On Growth and Fluctuation of k-Abelian Complexity*

Julien Cassaigne[1], Juhani Karhumäki[2], and Aleksi Saarela[2]

[1] Institut de Mathmatiques de Luminy, Case 907, F13288 Marseille Cedex 9, France
cassaigne@iml.univ-mrs.fr
[2] Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland, karhumak,amsaar@utu.fi

Abstract. An extension of abelian complexity, so called k-abelian complexity, has been considered recently in a number of articles. This paper considers two particular aspects of this extension: First, how much the complexity can increase when moving from a level k to the next one. Second, how much the complexity of a given word can fluctuate. For both questions we give optimal solutions.

1 Introduction

Counting the factors of fixed lengths provides a natural measure of complexity of infinite words. Doing that modulo some equivalence relation gives other variants of complexity. For example, abelian complexity counts the number of factors of length n which are commutatively pairwise inequivalent. As an extension of abelian equivalence, k-abelian equivalence can be defined. Two words u and v are k-abelian equivalent if they possess the same number of each factor of length k (and as a technical requirement, start with the same prefix of length $k-1$). This then leads to the definition of the k-abelian complexity function $\mathcal{P}_w^k$, which counts the number of equivalence classes of factors of w of length n.

Among the first questions asked about k-abelian equivalence was the question of avoidability of repetitions. As is well known, and proved already by Thue [19, 20], the smallest alphabets avoiding squares (resp. cubes) in infinite words are of size three (resp. two). For abelian repetitions the corresponding values are four and three, as shown by Keränen [12] and Dekking [4].

Do k-abelian repetitions behave like ordinary words or like abelian words? This question was raised in the Oberwolfach minisymposium *Combinatorics on Words* in August 2010, and written down in [8]. It turned out that with respect to squares 2-abelian repetitions behave like abelian repetitions: There are only finitely many words avoiding 2-abelian squares over a ternary alphabet. However, the longest such word is of length 537, see [8]. The problem of avoiding cubes was more challenging. Step by step, it was shown that k-abelian cubes could be avoided over a binary alphabet for smaller and smaller values of k, see [7, 14, 13]. Finally, Rao [18] showed that 2-abelian cubes can be avoided over a binary

* Supported by the Academy of Finland under grant 257857

alphabet, closing the problem. Hence, the avoidability of cubes is similar in the k-abelian case as in the conventional case! The same is true for k-abelian squares if $k \geq 3$: These are avoidable over a ternary alphabet, as proved in [18].

Another natural research area is factor complexity. How are factor complexity, abelian complexity and k-abelian complexity related? For factor complexity, two fundamental results are as follows. First, the smallest complexity achieved among aperiodic words is $n+1$, see [15, 16], which characterizes so-called Sturmian words. Second, there is a complexity gap from bounded complexity to the complexity of Sturmian words. In other words, if the complexity of a word is lower than the complexity of Sturmian words, then it is bounded by a constant. For abelian complexity, there also exists a minimal complexity for aperiodic words, namely the constant complexity 2. This follows from the results in [16], see also [3]. Again this characterizes Sturmian words (among aperiodic words), but there does not exist a similar complexity gap above bounded complexities. In other words, there are arbitrarily slowly growing but unbounded complexity functions.

For k-abelian complexity the situation is more challenging. It is shown in [10] that there exists a minimal complexity among the aperiodic words. This is given over binary words by the function $f(n) = \min(n+1, 2k)$, and again the Sturmian words are exactly those aperiodic words which reach this. On the other hand, no gap, whatsoever, exists above bounded complexities. Indeed for any monotonic unbounded function $g(n)$ there exists an infinite word of unbounded complexity such that its complexity is bounded by $g(n)$, for all large n, see [11].

We continue research on k-abelian complexity concentrating on the following two questions:

Question 1. How much higher can the $(k+1)$-abelian complexity of an infinite word be compared to its k-abelian complexity? In particular, if the latter is bounded, how large can the former be?

As shown in [11], this question is motivated by the properties of the Thue–Morse word, whose abelian complexity is bounded by a constant (in fact, it takes only the values 2 and 3), while its 2-abelian complexity is unbounded, fluctuating between an upper limit of $O(\log n)$ and a lower limit of $\Omega(1)$. The 2-abelian complexity of the Thue-Morse word is also known to be 2-regular, see [5] and [17].

Actually, we can find much bigger fluctuations. Let $\mathrm{Max}_{m,k}(n)$ be the function which gives the number of k-abelian equivalence classes over m-letter alphabet. Then we can find an infinite word w such that its k-abelian complexity is bounded but its $(k+1)$-abelian complexity is $\Theta(\mathrm{Max}_{m,k+1}(n)/\mathrm{Max}_{m,k}(n))$.

Our other question asks about the fluctuation of the k-abelian complexity of a given word. As we already said, for the Thue–Morse word 2-abelian complexity, or in fact also k-abelian complexity, for $k \geq 2$, takes a constant value infinitely often, and infinitely often a value of order $\log n$. Hence its complexity values fluctuate from $O(1)$ to $\Omega(\log n)$. For ordinary factor complexity, the fluctuation can be very high, see Theorem 9 in [1].

Question 2. How much can the k-abelian complexity of a word fluctuate?

We are able to give an exhaustive answer to this question. Our results are as follows. Let $g(n) = o(\mathrm{Max}_{m,k}(n))$. We can construct words w_1 and w_2 such that their k-abelian complexity functions $\mathcal{P}^k_{w_1}$ and $\mathcal{P}^k_{w_2}$ satisfy

$$\mathcal{P}^k_{w_1}(a_n) = \Omega(g(a_n)), \qquad \mathcal{P}^k_{w_1}(b_n) = O(1)$$

and

$$\mathcal{P}^k_{w_2}(c_n) = \Omega(\mathrm{Max}_{m,k}(c_n)), \qquad \mathcal{P}^k_{w_2}(d_n) = O(d_n)$$

for infinite strictly increasing sequences $a_1, a_2, a_3, \dots$, $b_1, b_2, b_3, \dots$, $c_1, c_2, c_3, \dots$ and $d_1, d_2, d_3, \dots$. Moreover, we show that the above $g(n)$ cannot be chosen from $\Omega(\mathrm{Max}_{m,k}(n))$, and $O(d_n)$ cannot be replaced with $o(d_n)$. In other words, we show that the fluctuation can go from minimal to almost maximal, or from maximal to almost minimal, but cannot go all the way from minimal to maximal.

