On a group theoretic generalization of the Morse-Hedlund theorem

Émilie Charlier[a], Svetlana Puzynina[b,1], Luca Q. Zamboni[b,c]

[a]*Département de Mathématique, Université de Liège, Belgium*
[b]*LIP, ENS de Lyon, Université de Lyon, France and Sobolev Institute of Mathematics, Novosibirsk, Russia*
[c]*Institut Camille Jordan, Université Lyon 1, France*

Abstract

In this paper we give a broad unified framework via group actions for constructing complexity functions of infinite words $x = x_0x_1x_2\cdots \in \mathbb{A}^{\mathbb{N}}$ with values in a finite set $\mathbb{A}$. Factor complexity, Abelian complexity and cyclic complexity are all particular cases of this general construction. We consider infinite sequences of permutation groups $\omega = (G_n)_{n\geq 1}$ with each $G_n \subseteq S_n$. Associated with every such sequence is a complexity function $p_{\omega,x} : \mathbb{N} \to \mathbb{N}$ which counts, for each length n, the number of equivalence classes of factors of x of length n under the action of G_n on $\mathbb{A}^n$ given by $g * (u_1u_2\cdots u_n) = u_{g^{-1}(1)}u_{g^{-1}(2)}\cdots u_{g^{-1}(n)}$. Each choice of $\omega = (G_n)_{n\geq 1}$ defines a unique complexity function which reflects a different combinatorial property of a given infinite word. For instance, an infinite word x has bounded Abelian complexity if and only if x is k-balanced for some positive integer k, while bounded cyclic complexity is equivalent to x being ultimately periodic. A celebrated result of G.A. Hedlund and M. Morse states that every aperiodic infinite word $x \in \mathbb{A}^{\mathbb{N}}$ contains at least $n+1$ distinct factors of each length n. Moreover $x \in \mathbb{A}^{\mathbb{N}}$ has exactly $n+1$ distinct factors of each length n if and only if x is a Sturmian word, i.e., binary, aperiodic and balanced. We prove that this characterisation of aperiodicity and Sturmian words extends to this general framework.

Keywords: Symbolic dynamics, complexity, Sturmian words, discrete interval exchange transformations.
2010 MSC: 37B10

1. Introduction

For each infinite word $x = x_0x_1x_2\cdots \in \mathbb{A}^{\mathbb{N}}$, with values in a finite set $\mathbb{A}$, the factor complexity function $p_x : \mathbb{N} \to \mathbb{N}$ counts the number of distinct blocks (or factors) of each length n occurring in x. First introduced by G.A. Hedlund and M. Morse in their 1938

Email addresses: `echarlier@ulg.ac.be` (Émilie Charlier), `s.puzynina@gmail.com` (Svetlana Puzynina), `lupastis@gmail.com` (Luca Q. Zamboni)

[1]Supported by the LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

seminal paper on Symbolic Dynamics under the name of *block growth*[2], the factor complexity provides a useful measure of the extent of randomness of x and more generally of the subshift it generates. They proved that every aperiodic (meaning not eventually periodic) infinite word contains at least $n+1$ distinct factors of each length n. They further showed that an infinite word x has exactly $n+1$ distinct factors of each length n if and only if x is binary, aperiodic and balanced, i.e., x is a Sturmian word. Thus Sturmian words are those aperiodic words of lowest factor complexity. They arise naturally in many different areas of mathematics including combinatorics, algebra, number theory, ergodic theory, dynamical systems and differential equations. Sturmian words also have implications in theoretical physics as 1-dimensional models of quasi-crystals, and in theoretical computer science where they are used in computer graphics as digital approximation of straight lines. Despite their simplicity, Sturmian words possess several deep and mysterious properties (see [15, 16, 17]).

There are several variations and extensions of the Morse-Hedlund theorem associated with other types of complexity functions of an infinite word $x \in \mathbb{A}^{\mathbb{N}}$ including for instance Abelian complexity [8, 24], which counts for the number of distinct Abelian classes of words of each length n occurring in x, palindrome complexity [3], which counts the number of distinct palindromes of each length n occurring in x, cyclic complexity [7] which counts the number of conjugacy classes of factors of each length n occurring in x, and maximal patterns complexity [19]. In most cases, these different notions of complexity may be used to detect (and in some cases characterize) ultimately periodic words. Generally, amongst all aperiodic words, Sturmian words have the lowest possible complexity, although in some cases they are not the only ones (for instance, a restricted class of Toeplitz words is found to have the same maximal pattern complexity as Sturmian words [19]). There have also been numerous attempts at extending the Morse-Hedlund theorem in higher dimensions. A celebrated conjecture of M. Nivat states that any 2-dimensional word having at most mn distinct $m \times n$ blocks must be periodic. In this case, it is known that the converse is not true. To this day the Nivat conjecture remains open although the conjecture has been verified for m or n less or equal to 3 (see [9, 25]). A very interesting higher dimensional analogue of the Morse-Hedlund theorem was recently obtained by Durand and Rigo in [12] in which they re-interpret the notion of periodicity in terms of Presburger arithmetic.

In this paper we give a broad unified framework via group actions for constructing complexity functions of infinite words. Factor complexity, Abelian complexity and cyclic complexity turn out to be particular cases of this general construction. We consider infinite sequences of permutation groups $\omega = (G_n)_{n \geq 1}$ with each $G_n \subseteq S_n$. Associated with every such sequence, and with every infinite word $x \in \mathbb{A}^{\mathbb{N}}$, is a complexity function $p_{\omega,x} : \mathbb{N} \to \mathbb{N}$ which counts, for each length n, the number of equivalence classes of factors of x of length n

[2] In [13], Ehrenfeucht, Lee, and Rozenberg adopted the term *subword complexity.*

under the action of G_n on $\mathbb{A}^n$ given by $g * (u_1u_2\cdots u_n) = u_{g^{-1}(1)}u_{g^{-1}(2)}\cdots u_{g^{-1}(n)}$. Thus the usual factor complexity is obtained by taking the infinite sequence $(Id_n)_{n\geq 1}$, where Id_n is the trivial subgroup of S_n consisting only of the identity, while Abelian complexity corresponds to the sequence $(S_n)_{n\geq 1}$, and finally cyclic complexity to the sequence $(C_n)_{n\geq 1}$, where C_n is the cyclic group generated by the n-cycle $(1, 2, \ldots, n)$. Each choice of $\omega = (G_n)_{n\geq 1}$ defines a unique complexity function which reflects some combinatorial property of an infinite word. For instance, the Morse-Hedlund theorem asserts that an infinite word x has bounded factor complexity if and only if x is ultimately periodic. Bounded cyclic complexity is also equivalent to x being ultimately periodic [7]. In contrast bounded Abelian complexity is equivalent to the condition that x is k-balanced for some positive integer k (see [18]). Two Sturmian words x and y have the same factor complexity and the same Abelian complexity. Instead they have the same cyclic complexity if and only if they belong to the same minimal subshift, i.e., they have the same slope (see Theorem 2 in [7]).

