PU.M.A. Vol. **17** (2006), No. 3–4, pp. 183–195

Properties of palindromes in finite words

Mira-Cristiana Anisiu
Tiberiu Popoviciu Institute of Numerical Analysis
Romanian Academy, Cluj-Napoca
e-mail: mira@math.ubbcluj.ro

and

Valeriu Anisiu
Department of Mathematics
Faculty of Mathematics and Computer Science
Babeş-Bolyai University of Cluj-Napoca
e-mail: anisiu@math.ubbcluj.ro

and

Zoltán Kása
Department of Computer Science
Faculty of Mathematics and Computer Science
Babeş-Bolyai University of Cluj-Napoca
e-mail: kasa@cs.ubbcluj.ro

(Received: July 12–15, 2006)

Abstract. We present a method which displays all palindromes of a given length from De Bruijn words of a certain order, and also a recursive one which constructs all palindromes of length $n+1$ from the set of palindromes of length n. We show that the palindrome complexity function, which counts the number of palindromes of each length contained in a given word, has a different shape compared with the usual (subword) complexity function. We give upper bounds for the average number of palindromes contained in all words of length n, and obtain exact formulae for the number of palindromes of length 1 and 2 contained in all words of length n.

Mathematics Subject Classifications (2000). 68R15

1 Introduction

The palindrome complexity of infinite words has been studied by several authors (see [1], [3], [14] and the references therein). Similar problems related to the number of palindromes are important for finite words too. One of the reasons is that palindromes occur in DNA sequences (over 4 letters) as well as in protein description (over 20 letters), and their role is under research ([9]).

Let an alphabet A with card $(A) = q \geq 1$ be given. The set of the words of length n over A will be denoted by A^n.

Given a word $w = w_1 w_2 ... w_n$, the *reversed* of w is $\widetilde{w} = w_n ... w_2 w_1$. Denoting by ε the empty word, we put by convention $\widetilde{\varepsilon} = \varepsilon$. The word w is a *palindrome* if $\widetilde{w} = w$. We denote by a^k the word $\underbrace{a \dots a}_{k \text{ times}}$. The set of the subwords of a word w which are nonempty palindromes will be denoted by PAL (w). The (infinite) set of all palindromes over the alphabet A is denoted by PAL (A), while $\text{PAL}_n(A) = \text{PAL}\,(A) \cap A^n$.

2 Storing and generating palindromes

An old problem asks if, given an alphabet A with card $(A) = q$, there exists a shortest word of length $q^k + k - 1$ containing all the q^k words of length k. The answer is affirmative and was given in [6], [10], [4]. For each $k \in \mathbb{N}$, these words are called De Bruijn words of order k. This property can be proved by means of the Eulerian cycles in the De Bruijn graph B_{k-1}. If a window of length k is moved along a De Bruijn word, at each step a different word is seen, all the q^k words being displayed.

We ask if it is possible to arrange all palindromes of length k in a similar way. The answer is in general no, excepting the case of the two palindromes $aba...a$ and $bab...b$ of odd length.

PROPOSITION 1 *Given a word $w \in A^n$ and $k \geq 2$, the following statements are equivalent:*

(1) *all the subwords of length k are palindromes;*

(2) *n is even, $k = n - 1$ and there exists $a, b \in A$, $a \neq b$ so that $w = (ab)^{n/2}$.*

Furthermore, in this case the only palindromes of w are $(ab)^{n/2-2}a$ and $(ba)^{n/2-2}b$.

Proof. Let us consider the first two palindromes $a_1a_2...a_k$ and $b_1b_2...b_k$ such that $a_2a_2...a_k = b_1b_2...b_{k-1}$, hence

$$a_{k-i+1} = a_i = b_{i-1} = b_{k-i+2}, \quad i = 2, \ldots, k.$$

It follows

$$\begin{array}{ll} i = 2 & a_{k-1} = a_2 = b_1 = b_k \\ i = 3 & a_{k-2} = a_3 = b_2 = b_{k-1} \\ i = 4 & a_{k-3} = a_4 = b_3 = b_{k-2} \\ & \cdots \\ i = k-1 & a_2 = a_{k-1} = b_{k-2} = b_3 \\ i = k & a_1 = a_k = b_{k-1} = b_2. \end{array}$$

If $k = 2l$, $(l \geq 1)$ we have $b_2 = a_1 = a_3 = ... = a_{k-1}$ and $b_3 = a_2 = ... = a_k$ and $a_1a_2...a_k$ is a palindrome if and only if $a_1 = a_2 = ... = a_k$, hence $a_1a_2...a_k = a^k$; it follows that $b_1b_2...b_k = a^k$ too, and the two palindromes are equal.

