FACTOR AND PALINDROMIC COMPLEXITY OF THUE-MORSE'S AVATARS

CHRISTIANE FROUGNY[b], KAREL KLOUDA[a,*]

[a] *Faculty of Information Technology, Czech Technical University in Prague, Thákurova 9, Praha 6, 160 00, Czech Republic*

[b] *LIAFA, CNRS UMR 7089, Case 7014, 75205 Paris Cedex 13, France*

[*] corresponding author: `karel.klouda@fit.cvut.cz`

ABSTRACT. Two infinite words that are connected with some significant univoque numbers are studied. It is shown that their factor and palindromic complexities almost coincide with the factor and palindromic complexities of the famous Thue-Morse word.

KEYWORDS: factor complexity, palindromic complexity, univoque numbers, Thue-Morse word.

1. INTRODUCTION

The main result of this paper is the computation of the factor and palindromic complexity of two infinite words which appear in [1] as a representation of some significant *univoque* numbers. A real number $\lambda > 1$ is said to be univoque if 1 admits a unique expansion in base λ of the form

$$1 = \sum_{i>0} a_i \lambda^{-i} \quad \text{with } a_i \in \{0, 1, \ldots, \lceil \lambda \rceil - 1\}.$$

Komornik and Loreti showed in [2] that there is a smallest univoque number γ in the interval $(1, 2)$. This number is transcendental [3] and is connected with the Thue-Morse word in this sense: if $1 = \sum_{i>0} a_i \gamma^{-i}$, then $a_1 a_2 a_3 \cdots = 11010011 \cdots = 0^{-1}\mathbf{u}_{\mathrm{TM}}$, i.e., the Thue-Morse word without the leading zero. There are two generalizations of this result. The first one is the work of the same authors [4], where they studied the univoque numbers $\lambda \in (1, b+1)$, $b \in \mathbb{N}$. The second one is the work of Allouche and Frougny [1]. They proved that there exists a smallest univoque number in $(b, b+1)$ (this is proved also in [5]) and they also found the corresponding unique expansion of 1. These expansions and some other significant words from [1] are studied in the sequel.

As explained in the concluding remark, at least the factor complexity could be computed using the common method employing special factors (see e.g. [6] for details). However, here we derive both complexities directly from the definition of words which really enlighten the connection between the studied words and the Thue-Morse word.

2. PRELIMINARIES

An *alphabet* $\mathcal{A}$ is a finite set of *letters*. A concatenation of n letters $v = v_0 v_1 \cdots v_{n-1}$ from $\mathcal{A}$ is a (finite) *word* over $\mathcal{A}$ of length n. An infinite sequence $\mathbf{u} = u_0 u_1 u_3 \cdots$ is an *infinite word* over $\mathcal{A}$. Any finite word v such that $v = u_k u_{k+1} \cdots u_{k+n-1}$ for some $k \in \mathbb{N}$ is called a *factor* of $\mathbf{u}$ and $u_k u_{k+1} \cdots u_{k+n-1}$

is its *occurrence* in it. The set of all factors of $\mathbf{u}$ is denoted by $\mathcal{L}(\mathbf{u})$, the set $\mathcal{L}_n(\mathbf{u})$ is the set of all factors of length n. The *factor complexity* of an infinite word $\mathbf{u}$ is the function $\mathcal{C}_{\mathbf{u}}(n)$ that returns for all $n \in \mathbb{N}$ the number of factors of $\mathbf{u}$ of length n. Given a word $v = v_0 v_1 \cdots v_{n-1}$, the word $\tilde{v}$ is defined as $v_{n-1}v_{n-2} \cdots v_1 v_0$. If $v = \tilde{v}$, v is called a *palindrome*. The *palindromic complexity* of an infinite word $\mathbf{u}$ is the function $\mathcal{P}_{\mathbf{u}}(n)$ that returns for all $n \in \mathbb{N}$ the number of factors of $\mathbf{u}$ of length n that are palindromes.

