Discrete Mathematics 310 (2010) 109–114

Contents lists available at ScienceDirect

Discrete Mathematics

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

Total palindrome complexity of finite words

Mira-Cristiana Anisiu [a,*], Valeriu Anisiu [b], Zoltán Kása [c]

[a] Tiberiu Popoviciu Institute of Numerical Analysis, Romanian Academy, Cluj-Napoca, Romania

[b] Babeş-Bolyai University of Cluj-Napoca, Faculty of Mathematics and Computer Science, Department of Mathematics, Romania

[c] Sapientia University Cluj-Napoca, Department of Mathematics and Informatics Tg. Mureş, Romania

ARTICLE INFO

Article history:
Received 14 July 2008
Received in revised form 3 August 2009
Accepted 6 August 2009
Available online 20 August 2009

Keywords:
Palindrome complexity
Finite words

ABSTRACT

The palindrome complexity function pal_w of a word w attaches to each $n \in \mathbb{N}$ the number of palindromes (factors equal to their mirror images) of length n contained in w. The number of all the nonempty palindromes in a finite word is called the total palindrome complexity of that word. We present exact bounds for the total palindrome complexity and construct words which have any palindrome complexity between these bounds, for binary alphabets as well as for alphabets with the cardinal greater than 2. Denoting by $M_q(n)$ the average number of palindromes in all words of length n over an alphabet with q letters, we present an upper bound for $M_q(n)$ and prove that the limit of $M_q(n)/n$ is 0. A more elaborate estimation leads to $M_q(n) = O(\sqrt{n})$.

1. Introduction and notations

Let an alphabet A with $q \geq 1$ letters be given. The free monoid A^* generated by A is the set of all finite words on A. Let $w = a_1 \ldots a_n$ be a word; the integer n is the *length* of w and is denoted by $|w|$. The empty word is denoted by ε and its length is 0. The word $u = a_i \ldots a_j$, $1 \leq i \leq j \leq n$ is a *factor* (or *subword*) of w; if $i = 1$ it is called a *prefix*, and if $j = n$ a *suffix* of w. The *reversal* (or the *mirror image*) of w is denoted by $\widetilde{w} = a_n \ldots a_1$. A word which is equal to its mirror image is called a *palindrome*.

For the q-letter alphabet A, let A^n be the set of all words of length n over A. We denote by PAL_w the set of all factors in the word w which are nonempty palindromes, and by $\mathrm{PAL}_w(n) = \mathrm{PAL}_w \cap A^n$ the set of the palindromes of length n contained in w. The (infinite) set of all palindromes over the alphabet A is denoted by PAL_A, while $\mathrm{PAL}_A(n) = \mathrm{PAL}_A \cap A^n$ is the set of all palindromes of length n over the alphabet A.

The *palindrome complexity function* pal_w of a finite or infinite word w attaches to each $n \in \mathbb{N}$ the number of palindrome factors of length n in w, hence

$$\mathrm{pal}_w(n) = \#\mathrm{PAL}_w(n).$$

Palindromes in infinite words are widely studied. A nonexhaustive list of these papers contains [3,9,4,1,8,6] and [7]. In [2] some properties related to the palindrome complexity of finite words are considered.

The *total palindrome complexity* of a finite word $w \in A^*$ is equal to the number of all nonempty palindrome factors of w, i.e.:

$$P(w) = \sum_{n=1}^{|w|} \mathrm{pal}_w(n).$$

This is similar to the total complexity of words (see [12–15] for finite words, [11] for infinite words).

* Corresponding author.
E-mail addresses: mira@math.ubbcluj.ro (M.-C. Anisiu), anisiu@math.ubbcluj.ro (V. Anisiu), kasa@ms.sapientia.ro (Z. Kása).

0012-365X/$ – see front matter
doi:10.1016/j.disc.2009.08.002

If A is an alphabet with q letters, we define the *average number* $M_q(n)$ *of palindromes* contained in all words of length n by

$$M_q(n) = \frac{\sum_{w \in A^n} P(w)}{q^n}.$$

In Section 2 we determine all the possible values of the total palindrome complexity in a constructive way. In Section 3 we show that $\lim_{n\to\infty} M_q(n)/n = 0$ and, moreover, $M_q(n) = O(n^{1/2})$.