If $k = 2l + 1$, we have $b_2 = a_1 = a_3 = ... = a_k$ and $b_3 = a_2 = ... = a_{k-1}$, hence $a_1a_2...a_k = abab...a$ $(a \neq b)$ and $b_1b_2...b_k = bab...b$. If another palindrome will follow, it must be again $(ab)^{n/2}$ (equal with the first one). □

REMARK 1 For $k = 1$, the maximum length of a word containing all distinct palindromes of length 1 (i.e. letters) exactly once is $n = q$.

It is obvious that for $k \geq 2$ it is not possible to arrange all palindromes of length k in the most compact way. But each palindrome is determined by the parity of its length and its first $\lceil k/2 \rceil$ letters, where $\lceil \cdot \rceil$ denotes the ceil function (which returns the smallest integer that is greater than or equal to a specified number).

PROPOSITION 2 *All palindromes of length k can be obtained from a De Bruijn word of length $q^{\lceil k/2 \rceil} + \lceil k/2 \rceil - 1$.*

Proof. The De Bruijn word contains all different words of length $\lceil k/2 \rceil$. Each such word $a_1 ... a_{\lceil k/2 \rceil}$ can be extended to a palindrome by symmetry, for k even, and by taking $a_{\lceil k/2 \rceil + 1} = a_{\lceil k/2 \rceil - 1}, \ldots, a_k = a_1$, for k odd. □

EXAMPLE 1 *Let $k = 3$, $q = 3$ and the De Bruijn word of order $\lceil k/2 \rceil = 2$ $w_1 = 0221201100$. From each word of length 2 which appears in the given De Bruijn word, we obtain the corresponding palindrome of length $k = 3$:*

$$\begin{array}{lcl} 02 & \rightarrow & 020 \\ 22 & \rightarrow & 222 \\ 21 & \rightarrow & 212 \\ 12 & \rightarrow & 121 \\ 20 & \rightarrow & 202 \\ 01 & \rightarrow & 010 \\ 11 & \rightarrow & 111 \\ 10 & \rightarrow & 101 \\ 00 & \rightarrow & 000. \end{array}$$

Let $k = 4$, $q = 2$ and the De Bruijn word of order $\lceil k/2 \rceil = 2$ $w_2 = 01100$. From each word of length 2 contained in 01100 we obtain by symmetry the corresponding palindrome of length $k = 4$:

$$\begin{array}{lcl} 01 & \rightarrow & 0110 \\ 11 & \rightarrow & 1111 \\ 10 & \rightarrow & 1001 \\ 00 & \rightarrow & 0000. \end{array}$$

There are several algorithms which construct De Bruijn words, for example, in [16], [18], [7] and [8].

We can generate recursively all palindromes of length n, $n \in \mathbb{N}$, using the difference representation. This is based on the following proposition.

PROPOSITION 3 *If $w_1, w_2, \ldots, w_p$ are all binary ($A = \{0, 1\}$) palindromes of length n, where $p = 2^{\lceil \frac{n}{2} \rceil}$, $n \geq 1$, then*

$$2w_1, 2w_2, \ldots, 2w_p, 2^{n+1} + 1 + 2w_1, 2^{n+1} + 1 + 2w_2, \ldots, 2^{n+1} + 1 + 2w_p$$

are all palindromes of length $n + 2$.

Proof. If w is a binary palindrome of length n, then $0w0$ and $1w1$ will be palindromes too, and the only palindromes of length $n+2$ which contains w as a subword, which proves the proposition. □

In order to generate all binary palindromes of a given length let us begin with an example considering all binary palindromes of length 3 and 4 and their decimal representation:

000	0	0000	0
010	2	0110	6
101	5	1001	9
111	7	1111	15

The sequence of palindromes in increasing order based on their decimal value for a given length can be represented by their differences. The difference representation of the sequence 0, 2, 5, 7 is 2, 3, 2 ($2-0=2$, $5-2=3$, $7-5=2$), and the difference representation of the sequence 0, 6, 9, 15 is 6, 3, 6. A difference representation is always a symmetric sequence and the corresponding sequence of palindromes in decimal can be obtained by successive addition beginning with 0: $\mathbf{0}+6=\mathbf{6}$, $\mathbf{6}+3=\mathbf{9}$, $\mathbf{9}+6=\mathbf{15}$. By direct computation we obtain the following difference representation of palindromes for length $n \leq 8$.