All infinite words in question are derived from the famous Thue-Morse word $\mathbf{u}_{\mathrm{TM}}$. The Thue-Morse word is the fixed point of the Thue-Morse morphism

$$\begin{aligned} \varphi_{\mathrm{TM}}(0) &= 01 \\ \varphi_{\mathrm{TM}}(1) &= 10 \end{aligned}$$

starting in the letter 0, i.e.,

$$\mathbf{u}_{\mathrm{TM}} = \lim_{n \to \infty} \varphi_{\mathrm{TM}}^n(0) =$$
$$= 0110100110 \cdots = \varepsilon_0 \varepsilon_1 \varepsilon_2 \varepsilon_3 \cdots.$$

We are interested in the factor and palindromic complexity of infinite words $\mathbf{w} = m_0 m_1 m_2 \cdots$ given by

$$m_n = \varepsilon_{n+1} - (2t - b - 1)\varepsilon_n + t - 1, \qquad (1)$$

where $2t > b \geq 1$. In particular, we want to determine both complexities for all the three cases stated in [1, Theorem 2], i.e., for $2t \geq b+3$, $2t = b+2$ and $2t = b+1$. If $2t = b+1$, then $m_n = \varepsilon_{n+1} + t - 1$ and so $\mathbf{w}$ equals the word $0^{-1}\mathbf{u}_{\mathrm{TM}}$ (after renaming the letters $0 \to t-1$ and $1 \to t$) having the same factor and palindromic complexity. Analogously, in the other two cases $2t \geq b+3$ and $2t = b+2$, it is sufficient to consider only one choice of parameters t and b satisfying the inequality and the equality, respectively. That is because the former formula implies that the word $\mathbf{w}$ consists of four distinct letters $b-t < b-t+1 < t-1 < t$ and the latter one that the word $\mathbf{w}$ consists of three distinct letters $b-t < t-1 < t$. If we choose $t = b = 3$ and

$t = b = 2$ respectively, all the other words given by (1) are (after renaming the letters) equal to the words corresponding to these two choices of parameters b and t. Thus, we can simplify the definition of the infinite words we study as follows.

Definition 1. *For $a = 1, 2$, the infinite word $\mathbf{w}_a = m_0 m_1 m_2 \cdots$ is defined by*

$$m_n = \varepsilon_{n+1} - a\varepsilon_n + a = \varepsilon_{n+1} + a(1 - \varepsilon_n). \quad (2)$$

Hence, we get

$$
\begin{aligned}
\mathbf{w}_1 &= 210201210120 \cdots \\
\mathbf{w}_2 &= 310302310230 \cdots
\end{aligned}
$$

As we will see, the factor and palindromic complexity of the word $\mathbf{w}_a$ will be expressed using the factor and palindromic complexity of the Thue-Morse word $\mathbf{u}_{\mathrm{TM}}$. Therefore we recall the following two theorems.

Theorem 2 ([7], [8]). *For the Thue-Morse sequence, $\mathcal{C}_{\mathbf{u}_{\mathrm{TM}}}(1) = 2$, $\mathcal{C}_{\mathbf{u}_{\mathrm{TM}}}(2) = 4$ and, for $n \geq 3$, if $n = 2^r + q + 1$, $r \geq 0$, $0 \leq q < 2^r$, then*

$$
\mathcal{C}_{\mathbf{u}_{\mathrm{TM}}}(n) = \begin{cases} 6 \times 2^{r-1} + 4q & 0 \leq q \leq 2^{r-1}, \\ 2^{r+2} + 2q & 2^{r-1} < q < 2^r. \end{cases}
$$

Theorem 3 ([9]). *Let $n \geq 3$ and $n = 2 \times 4^k + q$, $k \in \mathbb{N}$, $0 \leq q < 6 \times 4^k$, then*

$$
\mathcal{P}_{\mathbf{u}_{\mathrm{TM}}}(2n) = \begin{cases} 4 & 0 < q \leq 3 \times 4^k, \\ 2 & 3 \times 4^k < q < 3 \times 4^k \text{ or } q = 0. \end{cases}
$$

Furthermore, $\mathcal{P}_{\mathbf{u}_{\mathrm{TM}}}(1) = \mathcal{P}_{\mathbf{u}_{\mathrm{TM}}}(2) = \mathcal{P}_{\mathbf{u}_{\mathrm{TM}}}(3) = \mathcal{P}_{\mathbf{u}_{\mathrm{TM}}}(4) = 2$ and there are no palindromes of odd length greater than 3.

3. Factor complexity

The following lemma points out the similarity between the languages of the words $\mathbf{u}_{\mathrm{TM}}$ and $\mathbf{w}_a$.