2. Values of the total palindrome complexity

An upper bound for the total palindrome complexity was given in [10], which is one unit greater than in Proposition 1, due to the fact that the empty palindrome was counted there too. We remind that we consider in $P(w)$ only nonempty palindromes. For the sake of completeness we give a direct proof of this result.

Proposition 1. *The total palindrome complexity $P(w)$ of any finite word w satisfies $P(w) \leq |w|$.*

Proof. We proceed by induction on the length n of the word w. For $n = 1$ we have $P(w) = 1$.

We consider $n \geq 2$ and suppose that the assertion holds for all words of length $n - 1$. Let $w = a_1a_2 \ldots a_n$ be a word of length n and $u = a_1a_2 \ldots a_{n-1}$ its prefix of length $n - 1$. By the induction hypothesis it is true that $P(u) \leq n - 1$.

If $a_n \neq a_j$ for each $j \in \{1, 2, \ldots n - 1\}$, the only palindrome in w which is not in u is a_n, hence $P(w) = P(u) + 1 \leq n$.

If there is an index j, $1 \leq j \leq n - 1$ such that $a_n = a_j$, then $P(w) > P(u)$ if and only if w has suffixes which are palindromes. Let us suppose that there are at least two such suffixes $a_ia_{i+1} \ldots a_n$ and $a_{i+k}a_{i+k+1} \ldots a_n$, $1 \leq k \leq n - i$, which are palindromes. It follows that

$$\begin{aligned}
&a_i = a_n = a_{i+k}\\
&a_{i+1} = a_{n-1} = a_{i+k+1}\\
&\ldots\\
&a_{n-k} = a_{i+k} = a_n,
\end{aligned}$$

hence $a_{i+k} \ldots a_n = a_i \ldots a_{n-k}$. The last palindrome appears in u (because $k \geq 1$) and has been already counted in $P(u)$. It follows that $P(w) \leq P(u) + 1 \leq n$. ∎

This result shows that the total number of palindromes in a word cannot be larger than the length of that word. We examine now if there are words which are 'poor' in palindromes. In the next lemma we construct finite words w_n of arbitrary length $n \geq 9$, which contain precisely 8 palindromes. A general method to construct words whose palindrome factors are contained in a prescribed finite set is given in [6].

Let us denote by $w^{\frac{p}{q}}$ the fractional power of the word w of length q [5,14], which is the prefix of length p of w^p.

Lemma 1. *If* $w_n = (112122)^{\frac{n}{6}}$, $n \geq 9$, *then* $P(w_n) = 8$.

Proof. In w_n there are the following palindromes: 1, 2, 11, 22, 121, 212, 1221, 2112. Because 121 and 212 are situated in w_n between 1 on the left and 2 on the right, these cannot be continued to obtain any palindromes. The same is true for 2112 and 1221, which are situated between 2 on the left and 1 on the right, excepting the cases when 2112 is a suffix. So, there are no other palindromes in w_n. ∎

Remark 1. If u is a circular permutation of 112122 and $n \geq 9$ then $P(u^{\frac{n}{6}}) = 8$ too. Because we can interchange 1 with 2, for any n there will be at least 12 words of length n with total complexity equal to 8.

We shall give now, beside the upper delimitation from Proposition 1, lower bounds for the number of palindromes contained in finite binary words. (In the trivial case of a 1-letter alphabet it is obvious that, for any word w, $P(w) = |w|$.)

Remark 2. It can be easily checked that for all the short binary words (up to $|w| = 7$), the palindrome complexity takes always the maximal possible value given in Proposition 1; from the words with $|w| = 8$, only four (out of 2^8) have $P(w) = 7$, namely 11221211, 11212211 and their complemented words.