We prove that the celebrated theorem of Hedlund and Morse extents to this general framework. More precisely, if an infinite word $x \in \mathbb{A}^{\mathbb{N}}$ is aperiodic, then for every infinite sequence of permutation groups $\omega = (G_n)_{n\geq 1}$ we have $p_{\omega,x}(n) \geq \epsilon(G_n) + 1$ for each $n \geq 1$, where $\epsilon(G_n)$ is the number of distinct G_n-orbits of $\{1, 2, \ldots, n\}$ (see Theorem 1). Applied to the sequence $(Id_n)_{n\geq 1}$, it says that every aperiodic word contains at least $\epsilon(Id_n) + 1 = n + 1$ distinct factors of each length n. Similarly applied to the sequence $(S_n)_{n\geq 1}$ it states that every aperiodic word contains at least $\epsilon(S_n) + 1 = 2$ Abelian classes of factors of each length n. We further show that in this general setting, Sturmian words are characterised as those aperiodic words of minimal complexity. More precisely, we show that if $x \in \mathbb{A}^{\mathbb{N}}$ is aperiodic and $\omega = (G_n)_{n\geq 1}$ is such that $p_{\omega,x}(n) = \epsilon(G_n) + 1$ for each $n \geq 1$, then x is Sturmian. The converse is in general not true, that is if x is Sturmian and $\omega = (G_n)_{n\geq 1}$, then it is not always the case that $p_{\omega,x}(n) = \epsilon(G_n) + 1$ for each $n \geq 1$. For instance, if x is Sturmian and $\omega = (C_n)_{n\geq 1}$, where each C_n is the cyclic subgroup of S_n generated by the n-cycle $(1, 2, \ldots, n)$, then $\epsilon(C_n) = 1$ while $p_{\omega,x}(n)$ is unbounded (see Theorem 1 in [7]). However, we show that if x is Sturmian, then there exists a sequence $\omega' = (C'_n)_{n\geq 1}$, where each C'_n is a cyclic subgroup of S_n generated by an n-cycle, and $p_{\omega',x}(n) = 2$ for each $n \geq 1$ (see Corollary 2). Combined with the fundamental theorem of finite Abelian groups, we prove that if x is a Sturmian word, then for every infinite sequence $\omega = (G_n)_{n\geq 1}$ of Abelian permutation groups there exists $\omega' = (G'_n)_{n\geq 1}$ with G'_n isomorphic to G_n and $p_{\omega',x}(n) = \epsilon(G'_n) + 1$ for each $n \geq 1$ (see Theorem 2).

Our methods rely largely on the rich combinatorial properties of Sturmian words and in particular the structure of the bispecial factors. We use results from [18] on the Christoffel array associated with a bispecial factor w of a Sturmian word, in which the cyclic conjugates of $0w1$ are ordered lexicographically in a rectangular array. Another key feature is the use of discrete 3-interval exchange transformations in the sense of [23]. More precisely, we associate to each Abelian permutation group $G_n \subseteq S_n$ a system of discrete 3-interval exchange transformations which acts on the factors of a Sturmian word of length n.

2. Main Results

Let S_n denote the symmetric group on n-letters which we regard as the set of all bijections of $\{1, 2, \ldots, n\}$. Fix a subgroup $G \subseteq S_n$. We consider the G-action $G \times \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ given by $(g, i) \mapsto g(i)$ and let $\epsilon(G)$ denote the number of distinct orbits, i.e.,

$$\epsilon(G) = \mathrm{Card}(\{G(i) \mid i \in \{1, 2, \ldots, n\}\})$$

where $G(i) = \{g(i) \mid g \in G\}$ denotes the G-orbit of i. For instance if G is the trivial subgroup of S_n consisting only of the identity, then $\epsilon(G) = n$, while if G contains an n-cycle, then $\epsilon(G) = 1$. We note that $\epsilon(G)$ strongly depends on the embedding of G in S_n, and in fact is not a group isomorphism invariant, even for isomorphic subgroups of S_n. For instance, the subgroups $G_1 = \{e, (1, 2), (3, 4), (1, 2)(3, 4)\}$ and $G_2 = \{e, (1, 2)(3, 4), (1, 3)(2, 4), (1, 4)(2, 3)\}$ are two embeddings of the Klein four-group $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ in S_4, and yet $\epsilon(G_1) = 2$ while $\epsilon(G_2) = 1$. On the other hand, it is easily checked that $\epsilon(G)$ only depends on the conjugacy class of G in S_n.

Let $\mathbb{A}$ be a finite non-empty set. For each $n \geq 1$, let $\mathbb{A}^n$ denote the set of all words $u = u_1u_2 \cdots u_n$ with $u_i \in \mathbb{A}$. For $a \in \mathbb{A}$ we denote by $|u|_a$ the number of occurrences of the symbol a in u. Two words $u, v \in \mathbb{A}^n$ are *Abelian equivalent*, written $u \sim_{ab} v$, if $|u|_a = |v|_a$ for each $a \in \mathbb{A}$. It is convenient to consider elements of $\mathbb{A}^n$ as functions $u : \{1, 2, \ldots, n\} \to \mathbb{A}$ where $u(i) = u_i \in \mathbb{A}$ for $1 \leq i \leq n$. For each subset $S \subseteq \{1, 2, \ldots, n\}$ we denote by $u|_S$ the restriction of u to S. There is a natural G-action $G \times \mathbb{A}^n \to \mathbb{A}^n$ given by $g * u : i \mapsto u(g^{-1}(i))$ for each $i \in \{1, 2, \ldots, n\}$. In terms of the word representation we have $g * u = u_{g^{-1}(1)}u_{g^{-1}(2)} \cdots u_{g^{-1}(n)}$. In particular we have $g * u \sim_{ab} u$ for all $g \in G$.

Let $x = x_0x_1x_2 \cdots \in \mathbb{A}^{\mathbb{N}}$ be an infinite word. Then G defines an equivalence relation $\sim_G$ on $\mathrm{Fact}_x(n) = \{x_ix_{i+1} \cdots x_{i+n-1} \mid i \geq 0\}$, the set of factors of x of length n, given by $u \sim_G v$ if and only if $g * u = v$ for some $g \in G$, in other words if u and v are in the same G-orbit relative to the action of G on $\mathbb{A}^n$. We say that $\sim_G$ is *Abelian transitive* on x if for all $u, v \in \mathrm{Fact}_x(n)$ we have $u \sim_{ab} v$ if and only if $u \sim_G v$.

We are interested in the number of $\sim_G$ equivalence classes, i.e., $\mathrm{Card}(\mathrm{Fact}_x(n)/\sim_G)$. Unlike $\epsilon(G)$, this quantity is not a conjugacy invariant of G in S_n. For instance, consider the cyclic subgroups $G_1 = \langle\sigma_1\rangle$ and $G_2 = \langle\sigma_2\rangle$ of S_4 where $\sigma_1 = (1, 2, 3, 4)$ and $\sigma_2 = (1, 3, 2, 4)$. Let x denote the Fibonacci word fixed by the substitution $0 \mapsto 01, 1 \mapsto 0$. Then $\mathrm{Fact}_x(4) = \{0010, 0100, 0101, 1001, 1010\}$ and

$$\mathrm{Fact}_x(4)/\sim_{G_1} = \{[0100 \overset{\sigma_1}{\curvearrowright} 0010]; [0101 \overset{\sigma_1}{\curvearrowright} 1010]; [1001]\}$$

while

$$\mathrm{Fact}_x(4)/\sim_{G_2} = \{[0010 \overset{\sigma_2}{\curvearrowright} 0100]; [0101 \overset{\sigma_2}{\curvearrowright} 1001 \overset{\sigma_2}{\curvearrowright} 1010]\}.$$

We observe that the two equivalence classes relative to $\sim_{G_2}$ correspond to the two Abelian classes of $\mathrm{Fact}_x(4)$. Thus the equivalence relation $\sim_{G_2}$ is Abelian transitive on x while $\sim_{G_1}$ is not. On the other hand if $y \in \{0,1\}^{\mathbb{N}}$ is such that $\mathrm{Fact}_y(4) = \{0000, 0001, 0010, 0100, 1000\}$, then both $\sim_{G_1}$ and $\sim_{G_2}$ are Abelian transitive on y.

