon \mathbb{N} \to \mathbb{N}$, $r_a(n) = |\varphi^n(a)|$ is called *the growth rate* of a. Let us define *orders of letters* with respect to φ. We say that $a \in \Sigma$ has *order* k if $r_a(n) = \Theta(n^{k-1})$, and has *order* ∞ if $r_a(n) = \Omega(q^n)$ for some $q > 1$ ($q \in \mathbb{R}$).

Consider a directed graph G defined as follows. Vertices of G are letters of Σ. For every $a, b \in \Sigma$, for each occurrence of b in $\varphi(a)$, construct an edge $a \to b$. For instance, if $\varphi(a) = abbab$, we construct two edges $a \to a$ and three edges $a \to b$. Fig. 1 shows an example of graph G.

Using the graph G, one can prove the following

Lemma 1. *For every $a \in \Sigma$, either a has some order $k < \infty$, or has order ∞. For every a of order $k < \infty$, either a never appears in $\varphi^n(a)$ (and then a is called* pre-periodic*), or for each n a unique letter b_n of order k occurs in $\varphi^n(a)$, and the sequence $(b_n)_{n\in\mathbb{Z}_{\ge 0}}$ is periodic (then a is called* periodic*). If a is a periodic letter of order $k > 1$, then at least one letter of order $k-1$ occurs in $\varphi(a)$.*

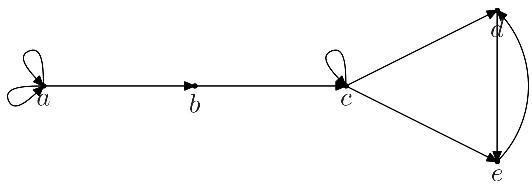

Fig. 1. An example of graph G for the following morphism φ: $\varphi(a) = aab$, $\varphi(b) = c$, $\varphi(c) = cde$, $\varphi(d) = e$, $\varphi(e) = d$. Here a is a letter of order ∞, b is a pre-periodic letter of order 2, c is a periodic letter of order 2, d and e are periodic letters of order 1.

2 Blocks and Semiblocks

A (possibly empty) occurrence $\alpha_{i\ldots j}$ is *a k-block* if it consists of letters of order $\leq k$, and the letters α_{i-1} and α_{j+1} both have order $> k$. The letter α_{i-1} is called the *left border* of this block and is denoted by $\mathrm{LB}(\alpha_{i\ldots j})$. The letter α_{j+1} is called the *right border* of this block and is denoted by $\mathrm{RB}(\alpha_{i\ldots j})$.

The image under φ of a letter of order $\leq k$ cannot contain letters of order $> k$. Let $\alpha_{i\ldots j}$ be a k-block. Then $\varphi(\alpha_{i\ldots j})$ is a suboccurrence of some k-block which is called the *descendant* of $\alpha_{i\ldots j}$ and is denoted by $\mathrm{Dc}(\alpha_{i\ldots j})$. The l-th superdescendant (denoted by $\mathrm{Dc}^l(\alpha_{i\ldots j})$) is the descendant of …of the descendant of $\alpha_{i\ldots j}$ (l times).

Let $\alpha_{s\ldots t}$ be a k-block in α. Then a unique k-block $\alpha_{i\ldots j}$ such that $\mathrm{Dc}(\alpha_{i\ldots j}) = \alpha_{s\ldots t}$, is called the *ancestor* of $\alpha_{s\ldots t}$ and is denoted $\mathrm{Dc}^{-1}(\alpha_{s\ldots t})$. The l-th superancestor (denoted by $\mathrm{Dc}^{-l}(\alpha_{s\ldots t})$) is the ancestor of …of the ancestor of $\alpha_{s\ldots t}$ (l times). If $\mathrm{Dc}^{-1}(\alpha_{s\ldots t})$ does not exist (it can happen only if α_{s-1} and α_{t+1} belong to an image of the same letter), then $\alpha_{s\ldots t}$ is called an *origin*. A sequence $\mathcal{E}$ of k-blocks, $\mathcal{E}_0 = \alpha_{i\ldots j}, \mathcal{E}_1 = \mathrm{Dc}(\alpha_{i\ldots j})$, $\mathcal{E}_2 = \mathrm{Dc}^2(\alpha_{i\ldots j}), \ldots, \mathcal{E}_l = \mathrm{Dc}^l(\alpha_{i\ldots j}), \ldots$, where $\alpha_{i\ldots j}$ is an origin, is called an *evolution*.

Lemma 2. *The set of all abstract words that can be origins in α, is finite.*

Corollary 1. *The set of all possible evolutions in α (considered as sequences of abstract words rather than sequences of occurrences in α), is finite.*

Now we define *atoms* inside k-blocks. The l-th left and right atoms exist in a k-block $\mathcal{E}_m$, where $\mathcal{E}$ is an evolution, iff $m \geq l > 0$. First, define the l-th atoms inside the k-block $\mathcal{E}_l$. Let $\mathcal{E}_l = \alpha_{i\ldots j}$. There is a suboccurrence $\alpha_{s\ldots t} = \varphi(\mathrm{Dc}^{-1}(\alpha_{i\ldots j}))$ inside of it. The occurrence $\alpha_{i\ldots s-1}$ that comes from the image of the left border of the ancestor, is called *the l-th left atom* of the block and is denoted by $\mathrm{LA}_l(\alpha_{i\ldots j})$. Similarly, the occurrence $\alpha_{t+1\ldots j} = \mathrm{RA}_l(\alpha_{i\ldots j})$ is called *the l-th right atom* of the block. Then, $\mathrm{LA}_l(\mathcal{E}_m) = \varphi^{m-l}(\mathrm{LA}_l(\mathcal{E}_l))$, and the same for right atoms. See Fig. 2.