Theorem 1. *If w is a finite word of length n on a 2-letter alphabet, then $P(w) = n$ for $1 \leq n \leq 7$; $7 \leq P(w) \leq 8$ for $n = 8$; $8 \leq P(w) \leq n$ for $n \geq 9$.*

Proof. Up to 8 the statement follows from direct computation as pointed out in Remark 2. Any word w of length 9 has the total palindrome complexity $P(w) \geq 8$. Indeed, adding a letter 1 or 2 before or after the palindromes of length 8 which have complexity 7 (mentioned in Remark 2) also add a new palindrome. For $n > 9$, Lemma 1 gives words v_n for which $P(v_n) = 8$. The maximal value is obtained for words of the form a^n, $a \in A$, $n \in \mathbb{N}$. ∎

In the following lemmas we construct binary words which have a given total palindrome complexity greater than or equal to 8.

Lemma 2. *If* $u_{k,\ell} = 1^k 21221^\ell 2$ *for* $k \geq 2$ *and* $1 \leq \ell \leq k-1$, *then* $P(u_{k,\ell}) = k+6$.

Proof. In the prefix of length k of $u_{k,\ell}$ there are always k palindromes $(1, \ldots, 1^k)$. The other palindromes different from these are 2, 22, 121, 212, 1221 and $21^\ell 2$ (for $\ell \geq 2$), respectively 212212 (for $\ell = 1$). In each case $P(u_{k,\ell}) = k+6$. ■

Lemma 3. *If* $v_{k,\ell} = (1^k 2122)^{\frac{k+\ell+5}{k+4}}$ *for* $k \geq 2$ *and* $k \leq \ell \leq n-k-5$, *then* $P(v_{k,\ell}) = k+6$.

Proof. Since $\ell \geq k$, the prefix of $u_{k,j}$ is at least $1^k 21221^k 2$, which includes the palindromes $1, \ldots, 1^k$, 2, 22, 121, 212, 1221 and $21^k 2$, hence $P(v_{k,\ell}) \geq k+6$. The palindromes 121 and 212 are situated between 1 and 2, while 1221 and $21^k 2$ are between 2 and 1 (excepting the cases when they are suffixes), no matter how large is ℓ. It follows that $v_{k,\ell}$ contains no other palindromes, hence $P(v_{k,\ell}) = k+6$. ■

Remark 3. If $k = 2$, then the word $v_{2,\ell}$ is equal to $w_{\ell+7}$, with w_n defined in Lemma 1.

We can determine now precisely the image of the restriction of the palindrome complexity function to A^n, $n \geq 1$.

Theorem 2. *Let A be a binary alphabet. For* $1 \leq n \leq 7$, $P(A^n) = \{n\}$; *for* $n = 8$, $P(A^n) = \{7, 8\}$; *for* $n \geq 9$, $P(A^n) = \{8, \ldots, n\}$.

Proof. Having in mind the result in Theorem 1, we have to prove only that for each n and i so that $8 \leq i \leq n$, there exists always a binary word $w_{n,i}$ of length n for which the total palindrome complexity is $P(w_{n,i}) = i$. Let n and i be given so that $8 \leq i \leq n$. We denote $k = i - 6 \geq 2$ and $\ell = n - k - 5$.

If $\ell \leq k-1$, we take $w_{n,i} = u_{k,\ell}$ (from Lemma 2); if $\ell \geq k$, $w_{n,i} = v_{k,\ell}$ (from Lemma 3). It follows that $|w_{n,i}| = n$ and $P(w_{n,i}) = k + 6 = i$. ■

Example 1. Let us consider $n = 25$ and $i = 15$. Then $k = 15 - 6 = 9$, $\ell = 25 - 9 - 5 = 11$. Because $\ell > k - 1$, we use $v_{9,11} = (1^9 2122)^{\frac{25}{13}} = 1^9 21221^9 212$, whose total palindrome complexity is 15.

We give similar results for the case of alphabets with $q \geq 3$ letters.

Corollary 1. *If* w *is a finite word of length n over a q-letter* ($q \geq 3$) *alphabet, then* $P(w) = n$ *for* $n \in \{1, 2\}$; $3 \leq P(w) \leq n$ *for* $n \geq 3$.

Proof. If w is written with at most two letters then the result follows from Theorem 1, otherwise w contains at least one occurrence of each of the three letters thus its total palindrome complexity is at least 3. ■

Theorem 3. *Let A be a q-letter* ($q \geq 3$) *alphabet. Then for* $1 \leq n \leq 3$, $P(A^n) = \{n\}$; *for* $n \geq 4$, $P(A^n) = \{3, \ldots, n\}$.

