On the additive complexity of a Thue-Morse like sequence

Jin Chen[a], Zhixiong Wen[b], Wen Wu[c,*]

[a] *College of Science, Huazhong Agricultural University, Wuhan 430070, China.*
[b] *School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, 430074, China.*
[c] *School of Mathematics, South China University of Technology, Guangzhou 510641, China.*

Abstract

In this paper, we study the additive complexity $\rho^+_{\mathbf{t}}(n)$ of a Thue-Morse like sequence $\mathbf{t} = \sigma^\infty(0)$ with the morphism $\sigma : 0 \to 01, 1 \to 12, 2 \to 20$. We show that $\rho^+_{\mathbf{t}}(n) = 2\lfloor \log_2(n) \rfloor + 3$ for all integers $n \geq 1$. Consequently, $(\rho_{\mathbf{t}}(n))_{n\geq 1}$ is a 2-regular sequence.

Keywords: Thue-Morse like sequence, Additive complexity, k-regular sequence
2010 MSC: 05D99, 11B85

1. Introduction

Recently the study of the abelian complexity of infinite words was initiated by G. Richomme, K. Saari, and L. Q. Zamboni [14]. For example, the abelian complexity functions of some notable sequences, such as the Thue-Morse sequence and all Sturmian sequences, were studied in [14] and [5] respectively. There are also many other works including the unbounded abelian complexity, see [3, 7, 8, 10, 13] and references therein. At the mean time, many authors had devoted to the generalizations of the abelian complexity. For instance, l-abelian complexity, cyclic complexity and binomial complexity are first presented in [9], [4] and [12] respectively. In 1994, G. Pirillo and S. Varricchio [11] raised the following question: do there exist infinite words avoiding additive squares or additive cubes? Based on this infamous problem, H. Ardal, T. Brown, V. Jungić and J. Sahasrabudhe proposed the additive complexity for infinite word on a finite subset of $\mathbb{Z}$ in [1]. It follows from the definition of additive equivalence in Section 2 that the additive complexity $\{\rho^+(n)\}$ coincides with the abelian complexity $\{\rho^{ab}(n)\}$ for every infinite word on the alphabet composed of two integers. For every infinite word on the alphabet composed of integers whose cardinality is not less than three, it is easy to know that $\rho^+(n) \leq \rho^{ab}(n)$ for every n.

Let σ be the morphism $0 \mapsto 01, 1 \mapsto 12, 2 \mapsto 20$ on $\{0, 1, 2\}$ and $\mathbf{t} := \sigma^\infty(0)$. The infinite sequence $\mathbf{t}$ is a Thue-Morse like sequence (see [2, 15]). Further, $\mathbf{t}$ is 2-automatic and uniformly recurrent (see [6]). A sequence $\mathbf{w} = w_0 w_1 w_2 \cdots$ is a *k-automatic* sequence if its *k-kernel* $\{(w_{k^e n+c})_{n\geq 0} \mid e \geq 0, 0 \leq c < k^e\}$ finite. If the $\mathbb{Z}$-module generated by its k-kernel is finitely generated, then $\mathbf{w}$ is a *k-regular* sequence.

In this paper, we investigate the additive complexity function $\rho^+_{\mathbf{t}}(n)$ of $\mathbf{t}$, where $\rho^+_{\mathbf{t}}(n)$ is the number of different digit sums of all words (of length n) that occur in $\mathbf{t}$. We give the explicit value of $(\rho^+_{\mathbf{t}}(n))_{n\geq 1}$.

*Wen Wu is the corresponding author.
Email addresses: cj@mail.hzau.edu.com (Jin Chen), zhi-xiong.wen@hust.edu.cn (Zhixiong Wen), wuwen@scut.edu.cn (Wen Wu)

Theorem 1. *For all integer $n \geq 1$,*

$$\rho_{\mathbf{t}}^{+}(n) = 2\lfloor \log_2 n \rfloor + 3.$$

Consequently, we know that the additive complexity function $(\rho_{\mathbf{t}}(n))_{n\geq 1}$ satisfying the recurrence relations: $\rho_{\mathbf{t}}^{+}(1) = 3$ and for all $n \geq 1$,

$$\rho_{\mathbf{t}}^{+}(2n) = \rho_{\mathbf{t}}^{+}(2n+1) = \rho_{\mathbf{t}}^{+}(n) + 2.$$

The above recurrence relations imply the regularity of $(\rho_{\mathbf{t}}(n))_{n\geq 1}$.

Corollary 1. *The additive complexity $(\rho_{\mathbf{t}}(n))_{n\geq 1}$ of $\mathbf{t}$ is a 2-regular sequence.*

It is natural to ask that whether the additive complexity of every k-automatic sequence be always a k-regular sequence.

This paper is organized as follows. In Section 2, we give some notations. In Section 3, we prove Theorem 1. The proof is separated into 3 steps. Each step gives a more specific result.

2. Preliminaries

An *alphabet* $\mathcal{A}$ is a finite and non-empty set (of symbols) whose elements are called *letters*. A (finite) *word* over the alphabet $\mathcal{A}$ is a concatenation of letters in $\mathcal{A}$. The concatenation of two words $u = u_0u_1\cdots u(m)$ and $v = v_0v_1\cdots v_n$ is the word $uv = u_0u_1\cdots u_m v_0v_1\cdots v_n$. The set of all finite words over $\mathcal{A}$ including the *empty word* ε is denoted by $\mathcal{A}^*$. An infinite word $\mathbf{w}$ is an infinite sequence of letters in $\mathcal{A}$. The set of all infinite words over $\mathcal{A}$ is denoted by $\mathcal{A}^{\mathbb{N}}$.