Let $\mathcal{E}$ be an evolution of k-blocks. Consider a sequence $\mathrm{LB}(\mathcal{E}_0), \ldots, \mathrm{LB}(\mathcal{E}_l), \ldots$ All these letters are of order $> k$, and the letter $\mathrm{LB}(\mathcal{E}_{l+1})$ is the rightmost letter of order $> k$ in $\varphi(\mathrm{LB}(\mathcal{E}_l))$. Hence, not later than starting from $\mathrm{LB}(\mathcal{E}_{|\Sigma|})$, this

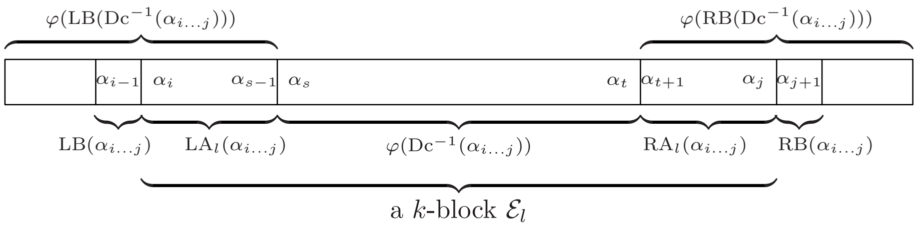

Fig. 2. Structure of a k-block

sequence (of abstract letters) is periodic. Its period length is denoted by $\mathrm{LB}\mathbf{P}(\mathcal{E})$. The same can be said about the sequence of right borders (the period length is $\mathrm{RB}\mathbf{P}(\mathcal{E})$). Their l. c. m. is denoted by $\mathrm{B}\mathbf{P}(\mathcal{E})$. The exact place of the evolution where both sequences become periodic (i. e. both of them have reached at least the first positions of their first periods) is denoted by $\mathrm{F}(\mathcal{E})$. The block $\mathcal{E}_{\mathrm{F}(\mathcal{E})}$ is called *first pre-stable*. The block $\mathcal{E}_{\mathrm{F}(\mathcal{E})}$ and all its superdescendants are called *pre-stable*.

Notice that $\mathrm{LA}_{l+1}(\mathcal{E}_{l+1})$ depends only on $\mathrm{LB}(\mathcal{E}_l)$, not on the whole $\mathcal{E}_l$. Hence, the sequence $\mathrm{LA}_{\mathrm{F}(\mathcal{E})+1}(\mathcal{E}_{\mathrm{F}(\mathcal{E})+1}),\ldots,\mathrm{LA}_l(\mathcal{E}_l),\ldots$ is periodic with a period of length $\mathrm{LB}\mathbf{P}(\mathcal{E})$ if considered as a sequence of abstract words. Consider one of its periods, e. g., $\mathrm{LA}_{\mathrm{F}(\mathcal{E})+1}(\mathcal{E}_{\mathrm{F}(\mathcal{E})+1}),\ldots,$ $\mathrm{LA}_{\mathrm{F}(\mathcal{E})+\mathrm{LB}\mathbf{P}(\mathcal{E})}(\mathcal{E}_{\mathrm{F}(\mathcal{E})+\mathrm{LB}\mathbf{P}(\mathcal{E})})$. There are three possible cases.

Case I. At least one of these words contains a letter of order k.

Case II. None of these words contain a letter of order k, but at least one of them is not empty.

Case III. All these words are empty.

Similarly, cases I, II and III are defined for right borders and atoms. These cases happen independently at right and at left, in any combination.

Now we define the *core* of a pre-stable k-block (notation: C). Consider the first pre-stable k-block of the evolution. Its core is the whole block. Then, the core of $\mathcal{E}_{l+1}$ is $\varphi(\mathrm{C}(\mathcal{E}_l))$. Thus, $\mathrm{C}(\mathcal{E}_{\mathrm{F}(\mathcal{E})+l}) = \varphi^l(\mathcal{E}_{\mathrm{F}(\mathcal{E})})$. The suboccurrence in the block between the core and its left (right) border is called its *left (right) component*.

We know that $\mathrm{F}(\mathcal{E}) \leq |\Sigma|$. Thus $|\,\mathrm{C}(\mathcal{E}_{\mathrm{F}(\mathcal{E})})| \leq D := 2|\varphi|^{|\Sigma|+1}$.

Lemma 3. *Consider a k-block and its evolution. If case I holds at right (at left), the right (left) component has growth rate $\Theta(n^k)$. If case II or III holds, it has growth rate $O(n^{k-1})$.*

Now we introduce semiblocks to consider evolutions of words that grow ‘at one side’ (at left or at right only), while evolutions of blocks represent sequences of words that grow (that may have atoms) ‘at both sides’.

