Theoretical Computer Science 412 (2011) 493–497

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Note

Complexity and palindromic defect of infinite words

S. Brlek, C. Reutenauer *

LaCIM, Université du Québec à Montréal, C. P. 8888 Succursale "Centre-Ville", Montréal (QC), Canada H3C 3P8

ARTICLE INFO

Article history:
Received 18 April 2010
Received in revised form 1 September 2010
Accepted 9 November 2010
Communicated by M. Crochemore

Keywords:
Complexity
Palindromic complexity
Palindromic defect

ABSTRACT

In this note, we state a conjecture, and prove it in the periodic case, which is an equality relating the number of factors and palindromic factors of infinite words. This equality establishes a link between two inequalities, one due to Droubay, Justin and Pirillo, and the other to Baláži, Masáková and Pelantová, by means of the palindromic defect.

1. Introduction

Let W be an infinite word, that is, W is a sequence over a finite alphabet A. We write $W = W_0W_1W_2\ldots$, where $W_i \in A$. The *complexity* of W is the function $C_W : \mathbb{N} \to \mathbb{N}$ with $C_W(n) =$ number of factors (that is, connected subwords) of length n in W. This function has been widely studied for a long time, an account of which may be found in [1]. More recently, the subclass of palindromic factors appears in various studies, and the *palindromic complexity* was introduced as the function $P_W : \mathbb{N} \to \mathbb{N}$, with $P_W(n) =$ number of palindromic factors of length n of W; see [2].

For a finite word w, there is a simple but deep inequality, due to Droubay, Justin and Pirillo ([11] Proposition 2), that relates the number of its palindromic factors P_w to its length:

$$P_w \leq |w| + 1. \tag{1}$$

The difference between the two is called *defect* [7] (or *palindromic defect*), that is, the defect $\mathcal{D}(w)$ of w is:

$$\mathcal{D}(w) = 1 + |w| - P_w. \tag{2}$$

It follows from [11] that if u is a factor of w, then $\mathcal{D}(u) \leq \mathcal{D}(w)$. Hence we may, as in [7] define the *defect* of an infinite word W as

$$\mathcal{D}(W) = \text{the maximum (finite or } \infty\text{) of the defect of the factors of } W.$$

Thus, the Droubay–Justin–Pirillo inequality means equivalently that the defect of a word (finite or infinite) is always nonnegative.

There is another nice inequality, due to Baláži et al. [4], which relates complexity and palindromic complexity: if W is an infinite word whose set of factors is closed under mirror image,[1] then for any $n \in \mathbb{N}$,

$$P_W(n) + P_W(n+1) \leq C_W(n+1) - C_W(n) + 2. \tag{3}$$

* Corresponding author. Tel.: +1 514 987 3000.
E-mail addresses: Brlek.Srecko@uqam.ca (S. Brlek), Reutenauer.Christophe@uqam.ca, Christophe.Reutenauer@uqam.ca (C. Reutenauer).

[1] In [4], the hypothesis "uniformly recurrent" is used; but in the proof, "recurrent" is enough. Moreover, if the set of factors of W is closed under mirror image, then W is recurrent, as follows from an argument that appears in the proof of [8, Proposition 9].

0304-3975/$ – see front matter
doi:10.1016/j.tcs.2010.11.025

It is useful to introduce the function

$$T_W(n) = C_W(n+1) - C_W(n) + 2 - P_W(n) - P_W(n+1). \qquad (4)$$

Thus the inequality of Baláži, Masáková and Pelantová means that $T_W(n) \geq 0$. Hereafter we often omit the subscript in functions C, P and T when the context is clear.

The conjecture given below relates the two Inequalities (1) and (3) reported above.

Conjecture 1. *Let W be an infinite word whose set of factors is closed under mirror image. Then*

$$2\mathcal{D}(W) = \sum_{n=0}^{\infty} T_W(n). \qquad (5)$$

Note that if $\mathcal{D}(W)$ is finite, then this equality means that the right summation is finite; that is, for n large enough, we have

$$T_W(n) = 0.$$

Theorem 2. *The conjecture is true if W is periodic.*

The theorem will be proved in Section 3.

