On the abelian complexity of generalized Thue-Morse sequences

Jin Chen$^a$, Zhi-Xiong Wen$^b$

$^a$College of Science, Huazhong Agricultural University, Wuhan 430070, China.
$^b$School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, 430074, China.

Abstract
In this paper, we study the abelian complexity $\rho_n^{ab}(t^{(k)})$ of generalized Thue-Morse sequences $t^{(k)}$. We obtain the exact value of $\rho_n^{ab}(t^{(k)})$ for every integer $n \geq k$. Consequently, $\rho_n^{ab}(t^{(k)})$ is ultimately periodic with the period $k$. Moreover, we show that the abelian complexities of a class of infinite sequences are $k$-automatic.

Keywords: generalized Thue-Morse sequences, abelian complexity, $k$-automatic sequence

1. Introduction

Recently the study of the abelian complexity of infinite words was initiated by G. Richomme, K. Saari, and L. Q. Zamboni [11]. For example, the abelian complexity functions of some notable sequences, such as the Thue-Morse sequence and all Sturmian sequences, were studied in [11] and literature therein.

Let $\sigma_k$ be the morphism $0 \mapsto 01 \cdots (k-1), 1 \mapsto 12 \cdots (k-1)(k-2), \ldots, k-1 \mapsto (k-1)0 \cdots (k-2)$ on $\{0, 1, \ldots, k-1\}$ and $t^{(k)} := \sigma_k^{\infty}(0)$. The infinite sequence $t^{(k)}$ is a generalized Thue-Morse sequence with respect to $k$. Trivially, $t^{(2)}$ is the infamous Thue-Morse sequence. Further, $t^{(k)}$ is $k$-automatic and uniformly recurrent (see [3]). Recall that a sequence $w = w_0w_1w_2 \cdots$ is a $k$-automatic sequence if its $k$-kernel $\{(w_{r+n}c)_{n \geq 0} \mid e \geq 0, 0 \leq c < k^e\}$ finite. If the $\mathbb{Z}$-module generated by its $k$-kernel is finitely generated, then $w$ is a $k$-regular sequence.

Adamczewski [1] obtained the sufficient and necessary condition of bounded abelian complexity. There is a natural question to find some optimal condition for periodic or automatic the abelian complexity. The abelian complexity of $t^{(3)}$ has been studied by Kaboré and Kientéga [8], in which they show that $\rho_n^{ab}(t^{(3)})$ is ultimately periodic with the period 3. In this paper, we are interested in the abelian complexity function $\rho_n^{ab}(t^{(k)})$ for every $n \geq k \geq 2$. The explicit value is obtained in the following theorem.

Theorem 1. For all integer $n \geq k \geq 2$, let $n \equiv r \mod k$, we have

$$\rho_n^{ab}(t^{(k)}) = \begin{cases} \frac{1}{k}(k^2-1) + 1 & \text{if } k \text{ is odd and } r = 0, \\ \frac{1}{k}(k-1)^2 + k & \text{if } k \text{ is odd and } r \neq 0, \\ \frac{1}{k}k^3 + 1 & \text{if } k \text{ is even and } r = 0, \\ \frac{1}{k}(k-1)^2 + \frac{1}{k}k & \text{if } k \text{ is even and } r \neq 0 \text{ is even}, \\ \frac{1}{k}k^2(k-2) + k & \text{if } k \text{ is even and } r \neq 0 \text{ is odd}. \end{cases}$$

Email addresses: cj@mail.hzau.edu.cn (Jin Chen), zhi-xiong.wen@ hust.edu.cn (Zhi-Xiong Wen)

Preprint submitted to Elsevier September 19, 2018
Consequently, we have the following corollary.