Let $\alpha_{i\ldots j}$ be a (possibly empty) occurrence in α consisting of letters of order $\leq k$, and suppose α_{j+1} has order $> k$. Suppose also that $j - i + 1 \leq D$. Then $\alpha_{i\ldots j}$ is called a *right k-semiblock*, and α_{j+1} is called its *right border* (a left border of a right k-semiblock does not exist). The image of the k-semiblock under φ prolonged right upto the leftmost letter of order $> k$ is also called a k-semiblock — the *descendant* of $\alpha_{i\ldots j}$. Evolution of right k-semiblocks, with arbitrary k-semiblock of length $\leq D$ as an origin, is defined analogously to how it was defined for k-blocks.

Left k-semiblocks and their evolutions are defined in a similar way.

The length of origins of k-semiblock evolutions is bounded by definition, so the set of all k-semiblock evolutions (considered as sequences of abstract words) is finite. Like an evolutional sequence of borders of k-block, an evolutional sequence of left (right) borders of left (right) k-semiblock is eventually periodic with pre-period not greater than $|\Sigma|$.

Pre-stable k-semiblocks and the first pre-stable k-semiblocks are defined analogously to those of k-blocks. All the notation we introduced for k-blocks, is used for k-semiblocks as well. Fig. 3 shows the structure of a k-semiblock.

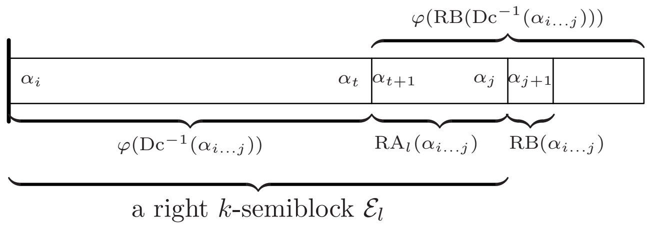

Fig. 3. Structure of a right k-semiblock

The *core* (C) of a k-semiblock is defined similarly too. Namely, let $\mathcal{E}$ be an evolution of k-semiblocks. Consider the first pre-stable k-semiblock of $\mathcal{E}$. Its core is the whole semiblock. Then, the core of $\mathcal{E}_{l+1}$ is $\varphi(\mathrm{C}(\mathcal{E}_l))$. Thus, $\mathrm{C}(E_{\mathrm{F}(\mathcal{E})+l}) = \varphi^l(E_{\mathrm{F}(\mathcal{E})})$. The suboccurrence in the semiblock between the core and its left border is called its left component.

3 1-Blocks and 1-Semiblocks

Now we will consider 1-blocks and 1-semiblocks more accurately.

From the fact that every 1-block or 1-semiblock consists of letters of order 1 only, and from Corollary 1, it follows that *all core lengths are bounded by a single constant that depends on* φ *and* Σ *only.* For an evolution $\mathcal{E}$, cores of the block $\mathcal{E}_{\mathrm{F}(\mathcal{E})+|\Sigma|}$ and its descendants consist of periodic letters only.

A core of 1-block or 1-semiblock is called its (unique) *central kernel.*

Consider a 1-block or a right 1-semiblock $\mathcal{E}_l$ where $\mathcal{E}$ is an evolution and $l \geq |\Sigma|$. The concatenation $\mathrm{RpreP}(\mathcal{E}_l) := \mathrm{RA}_{l-|\Sigma|+1}(\mathcal{E}_l) \ldots \mathrm{RA}_{l-1}(\mathcal{E}_l)\,\mathrm{RA}_l(\mathcal{E}_l)$ is called the *right pre-period* of $\mathcal{E}_l$. Left pre-periods are defined similarly. All letters in the right component of a 1-block or a 1-semiblock outside its right pre-period are periodic.

The left (right) component of a 1-block or a 1-semiblock except its left (right) pre-period is called the *left (right) repetition* (notation: LR, RR).

Lemma 4. *Let $\mathcal{E}$ be an evolution of 1-blocks or left 1-semiblock, $l > \mathrm{F}(\mathcal{E}) + |\Sigma|$. Then there are such abstract words $\mathrm{LP}(\mathcal{E}_l)$ (for each l), that $\mathrm{LR}(\mathcal{E}_l)$ is a left $\mathrm{LP}(\mathcal{E}_l)$-periodic word and:*

The sequence $\mathrm{LP}(\mathcal{E}_{\mathrm{F}(\mathcal{E})+|\Sigma|}), \ldots, \mathrm{LP}(\mathcal{E}_l), \ldots$ is periodic. In particular, lengths of these words are bounded by a constant that depends on Σ and φ only.

The lengths of the incomplete occurrences become periodic starting from $\mathrm{LR}(\mathcal{E}_{\mathrm{F}(\mathcal{E})+|\Sigma|})$. The sequence (of abstract words) $\mathrm{LpreP}(\mathcal{E}_{\mathrm{F}(\mathcal{E})+|\Sigma|}), \ldots, \mathrm{LpreP}(\mathcal{E}_l), \ldots$ is periodic. Lengths of these words are bounded by a constant that depends on Σ and φ only. The sequence (of abstract words) $\mathrm{C}(\mathcal{E}_{\mathrm{F}(\mathcal{E})+|\Sigma|}), \ldots, \mathrm{C}(\mathcal{E}_l), \ldots$ is also periodic.

