Ergod. Th. & Dynam. Sys. (2015), **35**, 142–151 doi:10.1017/etds.2013.51
© *Cambridge University Press, 2013*

Abelian maximal pattern complexity of words

TETURO KAMAE†, STEVEN WIDMER‡ and LUCA Q. ZAMBONI§¶

† *Advanced Mathematical Institute, Osaka City University, Osaka, 558-8585, Japan*
(e-mail: kamae@apost.plala.or.jp)
‡ *Department of Mathematics, General Academics Building 435,*
1155 Union Circle #311430, Denton, TX 76203-5017, USA
(e-mail: s.widmer1@gmail.com)
§ *Institut Camille Jordan, Université Claude Bernard Lyon 1,*
43 boulevard du 11 novembre 1918, F69622 Villeurbanne Cedex, France
(e-mail: lupastis@gmail.com)
¶ *Department of Mathematics and Turku Centre for Computer Science,*
University of Turku, 20014 Turku, Finland

(*Received* 24 *January* 2013 *and accepted in revised form* 3 *May* 2013)

Abstract. In this paper, we study the maximal pattern complexity of infinite words up to Abelian equivalence. We compute a lower bound for the Abelian maximal pattern complexity of infinite words which are both recurrent and aperiodic by projection. We show that in the case of binary words, the bound is actually achieved and gives a characterization of recurrent aperiodic words.

1. *Introduction*

Let $\mathbb{A}$ be a finite non-empty set. We denote by $\mathbb{A}^*$, $\mathbb{A}^{\mathbb{N}}$ and $\mathbb{A}^{\mathbb{Z}}$, respectively, the set of finite words, the set of (right) infinite words, and the set of bi-infinite words over the alphabet $\mathbb{A}$. Given an infinite word $\alpha = \alpha_0\alpha_1\alpha_2 \ldots \in \mathbb{A}^{\mathbb{N}}$ with $\alpha_i \in \mathbb{A}$, we denote by $\mathcal{F}_\alpha(n)$ the set of all *factors* of α of length n, that is, the set of all finite words of the form $\alpha_i\alpha_{i+1} \ldots \alpha_{i+n-1}$ with $i \geq 0$. We set

$$p_\alpha(n) = \#\left(\mathcal{F}_\alpha(n)\right).$$

The function $p_\alpha : \mathbb{N} \to \mathbb{N}$ is called the *factor complexity function* of α.

We recall that two words u and v in $\mathbb{A}^*$ are said to be *Abelian equivalent,* denoted by $u \sim_{\mathrm{ab}} v$, if and only if $|u|_a = |v|_a$ for all $a \in \mathbb{A}$, where $|u|_a$ denotes the number of occurrences of the letter a in u. It is readily verified that $\sim_{\mathrm{ab}}$ defines an equivalence relation on $\mathbb{A}^*$. We define

$$\mathcal{F}_\alpha^{\mathrm{ab}}(n) = \mathcal{F}_\alpha(n)/\sim_{\mathrm{ab}}$$

and set

$$p_\alpha^{\mathrm{ab}}(n) = \#\left(\mathcal{F}_\alpha^{\mathrm{ab}}(n)\right).$$

https://doi.org/10.1017/etds.2013.51 Published online by Cambridge University Press

The function $p_\alpha^{\mathrm{ab}} : \mathbb{N} \to \mathbb{N}$, which counts the number of pairwise non-Abelian equivalent factors of α of length n, is called the *Abelian complexity* of α (see [**7**]).

There are a number of similarities between the usual factor complexity of an infinite word and its Abelian counterpart. For instance, both may be used to characterize periodic bi-infinite words (see [**1, 6**]). A word α is *periodic* if there exists a positive integer p such that $\alpha_{i+p} = \alpha_i$ for all indices i, and it is *ultimately periodic* if $\alpha_{i+p} = \alpha_i$ for all sufficiently large i. An infinite word is *aperiodic* if it is not ultimately periodic. The factor complexity function also provides a characterization of ultimately periodic words. On the other hand, Abelian complexity does not yield such a characterization. Indeed, both Sturmian words and the ultimately periodic word $01^\infty = 0111\cdots$ have the same, constant 2, Abelian complexity.

As another example, both complexity functions give a characterization of *Sturmian* words amongst all aperiodic words.

THEOREM 1. *Let α be an aperiodic infinite word over the alphabet $\{0, 1\}$. The following conditions are equivalent.*

- *The word α is balanced, that is, Sturmian.*
- *(Morse and Hedlund [**6**]). The word α satisfies $p_\alpha(n_0) = n + 1$ for all $n \geq 0$.*
- *(Coven and Hedlund [**1**]). The word α satisfies $p_\alpha^{\mathrm{ab}}(n) = 2$ for all $n \geq 1$.*

In [**3**], the first and third authors introduced a different notion of the complexity of an infinite word, called the maximal pattern complexity.