The *length* of a finite word $w \in \mathcal{A}^*$, denoted by $|w|$, is the number of letters contained in w. We set $|\varepsilon| = 0$. For any word $u \in \mathcal{A}^*$ and any letter $a \in \mathcal{A}$, let $|u|_a$ denote the number of occurrences of a in u.

A word w is a factor of a finite (or an infinite) word v, written by $w \prec v$ if there exist a finite word x and a finite (or an infinite) word y such that $v = xwy$. When $x = \varepsilon$, w is called a *prefix* of v, denoted by $w \triangleleft v$; when $y = \varepsilon$, w is called a suffix of v, denoted by $w \triangleright v$.

For a real number x, let $\lfloor x \rfloor$ (resp. $\lceil x \rceil$) be the integer that is less (resp. larger) than or equal to x. For every natural number n and some positive integer $b \geq 2$, set $(n)_b$ be the regular b-ary expansion of n.

2.1. Additive complexity

Now we assume that $\mathcal{A} \subset \mathbb{Z}$. Let

$$\mathbf{w} = w_0w_1w_2\cdots \in \mathcal{A}^{\mathbb{N}}$$

be an infinite word. Denote by $\mathcal{F}_{\mathbf{w}}(n)$ the set of all factors of $\mathbf{w}$ of length n, i.e.,

$$\mathcal{F}_{\mathbf{w}}(n) := \{w_iw_{i+1}\cdots w_{i+n-1} : i \geq 0\}.$$

Write $\mathcal{F}_{\mathbf{w}} = \cup_{n\geq 1}\mathcal{F}_{\mathbf{w}}(n)$. The *subword complexity function* $\rho_{\mathbf{w}} : \mathbb{N} \to \mathbb{N}$ of $\mathbf{w}$ is defined by

$$\rho_{\mathbf{w}}(n) := \#\mathcal{F}_{\mathbf{w}}(n).$$

Denote the *digit sum* of $u = u_0\cdots u_{|u|-1} \in \mathcal{A}^*$ by

$$\mathrm{DS}(u) := \sum_{j=0}^{|u|-1} u_j.$$

Two finite words u, $v \in \mathcal{A}^*$ is *additive equivalent* if $\mathrm{DS}(u) = \mathrm{DS}(v)$. The additive equivalent induces an equivalent relation, denoted by $\sim_+$.

Definition 1. *The* additive subword complexity function $\rho_{\mathbf{w}}{}^{+}: \mathbb{N} \to \mathbb{N}$ *of* $\mathbf{w}$ *is defined by*

$$\rho_{\mathbf{w}}^{+}(n) := \#\{\mathcal{F}_{\mathbf{w}}(n)/\sim_{+}\}.$$

In fact,

$$\rho_{\mathbf{w}}^{+}(n) = \#\{\mathrm{DS}(u) : u \in \mathcal{F}_{\mathbf{w}}(n)\}. \tag{2.1}$$

3. Additive complexity of t

In this section, we prove Theorem 1. According to (2.1), the study of the additive complexity function turns out to be the study of digit sums of all factors. Our strategy in the proof of Theorem 1 is the following:

- (Proposition 1) give the upper and lower bounds of $\rho_{\mathbf{t}}^{+}(n)$ for all $n \geq 1$;
- (Proposition 2) show that the upper and lower bounds can be attained;
- (Proposition 3) study the all the accessible values of digit sums.

Then, Theorem 1 follows from Proposition 1, 2 and 3.

3.1. Upper and lower bounds of digit sums of factors

Proposition 1. *For every integer* $n \geq 1$,

$$n - \lfloor \log_2 n \rfloor - 1 \leq \mathrm{DS}(u) \leq n + \lfloor \log_2 n \rfloor + 1$$

for all $u \in \mathcal{F}_{\mathbf{t}}(n)$.

Note that for every $u = u_0 u_1 \cdots u_{n-1} \in \mathcal{F}_{\mathbf{t}}(n)$,

$$\begin{aligned}\mathrm{DS}(u) = \sum_{i=0}^{n-1} u_i &= 0 \cdot |u|_0 + 1 \cdot |u|_1 + 2 \cdot |u|_2 = |u|_0 + |u|_1 + 2|u|_2 - |u|_0 \\ &= n + |u|_2 - |u|_0. \end{aligned} \tag{3.1}$$

To prove Proposition 1, we only need to show that for all $u \in \mathcal{F}_{\mathbf{t}}(n)$,

$$-\lfloor \log_2 n \rfloor - 1 \leq |u|_2 - |u|_0 \leq \lfloor \log_2 n \rfloor + 1. \tag{3.2}$$

The following lemmas are aimed to analysis the quantity $|u|_2 - |u|_0$.

Lemma 1. *For every* $u \in \{0, 1, 2\}^*$,

$$\begin{aligned} |\sigma(u)|_2 - |\sigma(u)|_0 &= |u|_1 - |u|_0, \\ |\sigma(u)|_1 - |\sigma(u)|_0 &= |u|_1 - |u|_2. \end{aligned}$$

Proof. It follows from the definition of σ that

$$|\sigma(u)|_0 = |u|_0 + |u|_2, \quad |\sigma(u)|_1 = |u|_0 + |u|_1, \quad |\sigma(u)|_2 = |u|_1 + |u|_2.$$

The above equations give the required results. □

Let a, b, c be any arrangement of $0, 1, 2$. Define $\tau_c : a \mapsto b, b \mapsto a, c \mapsto c$. For every finite word $w = w_0 w_1 \cdots w_{n-1} w_n \in \{0, 1, 2\}^*$, let $w^R = w_n w_{n-1} \cdots w_1 w_0$ be the mirror of w. For every $x \in \{0, 1, 2\}$, write $\underline{x} := x - 1 \pmod 3$ and $\overline{x} := x + 1 \pmod 3$. The morphisms σ and τ have the following commutative property.