**Corollary 1.** \( \{ \rho_n^{ab}(t^{(k)}) \}_{n \geq 1} \) is ultimately periodic with the period \( k \).

In fact, if we need not to compute the exact value of the abelian complexity, we could generalize Corollary 1 in more wide case. Let \( A_1 \) and \( A_2 \) be two alphabets. Given any uniformly recurrent \( k \)-automatic sequence \( w \in A_1^\mathbb{N} \), define a projection \( \pi : A_1^* \to A_2^* \) satisfying that

\[
\partial F_n(w) \text{ will be defined as the boundary word of length } n \text{ for } w \text{ in Section 2.}
\]

**Theorem 2.** Fore every infinite sequence \( w \in A_1^\mathbb{N} \) satisfying that \( \{ \partial F_n(w) \}_{n \geq 2} \) is \( k \)-automatic, let \( u = \pi(w) \in A_2^\mathbb{N} \) defined as prescribed, then the abelian complexity \( \{ \rho_n^{ab}(u) \}_{n \geq 1} \) is a \( k \)-automatic sequence.

This paper is organized as follows. In Section 2, we give some preliminaries and notations. In Section 3, we prove Theorem 1. In Section 4, we prove Theorem 2.

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 \( uv = u_0u_1\cdots u_mv_0v_1\cdots v_n \). The set of all finite words over \( \mathcal{A} \) including the empty word \( \varepsilon \) is denoted by \( A^* \). An infinite word \( w \) is an infinite sequence of letters in \( \mathcal{A} \). The set of all infinite words over \( \mathcal{A} \) is denoted by \( A^\mathbb{N} \). Let \( \Sigma_k \) be \( \{0,1,\ldots,k-1\} \) for every \( k \geq 1 \).

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 \( w \preceq v \); when \( y = \varepsilon \), \( w \) is called a suffix of \( v \), denoted by \( w \succeq v \). For a finite word \( u = u_0\cdots u_{n-1} \), denote by \( u[i,j] \) the factor of \( u \) from the position \( i \) to \( j \), i.e., \( u[i,j] = u_{i-1}\cdots u_{j-1} \). Fixed an positive integer \( k \geq 2 \), let \( u \mod k := (u_0 \mod k)(u_1 \mod k)\cdots(u_{n-1} \mod k) \).

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 \( x \).

2.1. Abelian complexity

Given an alphabet \( \mathcal{A} \). Let \( w = w_0w_1w_2\cdots \in \mathcal{A}^\mathbb{N} \) be an infinite word. Denote by \( F_n(w) \) the set of all factors of \( w \) of length \( n \), i.e.,

\[
F_n(w) := \{ w_iw_{i+1}\cdots w_{i+n-1} : i \geq 0 \}.
\]

In fact, when \( w \) is a finite word, \( F_n(w) \) is still well defined. Write \( F_w := \cup_{n \geq 1} F_n(w) \). The subword complexity function \( \rho_n(w) \) of \( w \) is defined by

\[
\rho_n(w) := |F_n(w)|.
\]
Fixed the alphabet $\mathcal{A} = \{a_0, \ldots, a_{q-1}\}$ and a factor $v$ of an infinite word $w$, the Parikh vector of $v$ is the $q$-uplet

$$
\psi(v) := (|v|_{a_0}, |v|_{a_1}, \ldots, |v|_{a_{q-1}}).
$$

Denote by $\Psi_{u}(n)$, the set of the Parikh vectors of the factors of length $n$ of $u$:

$$
\Psi_n(w) := \{ \psi(v) : v \in F_n(w) \}.
$$

The abelian complexity function of $w$ is defined by the distinct number of the set of Parikh vectors of all the factors:

$$
\rho_{n_{ab}}(w) := \#\Psi_n(w).
$$

In fact, we could give another definition of the abelian complexity function induced by the abelian equivalence relation. For the infinite word $w$, given two factors $u, v$ we say $u$ is abelian equivalent to $v$ (write $u \sim_{ab} v$) if $\psi(u) = \psi(v)$. Now the abelian complexity function of $w$ can be defined as follows:

$$
\rho_{n_{ab}}^{ab}(w) := \#\{ F_n(w) / \sim_{ab} \}.
$$

In addition, before the proof of the main theorems, we give some notations which will be important through all the paper. Let $w \in \mathcal{A}^{\mathbb{N}}$ be an infinite word, given a factor $u = u_0 \cdots u_{n-1}$, denote by $\partial u$ the 2-length (boundary) word consist of the first and last letter of $u$,

