Annals of the Tiberiu Popoviciu Seminar
of Functional Equations, Approximation and Convexity
ISSN 1584-4536, vol 6, 2008, pp. 3–12.

# Two-dimensional Total Palindrome Complexity

Mira-Cristiana Anisiu Valeriu Anisiu
(Cluj-Napoca) (Vluj-Napoca)

Abstract. We initiate a comparative study of the properties of total palindrome complexity for binary words and arrays. From this point of view, the HV-palindrome complexity for arrays seems to be more appropriate than the C-palindrome one. We prove also a theorem for the average number of HV-palindromes in arrays.

Key Words: arrays, palindromes, total palindrome complexity
MSC 2000: 68R15

## 1 Introduction

Several authors have studied the palindrome complexity of infinite words (see [1], [5], [13] 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

◇Mira-Cristiana Anisiu, T. Popoviciu Institute of Numerical Analysis, Romanian Academy, email: mira@math.ubbcluj.ro
◇Valeriu Anisiu, Faculty of Mathematics and Computer Science, Babeş-Bolyai University, email: anisiu@math.ubbcluj.ro

is under research ([9]). The values of the total palindrome complexity and the properties of the average number of palindromes have been investigated in [2], [3].

In section 2 we remind some facts related to the palindrome complexity of finite words. Section 3 contains some numerical results concerning the palindrome complexity of two-rows binary arrays, which show that these arrays have similar properties to words; a theorem for the average number of palindromes in arrays is proved.

## 2 Palindromes in words

Let a positive integer $q \geq 1$ and an alphabet $A = \{0, 1, \ldots, q-1\}$ be given. For the word $w = a_1...a_n$ with $a_i \in A$, $1 \leq i \leq n$, the integer $n$ is the *length* of $w$ and is denoted by $|w|$. The length of the *empty word* $\varepsilon$ is 0. The set of the words of length $n$ over $A$ will be denoted by $A^n$. The word $u = a_i...a_j$, $1 \leq i \leq j \leq n$ 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 $\widetilde{w} = a_n...a_1$. A word which is equal to its mirror image is called a *palindrome*. The empty word is considered a palindrome. We denote by $a^k$ the word $a...a$ ($k$ times).

Let $\mathrm{PAL}_w$ be the set of all factors in the word $w$ which are nonempty palindromes, and $\mathrm{PAL}_w(n) = \mathrm{PAL}_w \cap A^n$ the set of the palindromes of length $n$ contained in $w$. The (infinite) set of all palindromes over the alphabet $A$ is denoted by $\mathrm{PAL}_A$, while $\mathrm{PAL}_A(n) = \mathrm{PAL}_A \cap A^n$ is the set of all palindromes of length $n$ over the alphabet $A$. It is known that the number of all palindromes of length $k$ is $q^{\lceil k/2 \rceil}$, where $\lceil \cdot \rceil$ denotes the ceil function (which returns the smallest integer that is greater than or equal to a specified number).

### 2.1 The total palindrome complexity

The *palindrome complexity function* $\mathrm{pal}_w$ 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

$$\mathrm{pal}_w(n) = \#\mathrm{PAL}_w(n). \tag{2.1}$$

The *total palindrome complexity* of a finite word $w$ is equal to the number of all nonempty palindrome factors of $w$, namely

$$P(w) = \sum_{n=1}^{|w|} \mathrm{pal}_w(n). \tag{2.2}$$

This is similar to the total complexity of words, which was extensively studied in [10], [12] for finite words and in [8] for infinite ones.

An upper bound for $P(w)$ was given in [7].

**Theorem 2.1** *The total palindrome complexity $P(w)$ of any finite word $w$ satisfies $P(w) \leq |w|$.*

This result shows that the total number of palindromes in a word cannot be larger than the length of that word. There are words of length $n$ with $P(w) = n$, e. g. $0^n$, but also some which have few palindromes.

Beside the upper delimitation from Theorem 2.1, lower bounds for the number of palindromes contained in finite binary words were found. (In the trivial case of a 1-letter alphabet it is obvious that, for any word $w$, $P(w) = |w|$.)

**Theorem 2.2** *[2] If $w$ is a finite word of length $n$ on a binary 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 $n \geq 9$.*

**Remark 2.1** *For all the short binary words (up to $|w| = 7$), the palindrome complexity takes the maximal possible value given in Theorem 2.1; from the words with $|w| = 8$, only four (out of $2^8$) have $P(w) = 7$, namely* 00110100, 00101100 *and their complemented words, and* 252 *have $P(w) = 8$. There are* 24 *words of length* 9 *and* 16 *of length* 10 *with $P(w) = 8$; based on several numerical results, we conjecture that there are precisely* 12 *words of length $n \geq 11$ with $P(w) = 8$.*

It can be proved that for each $n \geq 8$, the restriction of the total palindrome complexity function to $A^n$ takes all the values between 8 and $n$.