A brief summary of this paper is as follows. In Section 3 we show that k-abelian equivalence classes are actually defined by a suitably chosen subset of factors. This auxiliary lemma turns out to be very useful. Section 3 contains also another independent lemma which relates abelian equivalence of words to k-abelian equivalence of their much longer morphic images. With these lemmata, and some simple observations made on k-abelian equivalence in Section 4, we move to the main considerations of this paper. In Section 5 we deal with Question 1 and Section 6 contains results on Question 2. Some proofs have been omitted because of space constraints, but they can be found in the full version of this paper.

2 Preliminaries

For $m \geq 1$, let $\Sigma_m = \{0, 1, \dots, m-1\}$ be an alphabet of m letters. The empty word is denoted by ε. For $n \geq 0$ and a word u, let $\mathrm{pref}_n(u)$ be the prefix of u of length n and let $\mathrm{suff}_n(u)$ be the suffix of u of length n. If $n > |u|$, it is convenient to define $\mathrm{pref}_n(u) = \mathrm{suff}_n(u) = u$. For words u and v, we define $\delta(u, v) = 1$ if $u = v$ and $\delta(u, v) = 0$ if $u \neq v$.

The set of positive integers is denoted by $\mathbb{N}_{\geq 1}$. For functions $f, g : \mathbb{N}_{\geq 1} \to \mathbb{R}$, we use the usual definitions for $O(g(n)), \Omega(g(n)), \Theta(g(n)), o(g(n))$, and the following definitions that might be less common:

- $f(n) = O'(g(n))$ if $\exists \alpha > 0$ such that $f(n) < \alpha g(n)$ for infinitely many n.
- $f(n) = \Omega'(g(n))$ if $\exists \alpha > 0$ such that $f(n) > \alpha g(n)$ for infinitely many n.

For $k \geq 1$, words u and v are *k-abelian equivalent* if $|u|_t = |v|_t$ for all words t such that $|t| \leq k$ ($|u|_\varepsilon$ is defined to be $|u|+1$). Equivalently, u and v are k-abelian equivalent if $\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v)$, $\mathrm{suff}_{k-1}(u) = \mathrm{suff}_{k-1}(v)$, and $|u|_t = |v|_t$ for all words t such that $|t| = k$. The equivalence of these definitions, together with many other properties of the k-abelian equivalence, is proved in [10]. The k-abelian equivalence class of u is denoted by $[u]_k$.

For $n \geq 0$ and an infinite word w, let $F_n(w)$ be the set of factors of w of length n. The *factor complexity* of w is the function

$$\mathcal{P}_w : \mathbb{N}_{\geq 1} \to \mathbb{N}_{\geq 1}, \mathcal{P}_w(n) = \# F_n(w).$$

For $k \geq 1$, the *k-abelian complexity* of w is the function

$$\mathcal{P}_w^k : \mathbb{N}_{\geq 1} \to \mathbb{N}_{\geq 1}, \mathcal{P}_w^k(n) = \#\{[u]_k \mid u \in F_n(w)\}.$$

Now we give some background for the results in this article. Generalizations of the results of Morse and Hedlund form a starting point for our considerations. The well-known theorem of Morse and Hedlund [15] can be stated as follows.

Theorem 3. *If $\mathcal{P}_w(n) < n+1$ for some n, then w is ultimately periodic. If w is ultimately periodic, then $\mathcal{P}_w$ is bounded.*

This was generalized for k-abelian complexity in [10].

Theorem 4. *If $\mathcal{P}_w^k(n) < \min(2k, n+1)$ for some n, then w is ultimately periodic. If w is ultimately periodic, then $\mathcal{P}_w^k$ is bounded.*

A particular consequence of the theorem of Morse and Hedlund is that there is a gap between bounded complexity and complexity $n+1$. For k-abelian complexity there is no such gap above bounded complexity; this was proved in [11].

There are many equivalent ways to define *Sturmian words*. We give three such definitions (here $k \geq 2$):

- w is Sturmian if $\mathcal{P}_w(n) = n+1$ for all n.
- w is Sturmian if $\mathcal{P}_w^1(n) = 2$ for all n and w is aperiodic.
- w is Sturmian if $\mathcal{P}_w^k(n) = \min(2k, n+1)$ for all n and w is aperiodic.

The first two characterizations were proved in [16] and the third one in [10].

3 Characterizing an Equivalence Class

From now on, we assume that $m \geq 2$ is fixed. We mostly study words over the alphabet Σ_m. We ignore the unary case $m = 1$, although many of the theorems would trivially work also in this case.

The k-abelian equivalence class of a word $u \in \Sigma_m^*$ is determined by the numbers $|u|_x$, $x \in \bigcup_{i=0}^{k} \Sigma_m^i$, or equivalently by the words $\mathrm{pref}_{k-1}(u)$ and $\mathrm{suff}_{k-1}(u)$ and the numbers $|u|_x$, $x \in \Sigma_m^k$. However, both these characterizations contain a lot of redundant information. For example, if $m = 2$ and $\mathrm{pref}_1(u) = \mathrm{suff}_1(u)$, then $|u|_{01} = |u|_{10}$. In this section we give a set Y_k of minimal size such that the equivalence class of every u is determined by the words $\mathrm{pref}_{k-1}(u)$ and $\mathrm{suff}_{k-1}(v)$ and the numbers $|u|_y$, $y \in Y_k$. If it were possible to replace Y_k by a smaller set, it would easily lead to an upper bound for the number of equivalence classes that would contradict Theorem 8.