Lemma 2. *For every $u \in \mathcal{F}_{\mathbf{t}}$ and every $c = 0, 1, 2$,*

$$\sigma(\tau_c(u)^R) = \tau_{\underline{c}}(\sigma(u))^R. \tag{3.3}$$

Proof. It is easy to check (3.3) for all $u \in \mathcal{F}_{\mathbf{t}}(1) = \{0, 1, 2\}$. Assume that (3.3) holds for all $u \in \cup_{i=1}^{n-1} \mathcal{F}_{\mathbf{t}}(i)$. For any $u \in \mathcal{F}_{\mathbf{t}}(n)$, we have $u = va$ where $v \in \mathcal{F}_{\mathbf{t}}(n-1)$ and $a \in \{0, 1, 2\}$. Then

$$\begin{aligned} \sigma(\tau_c(u)^R) &= \sigma(\tau_c(va)^R) = \sigma(\tau_c(a)^R \tau_c(v)^R) \\ &= \sigma(\tau_c(a)^R)\sigma(\tau_c(v)^R) = \tau_{\overline{c}}(\sigma(a))^R \tau_{\overline{c}}(\sigma(v))^R \qquad \text{(by the assumption)} \\ &= \tau_{\overline{c}}(\sigma(va))^R = \tau_{\overline{c}}(\sigma(u))^R, \end{aligned}$$

which implies that (3.3) holds for all $u \in \mathcal{F}_{\mathbf{t}}(n)$ and $c = 0, 1, 2$. ☐

While σ maps every factor of $\mathbf{t}$ to a factor of $\mathbf{t}$, the morphism τ_c maps every factor of $\mathbf{t}$ to the mirror of some factor of $\mathbf{t}$.

Lemma 3. *If $u \in \mathcal{F}_{\mathbf{t}}$, then $\tau_c(u)^R \in \mathcal{F}_{\mathbf{t}}$ for $c = 0, 1, 2$.*

Proof. When $u \in \mathcal{F}_{\mathbf{t}}(1) \cup \mathcal{F}_{\mathbf{t}}(2)$, the result can be checked directly. Now, suppose the result holds for all $u \in \cup_{i=1}^{n-1} \mathcal{F}_{\mathbf{t}}(i)$ (where $n \geq 3$). Let $u \in \mathcal{F}_{\mathbf{t}}(n)$. If n is odd, then $u = a\sigma(v)$ or $\sigma(v)b$ where $v \in \mathcal{F}_{\mathbf{t}}(\lfloor n/2 \rfloor)$ and $a, b \in \{0, 1, 2\}$, which also imply that $\underline{a}u = \sigma(\underline{a}v)$ or $u\overline{b} = \sigma(vb)$ with $\underline{a}v, vb \in \mathcal{F}_{\mathbf{t}}(\frac{n+1}{2})$. By Lemma 2, for $c = 0, 1, 2$,

$$\tau_c(\underline{a}u)^R = \tau_c(\sigma(\underline{a}v))^R = \sigma(\tau_{\overline{c}}(\underline{a}v)^R).$$

Since $\underline{a}v \in \mathcal{F}_{\mathbf{t}}(\frac{n+1}{2})$, by the inductive hypothesis, $\tau_{\overline{c}}(\underline{a}v)^R \in \mathcal{F}_{\mathbf{t}}(\frac{n+1}{2})$. So $\tau_c(\underline{a}u)^R \in \mathcal{F}_{\mathbf{t}}(n+1)$ and $\tau_c(u)^R \in \mathcal{F}_{\mathbf{t}}(n)$. The same is true for the case $u = \sigma(v)b$.

If n is even, then $u = \sigma(w)$ or $a\sigma(v)b$ where $w \in \mathcal{F}_{\mathbf{t}}(n/2)$, $v \in \mathcal{F}_{\mathbf{t}}(\frac{n}{2} - 1)$ and $a, b \in \{0, 1, 2\}$. When $u = a\sigma(v)b$, we have $\underline{a}ub = \sigma(\underline{a}vb)$ with $\underline{a}vb \in \mathcal{F}_{\mathbf{t}}(\frac{n}{2} + 1)$. By Lemma 2, for $c = 0, 1, 2$,

$$\tau_c(\underline{a}ub)^R = \tau_c(\sigma(\underline{a}vb))^R = \sigma(\tau_{\overline{c}}(\underline{a}vb)^R).$$

By the inductive hypothesis, $\tau_{\overline{c}}(\underline{a}vb)^R \in \mathcal{F}_{\mathbf{t}}$. So $\tau_c(\underline{a}ub)^R \in \mathcal{F}_{\mathbf{t}}$ which implies $\tau_c(u) \in \mathcal{F}_{\mathbf{t}}$. When $u = \sigma(w)$, the result follows from Lemma 2 and the inductive hypothesis in the same way. ☐

