



HAL
open science

On abelian saturated infinite words

Sergey Avgustinovich, Julien Cassaigne, Juhani Karhumaki, Svetlana
Puzynina, Aleksii Saarela

► **To cite this version:**

Sergey Avgustinovich, Julien Cassaigne, Juhani Karhumaki, Svetlana Puzynina, Aleksii Saarela.
On abelian saturated infinite words. *Theoretical Computer Science*, 2019, 792, pp.154-160.
10.1016/j.tcs.2018.05.013 . hal-02375068

HAL Id: hal-02375068

<https://hal.science/hal-02375068>

Submitted on 29 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On Abelian Saturated Infinite Words

Sergey Avgustinovich* Julien Cassaigne† Juhani Karhumäki‡
Svetlana Puzynina§* Alekski Saarela‡

January 10, 2018

Abstract

Let $f : \mathbb{Z}_+ \rightarrow \mathbb{R}$ be an increasing function. We say that an infinite word w is abelian $f(n)$ -saturated if each factor of length n contains $\Theta(f(n))$ abelian nonequivalent factors. We show that binary infinite words cannot be abelian n^2 -saturated, but, for any $\varepsilon > 0$, they can be abelian $n^{2-\varepsilon}$ -saturated. There is also a sequence of finite words (w_n) , with $|w_n| = n$, such that each w_n contains at least Cn^2 abelian nonequivalent factors for some constant $C > 0$. We also consider saturated words and their connection to palindromic richness in the case of equality and k -abelian equivalence.

1 Introduction

Uniform recurrence is a well-studied property of infinite words requiring that each factor occurring in the word occurs with a bounded gap, or, equivalently, for any n , there exists a constant N such that all factors of length n occur in any factor of length N .

We define here a related notion asking each factor to contain a certain amount of different factors. We formulate this property for abelian factors as follows: Let $f : \mathbb{Z}_+ \rightarrow \mathbb{R}$ be an increasing function. We say that an infinite word w is abelian $f(n)$ -saturated if each of its factors of length n contains $\Theta(f(n))$ abelian nonequivalent factors. We consider the problem for binary words and search for as highly growing functions as possible. Clearly, a quadratic upper bound for such functions $f(n) = n^2$ is given by the total number of factors of a word of length n .

We show that this trivial upper bound cannot be reached for abelian factors in binary words. To prove this, we use the fact that almost abelian k -powers are unavoidable (in the sense that there are factors $u_1 \cdots u_k$ such that for every letter, the frequencies of

*Sobolev Institute of Mathematics, Russia, avgust@math.nsc.ru

†Institut de Mathématiques de Marseille, Case 907, 13288 Marseille Cedex 9, France, julien.cassaigne@math.cnrs.fr

‡Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland, karhumak,amsaar@utu.fi

§St. Petersburg State University, 14th Line V.O., 29B, Saint Petersburg, 199178, Russia, s.puzynina@gmail.com

that letter in the blocks u_i are close to each other). On the other hand, we give an example showing that for each $\varepsilon > 0$, there exist abelian $n^{2-\varepsilon}$ -saturated infinite words. Interestingly, these infinite words are obtained by iterating a morphism.

A natural restriction of the problem to finite words asks whether there exists a constant C and a sequence $(w_n)_{n \geq 0}$ of words such that $|w_n| = n$ and w_n contains at least $Cf(n)$ abelian nonequivalent factors. We give a simple example that in this finite setting the maximal bound $f(n) = n^2$ can be achieved. In fact, this example is a special case of a construction from [6] where k -abelian palindromes were considered. We also give the exact maximal number of abelian nonequivalent factors a binary word of length n can contain.

Of course, $f(n)$ -saturated infinite words can be defined with respect to ordinary factors instead of abelian factors. In this setting the maximal bound $f(n) = n^2$ can be achieved.

The motivation to this research comes from the articles [6, 3] where palindromicity was considered with respect to three types of equivalence relations: equality, abelian equivalence, and, as an in between case, k -abelian equivalence.

In the case of poor words, it turned out that the k -abelian palindromicity behaves like the ordinary palindromicity: There exist infinite words containing only finitely many (k -abelian, $k \geq 2$) palindromes.