n															
1	1														
2	3														
3	2	3	2												
4	6	3	6												
5	4	6	4	3	4	6	4								
6	12	6	12	3	12	6	12								
7	8	12	8	6	8	12	8	3	8	12	8	6	8	12	8
8	24	12	24	6	24	12	24	3	24	12	24	6	24	12	24

We easily can generalize and prove by induction that the difference representations can be obtained as follows.

For $n=2k$ we have the difference representation:

$$a_1, a_2, \ldots, a_{2^k-1},$$

from which the difference representation for $2k+1$ is:

$$2^k, a_1, 2^k, a_2, 2^k, \ldots, 2^k, a_{2^k-1}, 2^k.$$

For $n=2k+1$ we have the difference representation:

$$2^k, a_1, 2^k, a_2, 2^k, \ldots, 2^k, a_{2^k-1}, 2^k,$$

from which the difference representation for $2k+2$ is:

$$3\cdot 2^k, a_1, 3\cdot 2^k, a_2, 3\cdot 2^k, \ldots, 3\cdot 2^k, a_{2^k-1}, 3\cdot 2^k.$$

This representation can be generalized for $q \geq 2$. The number of palindromes in this case is $q^{\left\lceil \frac{n}{2} \right\rceil}$.

For $n = 2k$ we have the difference representation:

$$a_1, a_2, \ldots, a_{q^k-1},$$

from which the difference representation for $2k+1$ is:

$$\underbrace{q^k, \ldots, q^k}_{q-1 \text{ times}},\ a_1,\ \underbrace{q^k, \ldots, q^k}_{q-1 \text{ times}},\ a_2,\ \underbrace{q^k, \ldots, q^k}_{q-1 \text{ times}}, \ldots, \underbrace{q^k, \ldots, q^k}_{q-1 \text{ times}},\ a_{q^k-1},\ \underbrace{q^k, \ldots, q^k}_{q-1 \text{ times}}.$$

For $n = 2k+1$ we have the difference representation:

$$\underbrace{q^k, \ldots, q^k}_{q-1 \text{ times}},\ a_1,\ \underbrace{q^k, \ldots, q^k}_{q-1 \text{ times}},\ a_2, \ldots, a_{q^k-1},\ \underbrace{q^k, \ldots, q^k}_{q-1 \text{ times}},$$

from which the difference representation for $2k+2$ is:

$$\underbrace{(q+1)q^k, \ldots, (q+1)q^k}_{q-1 \text{ times}},\ a_1,\ \underbrace{(q+1)q^k, \ldots, (q+1)q^k}_{q-1 \text{ times}},\ a_2,$$

$$\ldots,\ \underbrace{(q+1)q^k, \ldots, (q+1)q^k}_{q-1 \text{ times}},\ a_{q^k-1},\ \underbrace{(q+1)q^k, \ldots, (q+1)q^k}_{q-1 \text{ times}}.$$

3 The shape of the palindrome complexity functions

For an infinite sequence U, the *(subword) complexity function* $p_U : \mathbb{N} \longrightarrow \mathbb{N}$ (defined in [17] as the *block growth*, then named *subword complexity* in [5]) is given by $p_U(n) = \text{card}\,(F(U) \cap A^n)$ for $n \in \mathbb{N}$, where $F(U)$ is the set of all finite subwords (factors) of U. Therefore the complexity function maps each nonnegative number n to the number of subwords of length n of U; it verifies the iterative equation

$$p_U(n+1) = p_U(n) + \sum_{j=2}^{q} (j-1)s(j,n), \tag{1}$$

$s(j,n)$ being the cardinal of the set of the subwords in U having the length n and the right valence j. A subword $u \in U$ *has the right valence* j if there are j and only j distinct letters x_i such that $ux_i \in F(U)$, $1 \leq i \leq j$.