$$
\partial u = u_0 u_{n-1}.
$$

For the completeness, let $\partial u$ be itself $u$ when $|u| \leq 1$. Define the boundary letters of the length $n$ of $w$ by

$$
\partial F_n(w) := \{ \partial u : u \in F_n(w) \}.
$$

### 3. Abelian complexity of $t^{(k)}$

First we give some notations and lemmas. For every $n \geq k \geq 2$, set $n = km + r$ for some $m \geq 1$ and $r = 0, \cdots k-1$. Let

$$
S_n(w) := \bigcup_{ab \in \partial F_{m+1}(w)} F_{r+k}(\sigma_k(ab))
$$

and

$$
G_n(w) := \bigcup_{ab \in \partial F_{m+2}(w)} F_r(\sigma_k(ab)[k-r+2, k+r-1])
$$

**Lemma 1.** For every $n \geq k \geq 2$, set $n = km + r$ for some $m \geq 1$ and $r = 0, \cdots k-1$. Then if $r \leq 1$, then we have

$$
\rho_n^{ab}(t^{(k)}) = \#\{ \psi(u) : u \in S_n(t^{(k)}) \}.
$$

Otherwise if $r \geq 2$, then we have

$$
\rho_n^{ab}(t^{(k)}) = \#\{ \psi(u) \cup (I_{k} + \psi(v)) : u \in S_n(t^{(k)}), v \in G_n(t^{(k)}) \}
$$

where $I_k = (1, 1, \ldots, 1)$ is the vector of length $k$ with all the element being the same $1$.  

3
Proof. For every \( n \geq k \geq 2 \), set \( n = km + r \) for some \( m \geq 1 \) and \( r = 0, \cdots k - 1 \). Given any factor \( u \) of length \( n \) of \( t^{(k)} \), there exists a factorization (maybe not unique) such that
\[
u = \alpha \sigma_k(v) \beta
\]
for some factors \( \alpha, v, \beta \) satisfying that \( |v| = m - 1 \), \( |
\alpha| + |\beta| = k + r \). Without loss of generality, we can assume \( \alpha \) or \( \beta \) could be empty word only if \( r = 0 \). In fact, if \( r > 0 \) and \( \alpha = \varepsilon \), then \( |\beta| = k + r \) and let \( \beta = \sigma(a) \beta' \) for some letter \( a \). Set \( v = v_0 \cdots v_{m-2} \), now we could set the new prefix \( \alpha' = \sigma(v_0) \) and \( v' = v_1 \cdots v_{m-2}a \), then
\[
u = \alpha' \sigma_k(v') \beta'.
\]
It is trivially that
\[
\psi(u) = \psi(\alpha) + \psi(\sigma_k(v)) + \psi(\beta) = \psi(\alpha \beta) + (m - 1)\|
\]
(1) If \( r \leq 1 \), the length of preimage of \( \alpha \) and \( \beta \) are both 1 except that the case \( \alpha \) or \( \beta \) is empty word. In other words, \( \alpha \triangleright \sigma_k(a) \), \( \beta \preceq \sigma_k(b) \) for some letters \( a, b \in \{0, \cdots, k - 1\} \). Consequently, \( \alpha \beta \) is a factor of length \( k + r \) of \( \sigma(ab) \) where \( ab \in \partial F_{m+1}(t^{(k)}) \). At the mean time, it is easy to know that all the possible \( \alpha \beta \) is exactly the factor set of length \( k + r \) of \( \sigma(ab) \) for the above \( ab \) even if \( r = 0 \). i.e.,
\[
\rho_{\alpha}^{ab}(t^{(k)}) = \#\{\psi(u) : u \in F_n(t^{(k)})\}
= \#\{\psi(\alpha \beta) : \alpha \sigma_k(v) \beta \in F_n(t^{(k)})\}
= \#\{\psi(\alpha \beta) : \alpha \beta \preceq \sigma_k(ab) \text{ with } ab \in \partial F_{m+1}(t^{(k)}) \text{, } |\alpha \beta| = k + r\}
= \#\{\psi(v) : v \in S_n(t^{(k)})\}
\]
which is the desired result.