We apply the above considerations to define a complexity function on infinite words. More precisely, we consider the category $\mathcal{G}$ whose objects are all infinite sequences $(G_n)_{n\geq 1}$ where G_n is a subgroup of S_n and $\mathrm{Hom}((G_n)_{n\geq 1}, (G'_n)_{n\geq 1})$ is the collection of all $(f_n)_{n\geq 1}$ where $f_n : G_n \to G'_n$ is a group homomorphism. Two elements $(G_n)_{n\geq 1}, (G'_n)_{n\geq 1} \in \mathcal{G}$ are said to be *conjugate* if there exists $(\sigma_n)_{n\geq 1}$ with $\sigma_n \in S_n$ such that $G'_n = \sigma_n G_n \sigma_n^{-1}$ for each $n \geq 1$, and *isomorphic* if there exists $(f_n)_{n\geq 1} \in \mathrm{Hom}((G_n)_{n\geq 1}, (G'_n)_{n\geq 1})$ such that $f_n : G_n \to G'_n$ is a group isomorphism for each $n \geq 1$. Associated with every $\omega = (G_n)_{n\geq 1} \in \mathcal{G}$ is a complexity function $p_{\omega,x} : \mathbb{N} \to \mathbb{N}$ which counts for each length n the number of $\sim_{G_n}$ equivalence classes of factors of length n of an infinite word x.

Theorem 1. *Let $x \in \mathbb{A}^{\mathbb{N}}$ be aperiodic. Then for every infinite sequence $\omega = (G_n)_{n\geq 1} \in \mathcal{G}$ we have $p_{\omega,x}(n) \geq \epsilon(G_n) + 1$ for each $n \geq 1$. Moreover if $p_{\omega,x}(n) = \epsilon(G_n) + 1$ for each $n \geq 1$, then x is Sturmian.*

Remark 2.1. In our proof of Theorem 1, we actually show that if x is any infinite word such that $p_{\omega,x}(n) = \epsilon(G_n) + 1$ (for each $n \geq 1$) for some infinite sequence $\omega = (G_n)_{n\geq 1} \in \mathcal{G}$, then x is binary and balanced. In other words the assumption that x is aperiodic is necessary to deduce that x is Sturmian. For instance, the complexity function associated with the sequence $\omega = (S_n)_{n\geq 1}$ does not distinguish between the eventually periodic word $01^{\omega} = 01111\cdots$ and any Sturmian word. In both cases the complexity is the constant function $p_{\omega,x}(n) = 2$. On the other hand, in view of the Morse-Hedlund theorem, the factor complexity distinguishes between these two words. The same is true of cyclic complexity (see Theorem 2 in [7]).

As an immediate corollary we have:

Corollary 1. *An aperiodic word $x \in \mathbb{A}^{\mathbb{N}}$ is Sturmian if and only if there exists a sequence $\omega = (G_n)_{n\geq 1} \in \mathcal{G}$ verifying $p_{\omega,x}(n) = \epsilon(G_n) + 1$ for each $n \geq 1$.*

One direction follows immediately from Theorem 1. For the other implication, if x is Sturmian, we may take the sequence $\omega = (Id_n)_{n\geq 1} \in \mathcal{G}$.

In general it is not true that if x is Sturmian and $\omega = (G_n)_{n\geq 1} \in \mathcal{G}$ then $p_{\omega,x}(n) = \epsilon(G_n)+1$ for each $n \geq 1$. For instance, if we take $\omega = (C_n)_{n\geq 1} \in \mathcal{G}$, where each C_n is the cyclic subgroup of S_n generated by the n-cycle $(1, 2, \ldots, n)$, then $\epsilon(C_n) = 1$ while $p_{\omega,x}(n)$ is unbounded (see Theorem 1 in [7]). On the other hand we show that if x is Sturmian, then there exists a sequence $\omega' = (C'_n)_{n\geq 1}$, where each C'_n is a cyclic subgroup of S_n generated by an n-cycle, such that $p_{\omega',x}(n) = 2$ for each $n \geq 1$ (see Corollary 2). Combined with the fundamental theorem of finite abelian groups, we are able to obtain a partial converse to Theorem 1 which is stronger than the characterisation given in Corollary 1. For this purpose we restrict to the sub category $\mathcal{G}_{\rm ab}$ of all infinite sequences $(G_n)_{n\geq 1}$ of Abelian subgroups of S_n.

Theorem 2. *Let x be a Sturmian word. Then for each infinite sequence $\omega = (G_n)_{n\geq 1} \in \mathcal{G}_{ab}$ of Abelian permutation groups there exists $\omega' = (G'_n)_{n\geq 1} \in \mathcal{G}_{ab}$ isomorphic to ω with $p_{\omega',x}(n) = \epsilon(G'_n) + 1$ for each $n \geq 1$.*

Remark 2.2. Let $\omega = (Id_n)_{n\geq 1} \in \mathcal{G}_{\rm ab}$, where Id_n denotes the trivial subgroup of S_n consisting only of the identity. Then $\epsilon(Id_n) = n$ for each $n \geq 1$. Moreover, for each infinite word x, we have that $p_{\omega,x}(n) = \mathrm{Card}(\mathrm{Fact}_x(n))$. Thus applying Theorem 1 to ω we deduce that every aperiodic word x contains at least $n+1$ distinct factors of length n and that if x has exactly $n+1$ distinct factors of each length n, then x is Sturmian. Conversely, if x is Sturmian, then Theorem 2 applied to ω implies that x contains exactly $n+1$ distinct factors of length n. Thus we recover the full Morse-Hedlund theorem. On the opposite extreme, taking $\omega = (S_n)_{n\geq 1}$, we get that $p_{\omega,x}$ is the abelian complexity function. Then applying Theorem 1 to ω we recover a classical result, namely that the abelian complexity of an aperiodic word is at least 2.

Before embarking on the proofs of Theorems 1& 2 we review a few key facts concerning aperiodic words in general and Sturmian words in particular. For all other definitions and basic notions in combinatorics on words we refer the reader to [20]. A factor u of an infinite word $x \in \mathbb{A}^{\mathbb{N}}$ is called *left special* (resp. *right special*) if there exist distinct symbols $a, b \in \mathbb{A}$ such that au and bu (resp. ua and ub) are factors of x. A factor u which is both left and right special is called *bispecial.* If x is aperiodic, then x admits at least one left and one right special factor of each length. Given u and v factors of x with u a prefix of v, we write $u \models_x v$ to mean that each occurrence of u in x is an occurrence of v. Clearly, if $u \models_x v$ and u is both a proper prefix and a proper suffix of v, then x is ultimately periodic.