For a finite word w of length n, the *complexity function* $p_w : \mathbb{N} \longrightarrow \mathbb{N}$ given by $p_w(k) = \text{card}\,(F(w) \cap A^k)$, $k \in \mathbb{N}$, has the property that $p_w(k) = 0$ for $k > n$. The corresponding iterative equation is

$$p_w(k+1) = p_w(k) + \sum_{j=2}^{q} (j-1)s(j,k) - s_0(k), \tag{2}$$

where $s_0(k) = s(0,k) \in \{0,1\}$ stands for the cardinal of the set of subwords v (suffixes of w of length k) which cannot be continued as $vx \in F(w)$, $x \in A$. We can write (2) in a condensed form

$$p_w(k+1) = p_w(k) + \sum_{j=0}^{q} (j-1)s(j,k). \tag{3}$$

The above relations have their correspondents in terms of left extensions of the subwords.

For an infinite sequence U, the complexity function p_U is nondecreasing; more than that, if there exists $m \in \mathbb{N}$ such that $p_U(m+1) = p_U(m)$, then p_U is constant for $n \geq m$.

The complexity function for a finite word w of length n has a different behaviour, because of $p_w(n) = 1$ (there is a unique subword of length n, namely w). It was proved ([12], [13], [15], [2]) that the shape of the complexity function is trapezoidal:

Theorem 1 *Given a finite word w of length n, there are three intervals of monotonicity for p_w: $[0,J]$, $[J,M]$ and $[M,n]$; the function increases at first, is constant and then decreases with the slope -1.*

The *palindrome complexity function* of a finite or infinite word w is given by $\text{pal}_w : \mathbb{N} \longrightarrow \mathbb{N}$, $\text{pal}_w(k) = \text{card}\,(\text{PAL}\,(w) \cap A^k)$, $k \in \mathbb{N}$. Obviously,

$$\text{pal}_w(k) \leq p_w(k), \ k \in \mathbb{N}, \tag{4}$$

and for finite words of length $|w| = n$,

$$\text{pal}_w(k) \leq \min\left\{q^{\lceil k/2 \rceil}, n-k+1\right\}, \ k \in \{0,\ldots,n\}. \tag{5}$$

The palindrome $u \in \text{PAL}\,(w)$ *has the palindrome valence* j if there are j and only j distinct letters x_i such that $x_i u x_i \in \text{PAL}\,(w)$, $1 \leq i \leq j$. We denote by

$$s_p(j,k) = \text{card}\ \left\{u \in (\text{PAL}\,(w) \cap A^k) : u \text{ has the palindrome valence } j\right\}, \tag{6}$$

and by $s_p(0,k)$ the cardinal of the set of subwords $v \in \text{PAL}\,(w) \cap A^k$ (not necessarily suffixes or prefixes of w) which cannot be continued as $xvx \in \text{PAL}\,(w)$, $x \in A$.

The palindrome complexity function of finite or infinite words satisfies the iterative equation

$$\text{pal}_w(k+2) = \text{pal}_w(k) + \sum_{j=0}^{q} (j-1)s_p(j,k). \tag{7}$$

Due to the fact that the number of even palindromes is not directly related to that of odd ones, we do not expect that pal_w is of trapezoidal shape, as it was the case for the subword complexity function p_w.

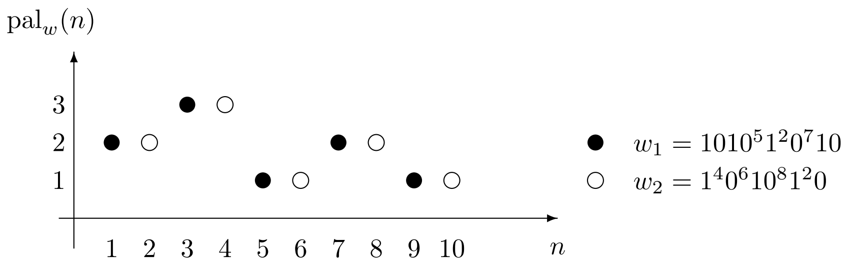

Figure 1: Odd and even palindrome complexity function

For this reason we define the *odd*, respectively *even palindrome complexity function* as the restrictions of pal_w to odd, respectively even integers: $\mathrm{pal}_w^o : 2\mathbb{N}+1 \to \mathbb{N}$, $\mathrm{pal}_w^o(k) = \mathrm{pal}_w(k)$; $\mathrm{pal}_w^e : 2\mathbb{N} \to \mathbb{N}$, $\mathrm{pal}_w^e(k) = \mathrm{pal}_w(k)$.