Proof. It remains to prove that for each n and i so that $3 \leq i \leq n$, there exists always a word $w_{n,i}$ of length n, for which the total palindrome complexity is $P(w_{n,i}) = i$. Such a word is $w_{n,i} = a_1^{i-3}(a_1 a_2 a_3)^{\frac{n-i+3}{3}}$, which has $i - 2$ palindromes in its prefix of length $i - 2$, and other two palindromes a_2 and a_3 in what follows. ■

Open problem. Find the number of words w of length n which satisfy $P(w) = n$ (named in [6] *full* words).

3. Average number of palindromes

We consider an alphabet A with $q \geq 2$ letters. We remind that the average number $M_q(n)$ of palindromes contained in all the words of length n over A is defined by

$$M_q(n) = \frac{\sum\limits_{w \in A^n} P(w)}{q^n},$$

where $P(w)$ is the total palindrome complexity of the word w.

In order to obtain upper bounds for $M_q(n)$, we write

$$\sum_{w \in A^n} P(w) = \sum_{w \in A^n} \sum_{\pi \in \mathrm{PAL}_w} 1 = \sum_{w \in A^n} \sum_{k=1}^{n} \sum_{\pi \in \mathrm{PAL}_w(k)} 1$$
$$= \sum_{k=1}^{n} \sum_{w \in A^n} \sum_{\pi \in \mathrm{PAL}_w(k)} 1$$

and denote

$$S_{n,k} = \sum_{w \in A^n} \sum_{\pi \in \mathrm{PAL}_w(k)} 1,$$

which represent the number of occurrences of the palindromes of length k in all words of length n (counted once if a palindrome appears in a word, and once again if it appears in another one). We use the fact that the number of palindromes of a given length k over an alphabet A with q letters is $q^{\lfloor (k+1)/2 \rfloor}$ in order to give upper bounds for $S_{n,k}$.

Lemma 4. *For each $1 \le k \le n$, the following inequalities hold:*

$$S_{n,k} \le q^{n+\lfloor (k+1)/2 \rfloor}, \tag{1}$$

$$S_{n,k} \le (n-k+1)q^{n-k+\lfloor (k+1)/2 \rfloor}. \tag{2}$$

Proof. For each $1 \le k \le n$, it follows that

$$S_{n,k} = \sum_{w \in A^n} \sum_{\pi \in \mathrm{PAL}_A(k)} 1 \le q^n q^{\lfloor (k+1)/2 \rfloor}.$$

The other upper bound is obtained by writing

$$\begin{aligned} S_{n,k} &= \sum_{\pi \in \mathrm{PAL}_A(k)} \sum_{\substack{w \in A^n \\ \pi \in \mathrm{PAL}_w(k)}} 1 \le \sum_{\pi \in \mathrm{PAL}_A(k)} (n-k+1)q^{n-k} \\ &= (n-k+1)q^{n-k+\lfloor (k+1)/2 \rfloor}. \quad \blacksquare \end{aligned}$$

Upper bounds for $M_q(n)$ have been obtained in [2] for n odd, respectively even. From this we could prove that $M_q(n) < q+2n/(q-1)$. However in the following theorem we make a more precise estimation and show that, in fact, the palindrome subwords are rather rare in long words, whatever $q \ge 2$ is.

Theorem 4. *For an alphabet A with $q \ge 2$ letters, the average number of palindromes $M_q(n)$ satisfies*

$$\lim_{n\to\infty} \frac{M_q(n)}{n} = 0.$$

Proof. For $p \in \mathbb{N}$, we apply the inequality (1) for $0 \le k \le p$ and (2) for $p < k \le n$ and obtain

$$\begin{aligned} M_q(n) &\le \frac{1}{q^n}\left(\sum_{k=1}^{p} q^{n+\lfloor (k+1)/2 \rfloor} + \sum_{k=p+1}^{n} (n-k+1)q^{n-k+\lfloor (k+1)/2 \rfloor}\right) \\ &= \sum_{k=1}^{p} q^{\lfloor (k+1)/2 \rfloor} + \sum_{k=p+1}^{n} (n-k+1)q^{-k+\lfloor (k+1)/2 \rfloor} \\ &\le \sum_{k=1}^{p} q^{\lfloor (k+1)/2 \rfloor} + n\sum_{k=p+1}^{n} q^{-k+\lfloor (k+1)/2 \rfloor}. \end{aligned} \tag{3}$$