For $n \geq 0$, let

$$X_n = (\Sigma_m^n \setminus 0\Sigma_m^*) \setminus \Sigma_m^* 0 \qquad \text{and} \qquad Y_n = \bigcup_{i=0}^{n} X_i.$$

In other words, X_n is the set of words of length n that do not begin with 0 and do not end with 0, and Y_n is the set of words of length at most n that do not begin with 0 and do not end with 0. These sets will be used in many proofs in this paper. The sizes of these sets are

$$\#X_n = \begin{cases} 1 & \text{if } n = 0, \\ m-1 & \text{if } n = 1, \\ (m-1)^2 m^{n-2} & \text{if } n \geq 2, \end{cases} \qquad \#Y_n = \begin{cases} 1 & \text{if } n = 0, \\ (m-1)m^{n-1} + 1 & \text{if } n \geq 1. \end{cases}$$

The following theorem gives another equivalent definition for k-abelian equivalence, that is extensively used in this paper

Theorem 5. *Let $k \geq 1$ and $u, v \in \Sigma_m^*$. If $\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v)$, $\mathrm{suff}_{k-1}(u) = \mathrm{suff}_{k-1}(v)$, and $|u|_y = |v|_y$ for all $y \in Y_k$, then u and v are k-abelian equivalent.*

Proof. We prove that $|u|_t = |v|_t$ for all $t \in \Sigma_m^k$. The proof is by induction on k. The case $k = 1$ is easy. Let $k \geq 2$. We already know that $|u|_t = |v|_t$ for $t \in X_k$, so we have to consider the two cases $t = 0rb$, $r \in \Sigma_m^{k-2}$, $b \in \Sigma_m \setminus \{0\}$, and $t = s0$, $s \in \Sigma_m^{k-1}$.

For all $r \in \Sigma_m^{k-2}$ and $b \in \Sigma_m \setminus \{0\}$,

$$|u|_{rb} = \sum_{a \in \Sigma_m} |u|_{arb} + \delta(rb, \mathrm{pref}_{k-1}(u)).$$

It follows that

$$|u|_{0rb} = |u|_{rb} - \sum_{a \in \Sigma_m, a \neq 0} |u|_{arb} - \delta(rb, \mathrm{pref}_{k-1}(u))$$

and a similar equation holds for v. It follows from the assumptions of the theorem and the induction hypothesis that the right-hand side remains the same if every u is replaced by v. Thus $|u|_{0rb} = |v|_{0rb}$. For $s \in \Sigma_m^{k-1}$, the equality $|u|_{s0} = |v|_{s0}$ can be proved in a similar way. ☐

Example 6. Consider the case $m = 2$. Then $Y_2 = \{\varepsilon, 1, 11\}$. Words $u, v \in \Sigma_m^*$ are 2-abelian equivalent if and only if

$$\mathrm{pref}_1(u) = \mathrm{pref}_1(v),\ \mathrm{suff}_1(u) = \mathrm{suff}_1(v),\ |u|_\varepsilon = |v|_\varepsilon,\ |u|_1 = |v|_1,\ |u|_{11} = |v|_{11}.$$

We get the following formulas:

$$|u|_0 = |u|_\varepsilon - |u|_1 - 1 = |u| - |u|_1, \qquad |u|_{01} = |u|_1 - |u|_{11} - \delta(1, \mathrm{pref}_1(u)),$$
$$|u|_{10} = |u|_1 - |u|_{11} - \delta(1, \mathrm{suff}_1(u)), \qquad |u|_{00} = |u|_0 - |u|_{01} - \delta(0, \mathrm{suff}_1(u)).$$

Sometimes we are studying factors of length n of an infinite word that does not contain 11 as a factor. If u, v are such factors, then they are 2-abelian equivalent if and only if

$$\mathrm{pref}_1(u) = \mathrm{pref}_1(v),\ \mathrm{suff}_1(u) = \mathrm{suff}_1(v),\ |u|_1 = |v|_1.$$

The construction in the following lemma is essential for our results. It will be used to relate the abelian complexity of a word to the k-abelian complexity of its image under a certain morphism.

Lemma 7. *Let $k \geq 1$, $M = (m-1)m^{k-1}+1$, and $y_0, \dots, y_{M-1}$ be the elements of the set Y_k. Let $h : \Sigma_M^* \to \Sigma_m^*$ be the morphism defined by*

$$h(i) = y_i 0^{2k-1-|y_i|} \quad \text{for } i \in \{0, \dots, M-1\}.$$

If $u, v \in \Sigma_M^+$, then $h(u)$ and $h(v)$ are k-abelian equivalent if and only if u and v are abelian equivalent and $\mathrm{pref}_{k-1}(h(u)) = \mathrm{pref}_{k-1}(h(v))$.

Proof. If u and v are abelian equivalent and $\mathrm{pref}_{k-1}(h(u)) = \mathrm{pref}_{k-1}(h(v))$, then

$$\mathrm{suff}_{k-1}(h(u)) = 0^{k-1} = \mathrm{suff}_{k-1}(h(v)), \qquad |h(u)|_\varepsilon = |h(v)|_\varepsilon, \qquad \text{and}$$
$$|h(u)|_y = \sum_{i=0}^{M-1} |u|_i |y_i|_y = \sum_{i=0}^{M-1} |v|_i |y_i|_y = |h(v)|_y$$

for all $y \in Y_k \setminus \{\varepsilon\}$, so $h(u)$ and $h(v)$ are k-abelian equivalent.

If $\mathrm{pref}_{k-1}(h(u)) \neq \mathrm{pref}_{k-1}(h(v))$, then $h(u)$ and $h(v)$ are not k-abelian equivalent. If u and v are not abelian equivalent, then let $|y_i| \leq |y_{i+1}|$ for all $i \in \{0, \dots, M-2\}$, let j be the largest index such that $|u|_j \neq |v|_j$, and let $y = y_j$. Then $j > 0$, $|y_i|_y = 0$ for $i < j$, and $|y_j|_y = 1$, so

$$\begin{aligned}|h(u)|_y &= \sum_{i=0}^{M-1} |u|_i |y_i|_y = |u|_j + \sum_{i=j+1}^{M-1} |u|_i |y_i|_y \\ &\neq |v|_j + \sum_{i=j+1}^{M-1} |u|_i |y_i|_y = |v|_j + \sum_{i=j+1}^{M-1} |v|_i |y_i|_y = \sum_{i=0}^{M-1} |v|_i |y_i|_y = |h(v)|_y.\end{aligned}$$

Thus $h(u)$ and $h(v)$ are not k-abelian equivalent. □

4 Lemmas About k-Abelian Equivalence

It was proved in [10] that if m and k are fixed, then the number of k-abelian equivalence classes of words in Σ_m^n is $\Theta(n^{(m-1)m^{k-1}})$. Here, and also later in this article, the hidden constants in the Θ-notation can depend on the parameters m and k. A shorter proof could be obtained in a fairly straightforward way by using Theorem 5 and Lemma 7. The exact numbers of k-abelian equivalence classes of words in Σ_m^n were calculated in [6] for small values of k, m, n.