Lemma 4. *There exists a bijective mapping from $\mathcal{L}_{\mathbf{u}_{\mathrm{TM}}}(n+1)$ to $\mathcal{L}_{\mathbf{w}_a}(n)$ for all $n \geq 2$.*

Proof. The mapping is defined by (2). We just have to prove that it is injective. Let $m_q \cdots m_{q+k}$ and $m_p \cdots m_{p+k}$ be two occurrences of the same factor of $\mathbf{w}_a$, $k \geq 1$, $q \neq p$. We prove that the factors $\varepsilon_q \varepsilon_{q+1} \cdots \varepsilon_{q+k+1}$ and $\varepsilon_p \varepsilon_{p+1} \cdots \varepsilon_{p+k+1}$ are the same as well. Obviously, it suffices to prove it for the case of $k = 1$. Let

$$m_q = \varepsilon_{q+1} + a(1 - \varepsilon_q) =$$
$$= m_p = \varepsilon_{p+1} + a(1 - \varepsilon_p)$$

and

$$m_{q+1} = \varepsilon_{q+2} + a(1 - \varepsilon_{q+1}) =$$
$$= m_{p+1} = \varepsilon_{p+2} + a(1 - \varepsilon_{p+1}).$$

Since there are only 8 possible three letter binary words $\varepsilon_p \varepsilon_{p+1} \varepsilon_{p+2}$ and $\varepsilon_q \varepsilon_{q+1} \varepsilon_{q+2}$, it is easy to find all solutions of these two equations. If $a = 2$, then $\varepsilon_p \varepsilon_{p+1} \varepsilon_{p+2} = \varepsilon_q \varepsilon_{q+1} \varepsilon_{q+2}$ is the unique solution of this system of two equations. If $a = 1$, $\varepsilon_q = \varepsilon_{q+1} = \varepsilon_{q+2} \neq \varepsilon_p = \varepsilon_{p+1} = \varepsilon_{p+2}$ is the only other solution, but it is not admissible since neither 000 nor 111 are factors of $\mathbf{u}_{\mathrm{TM}}$. ∎

This lemma allows us to determine the factor complexity $\mathcal{C}_{\mathbf{w}_a}(n)$ for $n \geq 2$. The case $n = 1$ is trivial, $\mathcal{C}_{\mathbf{w}_a}(1)$ is equal to the number of letters occurring in $\mathbf{w}_a$.

Corollary 5. *For both $a = 1$ and $a = 2$ and for all $n \geq 2$, it holds*

$$\mathcal{C}_{\mathbf{w}_a}(n) = \mathcal{C}_{\mathbf{u}_{\mathrm{TM}}}(n+1).$$

Furthermore, $\mathcal{C}_{\mathbf{w}_1}(1) = 3$ and $\mathcal{C}_{\mathbf{w}_2}(1) = 4$.

Corollary 6. *For both $a = 1$ and $a = 2$, $\mathbf{w}_a$ is square-free.*

Proof. Let ww be a factor of $\mathbf{w}_a$, w is of length n, and let $ww = m_i \cdots m_{i+2n-1}$. Then, according to the previous lemma, there exists a unique factor v of length n having b as its first letter such that $vvb = \varepsilon_i \cdots \varepsilon_{i+2n}$ is a factor of $\mathbf{u}_{\mathrm{TM}}$. But this is not possible since $\mathbf{u}_{\mathrm{TM}}$ is overlap-free (see e.g. [10]), which means exactly that it does not contain factors of this form. ∎

4. Palindromic complexity

As for the palindromic complexity, the difference between the cases $a = 1$ and $a = 2$ is more significant than it is for the factor complexity. However, the result still remains strongly related to the palindromic complexity of $\mathbf{u}_{\mathrm{TM}}$. First simple observation is that, since $\mathbf{w}_a$ is square-free for both values of a, it cannot contain palindromes of even length since such palindrome contains the square of a letter in its middle.

Definition 7. *Let $\mathcal{A} = \{0, 1, \ldots, n\}, a \in \mathcal{A}$ and $v = v_1 \cdots v_m \in \mathcal{A}^*, n, m \geq 1$. Set $\bar{a} = n - a$ and $\bar{v} = \bar{v}_1 \cdots \bar{v}_m$.*

Lemma 8. *Let $p \geq 2$ be even.*
A word $m_n m_{n+1} \ldots m_{n+p}$ is a palindrome of $\mathbf{w}_2$ if and only if

$$
\begin{aligned}
\varepsilon_n = \varepsilon_{n+2} = \cdots = \varepsilon_{n+p-2} = \varepsilon_{n+p}, \\
\varepsilon_{n+1} = \varepsilon_{n+3} = \cdots = \varepsilon_{n+p-1} = \varepsilon_{n+p+1},
\end{aligned} \quad (3)
$$

where $\varepsilon_{n+1} \neq \varepsilon_n$.
A word $m_n m_{n+1} \ldots m_{n+p}$ is a palindrome of $\mathbf{w}_1$ if and only if

$$
\begin{aligned}
\varepsilon_n &= \overline{\varepsilon_{n+p+1}} \\
&\vdots \qquad \vdots \\
\varepsilon_{n+\frac{p}{2}} &= \overline{\varepsilon_{n+\frac{p}{2}+1}}.
\end{aligned} \quad (4)
$$