In the case of rich words and ordinary palindromes, the situation is clear: A finite word of length n can contain at most $n + 1$ palindromes and this upper bound can be attained, see [4, 5]. Moreover, there exist infinite words all of whose factors contain the maximal number of palindromic factors. In the abelian case, any finite word is an abelian palindrome. So the question of how many nonequivalent abelian palindromes a word can contain is actually our earlier restriction of $f(n)$ -saturation for finite words. Hence, as we mentioned, a word of length n can contain $\Theta(n^2)$ nonequivalent abelian palindromes as factors. This extends to the k -abelian palindromicity. So, with respect to richness, the k -abelian palindromicity behaves as the abelian palindromicity.

A question about extensions of k -abelian richness for infinite words was asked in [6]. The goal and motivation of this note is to give partial answers to this question. Indeed, our main results deal directly with the palindromic richness of infinite words in the abelian case. Moreover, our Theorem 5.1 and Corollary 5.2 translate results from the abelian case to the k -abelian case. Although the motivation for our research came from palindromicity research, the notion of saturated words is quite independent and evokes a series of questions on its own right.

2 Preliminaries

Let Σ be an alphabet. As usual, the set of all finite words over Σ is denoted by Σ^* , the set of all (right) infinite words by Σ^ω , the set of words of length n by Σ^n , and the set of words of length at most n by $\Sigma^{\leq n}$.

The length of a word $u \in \Sigma^*$ is denoted by $|u|$ and the number of occurrences of a factor x in u by $|u|_x$. The reversal of u is denoted by u^R .

Words $u, v \in \Sigma^*$ are *abelian equivalent* if $|u|_x = |v|_x$ for all $x \in \Sigma$, that is, if they have the same letters with the same multiplicities but possibly in a different order.

Let k be a positive integer. Words $u, v \in \Sigma^*$ are *k -abelian equivalent* if $|u|_x = |v|_x$ for all $x \in \Sigma^{\leq k}$. Of course, 1-abelian equivalence is the same as abelian equivalence.

For words $u, v \in \Sigma^*$ of the same length, the following conditions are equivalent:

- u and v are k -abelian equivalent.
- $|u|_x = |v|_x$ for all $x \in \Sigma^k$ and u and v have the same prefix of length $\min\{k-1, |u|\}$.
- $|u|_x = |v|_x$ for all $x \in \Sigma^k$ and u and v have the same suffix of length $\min\{k-1, |u|\}$.

The proof of this and other basic properties of k -abelian equivalence can be found in [7].

It follows immediately from the definition that if u and v are k -abelian equivalent, then they are also k' -abelian equivalent for all $k' \leq k$, and if they are k -abelian equivalent for all $k \in \mathbb{Z}_+$, then they are equal.

Example 2.1. The words $aabab$ and $abaab$ are 2-abelian equivalent because they begin with the same letter, $|aabab|_x = 1 = |abaab|_x$ for $x \in \{aa, ba\}$, $|aabab|_{ab} = 2 = |abaab|_{ab}$, and $|aabab|_{bb} = 0 = |abaab|_{bb}$.

The words aba and bab are not 2-abelian equivalent even though $|aba|_x = |bab|_x$ for all $x \in \{a, b\}^2$. Their inequivalence follows, for example, from the fact that $|aba|_a = 2 \neq 1 = |bab|_a$, or from the fact that the first letters are different.

A word u is a *palindrome* if $u = u^R$, and a *k -abelian palindrome* if u and u^R are k -abelian equivalent. The case $k = 1$ is not interesting: u and u^R are always abelian equivalent, so every word is an abelian palindrome. If the alphabet is binary, also the case $k = 2$ is very simple: A nonempty binary word u is 2-abelian equivalent to u^R if and only if it begins and ends with the same letter.

Example 2.2. The word $aabacaa$ is a 2-abelian palindrome, but not a 3-abelian palindrome.

Let $k \geq 1$ and let $f : \mathbb{Z}_+ \rightarrow \mathbb{R}$ be an increasing function. An infinite word w is

- *$f(n)$ -saturated* if there exists a constant $C > 0$ such that for all sufficiently large n , every factor of w of length n has at least $Cf(n)$ factors,
- *palindromic $f(n)$ -saturated* if there exists a constant $C > 0$ such that for all sufficiently large n , every factor of w of length n has at least $Cf(n)$ palindromic factors,
- *k -abelian $f(n)$ -saturated* if there exists a constant $C > 0$ such that for all sufficiently large n , every factor of w of length n has at least $Cf(n)$ k -abelian nonequivalent factors,
- *k -abelian palindromic $f(n)$ -saturated* if there exists a constant $C > 0$ such that for all sufficiently large n , every factor of w of length n has at least $Cf(n)$ k -abelian nonequivalent k -abelian palindromic factors.