(2) If \( r \geq 2 \), the length vector set of preimage of \( \alpha \) and \( \beta \) is \( \{(1, 1), (2, 1), (1, 2)\} \). When the length vector is \( (1, 1) \), we have the same argument with the above case \( r \leq 1 \). It is suffice to consider the case that the length vector is \( (1, 2) \) since \( (2, 1) \) and \( (1, 2) \) is symmetric.

Now assume that \( |\beta| > k \) and \( 1 \leq |\alpha| < r \), and
\[
\alpha \triangleright \sigma_k(a), \beta = \sigma_k(b) \beta' \text{ with } \beta' \preceq \sigma_k(c)
\]
for some letters \( a, b, c \in \{0, \cdots, k - 1\} \). Note that here the letter \( b \) cloud not affect the Parikh vector of \( \alpha \beta \), the letters \( a, c \) is exactly the boundary letter of length \( m + 2 \) of \( t^{(k)} \). At the same time, we have that \( |\alpha| + |\beta'| = r \), the length of \( \beta' \) ranges from 1 to \( r - 1 \). In other words,
\[
\alpha \beta' \preceq \sigma_k(ac)[k - r + 2, k + r - 1].
\]
Moreover, all the possible \( \alpha \beta' \) is the set of factor of lenth \( r \) of \( \sigma_k(ac)[k - r + 2, k + r - 1] \). For the length vector \( (2, 1) \), we can apply the same statement like this. Overall, we have that
\[
\rho_{\alpha}^{ab}(t^{(k)}) = \#\{\psi(u) : u \in F_n(t^{(k)})\}
= \#\{\psi(v) \cup \psi(0u) : v \in S_n(t^{(k)}) \text{, } u \in G_n(t^{(k)})\}
= \#\{\psi(v) \cup (\psi(u) + \psi(u)) : v \in S_n(t^{(k)}) \text{, } u \in G_n(t^{(k)})\}
\]
This complete the proof. \( \square \)

According to the Lemma \[\partial F_m(w)\] \( m \geq 2 \) is the key point to prove Theorem \[\]
Lemma 2. For the generalized Thue-Morse sequence $t^{(k)}$ and every integer $n \geq 2$, we have

$$\partial F_n(t^{(k)}) = \Sigma_k^2.$$ 

Proof. We shall prove by induction by $n$. Obviously this lemma holds for $n = 2, \cdots, k$. Suppose it holds for all $i \leq n$, we will show it holds for $n + 1$.

- If $n + 1 \equiv 0 \pmod{k}$, with loss of generality, set $n + 1 = mk$ ($m \geq 1$), then there exists factors of length $m$ in the form of $aub$ where $ab$ is non-empty word with $|ab| \leq 2$. By the assumption, $ab$ could go through all the words in $\Sigma_k^{[ab]}$. However, we know that $\sigma_k(ab)$ begins with the letter $a$, and ends with the letter $(b - 1) \pmod{k}$. This implies that the lemma holds.

- If $n + 1 \equiv r \pmod{k}$ ($r \geq 1$), set $n + 1 = mk + r$ ($m \geq 1$), then there exists factors of length $m + 1$ in the form of $aub$ where $ab$ is non-empty word with $|ab| \leq 2$, $ab$ could go through all the words in $\Sigma_k^{[ab]}$. We know that for every letter (integer) $i \in \{0, 1, \cdots, k - 1\}$,

$$\sigma_k(i) = i(i + 1) \cdots (i + k - 1) \pmod{k}.$$ 

Now we consider the finite word $v = \sigma_k(ab)[1, n + 1]$, which begins with $a$ and ends with $(b + r - 1) \pmod{k}$. It follows that this lemma holds for $n + 1$ in this case.

For simplicity, let $|V|_i$ be the occurrence time of $i$ in some finite dimensional vector $V$.