Proof. We have for all $i = 0, 1, \ldots, \frac{p}{2} - 1$

$$m_{n+i} = \varepsilon_{n+i+1} + a(1 - \varepsilon_{n+i}) =$$
$$= m_{n+p-i} = \varepsilon_{n+p-i+1} + a(1 - \varepsilon_{n+p-i})$$
$$m_{n+i+1} = \varepsilon_{n+i+2} + a(1 - \varepsilon_{n+i+1}) = \qquad (5)$$
$$= m_{n+p-i-1} = \varepsilon_{n+p-i} + a(1 - \varepsilon_{n+p-i-1}),$$

where $m_{n+i} \neq m_{n+i+1}$ due to the square-freeness of $\mathbf{w}_a$. These two equations have a trivial solution

$$\varepsilon_{n+i} = \varepsilon_{n+i+2} = \varepsilon_{n+p-i} \neq \varepsilon_{n+i+1} =$$
$$= \varepsilon_{n+p-i+1} = \varepsilon_{n+p-i-1}$$

for $i = 0, 1, \ldots, \frac{p}{2} - 2$. For the case $a = 2$, it is the only solution.

If $a = 1$, we can rewrite (5) as

$$\begin{aligned}
\varepsilon_{n+1} + \varepsilon_{n+p} &= \varepsilon_n + \varepsilon_{n+p+1} \\
\varepsilon_{n+2} + \varepsilon_{n+p-1} &= \varepsilon_{n+1} + \varepsilon_{n+p} \\
&\vdots \\
\varepsilon_{n+\frac{p}{2}} + \varepsilon_{n+\frac{p}{2}+1} &= \varepsilon_{n+\frac{p}{2}-1} + \varepsilon_{n+\frac{p}{2}+2}.
\end{aligned}$$

Now, considering that $\varepsilon_n = \varepsilon_{n+p+1}$ leads to inadmissible solution $\varepsilon_n = \varepsilon_{n+1} = \cdots = \varepsilon_{n+p+1}$, therefore, the factor $\varepsilon_n \varepsilon_{n+1} \cdots \varepsilon_{n+p+1}$ is a solution if and only if (4) is satisfied. ∎

Thus, in the case of $a = 2$, the existence of a palindrome of odd length $p + 1$, $p \geq 2$ is equivalent to the existence of the factors $1010 \cdots 10$ or $0101 \cdots 01$ in $\mathbf{u}_{\mathrm{TM}}$ of length $p + 2$. But such words are factors of $\mathbf{u}_{\mathrm{TM}}$ only for $p = 2$.

Theorem 9.

$$\mathcal{P}_{\mathbf{w}_2}(n) = \begin{cases} 4 & n = 1, \\ 2 & n = 3, \\ 0 & otherwise. \end{cases}$$

In order to describe the relation between the palindromic complexities of $\mathbf{w}_1$ and $\mathbf{u}_{\mathrm{TM}}$, we need to introduce the following definition.

Definition 10. *A factor v of an infinite word $\mathbf{u}$ is said to be a C-palindrome if $\overline{v} = \widetilde{v}$. Denote $\mathcal{CP}_{\mathbf{u}}(n)$ the number of C-palindromes of length n in $\mathbf{u}$.*

Lemma 8 says that there exists a bijective mapping between the set of palindromes in $\mathbf{w}_1$ of odd length $p + 1$, $p \geq 2$, and the set of C-palindromes in $\mathbf{u}_{\mathrm{TM}}$ of length $p + 2$.

Corollary 11. *For $n \geq 1$*

$$\mathcal{P}_{\mathbf{w}_1}(2n + 1) = \mathcal{CP}_{\mathbf{u}_{\mathrm{TM}}}(2n + 2).$$

Lemma 12. *For all positive integers n it holds that*

$$\mathcal{CP}_{\mathbf{u}_{\mathrm{TM}}}(2n) = \mathcal{P}_{\mathbf{u}_{\mathrm{TM}}}(4n).$$

Proof. It is readily seen that if v is a C-palindrome in $\mathbf{u}_{\mathrm{TM}}$ of length $2n$, then $\varphi_{\mathrm{TM}}(v)$ is a palindrome of length $4n$. Similarly, if v' is a palindrome of length $4n$, then there exists a unique C-palindrome v of length $2n$ such that $\varphi_{\mathrm{TM}}(v) = v'$. ∎