A finite word of length n is *rich* if it has $n + 1$ distinct palindromic factors. An infinite word is *rich* if all of its factors are rich. For more on both finite and infinite rich words, see [4, 5]. It follows from the definitions that a rich infinite word is palindromic n -saturated. A word of length n cannot have more than $n + 1$ palindromic factors, so there are no palindromic $f(n)$ -saturated words if f is superlinear.

In [6] a related question has been posed for the k -abelian case: Do there exist infinite words such that each factor contains the maximal number, up to a constant, of k -abelian nonequivalent k -abelian palindromic factors? This maximal number is known to be quadratic with respect to the length of the factor, see [6]. We can then call such words k -abelian rich. This motivates our study of k -abelian palindrome saturated words.

3 A Characterization of n^2 -saturated words

In this section, we will consider saturated words in the ordinary case of equality.

Every word of length n has at least $n + 1$ factors (namely, one of each length from 0 to n), so every infinite word is n -saturated. If an infinite word contains arbitrarily high powers of some fixed word, then it is not $f(n)$ -saturated for any superlinear function f .

A word of length n can have at most $n(n + 1)/2 + 1$ factors, and this bound is reached only in the trivial case where the word contains n different letters. This obviously means that there are no infinite words such that every factor has exactly this maximal number of factors, but there are many infinite words such that every factor has the maximal number of factors up to a constant.

Theorem 3.1. *An infinite word w is n^2 -saturated if and only if there exists $p \geq 2$ such that w is p -power-free.*

Proof. Assume first that for all $p \geq 2$, there exists a nonempty word u such that u^p is a factor of w . The word u^p has at most $|u| \cdot |u^p| + 1 = |u^p|^2/p + 1$ factors. Therefore, w is not n^2 -saturated.

Assume then that w is p -power-free. Let $n \geq 1$ and let $u = a_0 \cdots a_{n-1}$ be an arbitrary factor of w of length n . Let $l \leq \lceil n/2 \rceil$ and consider the factors $v_i = a_i \cdots a_{i+l-1}$ of u of length l for $i < l/(p - 1)$. If two of these were equal, say, $v_i = v_j$ for $i < j$, then $a_i \cdots a_{j+l-1}$ would have the prefix $(a_i \cdots a_{j-1})^p$, which is a contradiction. Therefore, the factors v_i are all distinct, and u has at least

$$\sum_{l=1}^{\lceil n/2 \rceil} \frac{l}{p-1} \geq \frac{n(n+2)}{8(p-1)}$$

factors. This proves that w is n^2 -saturated. □

4 Abelian saturated binary words

In this section, we analyze binary words from the point of view of abelian equivalence. We show that there are abelian $n^{2-\varepsilon}$ -saturated binary words, but there are no abelian n^2 -saturated binary words. The case of ternary and larger alphabets remains open.

Let $w \in \{a, b\}^* \cup \{a, b\}^\omega$, $n \geq 1$, and $n \leq |w|$ if w is finite. We use the notation

$$\alpha_n(w) = \min\{|u|_b \mid u \in F_n(w)\}, \quad \beta_n(w) = \max\{|u|_b \mid u \in F_n(w)\}.$$

A simple fact that is frequently used when studying abelian factors of binary words is that w has exactly

$$\beta_n(w) - \alpha_n(w) + 1$$

abelian nonequivalent factors of length n .

Before moving on to the results about infinite words, let us first look at the maximal and minimal numbers of abelian nonequivalent factors of finite words. Every word of length n has at least $n + 1$ abelian nonequivalent factors, so every infinite word is abelian n -saturated. If $u \in \{a, b\}^n$ and $|u|_b = i \leq n/2$, then u has at most $in - i^2 + n + 1$ abelian nonequivalent factors, and this bound is reached if $u = a^{n-i}b^i$. Thus the maximal number of abelian nonequivalent factors a binary word of length n can have is $\lfloor n^2/4 \rfloor + n + 1$, and this bound is reached by the word $a^{\lfloor n/2 \rfloor} b^{\lfloor n/2 \rfloor}$.