Theorem 8. *Let $k \geq 1$. The number of k-abelian equivalence classes of words in Σ_m^n is $\Theta(n^{(m-1)m^{k-1}})$.*

Every k-abelian equivalence class is a disjoint union of $(k+1)$-abelian equivalence classes. In other words, for every word u there is a number r and words $u_1, \dots, u_r$ such that

$$[u]_k = [u_1]_{k+1} \cup \cdots \cup [u_r]_{k+1} \tag{1}$$

and $[u_i]_{k+1} \neq [u_j]_{k+1}$ for all $i \neq j$. For some words u, the number r of equivalence classes in the union is one (for example, if u is unary or shorter than $2k$), but usually it is much larger. Because the number of k-abelian equivalence classes of words in Σ_m^n is $\Theta(n^{(m-1)m^{k-1}})$, it follows immediately that there are words $u \in \Sigma_m^n$ such that the number r in (1), interpreted as a function of n, is lower bounded by a function that is in

$$\Theta\left(\frac{n^{(m-1)m^k}}{n^{(m-1)m^{k-1}}}\right) = \Theta(n^{(m-1)^2 m^{k-1}}).$$

The next theorem proves that the value $n^{(m-1)^2 m^{k-1}}$ should only be multiplied by an alphabet-dependent constant to get an upper bound for the number r in (1).

Theorem 9. *Let $k, n \geq 1$ and $u \in \Sigma_m^n$. The number of $(k+1)$-abelian equivalence classes contained in $[u]_k$ is at most $m^2 n^{(m-1)^2 m^{k-1}}$.*

Proof. By Theorem 5, the $(k+1)$-abelian equivalence class of $v \in [u]_k$ is characterized by $\mathrm{pref}_k(v)$, $\mathrm{suff}_k(v)$, and $|v|_y$ for $y \in Y_{k+1}$. Because $\mathrm{pref}_{k-1}(v) = \mathrm{pref}_{k-1}(u)$ and $\mathrm{suff}_{k-1}(v) = \mathrm{suff}_{k-1}(u)$, there are at most m possible values for each of $\mathrm{pref}_k(v)$ and $\mathrm{suff}_k(v)$. Because $|v|_y = |u|_y$ for all $y \in Y_k$, there is one possible value for every $|v|_y$, $y \in Y_k$. There are at most n possible values for every $|u|_x$, $x \in Y_{k+1} \setminus Y_k = X_{k+1}$. Multiplying these numbers gives the claimed bound, because there are $(m-1)^2 m^{k-1}$ different words $x \in X_{k+1}$. □

We end this section by stating two lemmas about k-abelian complexity. The proof of the first one has been omitted to save space, but it is quite easy and can be found in the full version of this article.

Often it is easier to estimate the k-abelian complexity of a word for some particular values of n than for all n. In general, this is not sufficient for determining the growth rate of the complexity: If there is a strictly increasing sequence of positive integers $n_1, n_2, n_3, \dots$ such that $\mathcal{P}_w^k(n_i) = \Theta(f(n_i))$, then it does not necessarily follow that $\mathcal{P}_w^k(n) = \Theta(f(n))$, even if the function f is reasonably well-behaving. This is discussed in Section 6. However, if $n_{i+1} - n_i$ is bounded, then we have the following lemma.

Lemma 10. *Let $k \geq 1$ and $w \in \Sigma_m^\omega$. Let $n_1, n_2, n_3, \dots$ be a strictly increasing sequence of positive integers such that the difference $n_{i+1} - n_i$ is bounded from above by a constant. Let $f : \mathbb{N}_{\geq 1} \to \mathbb{R}$ be a function such that $f(n)/f(n+1) = O(1)$.*

– *If* $\mathcal{P}_w^k(n_i) = O(f(n_i))$, *then* $\mathcal{P}_w^k(n) = O(f(n))$.
– *If* $\mathcal{P}_w^k(n_i) = \Omega(f(n_i))$, *then* $\mathcal{P}_w^k(n) = \Omega(f(n))$.

If a construction works for abelian complexity on all alphabets, then it can often be generalized for k-abelian complexities by the following lemma.

Lemma 11. *Let* $k \geq 2$, $M = (m-1)m^{k-1} + 1$, *and* $W \in \Sigma_M^\omega$. *There exists a word* $w \in \Sigma_m^\omega$ *such that* $\mathcal{P}_w^k(n) = \Theta(\mathcal{P}_W^1(n/(2k-1)))$ *for* n *divisible by* $2k-1$.

Proof. We can let h be the morphism in Lemma 7 and $w = h(W)$. Let $n = (2k-1)n'$.

If $U_1, \dots, U_N \in F_{n'}(W)$ and no two of them are abelian equivalent, then

$$h(U_1), \dots, h(U_N) \in F_n(w)$$

and no two of them are k-abelian equivalent by Lemma 7. Thus $\mathcal{P}_w^k(n) \geq \mathcal{P}_W^1(n')$.

On the other hand, if u is a factor of w, then there are $p, q \in \Sigma_m^*$ and $U \in F_{n'-1}(W)$ such that $u = ph(U)q$ and $|pq| = 2k-1$. By Lemma 7, the k-abelian equivalence class of u depends only on p, q, $\mathrm{pref}_{k-1}(h(U))$, and the abelian equivalence class of U. The number of different possibilities for p, q, and $\mathrm{pref}_{k-1}(h(U))$ does not depend on n', while the number of different possibilities for the abelian equivalence class of U is $\mathcal{P}_W^1(n'-1) = \Theta(\mathcal{P}_W^1(n'))$. Thus $\mathcal{P}_w^k(n) = O(\mathcal{P}_W^1(n'))$. □

5 k-Abelian Complexities for Different k

In this section we study the relations between the functions $\mathcal{P}_w^1, \mathcal{P}_w^2, \mathcal{P}_w^3, \dots$. Bounds for the ratio $\mathcal{P}_w^{k+1}(n)/\mathcal{P}_w^k(n)$ follow directly from Theorem 9.