An infinite word $x \in \mathbb{A}^{\mathbb{N}}$ is said to be *balanced* if for every pair of factors u and v of x of equal length we have $||u|_a - |v|_a| \leq 1$ for every $a \in \mathbb{A}$. An infinite word is called *Sturmian* if it is aperiodic, binary and balanced. Equivalently, x is Sturmian if x admits precisely $n+1$ distinct factors of each length n. This implies that x admits exactly one left and one right special factor of each length. Moreover, the set of factors of a Sturmian word is closed under reversal, i.e., $u = u_1 u_2 \cdots u_n$ is a factor of x if and only if the reverse of $\bar{u} = u_n \cdots u_2 u_1$ is a factor of x (see for instance Chapter 2 in [20]). Thus, the right special factors of a Sturmian word are precisely the reversals of the left special factors and vice versa. In particular, the bispecial factors of a Sturmian word, also called *central words* (see Proposition 10 in [11] and Theorem 2.2.11 in [20]), are palindromes.

Let $x \in \{0,1\}^{\mathbb{N}}$ be Sturmian and fix $n \geq 1$. It follows from the above considerations that there exists a unique word u of length $n-1$ such that both $u0$ and $u1$ belong to $\mathrm{Fact}_x(n)$, and a unique word v of length $n-1$ such that both $0v$ and $1v$ belong to $\mathrm{Fact}_x(n)$. In other words u (resp. v) is the unique right (resp. left) special factor of x of length $n-1$. In case $u \neq v$, then u is a suffix of a unique factor w of length n and both $w0$ and $w1$ belong to $\mathrm{Fact}_x(n+1)$. Moreover, for each other factor $z \neq w$ of length n, let z' denote the suffix of z of length $n-1$. Then as z' is not right special, it follows that there exists a unique $a \in \{0,1\}$ such that $z' \models_x z'a$. Hence $z \models_x za$. In other words, in case $u \neq v$, we have that $\mathrm{Fact}_x(n)$ uniquely determines $\mathrm{Fact}_x(n+1)$. On the other hand, in case $u = v$ (i.e., u is a bispecial factor of x of length $n-1$), then each of $u0, u1, 0u, 1u$ belong to $\mathrm{Fact}_x(n)$. In this case, exactly one of the following two cases occurs: Either $0u$ is right special, in which case by the balance property we must have $1u \models_x 1u0$, or $1u$ is right special, in which case $0u \models_x 0u1$. Moreover, each of these two cases is possible, meaning that there exists a Sturmian word x' whose factors agree with those of x up to length n and differ at length $n+1$: One admits the factor $0u0$ and the other $1u1$.

Given a factor u of a Sturmian word $x \in \{0,1\}^{\mathbb{N}}$ and $a \in \{0,1\}$, we say that u is *rich* in a if $|u|_a \geq |v|_a$ for all factors v of x of length equal to that of u.

Proof of Theorem 1. We will make use of the following lemma:

Lemma 2.3. *Let $E_1, E_2, \ldots, E_k$ be a partition of $\{1,2,\ldots,n\}$ ordered so that $i < j \Rightarrow \max E_i < \max E_j$. For each $1 \leq j \leq k$, let $\sim_j$ denote the equivalence relation on $\mathbb{A}^n$ defined by $u \sim_j v$ if and only if $u|_{E_i} \sim_{ab} v|_{E_i}$ for each $1 \leq i \leq j$. Then for each aperiodic word $x \in \mathbb{A}^{\mathbb{N}}$ and for each $1 \leq j \leq k$ we have Card$(\mathrm{Fact}_x(n)/\sim_j) \geq j+1$.*

Proof. Let $x \in \mathbb{A}^{\mathbb{N}}$ be aperiodic. We will show that $\mathrm{Card}(\mathrm{Fact}_x(n)/\sim_1) \geq 2$ and that

$$\mathrm{Card}(\mathrm{Fact}_x(n)/\sim_{j+1}) \geq \mathrm{Card}(\mathrm{Fact}_x(n)/\sim_j) + 1$$

for each $1 \leq j \leq k-1$. Let $m_i = \max E_i$. Since x is aperiodic, x contains at least one right special factor of each length. In particular, there exists $u \in \mathbb{A}^*$, with $|u| = m_1 - 1$, and distinct letters $a, b \in \mathbb{A}$ such that ua and ub are factors of x. Let $U, V \in \mathrm{Fact}_x(n)$ with ua a prefix of U and ub a prefix of V. As $ua \nsim_{ab} ub$ and $|ua| = |ub| = m_1 \in E_1$, we have $U|_{E_1} \nsim_{ab} V|_{E_1}$, whence $U \nsim_1 V$. Thus $\mathrm{Card}(\mathrm{Fact}_x(n)/\sim_1) \geq 2$. Next fix $1 \leq j \leq k-1$. We will show the existence of two factors U and V of length n such that $U \sim_j V$ and $U \nsim_{j+1} V$. As above, since x is aperiodic, there exists $u \in \mathbb{A}^+$, with $|u| = m_{j+1} - 1$, and distinct letters $a, b \in \mathbb{A}$ such that ua and ub are factors of x. Let $U, V \in \mathrm{Fact}_x(n)$ with ua a prefix of U and ub a prefix of V. Then for each $1 \leq i \leq j$ we have $U|_{E_i} = u|_{E_i} = V|_{E_i}$ and hence $U \sim_j V$. On the other hand, as before, since $ua \nsim_{ab} ub$ and $|ua| = m_{j+1}$, we have $U \nsim_{j+1} V$. □

Fix $n \geq 1$, and put $G = G_n$ and $\epsilon(G) = k$. We will show that if $x \in \mathbb{A}^{\mathbb{N}}$ is aperiodic, then $\mathrm{Card}(\mathrm{Fact}_x(n)/\sim_G) \geq k+1$. Let $E_1, E_2, \ldots, E_k$ denote the full set of G-orbits of $\{1,2,\ldots,n\}$. Then $E_1, E_2, \ldots, E_k$ is a partition of $\{1,2,\ldots,n\}$ and we can order these sets so that $i < j \Rightarrow \max E_i < \max E_j$. For $1 \le j \le k$ let $\sim_j$ denote the equivalence relation on $\mathrm{Fact}_x(n)$ defined in the previous lemma. Then for all $u, v \in \mathrm{Fact}_x(n)$ we have $u \sim_G v$ implies $u \sim_k v$. Thus

$$\mathrm{Card}(\mathrm{Fact}_x(n)/\sim_G) \ge \mathrm{Card}(\mathrm{Fact}_x(n)/\sim_k) \ge k+1 = \epsilon(G)+1$$

as required. This concludes our proof of the first statement of Theorem 1.

Next suppose that $p_{\omega,x}(n) = \epsilon(G_n)+1$ for each $n \ge 1$. We will show that x is binary and balanced. Since x is already assumed aperiodic, it will follow that x is Sturmian. Since $\epsilon(G_1) = 1$, and hence $p_{\omega,x}(1) = 2$, it follows that x is on a binary alphabet which we can take to be $\{0,1\}$.