In [6], there is a construction showing that for all $k \geq 1$, there exists a constant $C > 0$ such that for all $n \geq 1$, there exist a binary word of length n having at least Cn^2 k -abelian nonequivalent k -abelian palindromic factors. This general construction is somewhat complicated, but in the binary case, it gives exactly the words $a^{\lfloor n/2 \rfloor} b^{\lfloor n/2 \rfloor}$.

Theorem 4.1. *Let $K \geq 3$ and let w be the fixed point of the morphism*

$$\sigma : \{a, b\}^* \rightarrow \{a, b\}^*, \quad \sigma(a) = a^{K-1}b, \quad \sigma(b) = ab^{K-1}.$$

Then w is abelian $n^{1+\log(K-2)/\log(K)}$ -saturated.

Proof. For all $j \geq 0$, let $A_j = \sigma^j(a)$ and $B_j = \sigma^j(b)$. Then $|A_j| = |B_j| = K^j$ and

$$|A_j|_b = \frac{K^j - (K-2)^j}{2}, \quad |B_j|_b = \frac{K^j + (K-2)^j}{2},$$

as can be verified by induction.

Let u be an arbitrary factor of w , $n = |u|$ and $K' = \max\{K, 5\}$. We will show that u has at least the required number of abelian nonequivalent factors. We can assume that $n \geq K' + 1$.

Let

$$j = \left\lfloor \frac{\log(n/(K'+1))}{\log(K)} \right\rfloor$$

so that $n \geq (K'+1)K^j$. Then u has a factor $\sigma^j(v)$, where v is a factor of w of length K' . Every factor of w of length K' has the factors aa, ba or the factors ab, bb or the factors aa, ab or the factors ba, bb . We assume that v has the factors aa, ba (the other cases are symmetric). Then u has the factors $A_j A_j, B_j A_j$.

For all $i \in \{K^j, \dots, 2K^j\}$, let p_i be the prefix of A_j of length $i - K^j$. Then u has the factors $A_j p_i$ and $B_j p_i$ of length i , so $\alpha_i(u) \leq |A_j p_i|_b$ and $\beta_i(u) \geq |B_j p_i|_b$. Thus

$$\beta_i(u) - \alpha_i(u) \geq |B_j p_i|_b - |A_j p_i|_b = |B_j|_b - |A_j|_b = (K-2)^j.$$

The number of abelian nonequivalent factors of u is

$$\begin{aligned} \sum_{i=0}^n (\beta_i(u) - \alpha_i(u) + 1) &\geq \sum_{i=K^j}^{2K^j} (\beta_i(u) - \alpha_i(u)) \geq K^j(K-2)^j = \frac{(K^2 - 2K)^{j+1}}{K^2 - 2K} \\ &\geq \frac{1}{K^2 - 2K} \cdot (K^2 - 2K)^{\log(n/(K'+1))/\log(K)} = \frac{1}{K^2 - 2K} \cdot \left(\frac{n}{K'+1}\right)^{\log(K^2 - 2K)/\log(K)} \\ &= \frac{1}{K^2 - 2K} \cdot \left(\frac{n}{K'+1}\right)^{1 + \log(K-2)/\log(K)}. \end{aligned}$$

The claim follows. \square

Corollary 4.2. *For every $\varepsilon > 0$, there exists an abelian $n^{2-\varepsilon}$ -saturated binary infinite word.*

Proof. Follows from Theorem 4.1, because $\lim_{K \rightarrow \infty} \log(K-2)/\log(K) = 1$. \square

Theorem 4.3. *There does not exist an abelian n^2 -saturated binary infinite word.*

We give two proofs of this result. The first one is self-contained, the second one uses a fact about unavoidability of "almost abelian powers".

First proof. We assume that every factor of $w \in \{a, b\}^\omega$ of length n has at least Cn^2 abelian nonequivalent factors and derive a contradiction. Let $K = \lceil 2/C \rceil + 1$.

Let us first give a rough idea of the proof. We will look at factors of w of length K^j for different values of j . From the assumption that a factor of length K^j has many abelian nonequivalent factors we can deduce that it has factors with "many" b 's and factors with "few" b 's, and this leads to bounds for the maximal and minimal numbers of b 's in factors of length K^j . We will then show that the difference of the maximal and the minimal proportion of b 's in factors of length K^j , that is, $\beta_{K^j}(w)/K^j - \alpha_{K^j}(w)/K^j$, decreases and eventually becomes negative as j increases, which is a contradiction.