These functions have a trapezoidal form for short words; nevertheless, this is not true in general, as the following examples show.

EXAMPLE 2 *The word* $w_1 = 1010^51^20^710$ *with* $|w_1| = 19$ *has* $\mathrm{pal}_{w_1}^o(1) = 2$, $\mathrm{pal}_{w_1}^o(3) = 3$, $\mathrm{pal}_{w_1}^o(5) = 1$, $\mathrm{pal}_{w_1}^o(7) = 2$, $\mathrm{pal}_{w_1}^o(9) = 1$. *(see Fig. 1.)*

EXAMPLE 3 *The word* $w_2 = 1^40^610^81^20$ *with* $|w_2| = 22$ *has* $\mathrm{pal}_{w_2}^e(2) = 2$, $\mathrm{pal}_{w_2}^e(4) = 3$, $\mathrm{pal}_{w_2}^e(6) = 1$, $\mathrm{pal}_{w_2}^e(8) = 2$, $\mathrm{pal}_{w_2}^e(10) = 1$. *(see Fig. 1.)*

REMARK 2 The palindrome complexity for infinite words is not nondecreasing, as the usual complexity function is. Indeed, we can continue the word in Example 2 with 11001100..., and its odd palindrome complexity function will be as that for w_1, and then equal to 0 for $k \geq 11$. Similarly, we can continue w_2 in Example 3 with 1010... to obtain an infinite word with the even palindrome complexity of w_2 till $k = 10$ and equal to 0 for $k \geq 12$.

4 Average number of palindromes

We consider an alphabet A with $q \geq 2$ letters.

DEFINITION 1 We define the *total palindrome complexity* P by

$$P(w) = \sum_{n=1}^{|w|} \mathrm{pal}_w(n), \tag{8}$$

where w is a word of length $|w|$, and $\mathrm{pal}_w(n)$ denotes the number of distinct palindromes of length n which are nonempty subwords of w.

Because the set of the nonempty palindromes in w is denoted by PAL (w), we can write also $P(w) = \mathrm{card}\,(\mathrm{PAL}\,(w))$.

Definition 2 The *average number of palindromes* $M_q(n)$ contained in all words of length n is defined by

$$M_q(n) = \frac{\sum_{w\in A^n} P(w)}{q^n}. \tag{9}$$

We can give the following upper estimate for $M_q(n)$.

Theorem 2 *For $n \in \mathbb{N}$, the average number of palindromes contained in the words of length n satisfies the inequalities*

$$\begin{aligned} M_q(n) &\le \frac{q^{-(n-1)/2}(q+3) + 2n(q-1) + q^3 - 2q^2 - 2q - 1}{(q-1)^2}, \ \textit{for } n \textit{ odd,} \\ M_q(n) &\le \frac{q^{-n/2}(3q+1) + 2n(q-1) + q^3 - 2q^2 - 2q - 1}{(q-1)^2}, \quad \textit{for } n \textit{ even.} \end{aligned} \tag{10}$$

Proof. We have

$$\begin{aligned} \sum_{w\in A^n} P(w) &= \sum_{w\in A^n} \sum_{\pi\in \mathrm{PAL}\,(w)} 1 = \sum_{w\in A^n} \sum_{k=1}^{n} \sum_{\pi\in \mathrm{PAL}\,(w)\cap A^k} 1 \\ &= \sum_{w\in A^n} \sum_{\pi\in \mathrm{PAL}\,(w)\cap A^1} 1 + \sum_{k=2}^{n} \sum_{\pi\in \mathrm{PAL}_k(A)} \sum_{\substack{w\in A^n \\ \pi\in \mathrm{PAL}\,(w)\cap A^k}} 1, \end{aligned}$$

and

$$\sum_{w\in A^n} \sum_{\pi\in \mathrm{PAL}\,(w)\cap A^1} 1 \le q q^n = q^{n+1}. \tag{11}$$

For a fixed palindrome π, with $|\pi| = k$, the number of the words of length n in which it appears as a subword at position i $(1 \le i \le n-k+1)$ is q^{n-k}. But the position i is arbitrary, so that there are at most $(n-k+1)q^{n-k}$ words in which π is a subword, these words being not necessarily distinct. It follows that

$$\sum_{w\in A^n} P(w) \le q^{n+1} + \sum_{k=2}^{n} \sum_{\pi\in \mathrm{PAL}_k(A)} (n-k+1)q^{n-k}.$$

The number of the palindromes of length k is $q^{\lceil k/2\rceil}$, therefore

$$\sum_{w\in A^n} P(w) \le q^{n+1} + \sum_{k=2}^{n} (n-k+1)q^{n-k+\lceil k/2\rceil}$$

and $M_q(n) \le q + \sum_{k=2}^{n} (n-k+1)q^{-k+\lceil k/2\rceil}$.