Lemma 2.4. *Let $x \in \{0,1\}^{\mathbb{N}}$ be aperiodic. Then either x is Sturmian or there exist an integer $n \ge 2$, a Sturmian word y and a bispecial factor $u \in \{0,1\}^{n-2}$ of y such that $Fact_x(n) = Fact_y(n) \cup \{0u0, 1u1\}$.*

Proof. Suppose x is not Sturmian. Let n be the least positive integer such that for all Sturmian words z we have $\mathrm{Fact}_x(n) \ne \mathrm{Fact}_z(n)$. As x is binary, $n \ge 2$. By minimality of n there exists a Sturmian word y such that $\mathrm{Fact}_x(n-1) = \mathrm{Fact}_y(n-1)$. We claim that there exists a factor $u \in \mathrm{Fact}_x(n-2) = \mathrm{Fact}_y(n-2)$ which is bispecial in both x and y. In fact, let u be the unique right special factor of x and y of length $n-2$. If u is not left special, then there exists a unique factor $v \in \mathrm{Fact}_x(n-1) = \mathrm{Fact}_y(n-1)$ ending in u, and this factor would necessarily be right special in both x and y. Moreover all other factors of x and y of length $n-1$ admit a unique extension to a factor of length n determined by their suffix of length $n-2$. Hence we would have $\mathrm{Fact}_x(n) = \mathrm{Fact}_y(n)$ contrary to the choice of n. Thus u is also left special (in both x and y) and hence bispecial.

Since x is aperiodic, at least one of $0u$ or $1u$ is right special in x. Without loss of generality we may assume $0u$ is right special. We now claim that $1u$ must also be right special. In fact, suppose to the contrary that $1u \models_x 1ua$ for some $a \in \{0,1\}$. If $a = 0$, then $\mathrm{Fact}_x(n)$ would coincide with the set of factors of length n of some Sturmian word, contrary to our choice of n. Thus $a = 1$. We will show that this implies that x is ultimately periodic, and hence gives rise to a contradiction. We consider two cases: First suppose no non-empty prefix of $1u$ is right special; in this case $1 \models_x 1u \models_x 1u1$ whence x is ultimately periodic. Thus we may assume that some prefix $1v$ of $1u$ is right special. Consider the longest such right special prefix $1v$. Since we are assuming that $1u$ is not right special, it follows that vb is a prefix of u for some $b \in \{0,1\}$. Since vb is left special (as vb is a prefix of u), and since $1v$ is right special, we deduce that vb is equal to the reverse of $1v$ from which it follows that $b = 1$. Thus as $1v$ is a suffix of u (because $1v$ and u are both right special), we have that $1v1$ is a proper suffix of $1u1$. Now since $1v1 \models_x 1u \models_x 1u1$, it follows that x is ultimately periodic. Thus we have shown that both $0u$ and $1u$ are right special. Since in y exactly one of $0u$ and $1u$ is right special, the result follows. $\square$

Returning to the proof of Theorem 1, let us suppose that $p_{\omega,x}(n) = \epsilon(G_n)+1$ for each $n \geq 1$ and that x is not Sturmian. By the previous lemma there exist an integer $n \geq 2$, a Sturmian word y and a bispecial factor $u \in \{0,1\}^{n-2}$ of y such that $\mathrm{Fact}_x(n) = \mathrm{Fact}_y(n) \cup \{0u0, 1u1\}$. Since y is Sturmian, exactly one of $0u0$ and $1u1$ is a factor of y. Thus by the first part of Theorem 1 applied to the aperiodic word y, we deduce that $p_{\omega,x}(n) \geq p_{\omega,y}(n)+1 \geq \epsilon(G_n)+2$, a contradiction. This concludes our proof of Theorem 1. $\square$

We next establish various lemmas leading up to the proof of Theorem 2. As is well known, every finite Abelian group G can be written multiplicatively as a direct product of cyclic groups $\mathbb{Z}/m_1\mathbb{Z} \times \mathbb{Z}/m_2\mathbb{Z} \times \cdots \times \mathbb{Z}/m_k\mathbb{Z}$ where the m_i are prime powers. The unordered sequence $(m_1, m_2, \ldots, m_k)$ completely determines G up to isomorphism and any symmetric function of the m_i is an isomorphic invariant of G. We consider the *trace* of G given by $T(G) = m_1 + m_2 + \cdots + m_k$, and recall the following result from [14].

Proposition 2.5. *If an Abelian group G is embedded in S_n, then $T(G) \leq n$.*

A partition $\{E_1, E_2, \ldots, E_k\}$ of $\{1, 2, \ldots, n\}$ is called an *interval partition* if for each $1 \leq r < s \leq n$, we have $r, s \in E_i \Rightarrow t \in E_i$ for all $r \leq t \leq s$.

Lemma 2.6. *Let $\{E_1, E_2, \ldots, E_k\}$ be an interval partition of $\{1, 2, \ldots, n\}$ ordered so that $i < j \Rightarrow \max E_i < \max E_j$. For each $1 \leq j \leq k$, let $\sim_j$ denote the equivalence relation on $\mathbb{A}^n$ defined by $u \sim_j v$ if and only if $u|_{E_i} \sim_{ab} v|_{E_i}$ for each $1 \leq i \leq j$. Then for each Sturmian word $x \in \{0,1\}^{\mathbb{N}}$ we have Card(Fact$_x(n)/\sim_j) = j+1$ for each $1 \leq j \leq k$.*

Proof. Let $x \in \{0,1\}^{\mathbb{N}}$ be a Sturmian word. In view of Lemma 2.3 it suffices to show that $\mathrm{Card}(\mathrm{Fact}_x(n)/\sim_j) \leq j+1$ for each $1 \leq j \leq k$. Since x is Sturmian, there are exactly two Abelian classes of factors of x of each length, thus $\mathrm{Card}(\mathrm{Fact}_x(n)/\sim_1) = 2$. It also follows from this that for $1 \leq j \leq k-1$, each $\sim_j$ class splits into at most two $\sim_{j+1}$ classes. So it suffices to show that for each $1 \leq j \leq k-1$, at most one $\sim_j$ class splits under $\sim_{j+1}$. So fix $1 \leq j \leq k-1$, and suppose to the contrary that two distinct $\sim_j$ classes split under $\sim_{j+1}$. Then, there exist $u, u', v, v' \in \mathrm{Fact}_x(n)$ such that $u \sim_j u'$, $v \sim_j v'$, $u \nsim_j v$, $u \nsim_{j+1} u'$ and $v \nsim_{j+1} v'$. Exchanging if necessary u and u' and/or v and v', we may assume $u|_{E_{j+1}}$ and $v|_{E_{j+1}}$ are rich in 0 while $u'|_{E_{j+1}}$ and $v'|_{E_{j+1}}$ are rich in 1. Since $u \nsim_j v$, there exists a largest integer $1 \leq i \leq j$ such that $u|_{E_i} \nsim_{ab} v|_{E_i}$. Exchanging if necessary u and v and u' and v', we may assume that $u|_{E_i}$ is rich in 0 and $v|_{E_i}$ is rich in 1. Since $v|_{E_i} \sim_{ab} v'|_{E_i}$, we have that $u|_{E_i \cup \cdots \cup E_{j+1}}$ has two more occurrences of 0 than $v'|_{E_i \cup \cdots \cup E_{j+1}}$, contradicting that x is balanced. $\square$

In the next lemma we consider a discrete 3-interval exchange transformation (a, b, c) defined on the set $\{1, 2, \ldots, n\}$ (where $n = a + b + c$) in which the numbers $1, 2, \ldots, n$ are divided into three subintervals of length a, b and c respectively which are then rearranged in the order c, b, a. In other words

$$1, 2, \ldots, n \mapsto c + b + 1, c + b + 2, \ldots, n, c + 1, c + 2, \ldots, c + b, 1, 2, \ldots, c.$$

This bijection is also called an abc-permutation in [23]. We also include here the degenerate case in which one of a, b or c equals 0. The following lemma asserts that for each Sturmian word x and for each positive integer m, there exists an m-cycle corresponding to a discrete 3-interval exchange transformation which identifies all factors of x of length m belonging to the same Abelian class.

