On periodicity and complexity of generalized pseudostandard words

Josef Florian[*] and Lubomíra Balková[†]

June 2014

Abstract

Generalized pseudostandard words have been introduced by De Luca et al. in [6]. Recently, they have been studied intensively [4, 7], nevertheless in comparison to the palindromic and pseudopalindromic closure (studied for instance in [1, 3, 5, 6]), only little is known about the generalized pseudopalindromic closure and the associated generalized pseudostandard words. We present two new results concerning these words. The first one is a necessary and sufficient condition for their periodicity. The second result is a counterexample to Conjecture 43 from [7] that estimated the complexity of binary generalized pseudostandard words as $\mathcal{C}(n) \leq 4n$ for all sufficiently large n.

1 Basics from combinatorics on words

We restrict ourselves to the binary alphabet $\{0,1\}$. A (finite) word w over $\{0,1\}$ is any finite binary sequence. Its length $|w|$ is the number of letters it contains. The empty word – the neutral element for concatenation of words – is denoted ε and its length is set $|\varepsilon| = 0$. An infinite word $\mathbf{u}$ over $\{0,1\}$ is any binary infinite sequence. A finite word w is a factor of the infinite word $\mathbf{u} = u_0u_1u_2\ldots$ with $u_i \in \{0,1\}$ if there exists an index $i \geq 0$ such that $w = u_iu_{i+1}\ldots u_{i+|w|-1}$. The symbol $\mathcal{L}(\mathbf{u})$ is used for the set of factors of $\mathbf{u}$ and is called the language of $\mathbf{u}$, similarly $\mathcal{L}_n(\mathbf{u})$ stands for the set of factors of $\mathbf{u}$ of length n. A left special factor of a binary infinite word $\mathbf{u}$ is any factor v such that both $0v$ and $1v$ are factors of $\mathbf{u}$. A right special factor is defined analogously. Finally, a factor of $\mathbf{u}$ that is both right and left special is called a bispecial. We distinguish the following types of bispecials over $\{0,1\}$:

- A weak bispecial w satisfies that only $0w1$ and $1w0$, or only $1w0$ and $0w1$ are factors of $\mathbf{u}$.
- A strong bispecial w satisfies that all $0w0$, $0w1$, $1w0$ and $1w1$ are factors of $\mathbf{u}$.
- We do not use a special name for bispecials that are neither weak nor strong.

Let $w \in \mathcal{L}(\mathbf{u})$. A left extension of w is any word $aw \in \mathcal{L}(\mathbf{u})$, where $a \in \{0,1\}$, and a right extension is defined analogously. The set of left, resp. right extensions of w is denoted $\mathrm{Lext}(w)$, resp. $\mathrm{Rext}(w)$. The (factor) complexity of $\mathbf{u}$ is the map $\mathcal{C}_\mathbf{u} : \mathbb{N} \to \mathbb{N}$ defined as

$$\mathcal{C}_\mathbf{u}(n) = \text{the number of factors of } \mathbf{u} \text{ of length } n.$$

[*]corresponding author, Czech Technical University in Prague, funded by the grant SGS14/205/OHK4/3T/14
[†]Czech Technical University in Prague, funded by Czech Science Foundation grant GAČR 13-03538S

An involutory antimorphism is a map $\vartheta : \{0,1\}^* \to \{0,1\}^*$ such that for every $v, w \in \{0,1\}^*$ it holds $\vartheta(vw) = \vartheta(w)\vartheta(v)$ and moreover ϑ^2 equals identity. It is clear that in order to define an antimorphism, it suffices to provide letter images. There are only two involutory antimorphisms over the alphabet $\{0,1\}$: the reversal (mirror) map R satisfying $R(0) = 0, R(1) = 1$, and the exchange antimorphism E given by $E(0) = 1, E(1) = 0$. We use the notation $\overline{0} = 1$ and $\overline{1} = 0$, $\overline{E} = R$ and $\overline{R} = E$. A finite word w is an R-palindrome if $w = R(w)$, and w is an E-palindrome if $w = E(w)$. The palindromic closure w^R of a word w is the shortest R-palindrome having w as prefix. Similarly, the pseudopalindromic closure w^E of w is the shortest E-palindrome having w as prefix.