Let $\mathcal{E}$ be an evolution of 1-blocks or of 1-semiblocks. If case I holds at left or at right, we say that $\mathcal{E}_l$ is stable if $l \geq \mathrm{F}(\mathcal{E}) + |\Sigma| + 2\,\mathrm{B}\mathbf{P}(\mathcal{E})|\Sigma|!$. If case III holds both at left and at right (case II is impossible for 1-blocks), $\mathcal{E}_l$ is stable if $l \geq \mathrm{F}(\mathcal{E}) + |\Sigma|$. If $\mathcal{E}_l$ is stable then:

1. It is pre-stable.
2. The pre-periods $\mathrm{LpreP}(\mathcal{E}_l)$ and $\mathrm{RpreP}(\mathcal{E}_l)$ and the core $\mathrm{C}(\mathcal{E}_l)$ belong to the periodic parts of the corresponding sequences from Lemma 4.
3. If case I holds at left or at right, $\mathrm{LR}(\mathcal{E}_l)$ and $\mathrm{RR}(\mathcal{E}_l)$ consist of at least two their periods (when considered as left $\mathrm{LP}(\mathcal{E}_l)$-periodic and right $\mathrm{RP}(\mathcal{E}_l)$-periodic words, respectively).

The least number l such that $\mathcal{E}_l$ is stable is denoted by $\mathrm{S}(\mathcal{E})$.

The following lemma is an easy corollary of Lemma 4.

Lemma 5. *Let $\mathcal{E}$ be an evolution of 1-blocks or 1-semiblocks. The sequence $|\mathrm{LR}(\mathcal{E}_{\mathrm{S}(\mathcal{E})})|, \ldots, |\mathrm{LR}(\mathcal{E}_l)|, \ldots$ considered modulo s, is periodic for all $s \in \mathbb{N}$.*

Fig. 4 shows the detailed structure of a 1-block.

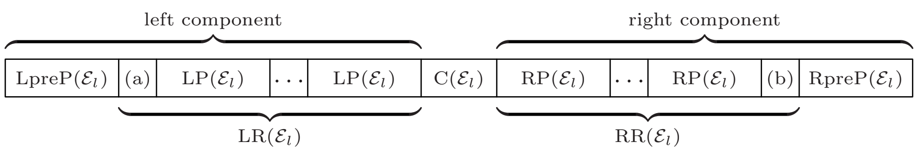

Fig. 4. Detailed structure of the 1-block $\mathcal{E}_l$: (a), (b) — the incomplete occurrences of periods

4 2-Blocks

Now consider 2-blocks more accurately. First, let us give definitions concerning 2-blocks that are necessary to define continuously periodic evolutions. Through this section, we will give examples based on the following morphism φ: $\varphi(s) = saca$, $\varphi(a) = bad$, $\varphi(b) = dbd$, $\varphi(c) = ece$, $\varphi(d) = d$, $\varphi(e) = e$. Then $\varphi^\infty(s) = \alpha = s\,aca\,badecebad\,dbdbaddeeceedbdbadd\ldots$. Here s is a letter of order 4, a is a letter of order 3, b and c are letters of order 2 and d and e are letters of order 1. Consider an evolution $\mathcal{E}$ of 2-blocks, whose origin is $\alpha_{3\dots3} = c$. A 2-block $\mathcal{E}_l$ where l is large enough looks as follows:

$$\underbrace{dd..d}_{\text{left component}}\underbrace{ee..ecee..e}_{\text{core}}\underbrace{dd..db\dots bdd..dbdd..db\dots ddbdddbdb}_{\text{right component}}.$$

Here case I holds at left and case II holds at right. Intervals denoted by ... may contain many intervals denoted by ..

First, let us define *stable 2-blocks*. We will impose requirements on the number of iterations to be made from the beginning of the evolution, to guarantee for the block to be stable.

If a 2-block is stable, it is required to be pre-stable. Other requirements depend on the case that holds at left and at right of the given block.

Case I (e. g., at right). Consider an alphabet Σ' which is Σ without all the letters of order 1. Let us define a morphism φ' as follows: for $a \in \Sigma'$, to obtain $\varphi'(a)$, we take the word $\varphi(a)$ and remove all the letters of order 1 from it. Let α' be a new pure morphic sequence generated by φ'. (If α is non-periodic, then α' is infinite.) In other words, α' is obtained from α by removing all letters of order 1. All the 2-blocks in α become 1-blocks in α'. In the example, the corresponding 1-block in α' is $cbb\dots b$. For a given 2-block in α to be stable, we require the corresponding 1-block in α' to be stable too.

Let $\mathcal{E}$ be an evolution of 2-blocks in α, $\alpha'_{s\dots t}$ be the 1-block corresponding to $\alpha_{i\dots j} \in \mathcal{E}$. Consider an occurrence of $\mathrm{RP}(\alpha'_{s\dots t})$ in $\alpha'_{s\dots t}$. We may assume that the number of atoms it consists of is divisible by $\mathrm{B}\mathbf{P}(\mathcal{E})$. A suboccurrence in $\alpha_{i\dots j}$ containing all the letters of $\mathrm{RP}(\alpha'_{s\dots t})$ and all the letters of order 1 to the right of $\mathrm{RP}(\alpha'_{s\dots t})$ upto the closest letter of order 2, is called a *pseudo-period*. In the example, both left and right borders are always a, and $\mathrm{B}\mathbf{P}(\mathcal{E}) = 1$. Thus, a right pseudo-period is an occurrence $bdd..d$ (the amount of d's may differ in different pseudo-periods).