Lemma 2.7. *Let $x \in \{0, 1\}^{\mathbb{N}}$ be a Sturmian word. Then for each positive integer m there exists a discrete 3-interval exchange transformation (a, b, c) on $\{1, 2, \ldots, m\}$ given by an m-cycle σ such that $\sim_{\langle\sigma\rangle}$ is Abelian transitive on x.*

Proof. The result is immediate in case $m = 1, 2,$ or 3. In fact, in this case we may take $\sigma = \mathrm{id}$, $(1, 2)$, or $(1, 2, 3)$ respectively. Thus we assume $m \geq 4$. Let w and w' be two consecutive (in length) bispecial factors of x such that $|w'| + 2 < m \leq |w| + 2$. Let r and s denote the number of occurrences of 1 and 0 in $0w1$, i.e., $r = |0w1|_1$, $s = |0w1|_0$, so that $r+s = |w|+2$. It is known that r and s are coprime (see Proposition 2.1 in [4] or Proposition 2.1 in [5]). Set $p = r^{-1} \bmod (r + s)$ and $q = s^{-1} \bmod (r + s)$. Then it is readily verified that $p + q = r + s$. It is shown that p and q are coprime periods of the central Sturmian word w (see Lemma 4 in [10] or Theorem 2.2.11 and Proposition 2.2.12 in [20]). Set $a = m - q$, $b = p + q - m$, $c = m - p$ and let $\sigma \in S_m$ denote the corresponding abc-permutation. We note that $|w'| + 2 = \max\{p, q\}$ (see Lemma 4 in [10] or Corollary 2.2.10 in [20]), whence a and c are both positive while $b \geq 0$. Since $\gcd(a + b, b + c) = \gcd(p, q) = 1$, it follows from Lemma 1 of [23] that σ is an m-cycle.

Now let u and v be two lexicographically consecutive factors of x of length m with $u < v$. Assume further that u and v are in the same Abelian class. We will show that $v = \sigma * u$. We consider the lexicographic Christoffel array $\mathcal{C}_{r,s}$ in which the cyclic conjugates of $0w1$ are ordered lexicographically in a rectangular array (see [18]). For instance, if $w = 010010$, the corresponding Christoffel array $\mathcal{C}_{3,5}$ is shown in Figure 1. Let U and V be two lexicographically consecutive factors of x of length $|w|+2$ with u a prefix of U and v a prefix of V. We recall that U and V differ in exactly two consecutive positions, more precisely we can write $U = X01Y$ and $V = X10Y$ for some $X, Y \in \{0, 1\}^*$ (see Corollary 5.1 in [6]). Writing $U = ABCB'$ where $|A| = a$, $|B| = |B'| = b$ and $|C| = c$, by Theorem C in [18] we have that $V = CB'AB$. Since u and v are distinct and belong to the same Abelian class, we have that $X01$ is a prefix of $u = ABC$ which in turn implies that $B = B'$. Whence $U = ABCB$ and $V = CBAB$ and hence $u = ABC$ and $v = CBA$ and $v = \sigma * u$ as required. $\square$

As an immediate consequence of Lemma 2.7 we have

$$\mathcal{C}_{3,5} = \begin{pmatrix} 0&0&1&0&0&1&0&1\\ 0&0&1&0&1&0&0&1\\ 0&1&0&0&1&0&0&1\\ 0&1&0&0&1&0&1&0\\ 0&1&0&1&0&0&1&0\\ 1&0&0&1&0&0&1&0\\ 1&0&0&1&0&1&0&0\\ 1&0&1&0&0&1&0&0 \end{pmatrix}$$

Figure 1: The Christoffel array $\mathcal{C}_{3,5}$.

Corollary 2. *Let $x \in \{0,1\}^{\mathbb{N}}$ be a Sturmian word. Then for each positive integer n there exists a cyclic group G_n generated by an n-cycle such that $Card(Fact_x(n)/\sim_{G_n}) = 2$.*

In contrast, if we set $G_n = \langle(1,2,\ldots,n)\rangle$, then $\limsup_{n\to\infty} \mathrm{Card}(\mathrm{Fact}_x(n)/\sim_{G_n}) = +\infty$ (see Theorem 1 of [7]), while $\liminf_{n\to\infty} \mathrm{Card}(\mathrm{Fact}_x(n)/\sim_{G_n}) = 2$ (see Lemma 9 of [7]).
As another consequence of Lemma 2.7 we have:

Lemma 2.8. *Let $x \in \{0,1\}^{\mathbb{N}}$ be a Sturmian word. Let $\{E_1, E_2, \ldots, E_k\}$ be an interval partition of $\{1,2,\ldots,n\}$, and put $m_i = Card(E_i)$. For each $1 \le i \le k$, there exists an m_i-cycle $\sigma_i = (a_1, a_2, \ldots, a_{m_i})$ such that $E_i = \{a_1, a_2, \ldots, a_{m_i}\}$ and, if G denotes the subgroup of S_n generated by $\sigma_1, \sigma_2, \ldots, \sigma_k$, then for all factors $u, v \in Fact_x(n)$ we have $u \sim_G v$ if and only if $u|_{E_i} \sim_{ab} v|_{E_i}$ for each $1 \le i \le k$.*

Proof. By Lemma 2.7 we know that, for each i there exists a cycle $\sigma_i = (a_1, a_2, \ldots, a_{m_i})$ with $E_i = \{a_1, a_2, \ldots, a_{m_i}\}$, such that for all factors $u, v \in \mathrm{Fact}_x(n)$ we have $u|_{E_i} \sim_{\langle\sigma_i\rangle} v|_{E_i}$ if and only if $u|_{E_i} \sim_{ab} v|_{E_i}$. In fact, $\{u|_{E_i} : u \in \mathrm{Fact}_x(n)\} = \mathrm{Fact}_x(m_i)$. Moreover as the sets E_i are pairwise disjoint, the cycles σ_i are also pairwise disjoint. Hence the σ_i commute with one another. Thus, given $u, v \in \mathrm{Fact}_x(n)$, if $u \sim_G v$, then there exists $g = \sigma_1^{r_1} \cdots \sigma_k^{r_k} \in G$ such that $v = g * u$. However, for each $1 \le i \le k$ we have $(g * u)|_{E_i} = (\sigma_i^{r_i} * u)|_{E_i}$, hence $u|_{E_i} \sim_{ab} v|_{E_i}$. Conversely if $u|_{E_i} \sim_{ab} v|_{E_i}$ for each $1 \le i \le k$, there exists r_i such that $v|_{E_i} = (\sigma_i^{r_i} * u)|_{E_i}$. Hence setting $g = \sigma_1^{r_1} \cdots \sigma_k^{r_k} \in G$ we have $v = g * u$. □

We now prove Theorem 2.