Lemma 4. *Let $n \geq 1$ be an integer and $u \in \mathcal{F}_{\mathbf{t}}(n)$.*

1. *There exists $x \in \mathcal{F}_{\mathbf{t}}(\lfloor n/2 \rfloor)$ such that*

$$|x|_1 - |x|_0 - 1 \leq |u|_2 - |u|_0 \leq |x|_1 - |x|_0 + 1. \tag{3.4}$$

2. *There exists $y \in \mathcal{F}_{\mathbf{t}}(\lfloor n/2 \rfloor)$ such that*

$$|y|_1 - |y|_2 - 1 \leq |u|_1 - |u|_0 \leq |y|_1 - |y|_2 + 1. \tag{3.5}$$

3. *There exists $z \in \mathcal{F}_{\mathbf{t}}(\lfloor n/2 \rfloor)$ such that*

$$|z|_0 - |z|_2 - 1 \leq |u|_1 - |u|_2 \leq |z|_0 - |z|_2 + 1.$$

Proof. (1) If n is odd, then $u = a\sigma(v)$ or $\sigma(v)b$ where $v \in \mathcal{F}_{\mathbf{t}}(\lfloor n/2 \rfloor)$ and $a, b \in \{0, 1, 2\}$. In either case,

$$|u|_2 - |u|_0 = \begin{cases} |\sigma(v)|_2 - |\sigma(v)|_0 - 1, & \text{if } a, b = 0, \\ |\sigma(v)|_2 - |\sigma(v)|_0, & \text{if } a, b = 1, \\ |\sigma(v)|_2 - |\sigma(v)|_0 + 1, & \text{if } a, b = 2. \end{cases}$$

$$= \begin{cases} |v|_1 - |v|_0 - 1, & \text{if } a, b = 0, \\ |v|_1 - |v|_0, & \text{if } a, b = 1, \\ |v|_1 - |v|_0 + 1, & \text{if } a, b = 2. \end{cases} \qquad \text{(by Lemma 1)}$$

Letting $x = v$, the result follows.

If n is even, then $u = \sigma(w)$ or $a\sigma(v)b$ where $w \in \mathcal{F}_{\mathbf{t}}(n/2)$, $v \in \mathcal{F}_{\mathbf{t}}(\frac{n}{2} - 1)$ and $a, b \in \{0, 1, 2\}$. When $u = \sigma(w)$, by Lemma 1, $|u|_2 - |u|_0 = |w|_1 - |w|_0$. Choosing $x = w$, we have the desired result. When $u = a\sigma(v)b$, let $\underline{a} = a - 1 \pmod 3$ and $\overline{b} = b + 1 \pmod 3$. Then $\underline{a}u\overline{b} = \sigma(\underline{a}vb)$. By Lemma 1,

$$|\underline{a}u\overline{b}|_2 - |\underline{a}u\overline{b}|_0 = |\sigma(\underline{a}vb)|_2 - |\sigma(\underline{a}vb)|_0 = |\underline{a}vb|_1 - |\underline{a}vb|_0,$$

which implies

$$|u|_2 - |u|_0 = |\underline{a}vb|_1 - |\underline{a}vb|_0 + |\underline{a}\overline{b}|_0 - |\underline{a}\overline{b}|_2.$$

When $ab \neq 00$ and 12,

$$|u|_2 - |u|_0 = |\underline{a}v|_1 - |\underline{a}v|_0 + \begin{cases} -1, & \text{if } ab = 01, 20, \\ 0, & \text{if } ab = 02, 10, 21, \\ 1, & \text{if } ab = 11, 22. \end{cases}$$

The result holds by choosing $x = \underline{a}v$. When $ab = 00$ or 12,

$$|u|_2 - |u|_0 = |vb|_1 - |vb|_0 + \begin{cases} -1, & \text{if } ab = 00, \\ 1, & \text{if } ab = 12. \end{cases}$$

Setting $x = vb$, we are done.

(2) Let $u \in \mathcal{F}_{\mathbf{t}}$. By Lemma 3, $\tau_0(u)^R \in \mathcal{F}_{\mathbf{t}}$ and

$$|u|_1 - |u|_0 = |\tau_0(u)^R|_2 - |\tau_0(u)^R|_0.$$

Applying (3.4) to $\tau_0(u)^R$, we have $x \in \mathcal{F}_{\mathbf{t}}$ such that

$$|x|_1 - |x|_0 - 1 \leq |\tau_0(u)^R|_2 - |\tau_0(u)^R|_0 \leq |x|_1 - |x|_0 + 1.$$

Let $y = \tau_1(x)^R$. Then, $y \in \mathcal{F}_{\mathbf{t}}$ and $|y|_1 - |y|_2 = |x|_1 - |x|_0$. We have the desired result.

(3) Applying (3.5) to $\tau_1(u)^R$ and letting $z = \tau_2(y)^R$, the result follows. □

Now we are ready to prove Proposition 1.

Proof of Proposition 1. For every $n \geq 1$, there exists $k \geq 1$ such that $2^k \leq n < 2^k + 1$. Suppose $u \in \mathcal{F}_{\mathbf{t}}(n)$. Let $n_1 = n$. By Lemma 4, we have $x^{(1)} \in \mathcal{F}_{\mathbf{t}}(\lfloor n_1/2 \rfloor)$ such that

$$|x^{(1)}|_1 - |x^{(1)}|_0 - 1 \leq |u|_2 - |u|_0 \leq |x^{(1)}|_1 - |x^{(1)}|_0 + 1.$$

Let $n_2 = \lfloor n_1/2 \rfloor$. Apply Lemma 4 to $x^{(1)}$, we have $x^{(2)} \in \mathcal{F}_{\mathbf{t}}(\lfloor n_2/2 \rfloor)$ such that

$$|x^{(2)}|_1 - |x^{(2)}|_2 - 1 \leq |x^{(1)}|_1 - |x^{(1)}|_0 \leq |x^{(2)}|_1 - |x^{(2)}|_2 + 1.$$