Theorem 13. *For $n \geq 1$*

$$\mathcal{P}_{\mathbf{w}_1}(2n + 1) = \mathcal{P}_{\mathbf{u}_{\mathrm{TM}}}(4n + 4),$$

$\mathcal{P}_{\mathbf{w}_1}(1) = 3$. *There are no palindromes of even length in $\mathbf{w}_1$.*

5. REMARKS

As remarked in [1, Remark 5], $\mathbf{w}_1 = 210201210120 \cdots$ is exactly the square-free Braunholtz sequence on three letters given in [11]. Moreover, this sequence is in fact the sequence $\mathbf{u}_I$ which can be defined as the fixed point of Istrail's substitution $1 \mapsto 102$, $0 \mapsto 12$, $2 \mapsto 0$ [12], thus we obtain $\mathbf{w}_1$ from $\mathbf{u}_I$ by exchanging letters $1 \leftrightarrow 2$, $2 \leftrightarrow 0$, $0 \leftrightarrow 1$. Then, of course, the factor complexity of $\mathbf{u}_I$ and $\mathbf{w}_1$ is the same. The word $\mathbf{u}_I$ was studied in [13], where its factor complexity is computed using the notion of (right) special factors. In [13] the sequence $\mathbf{u}_I$ is referred to as the Thue-Morse word on three symbols and as it is recalled there it was originally defined by Thue [14, 15] and later on rediscovered in various contexts by several authors, such as Morse [16]. Another relation between $\mathbf{u}_I$ and $\mathbf{u}_{\mathrm{TM}}$ is also pointed out there: if we define a (non-primitive) substitution $\delta(1) \mapsto 011, \delta(0) \mapsto 01, \delta(2) \mapsto 0$, we have $\delta(\mathbf{u}_I) = \mathbf{u}_{\mathrm{TM}}$. Consequently, $\delta'(\mathbf{w}_1) = \mathbf{u}_{\mathrm{TM}}$ for $\delta'(2) \mapsto 011$, $\delta'(1) \mapsto 01$, $\delta'(0) \mapsto 0$.

REFERENCES

[1] J.-P. Allouche, et al. Univoque numbers and an avatar of Thue-Morse. *Acta Arith* **136**:319–329, 2009.

[2] V. Komornik, et al. Unique developments in non-integer bases. *Amer Math Monthly* **105**:636–639, 1998.

[3] J.-P. Allouche, et al. The Komornik-Loreti constant is transcendental. *Amer Math Monthly* **107**:448–449, 2000.

[4] V. Komornik, et al. Subexpansions, superexpansions and uniqueness properties in non-integer bases. *Period Math Hungar* **44**:197–218, 2002.

[5] M. de Vries, et al. Unique expansions of real numbers. *Adv Math* **221**(2):390–427, 2009.

[6] J. Cassaigne. Special factors of sequences with linear subword complexity. In *Developments in Language Theory*, pp. 25–34. World Scientific, 1996.

[7] A. De Luca. On some combinatorial problems in free monoids. *Discrete Math* **38**:207–225, 1982.

[8] S. Brlek. Enumeration of factors in the Thue-Morse word. *Discrete Appl Math* **24**:83–96, 1989.

[9] A. Blondin-Massé, et al. Palindromic lacunas of the Thue-Morse word. *Proc GASCOM 2008* pp. 53–67, 2008.

Ch. Frougny, K. Klouda

[10] J.-P. Allouche, et al. Palindrome complexity. *Theor Comput Sci* **292**:9–31, 2003.

[11] C. H. Braunholtz. An infinite sequence on 3 symbols with no adjacent repeats, (Solution to Problem 439 posed by H. Noland). *Amer Math Monthly* **70**:675–676, 1963.

[12] S. Istrail. On irreductible [sic] languages and nonrational numbers. *Bull Math Soc Sci Math R S Roumanie* **21**:301–308., 1977.

[13] A. De Luca, et al. On the factors of the Thue-Morse word on three symbols. *Inform Process Lett* **27**(6):281–285, 1988.

[14] A. Thue. Über unendliche Zeichenreihen. *Norske Vid Skrifter I Mat-Nat Kl Chris* **7**:1–22, 1906.

[15] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Vid Skrifter I Mat-Nat Kl Chris* **8**:1–67, 1912.

[16] M. Morse, et al. Unending chess, symbolic dynamics and a problem in semigroups. *Duke Math J* **11**:1–7, 1944.