Lemma 3. For every $n \geq k \geq 2$, if $n = km$ for some $m \geq 1$, then we have

$$\rho_n^{(a,b)}(t^{(k)}) = 1 + \sum_{t=1}^{\lfloor \frac{t}{t} \rfloor} k(k - 2t + 1) = \begin{cases} 1 + \frac{1}{2}k^3 & \text{if } k \text{ is even}, \\ 1 + \frac{1}{2}k(k^2 - 1) & \text{otherwise}. \end{cases}$$

Proof. By Lemma 1 and Lemma 2 when $r = 0$, it suffice to consider $S_n(t^{(k)})$. In this case, we have

$$S_n(t^{(k)}) = \bigcup_{ab \in \Sigma_k^z} F_k(\sigma_k(ab)) = \bigcup_{ab \in \Sigma_k, 1 \leq t \leq k} \{\sigma_k(ab)[t, t + k - 1]\}.$$ 

For every $u \in S_n(t^{(k)})$, $|u|_i \leq 2$ for every $i \in \Sigma_k$. The only Parikh vector without 2 is $\emptyset$. It suffice to consider the number and the position of 2s in the elements of Parikh vector since the numbers of 2 and 0 are the same. In fact, for every $u \in S_n(t^{(k)})$, write

$$u = \alpha \beta = \alpha_1 \alpha_2 \cdots \alpha_t \beta_1 \beta_2 \cdots \beta_{k-t} = (\alpha_1 \cdots (\alpha_1 + t - 1) \beta_1 \cdots (\beta_1 + k - t - 1) \pmod{k})$$

where $\alpha \triangleright \sigma(a), \beta \triangleright \sigma(b)$ for some letters $a, b$, 

$$|\psi(u)|_0 + |\psi(u)|_1 + |\psi(u)|_2 = k$$

and

$$|\psi(u)|_1 + 2|\psi(u)|_2 = k,$$

which implies $|\psi(u)|_2 = |\psi(u)|_0$. We put the $k$ elements of Parikh vector into a circle, this is present in Figure 1. It is not hard to know that the position of 2 and 0 are both a consecutive block in this circle. The length of 2s’ block ranges from 1 to $\lfloor \frac{k}{2} \rfloor$. For every fixed length $t$ of 2s’ block there are $k$ positions to place the consecutive 2s’ block of length $t$. Next we only need to
place the consecutive block of 0s of length \( t \) in the left \( (k - t) \) positions. Hence the number of Parikh vector with at least one element 2 is

\[
\sum_{t=1}^{\lfloor \frac{k}{2} \rfloor} k(k - 2t + 1).
\]

Adding the Parikh vector \( I_k \), we complete the proof. 

Figure 1: Parikh vector circle over \( \Sigma_k \)

**Lemma 4.** For every \( n \geq k \geq 2 \), if \( n = km + r \) for some \( m \geq 1 \) and \( r = 1, \ldots, k - 1 \), then we have

\[
\rho_n^{ab}(t^{(k)}) = k + \frac{1}{2}(r - 1)(k - r + 1) + \sum_{t=1}^{\lfloor \frac{r}{2} \rfloor} k(r - 2t + 1) + \sum_{t=1+r}^{\lfloor \frac{r+1}{2} \rfloor} k(k + r - 2t + 1)
\]

\[= \begin{cases} 
\frac{1}{4}k(k-1)^2 + k & \text{if } k \text{ is odd,} \\
\frac{1}{4}k(k-1)^2 + \frac{5}{4}k & \text{if } k \text{ is even and } r \text{ is even,} \\
\frac{1}{4}k^2(k-2) + k & \text{if } k \text{ is even and } r \text{ is odd.}
\end{cases}
\]

**Proof.** By Lemma 1 and Lemma 2, there two subcases.

(1) when \( r = 1 \), we only need investigate \( S_n(t^{(k)}) \). We have that

\[S_n(t^{(k)}) = \bigcup_{ab \in \Sigma_k^2} \mathcal{F}_{k+1}(\sigma_k(ab)) = \bigcup_{ab \in \Sigma_k^2, 1 \leq t \leq k} \{\sigma_k(ab)[t, t+k]\}.
\]