Therefore,
$$|x^{(2)}|_1 - |x^{(2)}|_2 - 2 \leq |u|_2 - |u|_0 \leq |x^{(2)}|_1 - |x^{(2)}|_2 + 2.$$
Let $n_3 = \lfloor n_2/2 \rfloor$ and apply Lemma 4 to $x^{(2)}$. Then we have $x^{(3)} \in \mathcal{F}_{\mathbf{t}}(\lfloor n_3/2 \rfloor)$ satisfying

$$|x^{(3)}|_0 - |x^{(3)}|_2 - 3 \leq |u|_2 - |u|_0 \leq |x^{(3)}|_0 - |x^{(3)}|_2 + 3.$$

After applying Lemma 4 k times as above, we obtain that

$$-1 - \lfloor \log_2 n \rfloor = -1 - k \leq |u|_2 - |u|_0 \leq 1 + k = 1 + \lfloor \log_2 n \rfloor.$$

Hence (3.2) holds. □

3.2. Maximal and minimal digit sums

Let $(d_k)_{k\geq -1} \in \{0,1,2\}^\infty$ where

$$d_k = \begin{cases} 0 & \text{if } k \equiv 3,4 \mod 6, \\ 1 & \text{if } k \equiv 1,2 \mod 6, \\ 2 & \text{if } k \equiv 0,5 \mod 6. \end{cases}$$

Let $(c_\ell)_{\ell\geq 1} \in \{0,1,2\}^\infty$ given by $c_\ell = \ell + 1 \,(\text{mod } 3)$. Applying Lemma 1 several times, it follows that for $k = 0,1,2,3,4,5$,

$$|\sigma^k(d_k)|_2 - |\sigma^k(d_k)|_0 = 1 \quad \text{and} \quad |\sigma^k(c_k)|_2 - |\sigma^k(c_k)|_0 = 0. \tag{3.6}$$

In fact, these equalities hold for all $k \geq 1$.

Lemma 5. *For all $\ell \geq 1$,*

$$|\sigma^\ell(d_\ell)|_2 - |\sigma^\ell(d_\ell)|_0 = 1 \quad \textit{and} \quad |\sigma^\ell(c_\ell)|_2 - |\sigma^\ell(c_\ell)|_0 = 0.$$

Proof. Applying Lemma 1 six times, one obtain that for every $u \in \{0,1,2\}^*$,

$$|\sigma^6(u)|_2 - |\sigma^6(u)|_0 = |u|_2 - |u|_0. \tag{3.7}$$

For all $\ell \geq 1$, we have $\ell = 6j + k$ where $j \geq 1$ and $k = 0,1,2,3,4,5$. Then $d_\ell = d_k$ and $c_\ell = c_k$. By (3.6) and (3.7),

$$|\sigma^\ell(d_\ell)|_2 - |\sigma^\ell(d_\ell)|_0 = |\sigma^{6j+k}(d_k)|_2 - |\sigma^{6j+k}(d_k)|_0 = |\sigma^k(d_k)|_2 - |\sigma^k(d_k)|_0 = 1$$

and

$$|\sigma^\ell(c_\ell)|_2 - |\sigma^\ell(c_\ell)|_0 = |\sigma^{6j+k}(c_k)|_2 - |\sigma^{6j+k}(c_k)|_0 = |\sigma^k(c_k)|_2 - |\sigma^k(c_k)|_0 = 0.$$

We have the desired. □

Now we define a sequence of words $\{W(n)\}_{n\geq 1}$ whose digit sums will attain the upper bound in Proposition 1.

Let $W_1 := 2$. For $n \geq 2$, W_n is defined as follows: suppose $2^k \leq n < 2^{k+1}$ for some k and the 2-adic expansion of $n - 2^k$ is written as

$$(n - 2^k)_2 = m_{k-1} \cdots m_2 m_1 m_0$$

where $m_j \in \{0,1\}$ for $j = 0, 1, \cdots k-1$. Define

$$W_L(n) := \delta_{m_0} 2 \Big(\prod_{i=1}^{\lfloor \frac{k-1}{2} \rfloor} \sigma^{2i+m_{2i}}(d_{2i+m_{2i}}) \Big)$$

and

$$W_R(n) := \Big(\prod_{i=\lceil \frac{k-1}{2} \rceil}^{1} \sigma^{2i-1+m_{2i-1}}(d_{2i-1+m_{2i-1}}) \Big) 2$$

where $\delta_{m_0} = \varepsilon$ if $m_0 = 0$ and 1 if $m_0 = 1$. Let $W(n) := W_L(n)W_R(n)$.

Lemma 6. *For every integer n satisfying $2^k \leq n < 2^{k+1}$ for some $k \geq 0$, we have*

1. *if k is even, then $W_L(n) \triangleright \sigma^k(d_{k+1})$ and $W_R(n) \triangleleft \sigma^{k+1}(d_k)$,*
2. *if k is odd, then $W_L(n) \triangleright \sigma^{k+1}(d_{k+2})$ and $W_R(n) \triangleleft \sigma^k(d_{k-1})$.*

Proof. For $k = 0, 1, 2$, the result can be verified directly from the definition of W_L and W_R. Suppose the result hold for all $m \leq k$. Now we prove it for $m = k+1$. Let $2^{k+1} \leq n < 2^{k+2}$ with $(n - 2^{k+1})_2 = m_k m_{k-1} \cdots m_1 m_0$. Set $n' = 2^k + \sum_{i=0}^{k-1} m_i 2^i$.