Proof of Theorem 2. Let $x \in \{0,1\}^{\mathbb{N}}$ be a Sturmian word and let $(G_n)_{n\ge1}$ be a sequence of Abelian permutation groups. We show that for each $n \ge 1$ there exists a permutation group $G'_n \subseteq S_n$ isomorphic to G_n such that $\mathrm{Card}(\mathrm{Fact}_x(n)/\sim_{G'_n}) = \epsilon(G'_n) + 1$. Fix $n \ge 1$ and put $G = G_n$. By the fundamental theorem of finite Abelian groups, G is isomorphic to a direct product $\mathbb{Z}/m_1\mathbb{Z} \times \mathbb{Z}/m_2\mathbb{Z} \times \cdots \times \mathbb{Z}/m_k\mathbb{Z}$ where the m_i are prime powers. Let

$m = T(G) = m_1 + m_2 + \cdots + m_k$. By Proposition 2.5 we have $m \le n$. Thus, short of adding additional copies of the trivial cyclic group $\mathbb{Z}/1\mathbb{Z}$ or order 1, we may assume that $T(G) = n$. Let $E_1 = \{1, 2, \ldots, m_1\}$, $E_2 = \{m_1 + 1, \ldots, m_1 + m_2\}, \ldots, E_k = \{m_1 + \cdots m_{k-1} + 1, \ldots, n\}$. Then $\{E_1, E_2, \ldots, E_k\}$ is an interval partition of $\{1, 2, \ldots, n\}$. Pick cycles $\sigma_1, \sigma_2, \ldots, \sigma_k$ as in Lemma 2.8. Then the σ_i are pairwise disjoint (and hence commute with one another) and each σ_i is of order m_i. Hence, the subgroup G' of S_n generated by $\sigma_1, \sigma_2, \ldots, \sigma_k$ is isomorphic to G. Moreover, $E_1, E_2, \ldots, E_k$ is the full set of G'-orbits of $\{1, 2, \ldots n\}$ whence $\epsilon(G') = k$. Also by Lemma 2.8, for all $u, v \in \mathrm{Fact}_x(n)$ we have that $u \sim_{G'} v$ if and only if $u|_{E_i} \sim_{ab} v|_{E_i}$ for each $1 \le i \le k$. Thus the equivalence relation $\sim_{G'}$ on $\mathrm{Fact}_x(n)$ coincides with the equivalence relation $\sim_k$ given in Lemma 2.6. Thus by Lemma 2.6 we deduce that

$$\mathrm{Card}(\mathrm{Fact}_x(n)/\sim_{G'}) = \mathrm{Card}(\mathrm{Fact}_x(n)/\sim_k) = k + 1 = \epsilon(G') + 1$$

as required. This concludes our proof of Theorem 2. $\square$

As an immediate consequence of Theorem 2 and Cayley's theorem we have

Corollary 3. *Let G be an Abelian group of order n. Then for every Sturmian word x there exists a permutation group $G' \subseteq S_n$ isomorphic to G such that Card($Fact_x(n)/\sim_{G'}$) $=$ $\epsilon(G') + 1$.*

The following example illustrates that in Theorem 2, we cannot replace "isomorphic" by "conjugate". Let G be the cyclic subgroup of order 3 of S_6 generated by the permutation $\sigma = (1, 2, 3)(4, 5, 6)$. Then $\epsilon(G) = 2$. We will show that if x is the Fibonacci word, then

$$\mathrm{Card}\,(\mathrm{Fact}_x(6)/\sim_{G'}) \ge 4$$

for each subgroup G' of S_6 conjugate to G. To see this, let $G' \subseteq S_6$ be generated by the permutation $(a, b, c)(d, e, f)$ where $\{a, b, c, d, e, f\} = \{1, 2, 3, 4, 5, 6\}$. We claim that 100101 and 101001 belong to distinct equivalence classes under the action of G' on $\{0, 1\}^6$. In fact, suppose to the contrary that $g * 100101 = 101001$ for some $g \in G'$. Then $g(\{1, 4, 6\}) = \{1, 3, 6\}$ and $g(\{2, 3, 5\}) = \{2, 4, 5\}$. We claim that either $g(4) = 3$ or $g(3) = 4$. Otherwise, $g : 4 \mapsto x \mapsto y$ where $\{x, y\} = \{1, 6\}$. But $4 \notin g(\{1, 6\})$. Thus without loss of generality we can assume $g(4) = 3$. This means that $g(\{1, 6\}) = \{1, 6\}$, whence $g^2(1) = 1$, which implies that $g^2 = id$, a contradiction. Having established the claim, consider the induced equivalence relation $\sim_{G'}$ on the factors of length 6 of the Fibonacci word. One Abelian class is of size five $\{001001, 001010, 010010, 010100, 100100\}$ and the other of size two $\{100101, 101001\}$. Since $|G'| = 3$, there must be at least two distinct equivalence classes in the first Abelian class, and following the claim, two equivalence classes in the second. Thus at least 4 equivalence classes combined.

On the other hand:

Corollary 4. *Let $\sigma \in S_n$ and $G = \langle \sigma \rangle$. Writing $\sigma = \sigma_1 \cdots \sigma_k$ as a product of disjoint cycles, suppose $|\sigma_1|, \ldots, |\sigma_k|$ are pairwise relatively prime. Then for every Sturmian word x there exists $G' \subseteq S_n$ conjugate to G such that $Card(Fact_x(n)/\sim_{G'}) = \epsilon(G) + 1$.*

Proof. Since $|\sigma_1|, \ldots, |\sigma_k|$ are pairwise relatively prime, we have $G = \langle \sigma_1, \sigma_2, \ldots \sigma_k \rangle$. Adding if necessary additional σ_i of the form $\sigma_i = (a)$, we may assume that $\sum_{i=1}^{k} |\sigma_i| = n$. Let $\{E_1, E_2, \ldots, E_k\}$ be an interval partition of $\{1, 2, \ldots, n\}$ such that $\mathrm{Card}(E_i) = |\sigma_i|$. By Lemma 2.8, there exist disjoint cycles $\sigma'_1, \sigma'_2, \ldots, \sigma'_k$ such that $|\sigma_i| = |\sigma'_i|$ and, if G' denotes the subgroup of S_n generated by $\sigma'_1, \sigma'_2, \ldots, \sigma'_k$, then for all factors $u, v \in \mathrm{Fact}_x(n)$ we have $u \sim'_G v$ if and only if $u|_{E_i} \sim_{ab} v|_{E_i}$ for each $1 \leq i \leq k$. Thus G and G' are conjugate in S_n, and by Lemma 2.6 we have

$$\mathrm{Card}(\mathrm{Fact}_x(n)/\sim_{G'}) = \mathrm{Card}(\mathrm{Fact}_x(n)/\sim_k) = k + 1 = \epsilon(G') + 1 = \epsilon(G) + 1.$$

□

3. Further generalities and open questions

A first natural question, to which we do not know the answer, is whether Theorem 2 extends to sequences $\omega = (G_n)_{n \geq 1} \in \mathcal{G}$ in which the groups G_n are not necessarily Abelian. Our proof uses in an essential way that each G_n is a direct product of cyclic subgroups.