It was suggested to us by one of the two referees that the "conjecture could be extended without much pain to the case W is a finite palindrome". This is indeed the case, even for non-palindromic finite words. The only change that has to be made is that the summation in the formula of the conjecture has to be restricted from 0 to the length of the word. We have the following result, whose proof is easy.

Proposition 3. *For every finite word w of length k we have*

$$2\mathcal{D}(w) = \sum_{n=0}^{k} T_w(n).$$

Proof. First, observe that $C(k+1) = P(k+1) = 0$. Then we have

$$\begin{aligned}\sum_{n=0}^{k} T_w(n) &= C(k+1) - C(k) + 2 - P(k+1) - P(k) + C(k) - C(k-1) + 2 - P(k) - P(k-1)\\ &\quad \vdots\\ &\quad + C(2) - C(1) + 2 - P(2) - P(1) + C(1) - C(0) + 2 - P(1) - P(0)\\ &= -1 + 2(k+1) - 2P_w + 1\\ &= 2\mathcal{D}(w)\end{aligned}$$

where the last equality is equivalent to Eq. (2). □

Although each infinite word is the limit, in some sense, of its finite prefixes, it is not clear how one could deduce the conjecture from the finite case established by Proposition 3.

Note also that the positivity of $T(n)$ is not used in the proposition: the examples given below show that it is not necessarily positive for finite words.

Examples. Let $w = abbabaab$. Then we have $\mathcal{D}(w) = 0$ and

n	0	1	2	3	4	5	6	7	8	9
C	1	2	4	6	5	4	3	2	1	0
P	1	2	2	2	2	0	0	0	0	0
T	0	0	0	−3	−1	1	1	1	1	

For $w = aabcaacbc$ we have $\mathcal{D}(w) = 2$ and

n	0	1	2	3	4	5	6	7	8	9	10
C	1	3	6	7	6	5	4	3	2	1	0
P	1	3	1	1	1	0	1	0	0	0	0
T	0	1	1	−1	0	0	0	1	1	1	

2. A lemma on the product of two palindromes

It is well known and easy to verify that if a word v is equal to the product of two palindromes, then all its conjugates are. More precisely, this word v is conjugate to a word w which is one of the three following forms:

(1) w is an *even* palindrome (that is of even length);

(2) w is an *odd* palindrome;
(3) w is the product of a letter by an odd palindrome.

This fact was observed in [3] (see also [7] p. 299) and may be easily established by the inspection of the axial symmetry of the circular word associated with v. If w is primitive, these three cases are mutually exclusive.

Lemma 4. *Let w be a primitive word which is the product of two palindromes. Let $|w| = k$ and $W = w^\infty$.*

(i) *For $n \geq k$*

$$C_W(n) = k,$$

and

$$P_W(n) + P_W(n+1) = 2.$$

(ii) *Let α_i be the number of palindromic factors of length i in w^2. Suppose that w is of the form* (1), (2) *or* (3). *Then,*

$$2\sum_{i=k}^{2k} \alpha_i = k + P_W(k) + 2.$$

Remark. It will follow from the proof of the theorem, that the extra hypothesis in (ii) is actually not necessary.

Proof. We use the following observations in the proof: the circular word associated with w has a unique axial symmetry; moreover, each palindromic factor of length $\geq k$ of w^∞ is compatible with this axis and has a unique occurrence in this circular word.

(i) The first equality follows from the primitivity of w. For the second, we may replace w by any conjugate, since this does not change the set of factors of w^∞. We choose w of type (1), (2) or (3).

If w is of type (1), then $w = u\widetilde{u}$ and $w^\infty = u\widetilde{u}u\widetilde{u}u\widetilde{u}\cdots$. The palindromic factors of w^∞ of length $\geq k$ all have center $u\widetilde{u}$ or $\widetilde{u}u$. Therefore, for $n \geq k$, we have

$$P_W(n) = \begin{cases} 2 & \text{if } n \text{ is even,} \\ 0 & \text{if } n \text{ is odd.} \end{cases}$$

If w is of type (2), then $w = ua\widetilde{u}$, for some letter a. It follows that

$$w^\infty = ua\widetilde{u}ua\widetilde{u}ua\widetilde{u}\cdots$$

and the palindromic factors of w^∞ of length $\geq k$ all have $ua\widetilde{u}$ or $\widetilde{u}u$ as center. Therefore, for $n \geq k$, $P_W(n) = 1$.