Theorem 12. *Let* $k, n \geq 1$ *and* $w \in \Sigma_m^\omega$. *Then*

$$1 \leq \frac{\mathcal{P}_w^{k+1}(n)}{\mathcal{P}_w^k(n)} \leq m^2 n^{(m-1)^2 m^{k-1}}.$$

The bounds of Theorem 12 are optimal up to a constant. In fact, there are infinite words w such that

$$\mathcal{P}_w^{k+1}(n)/\mathcal{P}_w^k(n) = O(1) \tag{2}$$

for all k (for example, ultimately periodic words and Sturmian words). There are also infinite words w such that

$$\mathcal{P}_w^{k+1}(n)/\mathcal{P}_w^k(n) = \Theta(n^{(m-1)^2 m^{k-1}}) \tag{3}$$

for all k (words w that have every word in Σ_m^* as a factor satisfy (3)).

It is also possible to construct infinite words w such that for some k we have (2) and for some k we have (3). In fact, if we are considering only a finite number of different values of k, then the growth rates of the ratios $\mathcal{P}_w^{k+1}(n)/\mathcal{P}_w^k(n)$ can be chosen quite freely and independently of each other. This is made precise in the following theorem.

Theorem 13. *Let $K \geq 1$ and $0 \leq N_1 \leq m-1$ and $0 \leq N_k \leq (m-1)^2 m^{k-2}$ for $k \in \{2, \dots, K\}$. There exists $w \in \Sigma_m^\omega$ such that*

$$\mathcal{P}_w^k(n) = \Theta(n^{N_1 + \dots + N_k}) \qquad \text{for } k \in \{1, \dots, K\}.$$

Proof. Let Z be a subset of Y_K that contains ε and exactly N_k elements of X_k for $k \in \{1, \dots, K\}$. Let $M_k = N_1 + \dots + N_k + 1$ for all k, $M = M_K$, and $Z = \{z_0, \dots, z_{M-1}\}$. We can assume that $z_0 = \varepsilon$ and $|z_i| \leq |z_{i+1}|$ for all i. For $i \in \{1, \dots, M-1\}$, let

$$u_i = \begin{cases} 0^{5K-5} & \text{if } z_i = a,\ a \in \Sigma_m, \\ 0^{K-1} a s 0^{K-1} s b 0^{K-1+2(K-|z_i|)} & \text{if } z_i = asb,\ a, b \in \Sigma_m, \end{cases}$$

$$v_i = \begin{cases} 0^{K-1} a 0^{4K-5} & \text{if } z_i = a,\ a \in \Sigma_m, \\ 0^{K-1} a s b 0^{K-1} s 0^{K-1+2(K-|z_i|)} & \text{if } z_i = asb,\ a, b \in \Sigma_m. \end{cases}$$

Let $L = (M-1)(5K-5)$ and let $h : \Sigma_M^* \to \Sigma_m^*$ be the L-uniform morphism defined by

$$h(0) = \prod_{i=1}^{M-1} u_i \qquad \text{and} \qquad h(j) = \prod_{i=1}^{j-1} u_i \cdot v_j \cdot \prod_{i=j+1}^{M-1} u_i \quad (1 \leq j \leq M-1).$$

Let $W \in \Sigma_M^\omega$ be an infinite word that has a factor in every abelian equivalence class. We can show that we can take $w = h(W)$.

First we make some observations about the words u_i, v_i and the morphism h. If $1 \leq i \leq M-1$ and $y \in Y_K$, then $|v_i|_y - |u_i|_y = \delta(y, z_i)$. If $U \in \Sigma_M^n$ and $y \in Y_K \setminus \{\varepsilon\}$, then

$$|h(U)|_y = \sum_{i=0}^{M-1} ((n - |U|_i)|u_i|_y + |U|_i |v_i|_y) = \sum_{i=0}^{M-1} n|u_i|_y + \begin{cases} |U|_j & \text{if } y = z_j, \\ 0 & \text{if } y \notin Z. \end{cases} \tag{4}$$

For $U, V \in \Sigma_M^n$ and $k \in \{1, \dots, K\}$, $h(U)$ and $h(V)$ are k-abelian equivalent if and only if $|U|_j = |V|_j$ for all $j \in \{1, \dots, M_k - 1\}$. This follows from (4), Theorem 5, and the fact that $h(U)$ and $h(V)$ begin and end with 0^{k-1} and have the same length.

For the rest of the proof, let $k \in \{1, \dots, K\}$ be fixed. If $U_1, \dots, U_j \in F_n(W) \cap \Sigma_{M_k}^n$ and no two of them are abelian equivalent, then $h(U_1), \dots, h(U_j) \in F_{Ln}(w)$ and no two of them are k-abelian equivalent. We assumed that W has a factor in every abelian equivalence class, and the number of classes of words of length n is $\Theta(n^{M_k - 1})$, so we can assume that $j = \Theta(n^{M_k - 1})$. Thus $\mathcal{P}_w^k(Ln) = \Omega(n^{M_k - 1})$.

On the other hand, if u is a factor of w of length Ln, then there are $p, q \in \Sigma_m^*$ and $U \in F_{n-1}(W)$ such that $u = ph(U)q$ and $|pq| = L$. The k-abelian equivalence class of u depends only on p, q, and the numbers $|U|_i$ for $i \in \{1, \dots, M_k - 1\}$. The number of different possibilities for the pair (p, q) is at most $(L+1)m^L = O(1)$, while the number of different possibilities for each $|U|_i$ is n. Multiplying these numbers gives the upper bound $\mathcal{P}_w^k(Ln) = O(n^{M_k - 1})$.

We have proved $\mathcal{P}_w^k(Ln) = \Theta(n^{M_k - 1})$. The claim follows from Lemma 10. $\square$

The answer to Question 1 is given by Theorem 12 and the following special case of Theorem 13.

Corollary 14. *Let $k \geq 2$. There exists $w \in \Sigma_m^\omega$ such that*

$$\mathcal{P}_w^{k-1}(n) = O(1) \qquad \text{and} \qquad \mathcal{P}_w^k(n) = \Theta(n^{(m-1)^2 m^{k-2}}).$$

Theorem 13 cannot be generalized to the case where infinitely many k's are considered at the same time. For example, (3) holds either for all values of k or for only finitely many values k. This follows from the next theorem.