For every \( u \in S_n(t^{(k)}) \), \(|u_i| \leq 2 \) for every \( i \in \Sigma_k \). Moreover, there is at least one element being 2 in \( \psi(u) \). After putting the \( k \) elements of Parikh vector into a circle, the position of 2 and 0 are both a consecutive block. If there is only one 2 as the element in \( \psi(u) \), i.e., \(|\psi(u)|_2 = 1 \), then the possible number of Parikh vector is \( k \), since \( \sigma_k(0) \in S_n(t^{(k)}) \) for every \( a \in \Sigma_k \). Now set \(|\psi(u)|_2 = t \), we need to consider the \( t \) ranging from 2 to \([(k+1)/2]\). At the mean time, we have that \(|\psi(u)|_0 = t - 1 \). It is suffice to place the consecutive 0s’ block in the left \( (k - t) \) positions. i.e., the number of Parikh vector having \( \geq 2 \) consecutive 2s’ block is

\[
\sum_{t=2}^{\lfloor \frac{k+1}{2} \rfloor} k(k - 2t + 2).
\]
Adding the \( k \) Parikh vector with exactly one ‘2’ as the element, the lemma holds.

(2) when \( r > 1 \), \( S_n(t^{(k)}) \) and \( G_n(t^{(k)}) \) are both important to the result.

\[
S_n(t^{(k)}) = \bigcup_{ab \in \Sigma_k^2} \mathcal{F}_{k+r}(\sigma_k(ab)) = \bigcup_{ab \in \Sigma_k^2, 1 \leq t \leq k-r+1} \{\sigma_k(ab)[t, t+k+r-1]\}
\]

and

\[
G_n(t^{(k)}) = \bigcup_{ab \in \Sigma_k^2} \mathcal{F}_r(\sigma_k(ab)[k-r+2, k+r-1]).
\]

Following from Lemma 1 for every word \( u \) in \( G_n(t^{(k)}) \), when computing the last abelian complexity, \( \mathbb{I}_k \) should be added to \( \psi(u) \). For simplicity, we consider the following \( G_n'(t^{(k)}) \) instead of \( G_n(t^{(k)}) \).

\[
G_n'(t^{(k)}) = \bigcup_{ab \in \Sigma_k^2} \mathcal{F}_{k+r}(\sigma_k(abc)[k-r+2, k+r-1]).
\]

Note that the letter \( c \) can be arbitrary in \( \Sigma_k \). Given any \( u \in G_n'(t^{(k)}) \), write

\[
u = \alpha \sigma_k(c)\beta = \alpha_1 \alpha_2 \cdots \alpha_i \sigma_k(c)\beta_1 \beta_2 \cdots \beta_{r-t} = (\alpha_1 \cdots (\alpha_1 + t - 1)\sigma_k(c)\beta_1 \cdots (\beta_1 + r - t - 1) \mod k)
\]

where \( \alpha \triangleright \sigma(a), \beta \triangleright \sigma(b) \) for some letters \( a, b \).

(i) First we focus on the case that the maximal element of Parikh vector is 2. For every word \( u \in G_n'(t^{(k)}) \), then the length of consecutive 2's block is strictly less than \( r + 1 \). Obviously the number of Parikh vector with exactly \( r \) length of consecutive 2's block as the element is \( k \). For the word \( u = \alpha \sigma_k(c)\beta \), we have

\[
\{\alpha_i : 1 \leq i \leq t\} \cap \{\beta_i : 1 \leq i \leq r-t\} = \emptyset.
\]

First we put \( \alpha \) into the circle, the number of choice is \( k \). We can assume \( \beta_1 \) and \( \beta_{r-t-1} \) both are not belongs to \( \{\alpha_t + 1, \alpha_t - 1, \alpha_1 + 1, \alpha_1 - 1\} \mod k \), i.e., the gap between \( \alpha \) and \( \beta \) is positive. Otherwise, \( |\psi(u)|_2 = r \), which has been considered. There are \( (k - t - 2) \) positions for \( \beta \). Hence all the possible number of Parikh vector is

\[
\frac{1}{2} \sum_{t=1}^{r-1} k((k-t-2) - (r-t) + 1) = \frac{1}{2}(r-1)(k+r).\]