If w is of type (3), then $w = bua\widetilde{u}$ where a, b are letters, so that

$$w^\infty = bua\widetilde{u}bua\widetilde{u}bua\widetilde{u}\cdots.$$

The palindromic factors of w^∞ of length $\geq k$ all have center $ua\widetilde{u}$ or $\widetilde{u}bu$. Therefore, for $n \geq k$, we have

$$P_W(n) = \begin{cases} 2 & \text{if } n \text{ is odd,} \\ 0 & \text{if } n \text{ is even.} \end{cases}$$

Summarizing, we have in all the three cases $P_W(n) + P_W(n+1) = 2$, for $n \geq k$.

(ii) Observe first that w^2 cannot contain two distinct palindromic factors of length $> k$: indeed, otherwise, the circular word defined by w would have two distinct symmetry axes, and w would not be primitive (since the product of two axial symmetries is a rotation). We use this observation in all the three cases given below. Let $\alpha = \sum_{i=k}^{2k} \alpha_i$.

Suppose that w is of type (1), that is, $w = u\widetilde{u}$. Since w^2 and w^∞ have the same factors of length $\leq k$, by (i), $P_W(k) = 2$. Moreover, there is no palindromic factor in $w^2 = u\widetilde{u}u\widetilde{u}$ for any odd length $\geq k$, and only one for each length $k+2, k+4, \ldots, 2k$. Hence $\alpha = 2 + \frac{k}{2}$, so that $2\alpha = k+4$, and $k + P_W(k) + 2 = k+4$ as well.

When w is of type (2), $P_W(k) = 1$ by (i). Moreover, there is one palindromic factor of $w^2 = ua\widetilde{u}ua\widetilde{u}$ for each length $k+1, k+3, \ldots, 2k$, whose center is $a\widetilde{u}ua$. Thus, $\alpha = 1 + \frac{k+1}{2}$ so that $2\alpha = k+3$. Moreover $k + P_W(k) + 2 = k+3$, as well.

Finally, if w is of type (3), that is $w = bua\widetilde{u}$, then $P_W(k) = 0$ by (i). There are two palindromic factors of length $k+1$ in $w^2 = bua\widetilde{u}bua\widetilde{u}$, namely $bua\widetilde{u}b$ and $a\widetilde{u}bua$. Moreover, there is one palindromic factor of w^2 for each length $k+3, k+5, \ldots, 2k-1$, of center $a\widetilde{u}bua$. Thus $\alpha = 2 + \frac{k-2}{2}$, which implies that $2\alpha = k+2$. On the other hand, $k + P_W(k) + 2 = k+2$, too. □

3. Proof of theorem

Suppose first that W has infinitely many palindromes. Then $W = w^\infty$ for some (finite) primitive word w which is a product of two palindromes ([7] Theorem 4). We may replace w by one of its conjugates, since this does not change the factors of w^∞. So we may assume that w is of one of the three forms (1), (2) or (3), and apply the lemma. Now, it follows from [7] Corollary 8, that $\mathcal{D}(W) = \mathcal{D}(w^2)$. Since each factor of length $\leq k$ of W is also a factor of w^2, we have from Lemma 4 that twice the number of palindromic factors of w^2 is

$$2\sum_{n=0}^{k-1} P_W(n) + 2\sum_{n=k}^{2k} \alpha_n = 2\sum_{n=0}^{k-1} P_W(n) + k + P_W(k) + 2.$$

Thus twice the defect of w^2 is

$$\begin{aligned} 2\mathcal{D}(w^2) &= 2(|w^2| + 1) - 2\sum_{n=0}^{k-1} P_W(n) - k - P_W(k) - 2 \\ &= 3k - 2\sum_{n=0}^{k-1} P_W(n) - P_W(k). \end{aligned}$$