We divide by n and take $\limsup_{n\to\infty}$:

$$\begin{aligned} \limsup_{n\to\infty} \frac{M_q(n)}{n} &\le (0+\cdots+0)_{p\text{ times}} + \sum_{k>p} q^{-k+\lfloor (k+1)/2 \rfloor} \\ &= \sum_{k>p} q^{-k+\lfloor (k+1)/2 \rfloor}. \end{aligned}$$

The series

$$\sum_{k=1}^{\infty} q^{-k+\lfloor (k+1)/2 \rfloor} = \frac{q+1}{q-1}$$

being convergent, we get for its remainder $\lim_{p\to\infty} \sum_{k>p} q^{-k+\lfloor (k+1)/2 \rfloor} = 0$, hence we get

$$\limsup_{n\to\infty} \frac{M_q(n)}{n} \le 0$$

and the conclusion of the theorem follows. ■

A more elaborate estimation allows us to give the order of convergence for the sequence $M_q(n)/n$. The idea is to select, for a fixed n, the integer p from which we pass from the inequality (1) to the inequality (2) in an optimal way, in order to obtain the best result for this method.

Theorem 5. *The following inequality holds*

$$M_q(n) \le \frac{q+1}{q^{1/2}-1}\, q^{1/4}\, n^{1/2}. \tag{4}$$

Proof. From (3) we obtain for $0 \le p < n$ that

$$M_q(n) \le u_p + n v_p,$$

where

$$u_p = \sum_{k=1}^{p} q^{\lfloor (k+1)/2 \rfloor}, \qquad v_p = \sum_{k=p+1}^{\infty} q^{-k+\lfloor (k+1)/2 \rfloor}.$$

We have

$$u_p = \begin{cases} \left(q^{3/2} + q^{1/2} - 2q^{1-p/2}\right) q^{p/2}/(q-1), & \text{for } p \text{ odd} \\ 2\left(q - q^{1-p/2}\right) q^{p/2}/(q-1), & \text{for } p \text{ even,} \end{cases}$$

and

$$v_p = \begin{cases} 2q^{1/2} q^{-p/2}/(q-1), & \text{for } p \text{ odd} \\ (q+1)\, q^{-p/2}/(q-1), & \text{for } p \text{ even.} \end{cases}$$

It follows

$$u_p \le \left(q^{3/2} + q^{1/2} - 2q^{1-p/2}\right) q^{p/2}/(q-1) \le q^{1/2} \frac{q+1}{q-1} q^{p/2}$$

and

$$v_p \le \frac{q+1}{q-1} q^{-p/2},$$

therefore

$$M_q(n) \le \frac{q+1}{q-1} \left(q^{1/2} q^{p/2} + n q^{-p/2}\right). \tag{5}$$

We denote $p_0 := \left\lfloor \log_q \left(n q^{-1/2}\right) \right\rfloor$. The integer p_0 is actually the floor of the minimal point x_0 of the function $x \mapsto q^{1/2} q^{x/2} + n q^{-x/2}$. In fact, this minimal point x_0 satisfies $q^{1/2} q^{x_0/2} = n q^{-x_0/2}$, hence $x_0 = \log_q \left(n q^{-1/2}\right)$.

If $p_0 \ge 0$ it is easy to see that $p_0 < n$, and $q^{p_0} \le n q^{-1/2} < q^{p_0+1}$. For $p = p_0$ in (5) we get

$$M_q(n) \le \frac{q+1}{q-1} \left(q^{1/2} \frac{n^{1/2}}{q^{1/4}} + n \frac{q^{3/4}}{n^{1/2}}\right) = \frac{q+1}{q^{1/2}-1} q^{1/4} n^{1/2}.$$