For a given 2-block to be stable, its right component is required to contain at least two pseudo-periods. Moreover, all the 1-blocks inside two leftmost of them should be stable.

Now consider an occurrence of the empty word immediately at left of $\mathrm{RB}(\mathcal{E}_{\mathrm{F}(\mathcal{E})})$. This occurrence is a right 1-semiblock (an origin). It is called the *right outer central 1-semiblock* of $\mathcal{E}_{\mathrm{F}(\mathcal{E})}$ and is denoted by $\mathrm{RO}(\mathcal{E}_{\mathrm{F}(\mathcal{E})})$. Then, $\mathrm{RO}(\mathcal{E}_{l+1}) := \mathrm{Dc}(\mathrm{RO}(\mathcal{E}_l))$. For a 2-block $\mathcal{E}_l$ to be stable, we require $\mathrm{RO}(\mathcal{E}_l)$ to be stable too. In the example, the right outer central semiblock is the occurrence $dd..d$ between the letters e and b (the leftmost in the right component).

Case II (at right). If case II holds, the right component consists of the right outer central 1-semiblock (and noting else). For a 2-block to be stable, we require this 1-semiblock to be stable. In the example, case II holds at left, thus the left component of the block is its left outer semiblock.

Case III. No more requirements.

There are still other requirements concerning the core. If it contains letters of order 1 only, all these letters are required to become periodic. Let it contain some letters of order 2. Consider $\mathrm{C}(\mathcal{E}_{\mathrm{F}(\mathcal{E})})$. Its length is not greater than D, so its suboccurrence between its right end and the rightmost letter of order 2 may be considered as a left 1-semiblock. It is called the *right inner central 1-semiblock* (notation: RI). Then, $\mathrm{RI}(\mathcal{E}_{l+1}) := \mathrm{Dc}(\mathrm{RI}(\mathcal{E}_l))$. In the example, the right inner 1-semiblock is the word *ee..e* between the letter c and the right component. For the 2-block $\mathcal{E}_l$ to be stable, we require $\mathrm{LI}(\mathcal{E}_l)$ and $\mathrm{RI}(\mathcal{E}_l)$ to be stable. Moreover, the letters of order 2 inside $\mathrm{C}(\mathcal{E}_l)$ should be periodic, all the 1-blocks between them have to be stable too.

These are all the requirements we impose on 2-block to be stable. The least number l such that $\mathcal{E}_l$ is stable is denoted by $\mathrm{S}(\mathcal{E})$.

Now we define left and right pre-periods of 2-blocks. Let $\alpha_{i\ldots j}$ be a stable 2-block and $\alpha'_{s\ldots t}$ be the corresponding 1-block. If case I holds (e. g. at right), the occurrence $\alpha_{u\ldots j}$ is called the right pre-period (RpreP) if u is the maximal index such that

1. $\mathrm{RpreP}(\alpha'_{s\ldots t})$ is contained in $\alpha_{u\ldots j}$.
2. 1-blocks between letters of order 2 in the right component of $\alpha_{i\ldots j}$ outside $\alpha_{u\ldots j}$ are all stable.
3. α_u is a letter of order 2.

Notice that $\alpha_{u\ldots j}$ is then completely inside the last L atoms, where $L = |\Sigma| + \max\{S(\mathcal{F}) : \mathcal{F} \text{ is an evolution of 1-blocks}\}$. Hence, the sequence (of abstract words) $\mathrm{RpreP}(\mathcal{E}_{\mathrm{S}(\mathcal{E})}), \ldots, \mathrm{RpreP}(\mathcal{E}_l), \ldots$ is periodic. In the example, the right pre-period is really long since all the 1-blocks outside it are stable, and the length of a stable 1-block is greater than $|\Sigma|! = 120$. However, its structure can be written as follows: *bd..dbd..d . . . ddbdb*.

If case II holds, $\mathrm{RpreP}(\alpha_{i\ldots j}) := \mathrm{RpreP}(\mathrm{RO}(\alpha_{i\ldots j}))$. If case III holds, the right pre-period is the occurrence of the empty word immediately at left of α_{j+1}.

Now we define central kernels of stable 2-blocks. Let $\alpha_{i\ldots j}$ be a 2-block.

There may be two kinds of central kernels: ones that are suboccurrences of $\mathrm{C}(\alpha_{i\ldots j})$ and ones that are outside it. To define central kernels inside $\mathrm{C}(\alpha_{i\ldots j})$, consider two cases. If it consists of letters of order 1 only, it is called a central kernel itself. Let $\mathrm{C}(\alpha_{i\ldots j})$ contain letters of order 2. Then, if $\alpha_{i\ldots j}$ is stable, $\mathrm{C}(\alpha_{i\ldots j})$ can be decomposed into $\mathrm{LI}(\alpha_{i\ldots j}), \mathrm{RI}(\alpha_{i\ldots j})$ and some 1-blocks between them (they all evolute together with the 2-block). Central kernels of the 2-block are all the cores, left and right pre-periods of these 1-blocks and semiblocks and letters of order 2 themselves. In the example, central kernels inside the core are the occurrences of the word *eeeee* immediately at left and at right of c, the letter c itself and two occurrences of the empty word between the core and the left and the right components.