On the other hand, we have by the lemma

$$\begin{aligned} \sum_{n=0}^{\infty} T_W(n) &= \sum_{n=0}^{k-1} T_W(n) \\ &= \sum_{n=0}^{k-1} (C_W(n+1) - C_W(n) + 2 - P_W(n) - P_W(n+1)) \\ &= \sum_{n=0}^{k-1} (C_W(n+1) - C_W(n)) + 2k - \sum_{n=0}^{k-1} (P_W(n) + P_W(n+1)) \\ &= C_W(k) - C_W(0) + 2k - 2\sum_{n=0}^{k-1} P_W(n) + P_W(0) - P_W(k) \\ &= k - 1 + 2k - 2\sum_{n=0}^{k-1} P_W(n) + 1 - P_W(k) \\ &= 3k - 2\sum_{n=0}^{k-1} P_W(n) - P_W(k), \end{aligned}$$

which implies the theorem in this case.

Suppose that W has only finitely many palindromic factors. Then by [7] Theorem 4, the set of its factors is not closed under mirror image, since W is periodic. Hence this case does not happen. □

4. Other evidence for the conjecture

Besides the theorem, there is some evidence for the conjecture. Indeed, according to the terminology in [7], an infinite word is *full* if its defect is 0. It is shown in [9,12] (where full words are called *rich* words) that an infinite word is full if and only if $T(n) = 0$ for any n. Hence the conjecture is true for the class of full words (which includes Sturmian and episturmian words).

Infinite words of defect 0 and periodic infinite words of finite defect all satisfy the conjecture. We have unfortunately no other example of infinite words of positive finite defect. The conjecture in [5] states that there are no aperiodic infinite words which are fixed point of morphisms and which have positive finite defect, except the periodic ones. Note that, by Pansiot's theorem, the complexity $C(n)$ of an infinite word which is a fixed point of a morphism is either bounded, or grows like n, $n \log\log n$, $n \log n$ or n^2 [13] page 231 (see also [10]), so that, excluding the bounded and linear cases, the difference $C(n+1) - C(n)$ tends to infinity; and that, by the theorem of Damanik–Zare, see [2] Theorem 7, the palindromic complexity of these words is bounded. Putting all these together, we obtain (under the conjecture of [5]) that our conjecture still holds for morphic words whose factors are closed under mirror image.

Here is another example in favour of the conjecture. Consider the Thue–Morse word M. It is well known that the set of its factors is closed under mirror image; indeed, by definition $M = \mu^\infty(a)$, where μ is the morphism sending a onto ab and b onto ba. Since $\mu^2(a) = abba$ and $\mu^2(b) = baab$, μ^2 sends each palindrome onto a palindrome. Now M is the limit of the words $\mu^{2k}(a)$, hence M has arbitrarily long palindromic prefixes. Consequently, the set of its factors is closed by mirror image.

The complexity $C(n)$ of the Thue–Morse word and its palindromic complexity $P(n)$ satisfy: $C(0) = 1, C(1) = 2, C(2) = 4, C(3) = 6$, and if $n \geq 2$, $C(2n+1) = 2C(n+1)$, $C(2n) = C(n) + C(n+1)$, see [6]; moreover, for any $n \geq 0$, $P(2n+1) = 0$, $P(0) = 1$, $P(2) = 2$, $P(4) = 2$ and $P(2n) = P(n) + P(n+1)$ if $n \geq 3$, see [5]. It is then easy to verify that the function $T(n) = T_M(n)$ satisfies $T(0) = T(1) = T(2) = 0$, $T(3) = T(4) = 2$, for any $n \geq 2$, $T(2n+1) = T(n+1)$, and if $n \geq 3$, $T(2n) = T(n)$. This implies that $T(n) = 2$ for any $n \geq 3$. Moreover, the defect of the Thue–Morse word is infinite [5], which shows that it satisfies the conjecture.