2 Definition of generalized pseudostandard words

Generalized pseudostandard words form a generalization of infinite words obtained by the palindromic, resp. pseudopalindromic closure; such constructions were described and studied in [1, 3, 5, 6]. Generalized pseudostandard words are given by two directive sequences in the following way: Let $\Delta = \delta_1\delta_2\ldots$, where $\delta_i \in \{0,1\}$, and $\Theta = \vartheta_1\vartheta_2\ldots$, where $\vartheta_i \in \{E, R\}$, be two directive sequences. The prefixes w_n of the infinite word $\mathbf{u} = \mathbf{u}(\Delta, \Theta)$ are obtained from the recurrence relation:

$$w_{n+1} = (w_n\delta_n)^{\vartheta_n},$$

$$w_0 = \varepsilon.$$

The sequence (Δ, Θ) is called the bidirective sequence of the generalized pseudostandard word $\mathbf{u}(\Delta, \Theta)$. The sequence of prefixes w_n does not have to contain all R- and E-palindromic prefixes. However, if it is the case, we say that the bidirective sequence is normalized. In [7], the authors provide an algorithm for normalization of any bidirective sequence in such a way that the obtained generalized pseudostandard word remains unchanged.

Theorem 1. *Let $\Lambda = (\Delta, \Theta)$ be a bidirective sequence. Then there exists a normalized bidirective sequence $\widetilde{\Lambda} = (\widetilde{\Delta}, \widetilde{\Theta})$ such that $\mathbf{u}(\Delta, \Theta) = \mathbf{u}(\widetilde{\Delta}, \widetilde{\Theta})$.*

Moreover, in order to normalize the sequence Λ, it suffices firstly to execute the following changes of its prefix (if it is of the corresponding form):

- $(a\bar{a}, RR) \to (a\bar{a}a, RER)$,
- $(a^i, R^{i-1}E) \to (a^i\bar{a}, R^iE)$ *for* $i \geq 1$,
- $(a^i\bar{a}\bar{a}, R^iEE) \to (a^i\bar{a}\bar{a}a, R^iERE)$ *for* $i \geq 1$,

and secondly to replace step by step from left to right every factor of the form:

- $(ab\bar{b}, \vartheta\overline{\vartheta}\overline{\vartheta}) \to (ab\bar{b}b, \vartheta\overline{\vartheta}\vartheta\overline{\vartheta})$,

where $a, b \in \{0,1\}$ and $\vartheta \in \{E, R\}$.

3 Periodicity of generalized pseudostandard words

We have prepared everything to introduce the first result – a sufficient and necessary condition for periodicity of generalized pseudostandard words. Let us underline that such words cannot be eventually periodic since they are recurrent by construction.

Theorem 2. *Let $\Lambda = (\Delta, \Theta)$ be a bidirective sequence. The generalized pseudostadard word $\mathbf{u}(\Delta, \Theta)$ is periodic if and only if the following condition is satisfied:*

$$(\exists a \in \{0,1\})(\exists \vartheta \in \{E,R\})(\exists n_0 \in \mathbb{N})(\forall n > n_0)(\delta_{n+1} = a \Leftrightarrow \vartheta_n = \vartheta), \tag{1}$$

where $\Delta = \delta_1\delta_2\ldots$ and $\Theta = \vartheta_1\vartheta_2\ldots$.

We will not provide the proof here, however let us list two normalization properties that play an important role in the proof (and that can be easily verified by the reader):

1. Let Λ be a normalized bidirective sequence satisfying the condition (1), then this sequence is eventually periodic.
2. Moreover, if the bidirective sequence Λ satisfies (1), then the sequence $\widetilde{\Lambda}$ obtained by its normalization satisfies (1), too.

Example 1. *Let us show an example of a periodic generalized pseudostandard word. Assume $\Lambda = ((011)^\omega, (EER)^\omega)$, where ω denotes an infinite repetition. The condition (1) is met since E is always followed by 1 and R by 0. The bidirective sequence Λ is not normalized. For instance, the R-palindromic prefix 0110 is not equal to any prefix w_n:*

$$\begin{aligned} w_1 &= 01 \\ w_2 &= 011001 \\ w_3 &= 01100110. \end{aligned}$$

The sequence Λ can be normalized using Theorem 1: $\widetilde{\Lambda} = (01(10)^\omega, (RE)^\omega)$. Let us write down again the first prefixes $\widetilde{w_n}$:

$$\begin{aligned} \widetilde{w_1} &= 0 \\ \widetilde{w_2} &= 01 \\ \widetilde{w_3} &= 0110 \\ \widetilde{w_4} &= 011001 \\ \widetilde{w_5} &= 01100110. \end{aligned}$$

It can be easily verified by the reader that $\widetilde{\Lambda}$ satisfies the condition (1), too, and $\mathbf{u}((011)^\omega, (EER)^\omega) = \mathbf{u}(01(10)^\omega, (RE)^\omega) = (0110)^\omega$.