A second natural question concerns using this general framework to distinguish between two infinite words x and x' whose sets of factors are not word isomorphic. In general, each choice of $\omega = (G_n)_{n \geq 1} \in \mathcal{G}$ defines a unique complexity function which reflects some combinatorial property of a given infinite word. For instance, an infinite word x has bounded Abelian complexity if and only if x is k-balanced for some positive integer k (see Lemma 3 in [24]). In contrast, an infinite word x has bounded cyclic complexity if and only if x is ultimately periodic (see Theorem 1 in [7]). Given an infinite word $x = x_1 x_2 x_3 \cdots \in \mathbb{A}^{\mathbb{N}}$, let $\mathrm{Alph}(x) = \{x_n : n \geq 1\} \subseteq \mathbb{A}$. We say that two infinite words x and x' are *factor isomorphic* if there exists a bijection $\tau : \mathrm{Alph}(x') \to \mathrm{Alph}(x)$ such that x and $\tau(x')$ have exactly the same set of factors. Then given two non-factor isomorphic infinite words x and x', does there exist a sequence $\omega = (G_n)_{n \geq 1} \in \mathcal{G}$ of permutation groups which distinguishes them, i.e., for which $p_{\omega,x}(n) \neq p_{\omega,x'}(n)$ for some $n \geq 1$?

This question has an affirmative answer if one of the two words is Sturmian. In fact, Theorem 2 in [7] states that if x is Sturmian and x' is any infinite word whose cyclic complexity is equal to that of x, then up to renaming letters, x and x' have the same set of factors, i.e., are both Sturmian with the same slope. Thus each Sturmian subshift is completely characterised by the cyclic complexity of its set of factors.

Another instance in which this question admits an affirmative answer is in case x belongs to the subshift generated by the Thue-Morse infinite word $\mathbf{t} = 0110100110010110100 10 \cdots$

where the nth term of $\mathbf{t}$ (starting from $n = 0$) is defined as the sum modulo 2 of the digits in the binary expansion of n (see [26]). It is shown in [1] that if x' has the same factor complexity of the Thue-Morse infinite word, then either x' is in the subshift generated by $\mathbf{t}$ or in that generated by $\sigma(\mathbf{t})$ where σ is the letter doubling morphism $0 \mapsto 00$, $1 \mapsto 11$. However, if x' belongs to the subshift generated by $\sigma(\mathbf{t})$ then x' would contain the four factors $111, 110, 011, 000$ and hence the Abelian complexity of x and x' (for $n = 3$) would differ.

Let $\rho = 0010011000110110001001110 0\cdots$ be the regular paperfolding word given by the sequence of ridges and valleys obtained by unfolding a sheet of paper which has been folded in half infinitely many times in the same direction [2]. As in the case of the Thue-Morse word, ρ is a 2-automatic sequence. However, the paperfolding word is arbitrarily unbalanced while the Thue-Morse word is 2-balanced. Thus the Abelian complexity of ρ is unbounded while the Abelian complexity of $\mathbf{t}$ is bounded. As another example, consider the period doubling word $x = 01000101010001\cdots$ defined as the fixed point of the morphism $0 \mapsto 01$, $1 \mapsto 00$. Being a fixed point of a 2-uniform morphism, it is also 2-automatic. However, the limit infimum of the cyclic complexity of the period doubling word is equal to 2 (Example 1 in [7]) while for the Thue-Morse word it is unbounded (Proposition 23 in [7]).

It is likely that the subshift generated by $\mathbf{t}$ is completely characterised by the cyclic complexity, although we do not know how to show this. However, cyclic complexity alone does not in general distinguish between non-factor isomorphic words. For example, consider the periodic words $x = \tau((010011)^{\omega})$ and $x' = \tau((101100)^{\omega})$ where τ is the morphism: $0 \mapsto 010$, $1 \mapsto 011$. Then x and x' are not factor isomorphic yet have the same cyclic complexity. In fact, it is readily checked that x and x' have the same cyclic complexity up to $n \leq 17$. Since both words have period 18, it follows that the cyclic complexities of x and x' agree for all n.

Acknowledgments: We are very grateful to the two anonymous reviewers of the manuscript for their useful comments.

References

[1] A. Aberkane, S. Brlek, Suites de même complexité que celle de Thue–Morse, in Actes des Journées Montoises d'informatique théorique, Montpellier, France, (2002), 85–89.

[2] J.-P. Allouche, The number of factors in a paperfolding sequence, *Bull. Austral. Math. Soc.,* **46** (1992), 23–32.

[3] J.-P. Allouche, M. Baake, J. Cassaigne and D. Damanik, Palindrome complexity. Selected papers in honor of Jean Berstel. *Theoret. Comput. Sci.* **292** (2003), 9–31.

[4] J. Berstel, A. de Luca, Sturmian words, Lyndon words and trees. *Theoret. Comput. Sci.* **178** (1997), 171–203.

[5] V. Berthé, A. de Luca and C. Reutenauer, On an involution of Christoffel words and Sturmian morphisms. *European J. Combin.* **29** (2008), 535–553.

[6] J.-P. Borel and C. Reutenauer, On Christoffel classes. *RAIRO Theor. Inform. Appl.* **40** (2006), 15–27.

[7] J. Cassaigne, G. Fici, M. Sciortino and L.Q. Zamboni, Cyclic complexity of words. *J. Combin. Theory Ser. A* **145** (2017), 36–56.

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

[9] V. Cyr, B. Kra, Complexity of short rectangles and periodicity. *European J. Combin.* **52** (2016), 146–173.

[10] A. de Luca, Sturmian words: Structure, combinatorics and their arithmetics. *Theoret. Comput. Sci.* **183** (1997), 45–82.

[11] A. de Luca, F. Mignosi, Some combinatorial properties of Sturmian words. *Theoret. Comput. Sci.* **136** (1994), 361–385.

[12] F. Durand and M. Rigo, Multidimensional extension of the Morse-Hedlund theorem. *European J. Combin.* **34** (2013), 391–409.

[13] A. Ehrenfeucht, K.P. Lee, G. Rozenberg, Subword complexities of various deterministic developmental languages without interactions. *Theoret. Comput. Sci.* **1** (1975), 59–76.

[14] M. Hoffman, An invariant of finite abelian groups. *Amer. Math. Monthly* **94** (1987), 664–666.

[15] O. Jenkinson, Optimization and majorization of invariant measures. *Electronic Research Announcements of the American Mathematical Society* **13** (2007), 1–12.

[16] O. Jenkinson, Balanced words and majorization. *Discrete Mathematics Algorithms and Applications* **1** (2009), 463–483.

[17] O. Jenkinson and Vasso Anagnostopoulou, Which beta-shifts have a largest invariant measure? *J. London Math. Soc.* bf 79 (2009), 445–464.

[18] O. Jenkinson and L.Q. Zamboni, Characterisations of balanced words via orderings. *Theoret. Comput. Sci.* **310** (2004), 247–271.

[19] T. Kamae and L.Q. Zamboni, Sequence entropy and the maximal pattern complexity of infinite words. *Ergodic Theory Dynam. Systems* **22** (2002), 1191–1199.

[20] M. Lothaire, Algebraic Combinatorics on Words, Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge UK, 2002

[21] M. Morse and G. A. Hedlund, Symbolic dynamics. *Amer. J. Math.* **60** (1938), 815–866.

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

[23] I. Pak and A. Redlich, Long cycles in *abc*-permutations. *Funct. Anal. Other Math.* **2** (2008), 87–92.

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

[25] J. Sander and R. Tijdeman, The rectangle complexity of functions on two-dimensional lattices. *Theoret. Comput. Sci.* **270** (2002), 857–863.

[26] A. Thue, Über unendliche Zeichenreihen, Norske Vid. Selsk. Skr. I. Mat-Nat. Kl. 7 (1906), 1–22