We split the sum according to $k = 2j$, $j = 1, \ldots, \lfloor n/2 \rfloor$, respectively $k = 2j+1$, $j = 1, \ldots, \lfloor (n-1)/2 \rfloor$, and obtain

$$M_q(n) \le q + \sum_{j=1}^{\lfloor n/2 \rfloor} (n - 2j + 1) q^{-j} + \sum_{j=1}^{\lfloor (n-1)/2 \rfloor} (n - 2j) q^{-j}.$$

Making use of $\sum_{j=1}^{s} q^{-j} = (1 - q^{-s})/(q-1)$ and $\sum_{j=1}^{s} j q^{-j} = (q - q^{1-s}(s+1) + sq^{-s})/(q-1)^2$, it follows that $M_q(n)$ satisfies the inequalities in (10). □

COROLLARY 1 *The following inequality holds*

$$\limsup_{n\to\infty} \frac{M_q(n)}{n} \le \frac{2}{q-1}. \tag{12}$$

Proof.

$$\limsup_{n\to\infty} \frac{M_q(n)}{n} = \max\left\{ \limsup_{n\to\infty} \frac{M_q(2n+1)}{2n+1},\ \limsup_{n\to\infty} \frac{M_q(2n)}{2n} \right\}$$
$$\le \max\left\{ \lim_{n\to\infty} \left(\frac{q^{-n}(q+3) + 2\,(2n+1)\,(q-1) + q^3 - 2q^2 - 2q - 1}{(q-1)^2} \right) \frac{1}{2n+1}, \right.$$
$$\left. \lim_{n\to\infty} \left(\frac{q^{-n}(3q+1) + 4n(q-1) + q^3 - 2q^2 - 2q - 1}{(q-1)^2} \right) \frac{1}{2n} \right\} = \frac{2}{q-1}. \quad □$$

We are interested in finding how large is the average number of palindromes contained in the words of length n compared to the length n. The numerical estimations done for small values of n show that $M_q(n)$ is comparable to n, but Corollary 1 allows us to show that for $q \ge 4$ this does not hold.

COROLLARY 2 *For an alphabet with $q \ge 4$ letters,*

$$\limsup_{n\to\infty} \frac{M_q(n)}{n} < 1. \tag{13}$$

In the proof of Theorem 2 we have used the rough inequality (11), which was sufficient to prove the result. In fact, it is not difficult to calculate exactly

$$S_{n,p} = \sum_{w \in A^n} \sum_{\pi \in \mathrm{PAL}\,(w) \cap A^p} 1 \text{ for } p = 1, 2. \tag{14}$$

This result has intrinsic importance.

THEOREM 3 *The number of occurrences of the palindromes of length* 1*, respectively* 2*, in all words of length n (counted once if a palindrome appears in a word, and once again if it appears in another one) is given by*

$$S_{n,1} = q^{n+1} - q\,(q-1)^n\,, \tag{15}$$

respectively by

$$S_{n,2} = q^{n+1} - \frac{q}{(q-1)\sqrt{q^2+q-3}} \left(\left(\frac{q-1+\sqrt{q^2+q-3}}{2} \right)^{n+2} - \left(\frac{q-1-\sqrt{q^2+q-3}}{2} \right)^{n+2} \right). \qquad (16)$$

Proof. We use Iverson's convention [11]

$$[\alpha] = \begin{cases} 1, & \text{if } \alpha \text{ is true} \\ 0, & \text{if } \alpha \text{ is false} \end{cases}$$

and obtain

$$S_{n,1} = \sum_{w \in A^n} \sum_{a \in A} [a \text{ in } w] = q \sum_{w \in A^n} [a_1 \text{ in } w],$$

where a_1 is a fixed letter of the alphabet A. Then

$$S_{n,1} = q \sum_{w \in A^n} [a_1 \text{ in } w] = q \left(q^n - \sum_{w \in A^n} [a_1 \text{ not in } w] \right) = q^{n+1} - q\,(q-1)^n.$$