Let $j \geq 0$ and let u be a factor of w of length K^{j+1} . We can write $u = u_1 \cdots u_K$, where $|u_i| = K^j$ for all i . Because $|u|_b = |u_1|_b + \cdots + |u_K|_b$ and $\alpha_{K^j}(w) \leq |u_i|_b \leq \beta_{K^j}(w)$ for all i , we have $K\alpha_{K^j}(w) \leq |u|_b \leq K\beta_{K^j}(w)$, but our first goal is to prove better bounds than this for $|u|_b$. Let

$$m = \min\{|u_i|_b \mid i \in \{1, \dots, K\}\}, \quad M = \max\{|u_i|_b \mid i \in \{1, \dots, K\}\}.$$

Then

$$\begin{aligned} |u|_b &\geq (K-1)m + M \geq K\alpha_{K^j}(w) + (M-m), \\ |u|_b &\leq (K-1)M + m \leq K\beta_{K^j}(w) - (M-m), \end{aligned} \tag{1}$$

so we need to estimate the difference $M - m$. If for all $l \in \{1, \dots, |u|\}$, u has at most $C|u| - 1$ abelian nonequivalent factors of length l , then it has less than $C|u|^2$ abelian nonequivalent factors, which is a contradiction. Therefore there exists a number

$l \in \{1, \dots, |u|\}$ such that u has more than $C|u| - 1$ abelian nonequivalent factors of length l . There exists $k \in \{0, \dots, K - 1\}$ such that every factor of u of length l can be written in the form $pu_{i+1} \cdots u_{i+k}q$, where $|pq| \leq 2K^j - 2$. It follows that $\alpha_l(u) \geq km$ and $\beta_l(u) \leq kM + 2K^j - 2$. We get

$$CK^{j+1} - 1 = C|u| - 1 < \beta_l(u) - \alpha_l(u) + 1 \leq k(M - m) + 2K^j - 1 < K(M - m) + 2K^j - 1$$

and thus

$$M - m > \left(C - \frac{2}{K}\right) K^j = C'K^j,$$

where $C' = C - 2/K > 0$. From (1) it now follows that

$$K\alpha_{K^j}(w) + C'K^j < |u|_b < K\beta_{K^j}(w) - C'K^j. \quad (2)$$

Because (2) holds for all factors u of w of length K^{j+1} , we get

$$\alpha_{K^{j+1}}(w) > K\alpha_{K^j}(w) + C'K^j, \quad \beta_{K^{j+1}}(w) < K\beta_{K^j}(w) - C'K^j.$$

Then

$$\beta_{K^{j+1}}(w) - \alpha_{K^{j+1}}(w) < K(\beta_{K^j}(w) - \alpha_{K^j}(w)) - 2C'K^j$$

and thus

$$\frac{\beta_{K^{j+1}}(w) - \alpha_{K^{j+1}}(w)}{K^{j+1}} < \frac{\beta_{K^j}(w) - \alpha_{K^j}(w)}{K^j} - \frac{2C'}{K}.$$

Because this holds for all j , we get

$$\frac{\beta_{K^j}(w) - \alpha_{K^j}(w)}{K^j} < \beta_1(w) - \alpha_1(w) - j \cdot \frac{2C'}{K},$$

so for large enough j , $\beta_{K^j}(w) - \alpha_{K^j}(w) < 0$, which is a contradiction. \square

For the other proof we use the following fact basically saying that ‘‘almost abelian powers’’ (in the sense that frequencies of letters in each block differ by at most ε) are unavoidable.

For a finite word v and a letter $a \in \Sigma$, we let $\text{freq}_a(v)$ denote the frequency of a in v : $\text{freq}_a(v) = \frac{|v|_a}{|v|}$. The following lemma from [1] basically states that every infinite word contains ‘‘almost’’ abelian k -powers for any k and any ε :

Lemma 4.4. *Let w be an infinite word over an alphabet Σ . Then for any integers k and l and any $\varepsilon > 0$ there exists a factor $u = u_1 \cdots u_k$ of w , where $u_1, \dots, u_k \in \Sigma^*$, $|u_1| = |u_2| = \cdots = |u_k| \geq l$ and*

$$|\text{freq}_a(u_i) - \text{freq}_a(u_j)| < \varepsilon$$

for any $i, j \in \{1, \dots, k\}$.