For each positive integer k, let $\Sigma_k(\mathbb{N})$ denote the set of all k-element subsets of $\mathbb{N}$. An element $S = \{s_1 < s_2 < \cdots < s_k\} \in \Sigma_k(\mathbb{N})$ will be called a *k-pattern*. We put

$$\alpha[S] := \alpha(s_1)\alpha(s_2)\cdots\alpha(s_k) \in \mathbb{A}^k.$$

For each $n \in \mathbb{N}$, the word $\alpha[n + S]$ is called an *S-factor* of α, where $n + S := \{n + s_1, n + s_2, \ldots, n + s_k\}$. We denote by $\mathcal{F}_\alpha(S)$ the set of all S-factors of α. We define the *pattern complexity* $p_\alpha(S)$ by

$$p_\alpha(S) = \#\,\mathcal{F}_\alpha(S)$$

and the *maximal pattern complexity* $p_\alpha^*(k)$ by

$$p_\alpha^*(k) = \sup_{S \in \Sigma_k(\mathbb{N})} p_\alpha(S).$$

In [**3**], the authors show that maximal pattern complexity also gives a characterization of ultimately periodic words.

THEOREM 2. *Let $\alpha \in \mathbb{A}^{\mathbb{N}}$. Then the following are equivalent.*

(1) *α is eventually periodic.*
(2) *$p_\alpha^*(k)$ is uniformly bounded in k.*
(3) *$p_\alpha^*(k) < 2k$ for some positive integer k.*

In other words, α is aperiodic if and only if $p_\alpha^*(k) \geq 2k$ for each positive integer k. We say $\alpha \in \mathbb{A}^{\mathbb{N}}$ is *pattern Sturmian* if $p_\alpha^*(k) = 2k$ for each positive integer k. Two types of recurrent pattern Sturmian words are known: rotation words (see below) and a family of

https://doi.org/10.1017/etds.2013.51 Published online by Cambridge University Press

'simple' Toeplitz words (see [**3**]). Unfortunately, to date, there is no known classification of recurrent pattern Sturmian words (as in the case of Theorem 1).

The connection between items (1) and (3) in Theorem 2 was generalized by the first author and Rao in [**4**]. We say $\alpha \in \mathbb{A}^{\mathbb{N}}$ is *periodic by projection* if there exists a set $\emptyset \neq B \subsetneqq \mathbb{A}$ such that

$$\mathbf{1}_B(\alpha) := \mathbf{1}_B(\alpha(0))\mathbf{1}_B(\alpha(1))\mathbf{1}_B(\alpha(2)) \cdots \in \{0, 1\}^{\mathbb{N}}$$

is eventually periodic (where $\mathbf{1}_B$ denotes the characteristic function of B). We say α is *aperiodic by projection* if α is not periodic by projection. Then we get the following theorem.

THEOREM 3. *Let* $\#\mathbb{A} = r \geq 2$, *and* $\alpha \in \mathbb{A}^{\mathbb{N}}$ *be aperiodic by projection. Then* $p^*_\alpha(k) \geq rk$ *for each positive integer* k.

In other words, low pattern complexity (relative to the size of the alphabet) implies periodicity by projection. Notice that if $\#\mathbb{A} = 2$, then α is periodic by projection if and only if α is eventually periodic.

In this paper, we introduce and study an Abelian analog of maximal pattern complexity: given a k-pattern $S \in \Sigma_k(\mathbb{N})$, we define

$$\mathcal{F}^{\mathrm{ab}}_\alpha(S) = \mathcal{F}_\alpha(S)/\sim_{\mathrm{ab}}$$

and the associated *Abelian pattern complexity*

$$p^{\mathrm{ab}}_\alpha(S) = \#\,\mathcal{F}^{\mathrm{ab}}_\alpha(S),$$

which counts the number of pairwise non-Abelian equivalent S-factors of α. We define the *Abelian maximal pattern complexity*

$$p^{*\mathrm{ab}}_\alpha(k) = \sup_{S \in \Sigma_k(\mathbb{N})} p^{\mathrm{ab}}_\alpha(S).$$

It is clear that for each positive integer k, and for each pattern $S \in \Sigma_k(\mathbb{N})$, we have

$$p^{\mathrm{ab}}_\alpha(S) \leq p_\alpha(S) \quad \text{and} \quad p^{*\mathrm{ab}}_\alpha(k) \leq p^*_\alpha(k).$$

In this paper, we show the following theorem.

THEOREM 4. *Let* $\#\mathbb{A} = r \geq 2$ *and* $\alpha \in \mathbb{A}^{\mathbb{N}}$ *be recurrent and aperiodic by projection. Then, for each positive integer* k, *we have*

$$p^{*\mathrm{ab}}_\alpha(k) \geq (r-1)k + 1.$$

In the case $r = 2$, *equality always holds. Moreover, for* $k = 2$ *and general* r, *there exists* α *satisfying the equality.*

For example, if $\alpha \in \{0, 1\}^{\mathbb{N}}$ is a Sturmian word and $S \in \Sigma_k(\mathbb{N})$ is a k-block pattern, that is, $S = \{0, 1, 2, \ldots, k-1\}$, then we have $p^{\mathrm{ab}}_\alpha(S) = 2$ (since α is balanced), while $p_\alpha(S) = k + 1$. Since α is both recurrent and aperiodic, it follows from the above theorem that the Abelian maximal pattern complexity $p^{*\mathrm{ab}}_\alpha(k)$ takes the maximum value $k + 1$ for each positive integer k. Moreover, all recurrent pattern Sturmian words share this property.

https://doi.org/10.1017/etds.2013.51 Published online by Cambridge University Press

For a rotation word $\alpha \in \mathbb{A}^{\mathbb{N}}$ with $r = \#\mathbb{A} \geq 3$, we show that $p_\alpha^{*\mathrm{ab}}(k) = rk$ for each positive integer k (see Theorem 6). Since $p_\alpha^*(k) = rk$, the abelianization does not decrease the complexity in this case. On the other hand, in the proof of Theorem 4, we show that $p_\alpha^{*\mathrm{ab}}(2) = 2r - 1$ for any Toeplitz word $\alpha \in \mathbb{A}^{\mathbb{N}}$ with $\#\mathbb{A} = 2$.