Theorem 15. *If $z \in \Sigma_m^+$ is not a factor of $w \in \Sigma_m^\omega$, then*

$$\frac{\mathcal{P}_w^{k+1}(n)}{\mathcal{P}_w^k(n)} = O(n^{(m-1)^2 m^{k-1} - (m-1)m^{k-|z|}}) = o(n^{(m-1)^2 m^{k-1}})$$

for all $k \geq |z|$.

Proof. We can assume that the first letter of z is not 0. Let $u \in F_n(w)$. By Theorem 5, the $(k+1)$-abelian equivalence class of $v \in [u]_k \cap F_n(w)$ is characterized by $\mathrm{pref}_k(v)$, $\mathrm{suff}_k(v)$, and $|v|_y$ for $y \in Y_{k+1}$. The number of possible values for $\mathrm{pref}_k(v)$ and $\mathrm{suff}_k(v)$ is at most $m^{k-1} = O(1)$. Because $|v|_y = |u|_y$ for all $y \in Y_k$, there is one possible value for every $|v|_y$, $y \in Y_k$. There are at most n possible values for every $|v|_x$, $x \in Y_{k+1} \setminus Y_k = X_{k+1}$. However, if $x \in z\Sigma_m^{k-|z|}(\Sigma_m \setminus \{0\})$, then $|v|_x = 0$, and the number of these words x is $(m-1)m^{k-|z|}$. Thus we get the upper bound

$$\mathcal{P}_w^{k+1}(n)/\mathcal{P}_w^k(n) = O(n^{(m-1)^2 m^{k-1} - (m-1)m^{k-|z|}}). \qquad \square$$

6 Fluctuating Complexity

In [11], words w were given such that $\liminf \mathcal{P}_w^k < \infty$ and $\mathcal{P}_w^k(n) = \Omega'(\log n)$. For example, the Thue–Morse word has this property for $k \geq 2$. Thus the numbers $\mathcal{P}_w^k(n)$ can fluctuate between bounded and logarithmic values. In this section, we study how big these kinds of fluctuations can be. We give an "optimal" answer to Question 2. More specifically, we consider three questions:

1. If $\mathcal{P}_w^k$ is unbounded, then how small can $\liminf \mathcal{P}_w^k$ be?
2. If $\mathcal{P}_w^k = O'(1)$, then for how fast-growing functions f can we have $\mathcal{P}_w^k(n) = \Omega'(f(n))$?
3. If $\mathcal{P}_w^k = \Omega'(n^{(m-1)m^{k-1}})$, then for how slowly growing functions f can we have $\mathcal{P}_w^k(n) = O'(f(n))$?

Recall that the number of k-abelian equivalence classes of words in Σ_m^n is $\Theta(n^{(m-1)m^{k-1}})$, so $\mathcal{P}_w^k(n) = O(n^{(m-1)m^{k-1}})$ for all words $w \in \Sigma_m^\omega$.

For the first question, it was proved in [10] that if $\liminf \mathcal{P}_w^k < 2k$, then w is ultimately periodic and thus $\mathcal{P}_w^k$ is bounded. We prove in Theorem 16 that it is possible to have $\liminf \mathcal{P}_w^k = 2k$ but $\mathcal{P}_w^k$ unbounded. The constructed word is a morphic image of the period-doubling word. In [10] it was proved that an aperiodic word w is Sturmian if and only if $\mathcal{P}_w^k(n) = 2k$ for all $n \geq 2k-1$. A consequence of our result is that having $\mathcal{P}_w^k(n) = 2k$ for infinitely many n is not sufficient to guarantee that w is Sturmian, or even that $\mathcal{P}_w^k(n)$ is bounded.

For the second question, we prove in Theorems 17 and 19 that we can take any $f = o(n^{(m-1)m^{k-1}})$, but not $f = \Omega'(n^{(m-1)m^{k-1}})$. Here a Toeplitz-type construction is used. For Toeplitz words, see, e.g., [9] and [2].

For the third question, we prove in Theorems 18 and 19 that we can take $f(n) = n$, but not $f = o(n)$.

Theorem 16. *Let $k \geq 1$. There exists $w \in \Sigma_2^\omega$ such that*

$$\liminf \mathcal{P}_w^k = 2k \qquad \textit{and} \qquad \mathcal{P}_w^k(n) = \Omega'(\log n).$$

Proof. It was proved in [11] that the period-doubling word $S \in \Sigma_2^\omega$, defined as the fixed point of the morphism $0 \mapsto 01, 1 \mapsto 00$, satisfies the requirements for $k = 1$. For $k \geq 2$, we cannot use Lemma 11, because we want to prove $\liminf \mathcal{P}_w^k = 2k$ and not just $\liminf \mathcal{P}_w^k < \infty$. Instead, we prove that we can take $w = h(S)$, where $h : \Sigma_2^* \to \Sigma_2^*$ is the morphism defined by $h(0) = 0^{k-1}1$ and $h(1) = 0^k 1$. No factor of w of length k contains two 1's, so it follows from Theorem 5 that factors u and v of w are k-abelian equivalent if and only if $\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v), \mathrm{suff}_{k-1}(u) = \mathrm{suff}_{k-1}(v)$, and $|u|_1 = |v|_1$. In particular, this means that $\mathcal{P}_w^k(n) = \Theta(\mathcal{P}_w^1(n))$.

First we prove that $\liminf \mathcal{P}_w^k = 2k$. It was proved in [11] that for all l, $\mathcal{P}_S^1(2^l) = 2$, so there is a number n_l such that every factor of S of length 2^l has either n_l or n_l+1 occurrences of the letter 1. We prove that $\mathcal{P}_w^k(2^l k + n_l + k) = 2k$. Let u be a factor of w of length $2^l k + n_l + k$. Then u begins with $0^i 1$, where $0 \leq i \leq k$. In w, this is followed by $h(v)0^{k-1}$, where $|v| = 2^l$ and thus $|h(v)| = 2^l k + n_l + c$, $c \in \{0, 1\}$. There are the following possibilities:

