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 © 2009 Elsevier B.V. All rights reserved.
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 $P(w) \le |w|$.*

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

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

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

If there is an index j, $1 \le j \le n-1$ 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 $a_ia_{i+1}\dots a_n$ and $a_{i+k}a_{i+k+1}\dots a_n$, $1 \le k \le n-i$, which are palindromes. It follows that

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

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

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 $n \ge 9$, 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* $w_n = (112122)^{\frac{n}{6}}$, $n \ge 9$, *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 $n \ge 9$ 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 $1 \le n \le 7$; $7 \le P(w) \le 8$ for $n = 8$; $8 \le P(w) \le n$ for $n \ge 9$.*

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 $P(w) \ge 8$. 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.