4 Conjecture $4n$

The second result is a counterexample to Conjecture $4n$ (Conjecture 43 stated in [7]):

Conjecture 1. *For every binary generalized pseudostandard word $\mathbf{u}$ there exists $n_0 \in \mathbb{N}$ such that $\mathcal{C}_{\mathbf{u}}(n) \leq 4n$ for all $n > n_0$.*

We have found a counterexample $\mathbf{u}_p = \mathbf{u}(1^\omega, (EERR)^\omega)$ that verifies $\mathcal{C}_{\mathbf{u}_p}(n) > 4n$ for all $n > 10$. Let us write down its first prefixes w_n:

$$\begin{aligned} w_1 &= 10 \\ w_2 &= 1010 \\ w_3 &= 10101 \\ w_4 &= 1010110101 \\ w_5 &= 1010110101100101001010 \\ w_6 &= 1010110101100101001010110101100101001010 \end{aligned}$$

Again, we will not prove here that $\mathbf{u}_p$ is indeed a counterexample to Conjecture $4n$. We will instead mention the basic steps of the proof.

In order to get the complexity of $\mathbf{u}_p$, we used the well-known formula for the second difference of complexity [2].

$$\Delta^2\mathcal{C}_{\mathbf{u}}(n) = \Delta\mathcal{C}_{\mathbf{u}}(n+1) - \Delta\mathcal{C}_{\mathbf{u}}(n) = \sum_{w\in\mathcal{L}_n(\mathbf{u})} B(w),$$

where

$$B(w) = \#\{awb \mid a,b \in \{0,1\}, awb \in \mathcal{L}(\mathbf{u})\} - \#\mathrm{Rext}(w) - \#\mathrm{Lext}(w) + 1.$$

It is readily seen that for any factor of a binary infinite word $\mathbf{u}$ it holds:

- $B(w) = 1$ if and only if w is a strong bispecial.
- $B(w) = -1$ if and only if w is a weak bispecial.
- $B(w) = 0$ otherwise.

We managed to find all weak bispecial factors and enough strong bispecial factors so that it provided us with a lower bound on the second difference of complexity that lead to the strict lower bound equal to $4n$ on the complexity of $\mathbf{u}_p$.

We do not have enough observations to state a new conjecture, nevertheless in our computer experiments, we have on one hand several examples – including the word $\mathbf{u}_p$ – where $\limsup \frac{\mathcal{C}(n)}{n}$ seems to be greater than 4. On the other hand, in all our examples $\limsup \frac{\mathcal{C}(n)}{n}$ is likely less or equal to 5.

References

[1] M. Bucci, A. de Luca, A. De Luca, L. Zamboni, *On Some Problems Related to Palindrome Closure*, RAIRO – Theoretical Informatics and Applications **42** (2008), 679–700

[2] J. Cassaigne, *Complexité et facteurs spéciaux*, Bull. Belg. Math. Soc. Simon Stevin **4** (1997), 67–88

[3] X. Droubay, J. Justin, G. Pirillo, *Episturmian Words and Some Constructions of de Luca and Rauzy*, Theoretical Computer Science **255** (2001), 539–553

[4] D. Jamet, G. Paquin, G. Richomme, L. Vuillon, *On the Fixed Points of the Iterated Pseudopalindromic Closure*, Theoretical Computer Science **412** (2011), 2974–2987

[5] A. de Luca, *Sturmian Words: Structure, Combinatorics, and their Arithmetics*, Theoretical Computer Science **183** (1997), 45–82

[6] A. de Luca, A. De Luca, *Pseudopalindromic Closure Operators in Free Monoids*, Theoretical Computer Science **362** (2006), 282–300

[7] A. B. Massé, G. Paquin, H. Tremblay, L. Vuillon, *On Generalized Pseudostandard Words over Binary Alphabet*, Journal of Int. Sequences **16** (2013), Article 13.2.11

[8] W. A. Stein et al., *Sage Mathematics Software*, The Sage Developement Team (2010), www.sagemath.org