We define two classes of words, with $\mathbb{A} = \{0, 1, \ldots, r-1\}$ and $r \geq 2$. Let θ be an irrational number and $c_0 < c_1 < \cdots < c_{r-1} < c_r$ be real numbers such that $c_r = c_0 + 1$. Define $\alpha \in \mathbb{A}^k$ by $\alpha(n) = i$ if $n\theta \in [c_i, c_{i+1}) \pmod 1$ for any $i \in \mathbb{A}$ and $n \in \mathbb{N}$. We call such α a *rotation word*. Let $\mathbb{Z}_2$ be the 2-adic compactification of $\mathbb{Z}$ and $\gamma \in \mathbb{Z}_2$. For $n \in \mathbb{Z}_2$, let $\tau(n) \in \mathbb{N} \cup \{\infty\}$ be the supremum of $k \in \mathbb{N}$ such that 2^k divides n. Let B_i $(i \in \mathbb{A})$ be infinite subsets of $\mathbb{N} \cup \{\infty\}$ such that $B_i \cap B_j = \emptyset$ for any $i, j \in \mathbb{A}$ with $i \neq j$ and $\bigcup_{i \in \mathbb{A}} B_i = \mathbb{N} \cup \{\infty\}$. Define $\alpha \in \mathbb{A}^{\mathbb{N}}$ by $\alpha(n) = i$ if $\tau(n - \gamma) \in B_i$ for any $i \in \mathbb{A}$ and $n \in \mathbb{N}$. We call such α a *Toeplitz word*.

We do not know whether the inequality in Theorem 4 is tight when $r \geq 3$ and $k \geq 3$.

2. *Background and notation*

Given a finite non-empty set $\mathbb{A}$, we endow $\mathbb{A}^{\mathbb{N}}$ with the topology generated by the metric

$$d(x, y) = \frac{1}{2^n} \quad \text{where } n = \inf\{k : x_k \neq y_k\},$$

whenever $x = (x_n)_{n \in \mathbb{N}}$ and $y = (y_n)_{n \in \mathbb{N}}$ are two elements of $\mathbb{A}^{\mathbb{N}}$. For $\omega \in \mathbb{A}^{\mathbb{N}}$, let $\overline{O}(\omega)$ denote the closure of the orbit $O(\omega) := \{T^n\omega : n \in \mathbb{N}\}$ of ω with respect to the shift T on $\mathbb{A}^{\mathbb{N}}$, where $(T\omega)(n) = \omega(n+1)$ $(n \in \mathbb{N})$.

Given a finite word $u = a_1 a_2 \ldots a_n$ with $n \geq 1$ and $a_i \in A$, we denote the length n of u by $|u|$. For each $a \in A$, we let $|u|_a$ denote the number of occurrences of the letter a in u.

For each $u \in A^*$, we denote by $\Psi(u)$ the *Parikh vector* or *abelianization* of u, that is, the vector indexed by $\mathbb{A}$

$$\Psi(u) = (|u|_a)_{a \in \mathbb{A}}.$$

Given $\Xi \subset \mathbb{A}^*$, we set

$$\Xi^{\mathrm{ab}} := \Xi / \sim_{\mathrm{ab}}$$

and

$$\Psi(\Xi) := \{\Psi(\xi) \mid \xi \in \Xi\}.$$

There is an obvious bijection between the sets Ξ^{ab} and $\Psi(\Xi)$, where one identifies the Abelian class of an element $u \in \mathbb{A}^*$ with its Parikh vector $\Psi(u)$.

Given a non-empty set $\Omega \subset \mathbb{A}^{\mathbb{N}}$, $S \in \Sigma_k(\mathbb{N})$ and an infinite set $\mathcal{N} \subset \mathbb{N}$, we put

$$\Omega[S] := \{\omega[S] \mid \omega \in \Omega\} \subset \mathbb{A}^k$$

and

$$\Omega[\mathcal{N}] := \{\omega[\mathcal{N}] \mid \omega \in \Omega\} \subset \mathbb{A}^{\mathbb{N}},$$

where $\omega[\mathcal{N}] \in \mathbb{A}^{\mathbb{N}}$ is defined by $\omega[\mathcal{N}](n) = \omega(N_n)$ $(n \in \mathbb{N})$.