**Theorem 2.3** *[2] 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 $P(w_{n,i}) = i$.*

### 2.2 The average number of palindromes

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

$$E_q(n) = \frac{\sum_{w \in A^n} P(w)}{q^n}, \tag{2.3}$$

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

The following theorems proved in [2] show that, in fact, the palindrome subwords are rather rare in long words, whatever $q \geq 2$ is.

**Theorem 2.4** *For an alphabet $A$ with $q \geq 2$ letters, the average number of palindromes $E_q(n)$ satisfies*

$$\lim_{n \to \infty} \frac{E_q(n)}{n} = 0. \tag{2.4}$$

The order of convergence for the sequence $E_q(n)/n$ can be obtained with the aid of a more elaborate estimation of the upper bound of $E_q(n)$.

**Theorem 2.5** *The following inequality holds*

$$E_q(n) \leq \frac{q+1}{q^{1/2}-1} q^{1/4} n^{1/2}.$$

**Remark 2.2** *From Theorem 2.5 it follows that $E_q(n) = O(n^{1/2})$. For a binary alphabet ($q = 2$) we have $E_q(n) < 9n^{1/2}$. More generally, $E_q(n) < 6q^{3/4}n^{1/2}$.*

## 3 Palindromes in arrays

For a positive integer $q \geq 1$ and an alphabet $A = \{0, 1, \ldots, q-1\}$, let $A^{M \times N}$ denote the set of the $M \times N$ arrays with entries from $A$ ($M$, $N \geq 1$ positive integers).

Let $m$ and $n$ be positive integers with $1 \leq m \leq M$ and $1 \leq n \leq N$. An array $V = [v_{ij}] \in A^{m \times n}$ is a *subarray* of the array $W = [w_{k\ell}] \in A^{M \times N}$ if there exist indices $r$, $s$ such that $r+m-1 \leq M$, $s+n-1 \leq N$ and $v_{ij} = w_{r+i-1,s+j-1}$.

For arrays there are two definitions for palindromes, depending on the considered symmetry.

Let $W = [w_{k\ell}]$ with $1 \leq k \leq M$ and $1 \leq \ell \leq N$. The centrosymmetric image of $W$ is $\widetilde{W} = [\widetilde{w}_{k\ell}]$, where $\widetilde{w}_{k\ell} = w_{M-k+1,N-\ell+1}$. A *C-palindrome* is an array for which $W = \widetilde{W}$ ([4], [6]).

We define also the *horizontal and vertical reverse* of $W$ as $\overline{W}^H = [\overline{w}^H_{k\ell}]$, where $\overline{w}^H_{k\ell} = w_{M-k+1,\ell}$, respectively $\overline{W}^V = [\overline{w}^V_{k\ell}]$, where $\overline{w}^V_{k\ell} = w_{k,N-\ell+1}$. An *HV-palindrome* is an array for which $W = \overline{W}^H = \overline{W}^V$ (i. e. all its columns and rows are palindromes [11]). The number of C-palindromes is $q^{\lfloor MN/2 \rfloor}$, where $\lfloor \cdot \rfloor$ denotes the floor function (which returns the greatest integer that is smaller than or equal to a specified number). The number of HV-palindromes is $q^{\lceil M/2 \rceil \cdot \lceil N/2 \rceil}$.

For each type of two-dimensional palindromes, let $\mathrm{PAL}_W$ be the set of all factors in the array $W$ which are nonempty palindromes, and $\mathrm{PAL}_W(m,n)$ be the set of the palindromes which are $m \times n$ arrays contained in the array $W$. We shall also use the notation $\mathrm{PAL}_A(m,n)$ for the set of all palindromes which are $m \times n$ arrays over the alphabet $A$.

Let us define the functions $(M, N \geq 2)$ $c$, $h$, $v$:$A^{M \times N} \to A^{M \times N}$, $c(W) = \widetilde{W}$, $h(W) = \overline{W}^H$ and $v(W) = \overline{W}^V$.