We could apply the same argument of case (1) to obtain the number of Parikh vector having \( \geq r + 1 \) consecutive 2's block:

\[
\sum_{t=1+r}^{k+r-1} k(k+r-2t+1).
\]

Overall, the total number of Parikh vector in this case is

\[
k + \frac{1}{2}(r-1)(k+r+1) + \sum_{t=1+r}^{k+r-1} k(k+r-2t+1).
\]

(ii) Then the left case is that the maximal element of Parikh vector is 3. This case only happens for the word in \( G_n'(t^{(k)}) \). For the word \( u = \alpha \sigma_k(c)\beta \), we have

\[
\#(\{\alpha_i : 1 \leq i \leq t\} \cap \{\beta_i : 1 \leq i \leq r-t\}) := t > 0.
\]
Moreover, $|\psi(u)|_3 = t$. Clearly, $t$ ranges from 1 to $\left\lfloor \frac{r}{k} \right\rfloor$. At the mean time, the position of 3 and 1 are both a consecutive block. In the same manner, we can obtain the number of Parikh vector in this case is

$$\sum_{t=1}^{\left\lfloor \frac{r}{k} \right\rfloor} k(r - 2t + 1).$$

We know that the subcase (i) and subcase (ii) do not intersect, this implies the Lemma holds. \qed

**Proof of Theorem 7.** It follows from Lemma 3 and Lemma 4 that Theorem 1 holds. \qed

Richomme et al [11] obtain the sufficient and necessary condition for aperiodic sequence $w \in \Sigma_2^N$ satisfying $\rho_n^{ab}(w) = \rho_n(w^{(k)})$. It is a natural to ask the following question.

**Question 1.** For the aperiodic sequence $w \in \Sigma_2^N$ with $k \geq 3$, what is the sufficient and necessary condition for $w$ satisfying $\rho_n^{ab}(w) = \rho_n(w^{(k)})$?

4. Abelian complexities of a class of infinite sequences

Let $w = \{w_n\}_{n \geq 0}$ be an infinite sequence. Define the $k$-kernel of $w$ to be the set of subsequence

$$K_k(w) := \{(w_{kn+c})_{n \geq 0} \mid c \geq 0, 0 \leq c < k^e\}$$

**Definition 1.** Let $k \geq 2$, $w$ is $k$-automatic if and only if $K_k(w)$ is finite.

**Lemma 5.** [2, Theorem 6.8.2] Given an finite alphabet $A$ and $a \geq 0$. Let $\{w_n\}_{n \geq 0}$ be a sequence taking values in $A$ such that $\{w_{an+i}\}_{n \geq 0}$ is $k$-automatic for $0 \leq i < a$. Then $\{w_n\}_{n \geq 0}$ is itself $k$-automatic.

**Lemma 6.** [2, Theorem 6.9.2] Automatic sequences are closed under 1-uniform transducers.

Following from Lemma 6, we have the following corollary which will be useful.

**Corollary 2.** For every $k$-automatic sequence $w = \{w_n\}_{n \geq 0} \in A^N$, $\{(w_n, w_{n+1})\}_{n \geq 0} \in (A \times A)^N$ is also $k$-automatic.

Recall that $u = \pi(w)$ for some $k$-automatic sequence $w \in A_1^N$ with $\pi : A_1^* \rightarrow A_2^*$ satisfying that

for every pair of letters $a, b \in A_1$, $\pi(a) \sim_{ab} \pi(b)$

For every $n \geq k \geq 2$, set $n = km + r$ for some $m \geq 1$ and $r = 0, \cdots, k - 1$. Let

$$S'_n(w) := \bigcup_{ab \in \partial F_{m+1}(w)} F_{r+k}(\pi(ab))$$

and

$$C'_n(w) := \bigcup_{ab \in \partial F_{m+2}(w)} F_r(\pi(ab)[k - r + 2, k + r - 1])$$

We shall give the following lemma without the proof, since the method is the same with Lemma 4 with $t^{(k)}$ replaced by $w$. 