Analogously, we can define the maximal pattern complexity of Ω by

$$p_\Omega^*(k) = \sup_{S \in \Sigma_k(\mathbb{N})} p_\Omega(S),$$

https://doi.org/10.1017/etds.2013.51 Published online by Cambridge University Press

where

$$p_\Omega(S) = \#\,\Omega[S],$$

and the Abelian maximal pattern complexity of Ω by

$$p_\Omega^{*\mathrm{ab}}(k) = \sup_{S\in\Sigma_k(\mathbb{N})} p_\Omega^{\mathrm{ab}}(S),$$

where

$$p_\Omega^{\mathrm{ab}}(S) = \#\,\Omega[S]^{\mathrm{ab}}.$$

3. *Superstationary sets and Ramsey's infinitary theorem*

LEMMA 3.1. *Let $\omega \in \mathbb{A}^{\mathbb{N}}$ be a recurrent infinite word. Then there exists an infinite set $\mathcal{N} = \{N_0 < N_1 < N_2 < \cdots\} \subset \mathbb{N}$ satisfying the following condition:*

$$\textit{for all } i \geq 0 \textit{ for all } k \geq 0, \quad \omega_{i+N_0}\omega_{i+N_1}\cdots\omega_{i+N_{k-1}}\omega_{i+N_k}^{\infty} \in \overline{O}(\omega)[\mathcal{N}]. \qquad (*)$$

Proof. We show by induction on k that for each $k \geq 0$, there exist natural numbers $N_0 < N_1 < \cdots < N_k$ such that for each $j \leq k$, if $u_0u_1\cdots u_j \in \overline{O}(\omega)[\{N_0, N_1, \ldots, N_j\}]$, then $u_0u_1\cdots u_j^{k-j+1} \in \overline{O}(\omega)[\{N_0, N_1, \ldots, N_k\}]$. Clearly, we can take for N_0 any natural number in $\mathbb{N}$. Next, suppose we have chosen natural numbers $N_0 < N_1 < \cdots < N_k$ with the required property. Fix a positive integer L such that if $u \in \overline{O}(\omega)[\{N_0, N_1, \ldots, N_k\}]$, then there exists $i \leq L - N_k$ with $u = \omega_{i+N_0}\omega_{i+N_1}\cdots\omega_{i+N_k}$. Since ω is recurrent, there exists a positive integer $N_{k+1} > N_k$ such that $\omega_i = \omega_{i+N_{k+1}}$ for each $i \leq L$. We now verify that $N_0 < N_1 < \cdots < N_{k+1}$ satisfies the required property. So, assume $j \leq k+1$ and $u_0u_1\cdots u_j \in \overline{O}(\omega)[\{N_0, N_1, \ldots, N_j\}]$. We must show that $u_0u_1\cdots u_j^{k+1-j+1} \in \overline{O}(\omega)[\{N_0, N_1, \ldots, N_{k+1}\}]$. This is clear in the case $j = k+1$, thus we can assume $j \leq k$. Then, by an induction hypothesis, we have that $u = u_0u_1\cdots u_j^{k-j+1} \in \overline{O}(\omega)[\{N_0, N_1, \ldots, N_k\}]$. Fix $i \leq L - N_k$ such that $u = \omega_{i+N_0}\omega_{i+N_1}\cdots\omega_{i+N_k}$. Then

$$\begin{aligned} u_0u_1\cdots u_j^{k+1-j+1} &= u_0u_1\cdots u_j^{k-j+1}u_j \\ &= \omega_{i+N_0}\omega_{i+N_1}\cdots\omega_{i+N_k}\omega_{i+N_k} \\ &= \omega_{i+N_0}\omega_{i+N_1}\cdots\omega_{i+N_k}\omega_{i+N_{k+1}}. \end{aligned}$$

Hence, $u_0u_1\cdots u_j^{k+1-j+1} \in \overline{O}(\omega)[\{N_0, N_1, \ldots, N_{k+1}\}]$, as required. □

The following lemma is readily verified.

LEMMA 3.2. *Let $\mathcal{N} = \{N_0 < N_1 < N_2 < \cdots\} \subset \mathbb{N}$ be an infinite set satisfying the condition (*) above, and let $\mathcal{N}'$ be any infinite subset of $\mathcal{N}$. Then $\mathcal{N}'$ also satisfies (*).*

PROPOSITION 3.3. *Let $\Omega \subset \mathbb{A}^{\mathbb{N}}$ be non-empty and let $\mathcal{N} = \{N_0 < N_1 < N_2 < \cdots\} \subset \mathbb{N}$ be an infinite set. Then, for every positive integer k, there exists an infinite subset $\mathcal{N}'$ of $\mathcal{N}$ (depending on k) such that for any two finite subsets P and Q of $\mathcal{N}'$ with $1 \leq |P| = |Q| \leq k$, we have*

$$\Omega[P] = \Omega[Q].$$

https://doi.org/10.1017/etds.2013.51 Published online by Cambridge University Press

Proof. We will recursively construct a sequence of nested infinite patterns

$$\mathcal{N}' = \mathcal{N}_k \subset \cdots \subset \mathcal{N}_2 \subset \mathcal{N}_1 = \mathcal{N}$$

such that for each $1 \leq i \leq k$, we have

$$\Omega[P] = \Omega[Q]$$

for all finite subsets P and Q of $\mathcal{N}_i$ with $1 \leq |P| = |Q| \leq i$.