- If $i \leq k-2$, then $u = 0^i 1 h(v) 0^{k-i-1-c}$ and
$$(\mathrm{pref}_{k-1}(u), \mathrm{suff}_k(u), |u|_1) = (0^i 1 0^{k-2-i}, 0^{i+c} 1 0^{k-i-1-c}, n_l + 1).$$
- If $i = k-1$ and $c = 0$, then $u = 0^{k-1} 1 h(v)$ and
$$(\mathrm{pref}_{k-1}(u), \mathrm{suff}_{k-1}(u), |u|_1) = (0^{k-1}, 0^{k-2}1, n_l + 1).$$
- If $i = k-1$ and $c = 1$, then $u1 = 0^{k-1} 1 h(v)$ and
$$(\mathrm{pref}_{k-1}(u), \mathrm{suff}_{k-1}(u), |u|_1) = (0^{k-1}, 0^{k-1}, n_l).$$
- If $i = k$ and $c = 0$, then $u1 = 0^k 1 h(v)$ and
$$(\mathrm{pref}_{k-1}(u), \mathrm{suff}_{k-1}(u), |u|_1) = (0^{k-1}, 0^{k-1}, n_l).$$
- If $i = k$ and $c = 1$, then $u01 = 0^k 1 h(v)$. If it were $v = v'0$, then $1v'$ would be a factor of w of length 2^l with $|1v'|_1 = n_l + 2$, which is a contradiction, so $v = v'1$ and
$$(\mathrm{pref}_{k-1}(u), \mathrm{suff}_{k-1}(u), |u|_1) = (0^{k-1}, 0^{k-1}, n_l).$$

In total, there are $2k$ different possibilities for $(\mathrm{pref}_{k-1}(u), \mathrm{suff}_{k-1}(u), |u|_1)$, so $\mathcal{P}^1_w(2^l k + n_l + k) = 2k$.

We have already seen that $\mathcal{P}^k_w(n) = \Theta(\mathcal{P}^1_w(n))$, so it is sufficient to show $\mathcal{P}^1_w(n) = \Omega'(\log n)$. We will need the following simple fact, which is used frequently when studying abelian complexity of binary words: For any infinite binary word W,

$$\mathcal{P}^1_W(n) = \max\{|u|_1 \mid u \in F_n(W)\} - \min\{|u|_1 \mid u \in F_n(W)\} + 1. \tag{5}$$

We know that $\mathcal{P}^1_S(n) = \Omega'(\log n)$, so there is a strictly increasing sequence $n_1, n_2, n_3, \ldots$ such that $\mathcal{P}^1_S(n_i) = \Omega(\log n_i)$. By the definition of h and (5), for every i there are $u_i, v_i \in F_{n_i}(S)$ such that

$$|h(v_i)| - |h(u_i)| = |v_i|_1 - |u_i|_1 = \Omega(\log n_i).$$

Then $|h(v_i)|_1 = |v_i| = n_i$, and w has a factor $x = h(u_i)y$ such that $|x| = |h(v_i)|$ and

$$|x|_1 = |h(u_i)|_1 + |y|_1 \geq |u_i| + \lfloor |y|/k + 1 \rfloor = n_i + \Omega(\log n_i).$$

This means that $\mathcal{P}^1_w(|h(v_i)|) = \Omega(\log n_i)$, which proves that $\mathcal{P}^k_w(n) = \Omega'(\log n)$ because $kn_i \leq |h(v_i)| \leq (k+1)n_i$. □

Theorem 17. *Let $k \geq 1$. Let f be a function such that $f(n) = o(n^{(m-1)m^{k-1}})$. There exists $w \in \Sigma_m^\omega$ such that*

$$\mathcal{P}^k_w(n) = O'(1) \qquad \textit{and} \qquad \mathcal{P}^k_w(n) = \Omega'(f(n)).$$

Proof. If we prove the claim for $k = 1$, we can use Lemma 11 to get another word with similar k-abelian complexity for n divisible by $2k-1$. Then we can use Lemma 10 to prove that the complexity behaves in a similar way for all n (the sequence n_i of Lemma 10 is the sequence of numbers divisible by $2k-1$). Thus it is sufficient to prove the claim for $k = 1$.

We define w by a Toeplitz-type construction. Let $l_1, l_2, l_3, \ldots$ be a strictly increasing sequence of positive integers. For every i, let u_i be a word that has a factor in every abelian equivalence class of words in $\Sigma_m^{l_i}$. Let $v_0 = \diamond$ and, for $i \geq 1$, let v_i be the word obtained from $v_{i-1}^{|u_i|+1}$ by replacing the hole symbols $\diamond$ with the letters of $u_i\diamond$. Because $f(n) = o(n^{m-1})$ and $|v_{i-1}|$ depends only on $l_1, \ldots, l_{i-1}$, we can define the sequence $l_1, l_2, l_3, \ldots$ so that $f(|v_{i-1}|l_i) \leq l_i^{m-1}$ for all i. Let w be the limit of the sequence $v_0, v_1, v_2, \ldots$.

For every i, let $v_i = v_i'\diamond$. Then $w \in (v_i'\Sigma_m)^\omega$, so every factor of w of length $|v_i|$ is a conjugate of a word in $v_i'\Sigma_m$. Conjugates are abelian equivalent, so $\mathcal{P}^1_w(|v_i|) = \#v_i'\Sigma_m = m$. This proves that $\mathcal{P}^k_w(n) = O'(1)$.

If $a_1, \ldots, a_{l_i} \in \Sigma_m$ and $a_1 \cdots a_{l_i}$ is a factor of u_i, then $\prod_{j=1}^{l_i} v_{i-1}' a_j$ is a factor of w. If two factors of the form $a_1 \cdots a_{l_i}$ are not abelian equivalent, then the corresponding factors $\prod_{j=1}^{l_i} v_{i-1}' a_j$ are also not abelian equivalent. Thus $\mathcal{P}^1_w(|v_{i-1}|l_i) \geq \mathcal{P}^1_{u_i}(l_i) = \Omega(l_i^{m-1}) = \Omega(f(|v_{i-1}|l_i))$ for all i. This proves that $\mathcal{P}^k_w(n) = \Omega'(f(n))$. □

Theorem 18. *Let $k \geq 1$. There exists $w \in \Sigma_m^\omega$ such that*

$$\mathcal{P}_w^k(n) = O'(n) \qquad \text{and} \qquad \mathcal{P}_w^k(n) = \Omega'(n^{(m-1)m^{k-1}}).$$