To define central kernels of $\alpha_{i\ldots j}$ at right (or at left) of $\mathrm{C}(\alpha_{i\ldots j})$, consider cases I, II and III again. If case I holds at right, central kernels at right of $\mathrm{C}(\alpha_{i\ldots j})$ are $\mathrm{RpreP}(\mathrm{RO}(\alpha_{i\ldots j}))$ and $\mathrm{C}(\mathrm{RO}(\alpha_{i\ldots j}))$. If case II holds, the only central kernel at right is $\mathrm{C}(\mathrm{RO}(\alpha_{i\ldots j}))$. In case III there are no more cental kernels. In the example, central kernels outside the core are the occurrence of the word *ddddd* immediately at left of the leftmost letter b and two occurrences of the empty word between the core and the left and the right components once more.

Central kernels of a block or a semiblock, as well as its right and left pre-periods, are called *simple kernels*. A concatenation of consecutive simple kernels (with no simple kernel *immediately* at left and at right of it) is called *kernel*.

The following remarks are easy corollaries of Lemma 4 and the definition of central kernels. Every kernel of a stable 2-block $\mathcal{E}_l$ corresponds to some kernel of $\mathcal{E}_{l+1}$, and vice versa. Thus, we have evolutional sequences of kernels concerned with $\mathcal{E}$. (The amount of these sequences equals the amount of kernels in *each* $\mathcal{E}_l$). These sequences are periodic. A part of $\mathcal{E}_l$ between its (consecutive) central kernels is actually either an empty word or a (left or right) repetition of some 1-block or semiblock $\alpha_{s\ldots t}$. The part of $\mathcal{E}_{l+1}$ between the corresponding central kernels is the (left or right, respectively) repetition of $\mathrm{Dc}(\alpha_{s\ldots t})$.

The part of a 2-block between its rightmost (leftmost) central kernel and its right (left) pre-period is called the *right (left) pseudorepetition* of the 2-block (notation: $\mathrm{RpR}, \mathrm{LpR}$).

Now we can give an example of a 2-block with all its parts marked:

$$\underbrace{\underbrace{\overline{(ddddd)}}_{\text{LpreP}}\,\underbrace{d..d}_{\text{LpR}}}_{\text{left component = LO}}\overline{()()}\underbrace{\underbrace{e..\overline{(eeeee)}}_{\text{LI}}\overline{(c)}\underbrace{\overline{(eeeee)}..e}_{\text{RI}}}_{\text{core}}\overline{()()}\underbrace{\underbrace{d..\overline{(ddddd)}}_{\text{RO}}\underbrace{b\ldots\overbrace{bd..d}^{\text{a pseudo-period}}}_{\text{RpR}}\underbrace{\overline{(bd..db\ldots ddbdb)}}_{\text{RpreP}}}_{\text{right component}}.$$

Here parentheses denote simple kernels and lines above denote kernels. Fig. 5 shows a detailed structure of a 2-block in a more general case.

These are all the notions concerning 2-blocks that are necessary to define continuously periodic evolutions. We have to say another several words concerning left and right pseudorepetitions. Let $\alpha_{i\ldots j}$ be a 2-block that belongs to an evolution $\mathcal{E}$ and such that case I holds at right. Consider then all the 1-blocks between the letters of order 2 in $\mathrm{RpR}(\alpha_{i\ldots j})$. The sequence of evolutions they belong to is denoted by $\mathfrak{E}_R(\alpha_{i\ldots j})$. $\mathfrak{E}_R(\alpha_{i\ldots j})$ ($\mathfrak{E}_L(\alpha_{i\ldots j})$) is considered beginning at left (resp. at right).

Let l be divisible by $\mathrm{B}\mathbf{P}(\mathcal{E})$ and $l-\mathrm{B}\mathbf{P}(\mathcal{E}) > \mathrm{F}(\mathcal{E})$. Consider the concatenation of (right) atoms $\mathrm{RA}_{l-\mathrm{B}\mathbf{P}(\mathcal{E})}(\mathcal{E}_l)\ldots\mathrm{RA}_{l-1}(\mathcal{E}_l)\,\mathrm{RA}_l(\mathcal{E}_l)$. This concatenation does not depend on l. It may contain some pre-periodic letters of order 2, but the same atoms of the $|\Sigma|$-th and further superdescendants of $\mathcal{E}_l$ do not contain them. The same atoms of $\mathcal{E}_{l+|\Sigma|}$ form some sequence of periodic letters of order 2, separated by 1-blocks, and all of them do not depend on l. 1-blocks inside the same atoms of $\mathcal{E}_{l+|\Sigma|+m}$ are exactly the m-th superdescendants of ones inside $\mathcal{E}_{l+|\Sigma|}$, new 1-blocks will no longer arise in that atoms.