It should be noticed that closure under mirror image cannot be dropped, as shown by the following example. Indeed, consider a word $W \in A^*$ such that Eq. (5) is satisfied. Let $W' = \alpha W$ for some letter $\alpha \notin A$. Moreover, we have $C_{W'}(0) = C_W(0)$, $C_{W'}(n) = C_W(n) + 1$ for all $n > 0$, and $P_{W'}(n) = P_W(n)$ for all $n \neq 1$, with $P_{W'}(1) = P_W(1) + 1$. A straightforward computation shows that $T_{W'}(n) = T_W(n)$, $\forall n \geq 2$, while $T_{W'}(0) = T_W(0)$, $T_{W'}(1) = T_W(1) - 1$. Therefore

$$\sum_{n=0}^{\infty} T_{W'}(n) = -1 + \sum_{n=0}^{\infty} T_W(n).$$

On the other hand $\mathcal{D}(\alpha W) = \mathcal{D}(W)$, so that $2\mathcal{D}(W') - 1 = \sum_{n=0}^{\infty} T_{W'}(n)$.

Similarly, paperfolding sequences and generalized Rudin–Shapiro sequences satisfy our conjecture, for their palindromic complexity eventually vanishes (see [2] Theorem 4 page 14) and their complexity is more than linear.

As a final remark, note that the formula of the conjecture also holds for the famous Kolakoski word $K = 2211212212211211221 2\cdots$, for which it is not known whether its set of factors is closed under mirror image or not (see the table given below). The function $T(n)$ is positive for any $n \geq 3$, as may be deduced from the work in [8].

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	$\cdots$
C	1	2	4	6	10	14	18	26	34	42	50	62	78	94	110
P	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2
T	0	0	0	2	2	2	6	6	6	6	10	14	14	14	

The results and the examples in this article show that the function $T(n)$ (deduced from the Baláži, Masáková and Pelantová inequality) seems to be of some importance in the study of finite and infinite words.

Acknowledgements

The authors thank the two referees for useful suggestions. In particular, we have added more evidence for the conjecture. The first author's research was supported by an NSERC grant (Canada).

References

[1] J.-P. Allouche, Sur la complexité des suites infinies, Bull. Belg. Math. Soc. Simon Stevin 1 (1994) 133–143.
[2] J.-P. Allouche, M. Baake, J. Cassaigne, D. Damanik, Palindrome complexity, Theoret. Comput. Sci. 292 (2003) 9–31.
[3] P. Auger, Analyse d'équations combinatoires en théorie des espèces, Ph.D. Dissertation, Université du Québec à Montréal, Publications du LaCIM 26, 2000.
[4] P. Baláži, Z. Masáková, E. Pelantová, Factor versus palindromic complexity of uniformly recurrent infinite words, Theoret. Comput. Sci. 380 (2007) 266–275.
[5] A. Blondin Massé, S. Brlek, A. Garon, S. Labbé, Combinatorial properties of f-palindromes in the Thue-Morse sequence, Pure Math. Appl. (PU.M.A.) 19 (2–3) (2008) 39–52.
[6] S. Brlek, Enumeration of factors in Thue-Morse word, Discrete Appl. Math. 24 (1989) 83–96.
[7] S. Brlek, S. Hamel, M. Nivat, C. Reutenauer, On the palindromic complexity of infinite words, Internat. J. Found. Comput. Sci. 15 (2) (2004) 293–306.
[8] S. Brlek, A. Ladouceur, A note on differentiable palindromes, Theoret. Comput. Sci. 302 (1–3) (2003) 167–178.
[9] M. Bucci, A. De Luca, A. Glen, L.Q. Zamboni, A connection between palindromic and factor complexity using return words, Adv. in Appl. Math. 42 (2009) 60–74.
[10] C. Choffrut, J. Karhumäki, Combinatorics on words, in: A. Salomaa, G. Rozenberg (Eds.), Handbook of Formal Languages, vol. 1, Springer-Verlag, 1997, pp. 329–438.
[11] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, Theoret. Comput. Sci. 255 (2001) 539–553.
[12] A. de Luca, A. Glen, L. Zamboni, Rich Sturmian and Trapezoidal words, Theoret. Comput. Sci. 407 (2008) 569–573.
[13] J.-J. Pansiot, Bornes inférieures sur la complexité des facteurs des mots infinis engendrś par morphismes itérés, Lecture Notes in Comput. Sci. 166 (1984) 230–240.