We proceed similarly to calculate $S_{n,2} = \sum_{w \in A^n} \sum_{\pi \in \mathrm{PAL}\,(w) \cap A^2} 1$ and obtain

$$S_{n,2} = \sum_{w \in A^n} \sum_{a \in A} [aa \text{ in } w] = q \sum_{w \in A^n} [a_1 a_1 \text{ in } w],$$

where a_1 is again a fixed letter of the alphabet A. We denote $\varphi(n) := \sum_{w \in A^n} [a_1 a_1$ in $w]$, for which $\varphi(2) = 1$ and $\varphi(3) = 2q-1$. It is easier to establish a recurrence formula for $\psi(n) = q^n - \varphi(n) = \sum_{w \in A^n} [a_1 a_1 \text{ not in } w]$. The number $\psi(n)$ is obtained from:

- the number $(q-1)\psi(n-1)$ of words which do not end in a_1 and have not $a_1 a_1$ in their first $n-1$ positions;
- the number $(q-1)\psi(n-2)$ of words which end in a_1, have the $n-1$ position occupied by one of the other $q-1$ letters and have not $a_1 a_1$ in the first $n-2$ positions.

It follows that ψ satisfies the recurrence formula

$$\psi(n) = (q-1)(\psi(n-1) + \psi(n-2)), \qquad (17)$$

with $\psi(2) = q^2 - 1$ and $\psi(3) = q^3 - 2q + 1$. Its solution is

$$\psi(n) = \frac{1}{(q-1)\sqrt{q^2+q-3}} \left(\left(\frac{q-1+\sqrt{q^2+q-3}}{2} \right)^{n+2} - \left(\frac{q-1-\sqrt{q^2+q-3}}{2} \right)^{n+2} \right)$$

and (16) follows from the fact that

$$S_{n,2} = q\left(q^n - \psi(n)\right). \tag{18}$$

□

The expression of $S_{n,2}$ from (16) allows us to improve Corollary 1.

COROLLARY 3 *The following inequality holds*

$$\limsup_{n\to\infty} \frac{M_q(n)}{n} \le \frac{q+1}{q\,(q-1)}. \tag{19}$$

Proof. Taking into account the inequality

$$\sum_{w\in A^n} \sum_{\pi\in \mathrm{PAL}\,(w)\cap A^1} 1 \le q q^n = q^{n+1},$$

and (18), we get

$$\begin{aligned} M_q(n) &\le \frac{1}{q^n}\left(S_{n,1} + S_{n,2} + \sum_{k=3}^{n} \sum_{\pi\in \mathrm{PAL}_k(A)} (n-k+1)q^{n-k} \right) \\ &\le q\left(2 - \frac{\psi(n)}{q^n}\right) + \sum_{k=3}^{n}(n-k+1)q^{-k+\lfloor (k+1)/2 \rfloor}. \end{aligned}$$

But $0 < \left(q-1+\sqrt{q^2+q-3}\right)/2 < q$ and $-1 < \left(q-1-\sqrt{q^2+q-3}\right)/2 < 0$ for $q \ge 2$, hence $\lim\limits_{n\to\infty} \psi(n)/q^n = 0$. Then

$$\begin{aligned} \limsup_{n\to\infty} \frac{M_q(n)}{n} &\le \lim_{n\to\infty} \frac{1}{n} \sum_{k=3}^{n} (n-k+1)q^{-k+\lfloor (k+1)/2 \rfloor} \\ &\le \sum_{k=3}^{\infty} q^{-k+\lfloor (k+1)/2 \rfloor} = \sum_{i=1}^{\infty} q^{-2i-1+i+1} + \sum_{i=2}^{\infty} q^{-2i+i} \\ &= -\frac{1}{q} + 2\sum_{i=1}^{\infty} q^{-i} = \frac{q+1}{q\,(q-1)}. \end{aligned}$$

□

Corollary 4 *The inequality* (13) *holds for* $q = 3$ *too.*

It seems that (13) holds also for $q = 2$. Using a computer program we obtained some values for the terms of the sequence $M^*(n) = M_2(n)/n$, $n \geq 2$. The first values are: $M^*(n) = 1$, $n = 2, \ldots, 7$; $M^*(8) = 0.99750$; $M^*(9) = 0.98550$, which were close to 1. We tried for greater values of n and get

$$\begin{array}{lll} M^*(20) = 0.89975, & M^*(21) = 0.89002, & M^*(22) = 0.88043 \\ M^*(23) = 0.87101, & M^*(24) = 0.86177, \ \ldots, & M^*(30) = 0.81064. \end{array}$$