We begin with $\mathcal{N}_2$. Given two finite subpatterns P and Q of $\mathcal{N}_1$ with $|P| = |Q| = 2$, we write

$$P \sim_2 Q \iff \Omega[P] = \Omega[Q].$$

Then $\sim_2$ defines an equivalence relation on the set of all subpatterns of $\mathcal{N}_1$ of size two, and hence naturally defines a finite coloring on the set of all size-two subpatterns of $\mathcal{N}_1$, or, equivalently, on the set of all two-element subsets of the natural numbers $\mathbb{N}$, where two patterns P and Q are monochromatic if and only if $P \sim_2 Q$. We now recall the following well-known theorem of Ramsey.

THEOREM 5. ([**2**], Ramsey) *Let k be a positive integer. Then, given any finite coloring of the set of all k-element subsets of $\mathbb{N}$, there exists an infinite set $\mathcal{A} \subset \mathbb{N}$ such that any two k-element subsets of $\mathcal{A}$ are monochromatic.*

Thus, applying the above theorem, we deduce that there exists an infinite pattern $\mathcal{N}_2 \subset \mathcal{N}_1$ such that any two subpatterns P and Q of $\mathcal{N}_2$ of size 2 are $\sim_2$ equivalent.

Having constructed $\mathcal{N}_k \subset \mathcal{N}_{k-1} \subset \cdots \subset \mathcal{N}_2 \subset \mathcal{N}_1 = \mathcal{N}$ with the required properties, we next construct $\mathcal{N}_{k+1}$ as follows; given any two subpatterns P and Q of $\mathcal{N}_k$ of size $k+1$, we write

$$P \sim_{k+1} Q \iff \Omega[P] = \Omega[Q].$$

Again, this defines a finite coloring of the set of all size-$(k+1)$ subpatterns of $\mathcal{N}_k$, or, equivalently, on the set of all $(k+1)$-element subsets of $\mathbb{N}$. Hence, by Ramsey's theorem, we deduce that there exists an infinite pattern $\mathcal{N}_{k+1} \subset \mathcal{N}_k$ such that any two subpatterns of $\mathcal{N}_{k+1}$ of size $k+1$ are monochromatic, that is, $\sim_{k+1}$ equivalent. Moreover, since $\mathcal{N}_{k+1} \subset \mathcal{N}_k$, it follows that any two subpatterns P and Q of $\mathcal{N}_{k+1}$ of size $1 \leq |P| = |Q| \leq k$ are $\sim_{|P|}$ equivalent. □

Definition 3.4. Let $k \geq 2$. A non-empty set $\Omega \subset \mathbb{A}^{\mathbb{N}}$ is called a *k-superstationary* set if

$$\Omega[S] = \Omega[S']$$

for any S and $S' \in \Sigma_k(\mathbb{N})$ (see [**5**]).

As an immediate consequence of Proposition 3.3, we have the following corollary.

COROLLARY 3.5. *Let $\Omega \subset \mathbb{A}^{\mathbb{N}}$ be non-empty and let $\mathcal{N} \subset \mathbb{N}$ be an infinite set. Then, for every positive integer k, there exists an infinite subset $\mathcal{N}'$ such that $\Omega[\mathcal{N}']$ is k-superstationary.*

https://doi.org/10.1017/etds.2013.51 Published online by Cambridge University Press

LEMMA 3.6. *Let $\alpha \in \mathbb{A}^{\mathbb{N}}$ be aperiodic by projection and let $\mathcal{N} \subset \mathbb{N}$ be any infinite set. Put $\Omega := \overline{O}(\alpha)[\mathcal{N}]$. Then, for any $\{i < j\} \subset \mathbb{N}$, the directed graph $(\mathbb{A}, E_{i,j})$ is strongly connected, where*

$$E_{i,j} = \{(\omega(i), \omega(j)) \in \mathbb{A} \times \mathbb{A};\ \omega \in \Omega,\ \omega(i) \neq \omega(j)\}.$$

Proof. Fix $\{i < j\} \subset \mathbb{N}$ and $\mathcal{N} = \{N_0 < N_1 < \cdots\}$. For any $l = 0, 1, \ldots, N_j - N_i - 1$, let A_l be the set of $a \in \mathbb{A}$ such that $\alpha(n) = a$ holds for infinitely many $n \in \mathbb{N}$ with $n \equiv l \pmod{N_j - N_i}$. For any $a, b \in \mathbb{A}$, if $\{a, b\} \in A_l$ for some $l \in \{0, 1, \ldots, N_j - N_i - 1\}$, then a, b are two-way connected in the graph $(\mathbb{A}, E_{i,j})$. Hence, for $a, b \in \mathbb{A}$, a, b are two-way connected in the graph $(\mathbb{A}, E_{i,j})$ if there exist $a_0, a_1, \ldots, a_k \in \mathbb{A}$ and $l_1, \ldots, l_k \in \{0, 1, \ldots, N_j - N_i - 1\}$ such that: (i) $a_0 = a$, $a_k = b$; and (ii) $\{a_{i-1}, a_i\} \in A_{l_i}$ for any $i = 1, \ldots, k$.