**Proposition 3.1** *We have $c^2 = h^2 = v^2 = \mathrm{id}$, $h \circ v = c$, $v \circ c = h$, $h \circ c = v$ and $\circ$ is commutative, hence the group $G = \{\mathrm{id}, c, h, v\}$ is isomorphic with Klein's one. An array is a C-palindrome iff it is a fixed point of $c$, and is a HV-palindrome iff it is a common fixed point of $h$ and $v$.*

### 3.1 The total palindrome complexity

For each type of palindromes, we consider for $W \in A^{M\times N}$ the *palindrome complexity function* $p_W : \{1,2,\dots,M\}\times\{1,2,\dots,N\} \to \mathbb{N}$ of $W$ given by

$$p_W(m,n) = \#\mathrm{PAL}_W(m,n),\ m = 1,2,\dots,M,\ n = 1,2,\dots,N, \tag{3.1}$$

and the *total palindrome complexity function* $P(W)$ of $W$ as

$$P(W) = \sum_{m=1}^{M}\sum_{n=1}^{N} p_W(m,n). \tag{3.2}$$

We denote by $W \cdot V$ the *concatenation* of two arrays with the same number of rows. For a two-rows array $V \in A^{2\times N}$, let $|V| = N$ denote the *length* of $V$.

**Remark 3.1** *We list some maximal values of the two types of total palindrome complexity, obtained by using a computer program:*

| $M$ | $N$ | max $HV$-$P$ | max $C$-$P$ |
|---|---|---|---|
| 2 | 4 | 8 | 9 |
| 2 | 5 | 10 | 11 |
| 2 | 6 | 12 | 14 |
| 2 | 7 | 14 | 16 |
| 2 | 8 | 16 | 19 |
| 2 | 9 | 18 | 21 |
| 3 | 4 | 13 | 15 |
| 3 | 5 | 17 | 19 |

*A result similar to that in Theorem 2.1 seems to hold only for arrays with two lines (or columns)* and *for HV-palindrome complexity (which will be denoted from now on by $P(W)$).*

Based on the numerical results, we state the following

**Conjecture 3.1** *The total HV-palindrome complexity $P(W)$ of any finite word $W \in A^{2\times N}$ satisfies $P(W) \le 2\,|W|$, where $|W| = N$.*

We prove a result on the palindrome complexity of the concatenation of two-rows arrays.

**Theorem 3.1** *The following inequality holds*

$$P(W \cdot V) \leq P(W) + 3\,|V|\,, \tag{3.3}$$

*the constant* 3 *being the best possible.*

**Proof.** When a column is added to the array $W$, there appear at most three new palindromes: at most one on each row, and at most one two-dimensional, hence inequality (3.3) holds. For $W = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$ and $V = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$, the number of palindromes in $W \cdot V$ which do not appear in $W$ is equal to 9. These palindromes are: $0^4$, $0^5$, $0^3 1 0^3$, $0^2 1 0^2$, $10^5 1$, $10^3 1$, $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. ∎

**Remark 3.2** *If in (3.3) the constant would be* 2 *instead of* 3*, Conjecture 3.1 could be proved by induction, as Theorem 2.1 was.*

**Remark 3.3** *A result as in Theorem 2.3 does not hold for two-rows arrays, because for* $M = 2$ *and* $N = 9$ *there are* 12 *arrays with* 8 *palindromes, but no arrays with* 9 *palindromes. We conjecture that this is true for* $M = 2$ *and each* $N \geq 9$.

### 3.2 The average number of palindromes

The *average number of HV-palindromes* contained in all the arrays from $A^{M \times N}$, $A$ being an alphabet with $q \geq 2$ letters, is defined by

$$E_q(M,N) = \frac{\displaystyle\sum_{W \in A^{M \times N}} P(W)}{q^{MN}},$$

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

**Theorem 3.2** *For an alphabet $A$ with $q \geq 2$ letters, the average number of palindromes $E_q(M,N)$ satisfies*

$$\lim_{\substack{M\to\infty\\N\to\infty}} \frac{E_q(M,N)}{MN} = 0.$$

**Proof.** We calculate

$E_q(M,N) =$

$$\frac{1}{q^{MN}} \sum_{W\in A^{M\times N}} \sum_{\Pi\in \mathrm{PAL}_W} 1 = \frac{1}{q^{MN}} \sum_{W\in A^{M\times N}} \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{\Pi\in \mathrm{PAL}_W(m,n)} 1.$$