All we have said above and Lemma 4 imply the following lemma:

Lemma 6. *Let $\mathcal{E}_l$ be a 2-block such that case I holds at right, and let $\alpha'_{s\ldots t}$ be the corresponding 1-block. Consider all the letters of order 2 inside its right component. These letters, except for not more than Q rightmost ones that form* $\mathrm{RpreP}(\alpha'_{s\ldots t})$*, form a p_l-periodic sequence. Here Q is a constant that depends on Σ and φ only. p_l possibly depends on l, but the sequence $(p_l)_{l=\mathrm{S}(\mathcal{E})}^{\infty}$ is periodic.*

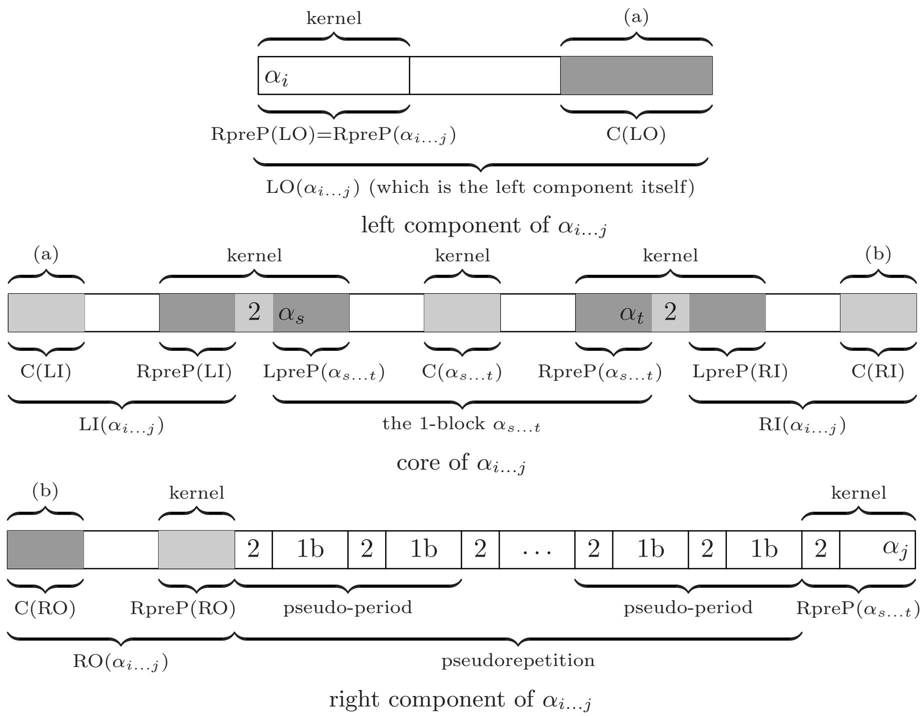

Fig. 5. Detailed structure of the 2-block $\alpha_{i...j}$, where case II holds at left and case I holds at right: 2 denotes a letter of order 2, 1b denotes a 1-block, two letters (a) denote two parts of a single kernel, two letters (b) denote another kernel. Central kernels are filled.

Everything we have said about components of 1-blocks, can be said about this sequence too.

During the evolution (as l grows), $\mathfrak{E}_R(\mathcal{E}_l)$ is prolonged to the right (and is not changed elsewhere). The amount of blocks added at right per iteration is periodic (the period length is $\mathrm{B}\mathbf{P}(\mathcal{E})$*). Thus we can get an infinite sequence, which is denoted by $\mathfrak{E}_R(\mathcal{E})$.*

$\mathfrak{E}_R(\mathcal{E})$ is a periodic sequence if it is considered as sequence of abstract words with known left and right borders. If two 1-blocks inside $\mathfrak{E}_R(\mathcal{E}_l)$ are at the same places in two consecutive periods of $\mathfrak{E}_R(\mathcal{E})$, then one of them is the $\mathrm{B}\mathbf{P}(\mathcal{E})$*-th superdescendant of another.*

The following lemma is a corollary of the periodicities we have noticed above.

Lemma 7. *Let $\mathcal{E}$ be an evolution of 2-blocks, $s \in \mathbb{N}$. The sequences $|\mathcal{E}_{\mathrm{S}(\mathcal{E})}|, \dots, |\mathcal{E}_l|, \dots$ and $|\operatorname{RpR}(\mathcal{E}_{\mathrm{S}(\mathcal{E})})|, \dots, |\operatorname{RpR}(\mathcal{E}_l)|, \dots$ modulo s are periodic.*

5 Continuously Periodic Evolutions

Let $\mathcal{E}$ be an evolution of k-blocks ($k = 1, 2$) or k-semiblocks ($k = 1$). Let case II or III hold at right. Let k_0 be the order of $a := \mathrm{RB}(\mathcal{E}_{\mathrm{S}(\mathcal{E})})$. Then $\varphi(a)$ contains a letter of order $k_0 - 1$. Hence, the part of $\varphi^{\mathrm{B}\mathbf{P}(\mathcal{E})}(\mathrm{RB}(\mathcal{E}_{\mathrm{S}(\mathcal{E})}))$ outside $\mathrm{Dc}^{\mathrm{B}\mathbf{P}(\mathcal{E})}(\mathcal{E}_{\mathrm{S}(\mathcal{E})})$ will be longer than one letter and will start with a again. Thus we can construct an abstract sequence (it is built like a pure morphic sequence and can be prolonged infinitely) starting with a. It is called *the right bounding sequence* of $\mathcal{E}_{\mathrm{S}(\mathcal{E})}$ (notation: RBS). If we consider $\mathcal{E}_{\mathrm{S}(\mathcal{E})+m}$ $(1 \leq m \leq \mathrm{B}\mathbf{P} - 1)$ instead of $\mathcal{E}_{\mathrm{S}(\mathcal{E})}$, we can build another sequence in the similar way. These sequences are also called *right bounding sequences*. These sequences are abstract, they are not occurrences in α. However, there is a beginning of one of these sequences in α at right of each block $\mathcal{E}_l$, and the lengths of these beginnings grow as evolution passes. Moreover, the length of this beginning is $\Theta(l^{k_0})$.