Proof. By Lemmas 11 and 10, it is sufficient to prove the claim for $k = 1$ (like in Theorem 17). We define a sequence $u_0, u_1, u_2, \dots$ of finite words and show that $w = u_0 u_1 u_2 \cdots$ satisfies the requirements of the theorem. Let $u_0 = 0$ and, for $j \geq 0$,

$$u_{j+1} = \prod_{(n_0,\dots,n_{m-1})} \prod_{i=0}^{m-1} i^{|u_j|+n_i},$$

where the outer product is taken over all sequences $(n_0, \dots, n_{m-1})$ of non-negative integers such that $\sum_{i=0}^{m-1} n_i = m|u_j|$ (the order in the product does not matter). It can be proved that $\mathcal{P}_w^1(2m|u_j|) = \Omega((m|u_j|)^{m-1})$ and $\mathcal{P}_w^k(|u_j|) = O(|u_j|)$. Details can be found in the full version of this article. □

Theorem 19. *Let $k \geq 1$. There does not exist $f(n) = o(n)$ and $w \in \Sigma_m^\omega$ such that*

$$\mathcal{P}_w^k(n) = O'(f(n)) \qquad \text{and} \qquad \mathcal{P}_w^k(n) = \Omega'(n^{(m-1)m^{k-1}}).$$

Proof. We assume that $\mathcal{P}_w^k(n) = O'(f(n))$ and $f(n) = o(n)$, and prove that $\mathcal{P}_w^k(n) = o(n^{(m-1)m^{k-1}})$. For every number n and word t, let

$$p_t(n) = \min\{|u|_t \mid u \in F_n(w)\} \qquad \text{and} \qquad q_t(n) = \max\{|u|_t \mid u \in F_n(w)\}.$$

Because $\mathcal{P}_w^k(n) = O'(f(n))$ and $f(n) = o(n)$, there is a strictly increasing sequence $n_1, n_2, n_3, \dots$ such that $q_t(n_i) - p_t(n_i) < \mathcal{P}_w^k(n_i) = o(n_i)$ for all t of length at most k. For $n > n_1^2$, let $g(n) = \max\{n_i \mid n_i < \sqrt{n}\}$. Every factor of w of length n can be written as $u = u_0 \cdots u_r$, where $u_0, \dots, u_{r-1} \in \Sigma_m^{g(n)}$, $r = \lfloor n/g(n) \rfloor$, and $|u_r| < g(n) < \sqrt{n}$. For every factor t of length at most k,

$$\begin{aligned} r p_t(g(n)) \leq \sum_{j=0}^{r-1} |u_j|_t \leq |u|_t &\leq \sum_{j=0}^{r} |u_j|_t + \sum_{j=0}^{r-1} |\mathrm{suff}_{k-1}(u_j)\mathrm{pref}_{k-1}(u_{j+1})|_t \\ &\leq r(q_t(g(n)) + 2k) + |u_r|, \end{aligned}$$

so

$$q_t(n) - p_t(n) \leq r(o(g(n)) + 2k) + |u_r| = o(n) + |u_r| = o(n).$$

By Theorem 5, there are $o(n^{(m-1)m^{k-1}})$ possible k-abelian equivalence classes for u. □

References

1. Balogh, J., Bollobás, B.: Hereditary properties of words. RAIRO Inform. Theor. Appl. 39(1), 49–65 (2005)

2. Cassaigne, J., Karhumäki, J.: Toeplitz words, generalized periodicity and periodically iterated morphisms. European J. Combin. 18(5), 497–510 (1997)
3. Coven, E.M., Hedlund, G.A.: Sequences with minimal block growth. Math. Systems Theory 7, 138–153 (1973)
4. Dekking, M.: Strongly nonrepetitive sequences and progression-free sets. J. Combin. Theory Ser. A 27(2), 181–185 (1979)
5. Greinecker, F.: On the 2-abelian complexity of Thue-Morse subwords (Preprint), arXiv:1404.3906
6. Harmaala, E.: Sanojen ekvivalenssiluokkien laskentaa (2010), manuscript
7. Huova, M., Karhumäki, J., Saarela, A.: Problems in between words and abelian words: k-abelian avoidability. Theoret. Comput. Sci. 454, 172–177 (2012)
8. Huova, M., Karhumäki, J., Saarela, A., Saari, K.: Local squares, periodicity and finite automata. In: Calude, C., Rozenberg, G., Salomaa, A. (eds.) Rainbow of Computer Science, LNCS, vol. 6570, pp. 90–101. Springer (2011)
9. Jacobs, K., Keane, M.: $0-1$-sequences of Toeplitz type. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 13(2), 123–131 (1969)
10. Karhumäki, J., Saarela, A., Zamboni, L.Q.: On a generalization of Abelian equivalence and complexity of infinite words. J. Combin. Theory Ser. A 120(8), 2189–2206 (2013)
11. Karhumäki, J., Saarela, A., Zamboni, L.Q.: Variations of the Morse-Hedlund theorem for k-abelian equivalence. In: Proceedings of the 18th DLT. LNCS, vol. 8633, pp. 203–214. Springer (2014)
12. Keränen, V.: Abelian squares are avoidable on 4 letters. In: Proceedings of the 19th ICALP. LNCS, vol. 623, pp. 41–52. Springer (1992)
13. Mercaş, R., Saarela, A.: 3-abelian cubes are avoidable on binary alphabets. In: Proceedings of the 17th DLT. LNCS, vol. 7907, pp. 374–383. Springer (2013)
14. Mercaş, R., Saarela, A.: 5-abelian cubes are avoidable on binary alphabets. RAIRO Inform. Theor. Appl. 48(4), 467–478 (2014)
15. Morse, M., Hedlund, G.A.: Symbolic dynamics. Amer. J. Math. 60(4), 815–866 (1938)
16. Morse, M., Hedlund, G.A.: Symbolic dynamics II: Sturmian trajectories. Amer. J. Math. 62(1), 1–42 (1940)
17. Parreau, A., Rigo, M., Rowland, E., Vandomme, E.: A new approach to the 2-regularity of the l-abelian complexity of 2-automatic sequences. Electron. J. Combin. 22(1), P1.27 (2015)
18. Rao, M.: On some generalizations of abelian power avoidability (Manuscript)
19. Thue, A.: Über unendliche zeichenreihen. Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. 7, 1–22 (1906)
20. Thue, A.: Über die gegenseitige lage gleicher teile gewisser zeichen-reihen. Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. 1, 1–67 (1912)