8
Lemma 7. For every $n \geq k \geq 2$, set $n = km + r$ for some $m \geq 1$ and $r = 0, \ldots, k - 1$. Then if $r \leq 1$, then we have
\[ \rho_n(\mathbf{u}) = \# \{ \psi(v) : v \in S_n(\mathbf{w}) \} \].
Otherwise if $r \geq 2$, then we have
\[ \rho_n(\mathbf{u}) = \# \{ \psi(v) \cup (\psi(\pi(a)v')) : v \in S_n(\mathbf{w}), v' \in G_n(\mathbf{w}) \} \]
where $a$ is an arbitrary letter in $A_1$.

Proof of Theorem 2 By Lemma 5 it suffices to show that $\{w_{km+r}\}_{m \geq 0}$ is $k$-automatic for $0 \leq r < k$. If $r \leq 1$, then it follow from Lemma 7 that for every $n = km + r$ with $k \geq 1$ and $r = 0, 1$, we have
\[ \rho_n(\mathbf{u}) = \rho_{km+r}(\mathbf{u}) = \# \left\{ \psi(v) : v \in \bigcup_{ab \in \partial F_m+1(\mathbf{w})} F_{r+k}(\pi(ab)) \right\} \].

This implies that the abelian complexity of length $km + r$ for $u$ is uniquely dependent on the set $\partial F_{m+1}(\mathbf{w})$. Following from Lemma 6 and Corollary 2, $\{\rho_{n}(\mathbf{u})\}_{m \geq 1}$ is $k$-automatic for $r = 0, 1$.

When $r \geq 2$, it follow from Lemma 7 that for every $n = km + r$ with $k \geq 1$ and $r \in [2, k - 1)$, we have
\[ \rho_{km+r}(\mathbf{u}) = \# \left\{ \psi(v) : v \in \bigcup_{ab \in \partial F_m+1(\mathbf{w})} F_{r+k}(\pi(ab)) \right\} \cup \{ (\psi(\pi(a)v')) : v' \in \bigcup_{ab \in \partial F_m+2(\mathbf{w})} F_r(\pi(ab)[k - r + 2, k + r - 1]) \} \].

This implies that the abelian complexity of length $km + r$ for $u$ is uniquely dependent on the set $\partial F_m+1(\mathbf{w})$ and $\partial F_m+2(\mathbf{w})$. By Lemma 6 and Corollary 2, $\{\rho_{km+r}(\mathbf{u})\}_{m \geq 1}$ is $k$-automatic for $r \geq 2$. This completes the proof since the leading $k - 1$ terms can not affect the $k$-automatic property.

Following from Lemma 2 $\partial F_{n}(\mathbf{t}(k))$ is periodic, which is a special $\ell$-automatic sequence for any $\ell \geq 2$. This implies that Corollary 1 holds by Theorem 2. We know that the key point for Theorem 2 is the boundary sequence $\partial F_{n}(\mathbf{w})$, it is significant to consider the infinite sequence $\mathbf{w}$ whose boundary sequence is $k$-automatic for some $k$. We give some simple $k$-automatic sequences, since it has been checked for some $k$-automatic sequences such as the above $\mathbf{t}(k)$, the Cantor sequence $\mathbf{c}$ which is the fixed point of the morphism $0 \rightarrow 000, 1 \rightarrow 101$.

Conjecture 1. For every $k$-automatic sequence $\mathbf{w} = \{w_n\}_{n \geq 0} \in A^\mathbb{N}$, $\{\partial F_n(\mathbf{w})\}_{n \geq 2} \in (2^A \times A)^\mathbb{N}$ is also $k$-automatic.

References

References

[1] B. Adamczewski, Balances for fixed points of primitive substitutions, Theoret. Comput. Sci. 307(1) (2003) 47-75.

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

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

[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] I. Kaboré and B. Kientéga *Abelian Complexity of Thue-Morse Word over a Ternary Alphabet*. InInternational Conference on Combinatorics on Words (pp. 132-143). Springer, Cham (2017 Sep).

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

[10] 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.

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