The last value was obtained in a very long time, so for greater values of n we generated some random words $w_1, w_2, \ldots, w_\ell$ of length 100, respectively 200, 300, 400 and 500 over $A = \{0, 1\}$ and get some roughly approximate values $M^*(n) \simeq \left(\mathrm{pal}_{w_1}(n) + \cdots + \mathrm{pal}_{w_\ell}(n)\right)/\ell$. For $\ell = 200$ we obtained

$$\begin{array}{lll} M^*(100) \simeq 0.53, & M^*(200) \simeq 0.39, & M^*(300) \simeq 0.32, \\ M^*(400) \simeq 0.29, & M^*(500) \simeq 0.26. & \end{array}$$

This method allows us to obtain the previous exactly computed values $M^*(20)$, $\ldots$, $M^*(30)$ with two exact digits. These numerical results allow us to formulate the following

Conjecture The sequence $M_q(n)/n$ is strictly decreasing for $n \geq 7$.

Acknowledgements. The first and the third author acknowledge the support of the Romanian Academy (Grant 13 GAR/2006) and the kind hospitality of Rényi Institute in the frame of the cooperation program between the Romanian Academy and the Hungarian Academy of Sciences.

References

[1] J.-P. Allouche, M. Baake, J. Cassaigne and D. Damanik, *Palindrome complexity*, Theoret. Comput. Sci., **292** (2003), 9–31.

[2] M.-C. Anisiu and J. Cassaigne, *Properties of the complexity function for finite words*, Rev. Anal. Num. Théor. Approx., **33** (2004), 123–139.

[3] J.-P. Borel and C. Reutenauer, *Palindromic factors of billiard words*, Theoret. Comput. Sci., **340** (2005), 334–348.

[4] N.G. De Bruijn, *A combinatorial problem*, Nederl. Akad. Wetensch. Proc., **49** (1946), 758–764 = Indag. Math., **8** (1946), 461–467.

[5] A. Ehrenfeucht, K.P. Lee and G. Rozenberg, *Subword complexities of various classes of deterministic developmental languages without interactions*, Theoret. Comput. Sci., **1** (1975), 59–75.

[6] C. Flye Sainte-Marie, *Solution to question nr. 48*, l'Intermédiaire des Mathématiciens, **1** (1894), 107–110.

[7] H. Fredricksen, *A survey of full length nonlinear shift register cycle algorithms*, SIAM Review, **24** (1982), 195–221.

[8] R.A. Games, *A generalized recursive construction for De Bruijn sequences*, IEEE Trans. Inform. Theory, **29** (1983), 843–850.

[9] M. Giel-Pietraszuk, M. Hoffmann, S. Dolecka, J. Rychlewski and J. Barciszewski, *Palindromes in proteins*, J. Protein Chem., **22** (2003), 109–113.

[10] I.J. Good, *Normal recurring decimals*, J. London Math. Soc., **21** (1946), 167–169.

[11] R.L. Graham, D.E. Knuth and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science*, 2nd edition (Reading, Massachusetts: Addison-Wesley), 1994.

[12] M. Heinz, *Zur Teilwortkomplexität für Wörter und Folgen über einem endlichen Alphabet*, EIK, **13** (1977), 27–38.

[13] A. de Luca, *On the combinatorics of finite words*, Theoret. Comput. Sci., **218** (1999), 13–39.

[14] A. de Luca and Al. de Luca, *Combinatorial properties of Sturmian palindromes*, Int. J. Found. Comput. Sci., **17** (2006), 557–573.

[15] F. Levé and P. Séébold, *Proof of a conjecture on word complexity*, Bull. Belg. Math. Soc. Simon Stevin, **8** (2001), 277–291.

[16] M.H. Martin, *A problem in arrangements*, Bull. American Math. Soc., **40** (1934), 859–864.

[17] M. Morse and G.A. Hedlund, *Symbolic dynamics*, Amer. J. Math., **60** (1938), 815–866.

[18] A. Ralston, *A new memoryless algorithm for De Bruijn sequences*, J. Algorithms, **2** (1981), 50–62.