When $k+1$ is odd, write $k = 2\ell$. Then $W_R(n) = W_R(n') \triangleleft \sigma^{k+1}(d_k)$ and

$$\begin{aligned} W_L(n) &= \delta_{m_0} 2 \Big(\prod_{i=1}^{\ell} \sigma^{2i+m_{2i}}(d_{2i+m_{2i}}) \Big) \\ &= \delta_{m_0} 2 \Big(\prod_{i=1}^{\ell-1} \sigma^{2i+m_{2i}}(d_{2i+m_{2i}}) \Big) \sigma^{2\ell+m_{2\ell}}(d_{2\ell+m_{2\ell}}) \\ &= W_L(n')\sigma^{k+m_k}(d_{k+m_k}). \end{aligned}$$

When $m_k = 0$, by the induction hypothesis,

$$\begin{aligned} W_L(n')\sigma^{k+m_k}(d_{k+m_k}) &\triangleright \sigma^k(d_{k+1})\sigma^k(d_k) = \sigma^k(d_{k+2})\sigma^k(d_k) \\ &= \sigma^{k+1}(d_{k+2}) \triangleright \sigma^{(k+1)+1}(d_{(k+2)+1}). \end{aligned}$$

When $m_k = 1$, by the induction hypothesis,

$$\begin{aligned} W_L(n')\sigma^{k+m_k}(d_{k+m_k}) &\triangleright \sigma^k(d_{k+1})\sigma^{k+1}(d_{k+1}) \triangleright \sigma^{k+1}(d_{k+3})\sigma^{k+1}(d_{k+1}) \\ &= \sigma^{(k+1)+1}(d_{(k+1)+2}). \end{aligned}$$

So, $W_L(n) \triangleright \sigma^{(k+1)+1}(d_{(k+1)+2})$.

When $k+1$ is even, write $k = 2\ell+1$. Then $W_L(n) = W_L(n') \triangleright \sigma^{k+1}(d_{k+2})$ and

$$\begin{aligned} W_R(n) &= \sigma^{k+m_k}(d_{k+m_k})W_R(n') \\ &\triangleleft \begin{cases} \sigma^k(d_k)\sigma^k(d_{k-1}) = \sigma^{k+1}(d_{k+1}), & \text{if } m_k = 0, \\ \sigma^{k+1}(d_{k+1})\sigma^k(d_{k-1}), & \text{if } m_k = 1, \end{cases} \\ &\triangleleft \sigma^{k+2}(d_{k+1}). \end{aligned}$$

This completes the induction. ∎

Proposition 2. *For all $n \geq 1$, $W(n) \in \mathcal{F}_{\mathbf{t}}$ and $\tau_1(W(n))^R \in \mathcal{F}_{\mathbf{t}}$. Moreover,*

$$\mathrm{DS}(W(n)) = n + \lfloor \log_2 n \rfloor + 1 \ \text{ and } \ \mathrm{DS}(\tau_1(W(n))^R) = n - \lfloor \log_2 n \rfloor - 1.$$

Proof. For all n with $2^k \leq n < 2^{k+1}$, by Lemma 6,

$$W(n) = W_L(n) W_R(n) \prec \sigma^k(d_{k+2} d_k d_{k-2)} \prec \sigma^{k+1}(d_{k+2} d_{k-2}).$$

Since $(d_\ell)_{\ell \geq 1}$ is periodic, $d_{k+2} d_{k-2} \in \{21, 10, 02\} \subset \mathcal{F}_{\mathbf{t}}(2)$. Hence $W(n) \in \mathcal{F}_{\mathbf{t}}$. By Lemma 3, we know $\tau_1(W(n))^R \in \mathcal{F}_{\mathbf{t}}$.

According to the definition of W_L and W_R,

$$\begin{aligned} |W(n)|_2 - |W(n)|_0 &= 2 + \sum_{i=1}^{k-1} (|\sigma^{i+m_i}(d_{i+m_i})|_2 - |\sigma^{i+m_i}(d_{i+m_i})|_0) \\ &= 2 + k - 1 \qquad\qquad \text{(by Lemma 5)} \\ &= \lfloor \log_2(n) \rfloor + 1. \end{aligned}$$

Then the results follow from (3.1) and the definition of τ_1. □

3.3. Accessible values of digit sums

We shall prove the following intermediate value property of digit sums of all the factors of length n.

Proposition 3. *For all $n \geq 1$ and all integer k satisfying $n - \lfloor \log_2 n \rfloor - 1 < k < n + \lfloor \log_2 n \rfloor + 1$, there exists $u \in \mathcal{F}_{\mathbf{t}}(n)$ such that $\mathrm{DS}(u) = k$.*

Before proving Proposition 3, we first study the behavior of digit sums during the shift (to the right). Denote by $\mathcal{I}(u)$ the set of all the indexes (or positions) of occurrences of u, i.e., for every $i \in \mathcal{I}(u)$, $t_i t_{i+1} \cdots t_{i+n-1} = u$. Since $\mathbf{t}$ is uniformly recurrent, $\mathcal{I}(u)$ is an infinite set for all $u \in \mathcal{F}_{\mathbf{t}}$. For every $i \in \mathcal{I}(u)$, set

$$r_i(u) = \min\{j > i : g_n(j) > \mathrm{DS}(u)\},$$

where $g_n(j) := \mathrm{DS}(t_j t_{j+1} \cdots t_{j+n-1})$. Set $\min \emptyset = -\infty$.

Lemma 7. *Let $u \in \mathcal{F}_{\mathbf{t}}(n)$. If $\mathrm{DS}(u) \neq \mathrm{DS}(W(n))$, then $r_i(u)$ is finite and*

$$g_n(r_i(u)) - \mathrm{DS}(u) = 1 \ \text{ or } \ 2.$$

Moreover, if $g_n(r_i(u)) - \mathrm{DS}(u) = 2$, then $g_n(r_i(u) - 1) = \mathrm{DS}(u)$.