Actually, a slightly stronger fact is implicitly contained in [2]. In the paper, the notion of ε -regular word is introduced: For a positive ε , $\varepsilon < 1/3$, a word w of length n over an alphabet Σ is called ε -regular if for every i , $\varepsilon n + 1 \leq i \leq n - 2\varepsilon n + 1$ and every $a \in \Sigma$ it holds that

$$|\text{freq}_a(w) - \text{freq}_a(w_i \dots w_{i+\varepsilon n-1})| < \varepsilon \quad (3)$$

(here we neglect integers in indices for readability). Further, a regularity lemma for words introduced in [2] states that any sufficiently long word admits an ε -regular partition, i.e. a factorization such that all factors in the factorization except for some of total length at most εn are ε -regular. In addition, the number of factors in the factorization can be bounded from above (by a certain function of ε). As a corollary, we obtain that

Lemma 4.5. *Let w be an infinite word over an alphabet Σ . Then for any integers k and N and any $\varepsilon > 0$ there exists a factor u of w , where $|u| \geq N$, and for any $a \in \Sigma$*

$$|\text{freq}_a(u) - \text{freq}_a(u_i \dots u_{i+\frac{|u|}{k}-1})| < \varepsilon$$

for each $i \in \{1, \dots, |u| - \frac{|u|}{k} + 1\}$.

Essentially, this is a strengthening of Lemma 4.4 saying that in fact we can choose a factor u such that all its factors of length $l = \frac{|u|}{k}$ differ by at most ε and not only those beginning in positions il .

In Theorem 3.1, we proved that if an infinite word contains high powers, it cannot be n^2 -saturated. In the following second proof of Theorem 4.3, we prove that because every infinite binary word contains high ‘‘almost abelian powers’’, it cannot be abelian n^2 -saturated.

Second proof. Let w be an infinite binary word and suppose that each its factor of length n has at least Cn^2 distinct abelian factors for some constant C . By Lemma 4.4, given an integer k and $\varepsilon > 0$, the word w contains a factor $u = u_1 \dots u_k$ of w , where $|u_1| = |u_2| = \dots = |u_k| \in \Sigma^l$ for some integer l and

$$|\text{freq}_a(u_i) - \text{freq}_a(u_j)| < \varepsilon$$

for any $i, j \in \{1, \dots, k\}$. Choose $k > \frac{6}{C}$, $\varepsilon < \frac{C}{2}$.

Now we count and estimate the number of its abelian factors.

First count ‘‘short’’ factors: notice that the number of abelian factors of u of length smaller than l is at most nl , where $n = |u|$. Indeed, for each length i we have $n - i + 1$ possible places for the initial position of a factor of length i , and a total number of lengths is l . Hence the total number of short distinct factors is at most nl , and the number of abelian factors can only be smaller or equal.

Now consider factors of length at least l . Take length $i \geq l$; each such factor consists of several full blocks of length l , where frequencies of letters differ by at most ε , and a prefix and a suffix of a total length smaller than $2l$. So the numbers of a given letter in different factors of length i differ by at most $\varepsilon i + 2l$, which gives an upper bound for the number of distinct abelian factors of length i .

Now the total number of abelian factors is bounded from above by

$$nl + \sum_{i=1}^n (\varepsilon i + 2l) < nl + \varepsilon n^2 + 2nl = \frac{3n^2}{k} + \varepsilon n^2 < Cn^2,$$

where the last inequality comes from the choice of k and ε . \square

We remark that in the proof either of the Lemmas 4.4 and 4.5 can be used, but neither seem to work directly for non-binary case.

5 From the abelian case to the k -abelian case

In this section, we show how abelian constructions can be turned into k -abelian constructions simply by mapping by suitable morphisms. Moreover, we can consider k -abelian palindromes instead of factors. We get k -abelian palindromic $n^{2-\varepsilon}$ -saturated words for all k and all nonunary alphabets. It remains open whether for some k there exists k -abelian palindromic n^2 -saturated words. If there are such words in the ternary abelian case, then there are such words for all $k \geq 2$ already in the binary case.