Suppose to the contrary that there exist $a, b \in \mathbb{A}$ such that a and b are not two-way connected in the graph $(\mathbb{A}, E_{i,j})$. Let A be the set of $a' \in \mathbb{A}$ such that a, a' are two-way connected in the graph $(\mathbb{A}, E_{i,j})$. Then, we have $\emptyset \neq A \subsetneqq \mathbb{A}$. Moreover, there exists S with $\emptyset \neq S \subset \{0, 1, \ldots, N_j - N_i - 1\}$ such that $A_l \subset A$ for any $l \in S$ and $A_l \cap A = \emptyset$ for any $l \in \{0, 1, \ldots, N_j - N_i - 1\} \backslash S$. Therefore,

$$1_A(\alpha(0))1_A(\alpha(1))1_A(\alpha(2)) \cdots$$

is periodic with period $N_j - N_i$, which contradicts our assumption that α is aperiodic by projection. Thus, the graph is strongly connected. □

Combining Lemmas 3.1, 3.2 and 3.6 with Proposition 3.3, we obtain the following proposition.

PROPOSITION 3.7. *Let $\omega \in \mathbb{A}^{\mathbb{N}}$ be recurrent and aperiodic by projection and $k \geq 2$. Then there exists an infinite set $\mathcal{N} \subset \mathbb{N}$ such that $\Omega := \overline{O}(\alpha)[\mathcal{N}]$ is a k-superstationary set and:*

(1) *for any $\omega \in \Omega$ and $i \in \mathbb{N}$,*

$$\omega(0)\omega(1) \cdots \omega(i-1)\omega(i)^{\infty} \in \Omega;$$

(2) *for any $\{i < j\} \subset \mathbb{N}$, the directed graph $(\mathbb{A}, E_{i,j})$ is strongly connected.*

4. *Main results*

Proof of Theorem 4. Fix a positive integer k. By Proposition 3.7, there exists an infinite set $\mathcal{N} \subset \mathbb{N}$ such that $\Omega = \overline{O}(\alpha)[\mathcal{N}] \subset \mathbb{A}^{\mathbb{N}}$ is $(k+1)$-superstationary and satisfies conditions (1) and (2) of Proposition 3.7. Since $p_{\alpha}^{*\mathrm{ab}}(k) \geq p_{\Omega}^{*\mathrm{ab}}(k)$, it is sufficient to prove that $\#\,{\Omega_k}^{\mathrm{ab}} \geq (r-1)k + 1$, where $\Omega_k := \Omega[\{0, 1, \ldots, k-1\}]$.

Let $(\mathbb{A}, E_{0,1})$ be the strongly directed graph where

$$E_{0,1} = \{(\omega(0), \omega(1)) \in \mathbb{A} \times \mathbb{A} \text{ with } \omega(0) \neq \omega(1);\ \omega \in \Omega\}.$$

Then there exists a sequence $a_0 a_1 \cdots a_l$ of elements in $\mathbb{A}$ containing all elements in $\mathbb{A}$ such that $(a_i, a_{i+1}) \in E_{0,1}$ $(i = 0, 1, \ldots, l-1)$.

Define a non-directed graph $(\mathbb{A}, F)$ by

$$F = \{\{a, b\} \subset \mathbb{A} \text{ with } a \neq b \text{ and either } a^k b^{\infty} \in \Omega \text{ or } b^k a^{\infty} \in \Omega\}.$$

https://doi.org/10.1017/etds.2013.51 Published online by Cambridge University Press

Since Ω is $(k+1)$-superstationary, for any $i = 0, 1, \ldots, l-1$, there exists $\omega \in \Omega$ such that $\omega[\{kr, kr+1\}] = a_i a_{i+1}$. Hence, by (1) of Proposition 3.7, there exists $\xi \in \mathbb{A}^{kr}$ such that $\xi a_i a_{i+1}^{\infty} \in \Omega$ and $\xi a_i^{\infty} \in \Omega$. Then there exists $b \in \mathbb{A}$ occurring in ξ at least k times. Since Ω is $(k+1)$-superstationary, this implies that $b^k a_i$ and $b^k a_{i+1}$ are in $\Omega[\{0, 1, \ldots, k\}]$. Therefore, $b^k a_i^{\infty} \in \Omega$ and $b^k a_{i+1}^{\infty} \in \Omega$, by (1) of Proposition 3.7. Hence, we have two cases according to whether $b \in \{a_i, a_{i+1}\}$ or not.

Case 1. $b \in \{a_i, a_{i+1}\}$. In this case, we have $\{a_i, a_{i+1}\} \in F$.

Case 2. $b \notin \{a_i, a_{i+1}\}$. In this case, we have two edges $\{b, a_i\}$ and $\{b, a_{i+1}\}$ in F, by which a_i and a_{i+1} are connected.

Thus, we have a connected graph $(\mathbb{A}, F)$. This implies there are at least $r-1$ edges. If $\{a, b\} \in F$, then either $a^k b^{\infty} \in \Omega$ or $b^k a^{\infty} \in \Omega$. Since Ω is k-superstationary, either $a^h b^{k-h} \in \Omega_k$ $(h = 0, 1, \ldots, k)$ or $b^h a^{k-h} \in \Omega_k$ $(h = 0, 1, \ldots, k)$. In any case, there are $k+1$ elements in ${\Omega_k}^{\mathrm{ab}}$ consisting only of a and b.

Since $\#F \geq r-1$, there are at least $(r-1)(k+1) - (r-2) = (r-1)k + 1$ elements in ${\Omega_k}^{\mathrm{ab}}$ consisting only of two elements, where we subtract $r-2$ since the number of overlaps counted for constant words is $2(r-1) - r = r-2$.