Proof. By Proposition 1 and 2, if $\mathrm{DS}(u) \neq \mathrm{DS}(W(n))$, then $\mathrm{DS}(u) < \mathrm{DS}(W(n))$. For any occurrence of u, say $t_i t_{i+1} \cdots t_{i+n-1} = u$, since $\mathbf{t}$ is uniformly recurrent, there exists $j > i$ such that $t_j t_{j+1} \cdots t_{j+n-1} = W(n)$. Thus, $r_i(u) < j$.

Now suppose $r_i(u)$ is finite. Write $k := r_i(u)$. Then, $g_n(k-1) \leq \mathrm{DS}(u)$. Since

$$g_n(k) - g_n(k-1) = t_{k+n-1} - t_{k-1} \in \{0, \pm 1, \pm 2\}$$

and $g_n(k) > \mathrm{DS}(u) \geq g_n(k-1)$, we know that $g_n(k) - \mathrm{DS}(u) = 1$ or 2. Moreover, if $g_n(k-1) < \mathrm{DS}(u)$, then $0 < g_n(k) - \mathrm{DS}(u) < g_n(k) - g_n(k-1) \leq 2$ which implies $g_n(k) - \mathrm{DS}(u) = 1$. □

The key to prove Proposition 3 is to figure out how many times we need to do the shift in order to increase the digit sum of a given factor by 1. The following two lemmas deal with the problem. The first one is a technical lemma. Let $u, v \in \mathcal{F}_{\mathbf{t}}(3)$. Write $\sigma^6(u) = u_0u_1 \cdots u_{191}$ and $\sigma^6(v) = v_0v_1 \cdots v_{191}$. For $64 \le i, j < 128$ satisfying $u_i = 0$ and $v_j = 2$, $0 < m < 192 - \max\{i, j\}$ and $0 < p \le \min\{i, j\}$, set

$$R(u, v, i, j, m) = \sum_{\ell=0}^{m} (v_{j+\ell} - u_{i+\ell}),$$

$$L(u, v, i, j, -p) = \sum_{\ell=1}^{p} (u_{i-\ell} - v_{j-\ell}).$$

Lemma 8. *For all* $u, v \in \mathcal{F}_{\mathbf{t}}(3)$ *and* $64 \le i, j < 128$ *satisfying* $u_i = 0$ *and* $v_j = 2$, $R(u, v, i, j, m) = 1$ *for some* $0 < m < 192 - \max\{i, j\}$ *or* $L(u, v, i, j, -p) = 1$ *for some* $0 < p \le \min\{i, j\}$.

Proof. Since the choices of variables of both L and R are finite, the result can be verified exhaustively. (This can be easily checked by a computer. We give the pseudocode for the corresponding procedures in Appendix A.) □

Lemma 9. *Let* $n > 128$. *For every* $u \in \mathcal{F}_{\mathbf{t}}(n)$ *with* $\mathrm{DS}(u) \neq \mathrm{DS}(W(n))$, *there exists* $z \in \mathcal{F}_{\mathbf{t}}(n)$ *satisfying* $\mathrm{DS}(z) - \mathrm{DS}(u) = 1$.

Proof. Let $i \in \mathcal{I}(u)$ with $i \ge 2^8$. Set $j = r_i(u) - 1$. By Lemma 7, if $g_n(r_i(u)) - \mathrm{DS}(u) = 1$, then we are done. If $g_n(r_i(u)) - \mathrm{DS}(u) = 2$, then $g_n(j) = \mathrm{DS}(u)$ which also implies $t_j = 0$ and $t_{j+n} = 2$. Write $w = t_jt_j \cdots t_{j+n-1}$.

The word w has the following decomposition:

$$w = (t_jt_{j+1} \cdots t_{j+\ell-1})\sigma^6(v)(t_{j+n-r} \cdots t_{j+n-1}) \prec \sigma^6(xvy)$$

where $v \in \mathcal{F}_t$, $t_jt_{j+1} \cdots t_{j+\ell-1} \triangleright \sigma^6(x)$ and $t_{j+n-r} \cdots t_{j+n-1} \triangleleft \sigma^6(y)$ for some $x, y \in \{0, 1, 2\}$. Note that $\ell, r \le 64$. Further, we have

$$w \prec \sigma^6(bxvyd)$$

where $v \in \mathcal{F}_t$, $bx, yd \in \mathcal{F}_{\mathbf{t}}(2)$ and $bxvyd \in \mathcal{F}_{\mathbf{t}}$. Let $\tilde{j} = j \pmod{64}$ and $j' = j + n - 1 \pmod{64}$. Let $a, c \in \{0, 1, 2\}$ with $a \triangleleft v$ and $c \triangleright v$. Then,

$$g_n(j + m) - \mathrm{DS}(w) = R(bxa, cyd, \tilde{j} + 64, j' + 64, m),$$

$$g_n(j - p) - \mathrm{DS}(w) = L(bxa, cyd, \tilde{j} + 64, j' + 64, p).$$

By Lemma 8, one of the following is true:

1. $g_n(j + m) - \mathrm{DS}(w) = 1$ for some $0 < m < 192 - \max\{i, j\}$;
2. $g_n(j - p) - \mathrm{DS}(w) = 1$ for some $0 < p \le \min\{i, j\}$.

Setting $z = t_{j+m}t_{j+m+1} \cdots t_{j+m+n-1}$ or $z = t_{j-p}t_{j-p+1} \cdots t_{j-p+n-1}$, we have the desired. □

Now we prove the intermediate value property of digit sums.