Theorem 5.1. *Let $k \geq 2$. If there exists an abelian $f(n)$ -saturated ternary infinite word, then there exists a k -abelian palindromic $f(\lfloor n/(2k) \rfloor - 2)$ -saturated binary infinite word.*

Proof. Let C be a constant and $w \in \{a, b, c\}^\omega$ an infinite word such that for all sufficiently large n , every factor of w of length n has at least $Cf(n)$ abelian nonequivalent factors. Let $h : \{a, b, c\}^* \rightarrow \{a, b\}^*$ be the morphism defined by

$$h(a) = a^{2k}, \quad h(b) = a^{k-1}ba^k, \quad h(c) = a^{k-1}bba^{k-1}$$

and let $w' = h(w)$. It is easy to see that the image of every word under h is a k -abelian palindrome. We can also see that if $u, v \in \{a, b, c\}^*$ are not abelian equivalent, then $h(u), h(v)$ are not 2-abelian equivalent: If u, v are not abelian equivalent, then one of the following three conditions is satisfied: $|u| \neq |v|$, or $|u|_c \neq |v|_c$, or $|u|_c = |v|_c$ and $|u|_b \neq |v|_b$. In the first case, $|h(u)| \neq |h(v)|$. In the second case, $|h(u)|_{bb} = |u|_c \neq |v|_c = |h(v)|_{bb}$. In the third case, $|h(u)|_b = |u|_b + 2|u|_c \neq |v|_b + 2|v|_c = |h(v)|_b$. In all cases, $h(u)$ and $h(v)$ are 2-abelian nonequivalent.

If v is a factor of w' of sufficiently large length n , then v has a factor $h(u)$, where u is a factor of w of length $m = \lfloor n/(2k) \rfloor - 2$ and the word u has at least $Cf(m)$ abelian nonequivalent factors, and their images under h are k -abelian nonequivalent k -abelian palindromic factors of v . \square

Corollary 5.2. *Let $k \geq 2$. For every $\varepsilon > 0$, there exists a binary infinite word w and a constant $C > 0$ such that for all $n \geq 0$, every factor of w of length n has at least $Cn^{2-\varepsilon}$ k -abelian nonequivalent k -abelian palindromic factors,*

Proof. Follows from Corollary 4.2 and Theorem 5.1. \square

6 Conclusion

In this article, we have characterized n^2 -saturated infinite words, and proved that there are k -abelian palindromic $n^{2-\varepsilon}$ -saturated binary infinite words for all $k \geq 1$, but there are no abelian n^2 -saturated binary infinite words.

We can ask the following question: Given $k \geq 1$ and $m \geq 2$, are there k -abelian (palindrome) n^2 -saturated m -ary infinite words? As stated above, the answer is negative for $k = 1$, $m = 2$, but all other cases remain open. If the answer is positive for $k = 1$, $m = 3$, then it is positive in all cases except $k = 1$, $m = 2$.

If there are k -abelian palindromic n^2 -saturated words, then it would make sense to call them *k-abelian rich*. In any case, studying the different variants of saturated words seems like an interesting topic for further research.

References

- [1] Sergey V. Avgustinovich and Anna E. Frid. Words avoiding abelian inclusions. *Journal of Automata, Languages and Combinatorics*, 7(1):3–9, 2002.
- [2] Maria Axenovich, Yury Person, and Svetlana Puzynina. A regularity lemma and twins in words. *J. Comb. Theory, Ser. A*, 120(4):733–743, 2013.
- [3] Julien Cassaigne, Juhani Karhumäki, and Svetlana Puzynina. On k -abelian palindromes, Submitted.
- [4] Xavier Droubay, Jacques Justin, and Giuseppe Pirillo. Episturmian words and some constructions of de Luca and Rauzy. *Theoret. Comput. Sci.*, 255(1–2):539–553, 2001.
- [5] Amy Glen, Jacques Justin, Steve Widmer, and Luca Q. Zamboni. Palindromic richness. *European J. Combin.*, 30(2):510–531, 2009.
- [6] Juhani Karhumäki and Svetlana Puzynina. On k -abelian palindromic rich and poor words. In *Proceedings of the 18th DLT*, volume 8633 of *LNCS*, pages 191–202. Springer, 2014.
- [7] Juhani Karhumäki, Aleksi Saarela, and Luca Q. Zamboni. On a generalization of Abelian equivalence and complexity of infinite words. *J. Combin. Theory Ser. A*, 120(8):2189–2206, 2013.