Thus, $\#\,\Omega_k^{\mathrm{ab}} \geq (r-1)k + 1$.

If $\#\mathbb{A} = 2$, then $p_{\alpha}^{*\mathrm{ab}}(k) \leq k+1$ $(k = 1, 2, \ldots)$ for any $\alpha \in \mathbb{A}^{\mathbb{N}}$, since the number of vectors $(|\xi|_0, |\xi|_1)$ over all $\xi \in \{0, 1\}^k$ is $k+1$.

Let $\mathbb{A} = \{0, 1, \ldots, r-1\}$ with $r \geq 3$. For $n \geq 1$, let $\tau(n)$ be the maximum $\tau \in \mathbb{N}$ such that 2^{τ} is a factor of n. Define $\alpha \in \mathbb{A}^{\mathbb{N}}$ by $\alpha(n) = \tau(n+1) \pmod r$. Then α is one of the Toeplitz words defined in the Introduction. It is clearly recurrent and aperiodic by projection.

Take any 2-pattern $S = \{s < t\} \subset \mathbb{N}$. Let $d = \tau(t-s)$. Then there exists $u \in \mathbb{N}$ with $0 \leq u < 2^d$ such that either $\tau(s-u) = d$, $\tau(t-u) > d$ or $\tau(s-u) > d$, $\tau(t-u) = d$. Assume, without loss of generality, that the latter holds. Let $c \in \mathbb{A}$ be such that $c \equiv d \pmod r$ and denote by $\mathbb{E}_a \in \mathbb{R}^{\mathbb{A}}$ the unit vector at $a \in \mathbb{A}$. There are three cases for $n \in \mathbb{Z}$.

Case 1. $\tau(n+u+1) > d$. In this case, $\tau(n+t+1) = d$ holds. Hence, $\Psi(\alpha[n+S]) = \mathbb{E}_a + \mathbb{E}_c$ for some $a \in \mathbb{A}$.

Case 2. $\tau(n+u+1) = d$. In this case, $\tau(n+s+1) = d$ holds. Hence, $\Psi(\alpha[n+S]) = \mathbb{E}_a + \mathbb{E}_c$ for some $a \in \mathbb{A}$.

Case 3. $\tau(n+u+1) < d$. In this case, $\tau(n+s+1) = \tau(n+t+1) < d$. Hence, $\Psi(\alpha[n+S]) = 2\mathbb{E}_a$ for some $a \in \mathbb{A}$.

Therefore,

$$\{\Psi(\alpha[n+S]) : n \in \mathbb{N}\} \subset \{\mathbb{E}_a + \mathbb{E}_c : a \in \mathbb{A}\} \cup \{2\mathbb{E}_a : a \in \mathbb{A}\},$$

and, hence, $p_{\alpha}^{*\mathrm{ab}}(2) \leq 2r-1$. Thus, $p_{\alpha}^{*\mathrm{ab}}(2) = 2r-1$ since we already have $p_{\alpha}^{*\mathrm{ab}}(2) \geq 2r-1$. Note that this proof remains true for any of the general Toeplitz words defined in the Introduction. □

https://doi.org/10.1017/etds.2013.51 Published online by Cambridge University Press

Remark 4.1. Theorem 4 is not true without the assumption of recurrency. In fact, let $\alpha = 10^3 10^{3^2} 10^{3^3} 1 \cdots \in \{0, 1\}^{\mathbb{N}}$. Then $p_\alpha^{*\mathrm{ab}}(3) = 3$. To see this, suppose $\alpha[n + S] = 111$ for some $n \in \mathbb{N}$, and some 3-pattern $S = \{i < j < k\}$. Then $j - i = 3^b - 3^a$ and $k - j = 3^c - 3^b$ for some positive integers $a < b < c$. Moreover, this happens when $n = 3^a - i$.

Suppose $\alpha[m + S] = 110$ for some m. Then, since there exist positive integers $d < e$ such that $m + i = 3^d$ and $m + j = 3^e$, we have $j - i = (m + j) - (m + i) = 3^e - 3^d = 3^b - 3^a$. This implies that $3^e + 3^a = 3^b + 3^d$, concluding $e = b$ and $a = d$ by the uniqueness of the 3-adic representation. Hence, $m = 3^d - i = 3^a - i$, a contradiction.

If $\alpha[m + S] = 101$ for some m, then, since there exist positive integers $d < e$ such that $m + i = 3^d$ and $m + k = 3^e$, we have $k - i = (m + k) - (m + i) = 3^e - 3^d = 3^c - 3^a$. This implies that $3^e + 3^a = 3^c + 3^d$, concluding $e = c$ and $a = d$ by the uniqueness of the 3-adic representation. Hence, $m = 3^d - i = 3^a - i$, a contradiction.

Finally, if $\alpha[m + S] = 011$ for some m, then, since there exist positive integers $d < e$ such that $m + j = 3^d$ and $m + k = 3^e$, we have $k - j = (m + k) - (m + j) = 3^e - 3^d = 3^c - 3^b$. This implies that $3^e + 3^b = 3^c + 3^d$, concluding $e = c$ and $b = d$ by the uniqueness of the 3-adic representation. Hence, $m = 3^d - j = 3^b - j = 3^a - i$, again a contradiction.