If $p_0 < 0$ (which happens in the case $n < q^{1/2}$), we apply (5) for $p = 0$ and get

$$M_q(n) \le \frac{q+1}{q-1} \left(q^{1/2} + n\right).$$

The inequality $q^{1/2} + n \le q^{1/4} \left(q^{1/2} + 1\right) n^{1/2}$ is equivalent to

$$\frac{q^{1/2}}{n^{1/2}} + n^{1/2} \le q^{1/4} \left(q^{1/2} + 1\right). \tag{6}$$

The function $n \mapsto q^{1/2}/n^{1/2} + n^{1/2}$ is strictly decreasing for $1 \le n < q^{1/2}$, hence (6) holds and (4) follows in this case too. ∎

Remark 4. From Theorem 5 it follows that $M_q(n) = O(n^{1/2})$. For a binary alphabet ($q = 2$) we have $M_2(n) < 9\, n^{1/2}$. More generally, $M_q(n) < 6\, q^{3/4}\, n^{1/2}$.

Remark 5. Even for relatively small numbers (e.g. $q = 2$ and $n = 100$) it is practically impossible to compute $M_q(n)$ using a computer. That is why, in order to check numerically the result in Theorem 5, we approximated $M_q(n)$ with $\sum_{w \in K} P(w)/(\#K)$, where $K \subset A^n$ is a set of at least 100 random words.

For $q = 2$, ignoring the decimal digits, the inequality (4) becomes $53 < 86$ for $n = 100$, $79 < 121$ for $n = 200$ and $182 < 272$ for $n = 1000$. Similarly, for $q = 3$ we get $38 < 71$ for $n = 100$, $57 < 111$ for $n = 200$ and $132 < 227$ for $n = 1000$.

The numerical values in Remark 5 allow us to infer that the bound in Theorem 5 cannot be essentially improved and to state the following

Open problem. It would be interesting to study if the sequence $M_q(n) n^{-1/2}$ is convergent and even to find its asymptotic development.

Acknowledgements

The authors thank the anonymous referee for the suggestions which contributed to a better presentation of the results. The first and the third authors acknowledge the kind hospitality of the Alfréd Rényi Institute of Mathematics and the Luminy Institute of Mathematics (cooperation programs between the Romanian Academy and the Hungarian Academy of Sciences, respectively CNRS France).

References

[1] J.-P. Allouche, M. Baake, J. Cassaigne, D. Damanik, Palindrome complexity, Theoret. Comput. Sci. 292 (2003) 9–31.
[2] M-C. Anisiu, V. Anisiu, Z. Kása, Properties of palindromes in finite words, Pure Math. Appl. 17 (2006) 183–195.
[3] M. Baake, A note on palindromicity, Lett. Math. Phys. 49 (1999) 217–227.
[4] V. Berthé, L. Vuillon, Palindromes and two-dimensional Sturmian sequences, J. Autom. Lang. Comb. 6 (2001) 121–138.
[5] F.J. Brandenburg, Uniformly growing *k*-th powerfree homomorphisms, Theoret. Comput. Sci. 23 (1) (1983) 69–82.
[6] 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.
[7] B Brešar, J Grytczuk, S Klavžar, S Niwczyk, I. Peterin, Nonrepetitive colorings of trees, Discrete Math. 307 (2007) 163–172.
[8] D. Damanik, L.Q. Zamboni, Combinatorial properties of Arnoux-Rauzy subshifts and applications to Schrödinger operators, Rev. Math. Phys. 15 (2003) 745–763.
[9] D. Damanik, D. Zare, Palindrome complexity bounds for primitive substitution sequences, Discrete Math. 222 (1–3) (2000) 259–267.
[10] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, Theoret. Comput. Sci. 255 (2001) 539–553.
[11] S. Ferenczi, Z. Kása, Complexity for finite factors of infinite sequences, Theoret. Comput. Sci. 218 (1999) 177–195.
[12] A. Iványi, On the *d*-complexity of words, Ann. Univ. Sci. Budapest. Sect. Comput. 8 (1987) 69–90.
[13] Z. Kása, On the *d*-complexity of strings, Pure Math. Appl. 9 (1998) 119–128.
[14] F. Levé, P. Séébold, Proof of a conjecture on word complexity, Bull. Belg. Math. Soc. Simon Stevin 8 (2001) 277–291.
[15] J. Shallit, On the maximum number of distinct factors in a binary string, Graphs Combin. 9 (1993) 197–200.