Proof of Proposition 3. For every $n > 128$, starting from $\tau_1(W(n))^R$ and applying Lemma 9 $2\lfloor \log_2 n \rfloor + 1$ times, the result follows. For $1 \le n \le 128$, the result can be verified exhaustively. (This has been done by a computer. We provide the pseudocode for the corresponding procedure in Appendix B.) □

Appendix A: Pseudocode for Lemma 8.

Algorithm 1 For every input u, v, i, j which is present in Lemma 8, the outputs of two following procedures can not be both false.

```
procedure RightShiftTimes
Input: u, v, i, j with 64 ≤ i, j < 128 satisfying u_i = 0 and v_j = 2
Output: m or false
    lword ← σ^6(u)
    rword ← σ^6(v)
    m ← 0
    s ← 0
    while m ≤ 192 − max(i, j) do
        s ← s + rword(j + m) − lword(i + m)
        if s = 1 then
            return m
        m ← m + 1
    return false
procedure LeftShiftTimes
Input: u, v, i, j with 64 ≤ i, j < 128 satisfying u_i = 0 and v_j = 2
Output: −p or false
    lword ← σ^6(u)
    rword ← σ^6(v)
    p ← 1
    s ← 0
    while p ≤ min(i, j) do
        s ← s + lword(i − p) − rword(j − p)
        if s = 1 then
            return −p
        p ← p + 1
    return false
```

Appendix B: Pseudocode for Proposition 3 for $1 \leq n \leq 128$.

Since $\mathbf{t}$ is uniformly recurrent, for every positive integer n, there exits an integer $R(n) > n$ such that for every $u \in \mathcal{F}_{\mathbf{t}}(n)$, we have $u \prec t_0 \cdots t_{R(n)}$. At the mean time, using the analogue of the proof of [6, Proposition 5.1.9], we can show the subword complexity function $\rho_{\mathbf{t}}(n)$: $\rho_{\mathbf{t}}(1) = 3, \rho_{\mathbf{t}}(2) = 9$, and for $n \geq 3$,

$$\begin{cases} \rho_{\mathbf{t}}(2n) = \rho_{\mathbf{t}}(n) + \rho_{\mathbf{t}}(n+1), \\ \rho_{\mathbf{t}}(2n+1) = 2\rho_{\mathbf{t}}(n+1). \end{cases}$$

Hence it is possible to find the index $R(n)$ for every $1 \leq n \leq 128$ with the help of a computer.

Algorithm 2 For every input n, k, the output of the following procedure always be true.

procedure HAVEDESIREDDIGITSUM
Input: n, k with $1 \le n \le 128$ and $n - \lfloor \log_2 n \rfloor - 1 < k < n + \lfloor \log_2 n \rfloor + 1$
Output: true or false
    $i \leftarrow 0$
    while $i \le R(n) - n + 1$ **do**
        $ds \leftarrow 0$
        $j \leftarrow i$
        while $j \le i + n - 1$ **do**
            $ds \leftarrow ds + t_j$
            $j \leftarrow j + 1$
        if $ds = k$ **then**
            return true
        $i \leftarrow i + 1$
    return false

References

References

[1] H. Ardal, T. Brown, V. Jungić and J. Sahasrabudhe, *On abelian and additive complexity in infinite words*, INTEGERS - Elect. J. Combin. Number Theory 12 (2012), #A21.

[2] J. P. Allouche and J. Shallit, Automatic Sequences, Theory, Applications, Generalizations, Cambridge, 2003.

[3] L. Balková, K. Břinda and O. Turek. *Abelian complexity of infinite words associated with quadratic Parry numbers,* Theoret. Comput. Sci. 412 (45) (2011) 6252-6260.

[4] J. Cassaigne, G. Fici, M. Sciortino and L. Q. Zamboni, *Cyclic complexity of words,* in International Symposium on Mathematical Foundations of Computer Science (pp. 159-170). Springer Berlin Heidelberg (2014, August).

[5] E. M. Coven and G. A. Hedlund, *Sequences with minimal block growth*, Math. Systems Theory 7 (1973) 138-153.

[6] N. P. Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics (Lecture Notes in Mathematics, 1794)* . Eds. Berth V., Ferenczi S., Mauduit C. and Siegel A.. Springer, Berlin, 2002

[7] X.-T. Lü, J. Chen, Z.-X. Wen and W. Wu, *On the abelian complexity of the Rudin-Shapiro sequence.* J. Math. Anal. Appl. 451 (2017) 822838.

[8] X.-T. Lü, J. Chen and W. Wu, *On the k-abelian complexity of the Cantor sequence.* J. Combin. Theory Ser. A 155 (2018) 287302.

[9] J. Karhumaki, A. Saarela and L. Q. Zamboni, *On a generalization of Abelian equivalence and complexity of infinite words,* J. Combin. Theory Ser. A 120(8) (2013)2189-2206.

[10] B. Madill and N. Rampersad. *The abelian complexity of the paperfolding word,* Discrete Math. 313 (7) (2013) 831-838.

[11] G. Pirillo and S. Varricchio. *On uniformly repetitive semigroups*, Semigroup Forum 49 (1994) 125129.

[12] M. Rigo and P. Salimov, *Another generalization of abelian equivalence: Binomial complexity of infinite words,* Theoret. Comput. Sci. 601 (2015) 47-57.

[13] G. Richomme, K. Saari and L. Q. Zamboni, *Balance and Abelian complexity of the Tribonacci word,* Adv. in Appl. Math. 45(2) (2010) 212-231.

[14] G. Richomme, K. Saari and L. Zamboni, *Abelian complexity of minimal subshifts*, J. Lond. Math. Soc. 83 (1) (2011) 79-95.

[15] N. J. A. Sloane. *The On-Line Encyclopedia of Integer Sequences*, http://oeis.org. Sequence A071858.