Thus, if $111 \in \{\alpha[n + S] : n \in \mathbb{N}\}$, then $\{110, 101, 011\} \cap \{\alpha[n + S] : n \in \mathbb{N}\} = \emptyset$. Thus, $p_\alpha^{*\mathrm{ab}}(3) \le 3$. Since it is clear that $p_\alpha^{*\mathrm{ab}}(3) \ge 3$, we have $p_\alpha^{*\mathrm{ab}}(3) = 3$.

Remark 4.2. We do not know whether there exists $\alpha \in \mathbb{A}^{\mathbb{N}}$ (with $\#\mathbb{A} = r \ge 3$) which is both recurrent and aperiodic by projection and such that

$$p_\alpha^{*\mathrm{ab}}(k) = (r - 1)k + 1 \quad (k = 1, 2, \ldots).$$

Let $\mathbb{A} = \{0, 1, \ldots, r - 1\}$ and $\Omega = \bigcup_{i=0}^{r-2} \{i, i + 1\}^{\mathbb{N}}$. Then it is readily verified that $p_\Omega^{*\mathrm{ab}}(k) = (r - 1)k + 1$ $(k = 1, 2, \ldots)$, but Ω is not equal to $\overline{O}(\alpha)$ for any choice of $\alpha \in \mathbb{A}^{\mathbb{N}}$.

THEOREM 6. *Let $\alpha \in \mathbb{A}^N$ be a rotation word with $\#\mathbb{A} = r$. Then we have $p_\alpha^{*\mathrm{ab}}(k) = rk$ $(k = 1, 2, \ldots)$.*

Remark 4.3. For a rotation word $\alpha \in \mathbb{A}^N$ with $\#\mathbb{A} = r$, it is known [**4**] that $p_\alpha^*(k) = rk$ $(k = 1, 2, \ldots)$. Hence, Theorem 6 shows that the abelianization does not decrease the complexity in the case of rotation words on more than two letters.

Proof of Theorem 6. Since $p_\alpha^{*\mathrm{ab}}(k) \le p_\alpha^*(k) = rk$ $(k = 1, 2, \ldots)$, it is sufficient to prove that $p_\alpha^{*\mathrm{ab}}(k) \ge rk$ $(k = 1, 2, \ldots)$. Let θ be an irrational number and $c_0 < c_1 < \cdots < c_{r-1} < c_r$ be real numbers with $c_r = c_0 + 1$. Let $\mathbb{A} = \{0, 1, \ldots, r - 1\}$. We may assume that $\alpha \in \mathbb{A}^{\mathbb{N}}$ is such that $\alpha(n) = i$, whenever $n\theta \in [c_i, c_{i+1}) \pmod 1$.

Fix $0 < \varepsilon < \min_i (c_{i+1} - c_i)$. Set $\mathcal{N} = \{N_0 < N_1 < \cdots\} \subset \mathbb{N}$ such that

$$\varepsilon > \{N_0\theta\} > \{N_1\theta\} > \cdots > 0$$

and $\lim_{n\to\infty} \{N_n\theta\} = 0$. Here, $\{\cdot\}$ denotes the fractional part. Then it is easy to see that

$$\{\alpha[n + \mathcal{N}] : n \in \mathbb{N}\} = \bigcup_{i=0}^{r-1} \{(i + 1)^n i^\infty : n \in \mathbb{N}\},$$

https://doi.org/10.1017/etds.2013.51 Published online by Cambridge University Press

where we identify r with 0 as letters. Thus, for any $k = 1, 2, \ldots$, we have

$$\{\alpha[n + \mathcal{N}_k] : n \in \mathbb{N}\} = \bigcup_{i=0}^{r-1} \{(i+1)^n i^{k-n} : 0 \le n \le k\},$$

where $\mathcal{N}_k = \{N_0 < N_1 < \cdots < N_{k-1}\}$. There are exactly rk words, as above. Thus, $p_\alpha^{*\mathrm{ab}}(k) \ge rk$ $(k = 1, 2, \ldots)$, which completes the proof. □

Acknowledgement. The third author is supported by a FiDiPro grant from the Academy of Finland.

References

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

[2] R. McCutcheon. *Elementary Methods in Ergodic Ramsey Theory (Lecture Notes in Mathematics, 1722)*. Springer, Berlin, 1999 (in Ch. 2).

[3] T. Kamae and L. Q. Zamboni. Sequence entropy and the maximal pattern complexity of infinite words. *Ergod. Th. & Dynam. Sys.* **22** (2002), 1191–1199.

[4] T. Kamae and H. Rao. Maximal pattern complexity over ℓ letters. *European J. Combin.* **27** (2006), 125–137.

[5] T. Kamae. Uniform sets and super-stationary sets over general alphabets. *Ergod. Th. & Dynam. Sys.* **31** (2011), 1445–1461.

[6] M. Morse and G. A. Hedlund. Symbolic dynamics II: Sturmian trajectories. *Amer. J. Math.* **62**(1) (1940), 1–42.

[7] G. Richomme, K. Saari and L. Q. Zamboni. Abelian complexity of minimal subshifts. *J. Lond. Math. Soc.* (2) **83** (2011), 79–95.

https://doi.org/10.1017/etds.2013.51 Published online by Cambridge University Press