Left bounding sequences are defined in a similar way.

Consider all the evolutions $\mathcal{E}$ of 1-blocks and 1-semiblocks. Consider all the words $\psi(\mathrm{LP}(\mathcal{E}_l))$ and $\psi(\mathrm{RP}(\mathcal{E}_l))$ (where $\mathcal{E}_l$ is stable and case I holds at left and at right, respectively). If one of these encoded periods is *completely* p-periodic itself (where p is less than its length), consider the smallest its complete period instead of it. This (finite due to Lemma 4) set of words is called the set of *admissible periods*. A cyclic shift of an admissible period is also called an admissible period. Since the set of admissible periods is finite, we assume they are enumerated in some way.

The following definition is very significant for the proof.

Definition 1. *Let $\mathcal{E}$ be an evolution of k-blocks ($k = 1, 2$) or k-semiblocks ($k = 1$) such that $|\mathcal{E}_l| \to \infty$ as $l \to \infty$ (it means that either the core contains letters of order > 1 or case I or II holds at left or at right). $\mathcal{E}$ is called* continuously periodic *if for every stable block $\mathcal{E}_l = \alpha_{i\ldots j}$ it is possible to choose some (actually, not more than three) of its kernels $\alpha_{i_1\ldots j_1}, \ldots, \alpha_{i_s\ldots j_s}$ so that:*

1. The total amount of letters of order k in any of the words $\alpha_{j_t\ldots i_{t+1}}$, grows unboundedly as $l \to \infty$ (thus, in particular, no two central kernels can be chosen);

2. All the words $\psi(\alpha_{j_t\ldots i_{t+1}})$, are (left or right) γ-periodic for some admissible periods γ;

3. If $j_s < j$, then case II or III should hold at right (and if $\mathcal{E}$ is an evolution of semiblocks, they should be right *ones), the infinite word $\psi(\alpha_{j_s\ldots j}\,\mathrm{RBS}(\mathcal{E}_l))$ is periodic and its period is admissible,*
And similar condition at left.

In fact, Lemma 4 implies that all the evolutions of 1-blocks or 1-semiblocks are continuously periodic.

After we gave the definition of a continuously periodic evolution, the statements of Propositions 1 and 2 are completely formulated.

6 An Example

Here we give an example of a sequence with complexity $\Theta(n^{3/2})$.

Let $\Sigma = \{1, 2, 3, 4\}$. Consider the following morphism φ: $\varphi(4) = 43, \varphi(3) = 32, \varphi(2) = 21, \varphi(1) = 1$ and the following coding ψ: $\psi(4) = 4, \psi(3) = 3, \psi(2) = \psi(1) = 1$. Then $\varphi^\infty(4) = \alpha = 4332322132212113221211211113\ldots$, and $\psi(\varphi^\infty(4)) = \beta = 4331311131111113111111111113\ldots$ In α there is a unique evolution $\mathcal{E}$ of 2-blocks, whose origin is the occurrence of the empty word between 4 and 3. Case I holds at left, so $|\mathcal{E}_l| = \Theta(l^2)$. It is clear that $\mathcal{E}$ is continuously periodic evolution, since $\psi(\mathcal{E}_l)$ consists of repeating letters 1. Proposition 2 asserts $p_\beta = O(n^{3/2})$. Let us show $p_\beta = \Omega(n^{3/2})$.

Indeed, β consists of alternating letters 3 and 2-blocks $\mathcal{E}_l$. Given a subword $\beta_{i\ldots j}$ of length n, consider the longest 2-block $\beta_{s\ldots t}$ that is completely inside it. If the longest 2-block in another subword $\beta_{i'\ldots j'}$ is of another length or if it is of the same length but starts from the different location in $\beta_{i'\ldots j'}$ than $\beta_{s\ldots t}$ in $\beta_{i\ldots j}$, the words $\beta_{i\ldots j}$ and $\beta_{i'\ldots j'}$ are obviously distinct.

Assume $3n/4 > |\beta_{s\ldots t}| > n/2$. Then there are $\Theta(\sqrt{n})$ possibilities for its length. There are $\Theta(n)$ possibilities for its location in $\beta_{i\ldots j}$, thus we obtain $\Theta(n\sqrt{n})$ different subwords of length n in β. Therefore, $p_\beta(n) = \Omega(n^{3/2})$, and hence $p_\beta(n) = \Theta(n^{3/2})$.

Acknowledgements

The author is kindly grateful to Y. Pritykin for a useful discussion.

References

1. Allouche, J.-P., Shallit, J.: Automatic Sequences. Cambridge University Press, Cambridge (2003)
2. Ferenczi, S.: Complexity of sequences and dynamical systems. Discrete Math. 206(1-3), 145–154 (1999)
3. Nicolas, F.: Master's thesis
4. Pansiot, J.-J.: Complexité des facteurs des mots infinis engendrés par morphimes itérés. In: Paredaens, J. (ed.) ICALP 1984. LNCS, vol. 172, pp. 380–389. Springer, Heidelberg (1984)