We fix $M_0 \leq M$ and $N_0 \leq N$ two natural numbers. Then

$$E_q(M,N) = \frac{1}{q^{MN}} \sum_{W\in A^{M\times N}} \sum_{m=1}^{M_0} \sum_{n=1}^{N_0} \sum_{\Pi\in \mathrm{PAL}_W(m,n)} 1$$

$$+\frac{1}{q^{MN}} \sum_{W\in A^{M\times N}} \sum_{m>M_0 \text{ or } n>N_0} \sum_{\Pi\in \mathrm{PAL}_W(m,n)} 1 =$$

$$\frac{1}{q^{MN}} \sum_{m=1}^{M_0} \sum_{n=1}^{N_0} \sum_{W\in A^{M\times N}} \sum_{\Pi\in \mathrm{PAL}_W(m,n)}$$

$$+\frac{1}{q^{MN}} \sum_{m>M_0 \text{ or } n>N_0} \sum_{\Pi\in \mathrm{PAL}_W(m,n)} \sum_{\substack{W\in A^{M\times N}\\ \Pi\subseteq W}} \leq$$

$$\frac{1}{q^{MN}} \sum_{m=1}^{M_0} \sum_{n=1}^{N_0} q^{MN} \cdot q^{\lceil m/2\rceil\cdot\lceil n/2\rceil}$$

$$+\frac{1}{q^{MN}} \sum_{m>M_0 \text{ or } n>N_0}^{M,N} \sum_{\Pi\in \mathrm{PAL}_A(m,n)} (M-m+1)(N-n+1)q^{MN-mn} \leq$$

$$\sum_{m=1}^{M_0} \sum_{n=1}^{N_0} q^{\lceil m/2\rceil\cdot\lceil n/2\rceil} + \sum_{m>M_0 \textit{ or } n>N_0}^{M,N} (M-m+1)(N-n+1)q^{\lceil m/2\rceil\cdot\lceil n/2\rceil-mn}.$$

We obtain

$$\frac{E_q(M,N)}{MN} \leq \frac{1}{MN} \sum_{m=1}^{M_0} \sum_{n=1}^{N_0} q^{\lceil m/2\rceil\cdot\lceil n/2\rceil} + \sum_{m>M_0 \text{ or } n>N_0}^{\infty,\infty} q^{\lceil m/2\rceil\cdot\lceil n/2\rceil-mn},$$

therefore $\displaystyle\limsup_{\substack{M\to\infty\\N\to\infty}} \frac{E_q(M,N)}{MN} \leq 0 + \sum_{m>M_0 \text{ or } n>N_0}^{\infty,\infty} q^{\lceil m/2\rceil\cdot\lceil n/2\rceil-mn}.$

The double series $\sum_{m,n} q^{\lceil m/2\rceil\cdot\lceil n/2\rceil - mn}$ is convergent (actually the inequality $q^{\lceil m/2\rceil\cdot\lceil n/2\rceil - mn} \leq q^{\frac{1-mn}{4}}$ holds).

So, $\sum\limits_{m>M_0 \text{ or } n>N_0}^{\infty,\infty} q^{\lceil m/2\rceil\cdot\lceil n/2\rceil - mn}$ tends to 0 for $M_0, N_0 \to \infty$. Hence

$$\limsup_{\substack{M\to\infty\\ N\to\infty}} \frac{E_q(M,N)}{MN} \leq 0. \quad \blacksquare$$

## 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, V. Anisiu and Z. Kása. Total palindrome complexity of finite words, Preprint.

[3] M.-C. Anisiu, V. Anisiu and Z. Kása. Properties of palindromes in finite words, *Pure Math. Appl.,* **17** (2006) 183-195.

[4] V. Berthé and L. Vuillon. Palindromes and two-dimensional Sturmian sequences, *J. Autom. Lang. Comb.,* **6** (2001), 121-138.

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

[6] E. B. Dagum and A. Luati. Asymmetric and symmetric weights of kernel smoothers and their spectral properties, *Estadistica, J. Interamerican Statistical Institute*, **53** (2001), 215-258.

[7] X. Droubay, J. Justin and G. Pirillo. Episturmian words and some constructions of de Luca and Rauzy, *Theoret. Comput. Sci.,* **255** (2001), 539-553.

[8] S. Ferenczi and Z. Kása. Complexity for finite factors of infinite sequences, *Theoret. Comput. Sci.,* **218** (1999) 177–195.

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

[10] A. Ivány. On the $d$-complexity of words, *Ann. Univ. Sci. Budapest. Sect. Comput.,* **8** (1987) 69–90.

[11] J. Jeuering. The derivation of on-line algorithms, with an application to finding palindromes, *Algorithmica*, **11** (1994), 146-184.

[12] Z. Kása. On the $d$-complexity of strings, *Pure Math. Appl.,* **9** 1–2 (1998) 119–128.

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