RAIRO-Theor. Inf. Appl. 46 (2012) 3–15
DOI: 10.1051/ita/2011121

Available online at:
www.rairo-ita.org

ON ABELIAN VERSIONS OF CRITICAL FACTORIZATION THEOREM*

Sergey Avgustinovich[1], Juhani Karhumäki[2, 3] and Svetlana Puzynina[1, 2, 3]

Abstract. In the paper we study abelian versions of the critical factorization theorem. We investigate both similarities and differences between the abelian powers and the usual powers. The results we obtained show that the constraints for abelian powers implying periodicity should be quite strong, but still natural analogies exist.

Mathematics Subject Classification. 68R15.

1. Introduction

One of the main results of combinatorics on words, the critical factorization theorem, relates local periodicities of a word to its global periodicity. It was first proved by Césari and Vincent, and in the present form it is due to Duval [2,5]. This theorem states, roughly speaking, a connection between local and global periods of a word; the local period at any position of the word is defined as the shortest repetition centered in this position. The theorem says that the global period of a word is the maximum of its local periods.

In [9] Mignosi *et al.* proposed a different notion of a local period: a local period at a position related to order ρ is defined as the length of the shortest repetition of order ρ to the left from this position. In such a definition of local periods squares are not enough to ensure the global periodicity, but the threshold is surprisingly given by the golden ratio φ. Namely, if every sufficiently long prefix of an infinite

Keywords and phrases. Combinatorics on words, periodicity, central factorization theorem, abelian properties of words.

* *Supported in part by the Academy of Finland under the grant 121419, by Finnish Cultural Foundation, by Russian Foundation of Basic Research (grants 10-01-00424, 09-01-00244).*

[1] Sobolev Institute of Mathematics, 630090 Novosibirsk, Russia. avgust@math.nsc.ru

[2] University of Turku, Department of Mathematics, 20014 Turku, Finland. svepuz@utu.fi

[3] Turku Centre for Computer Science, 20014 Turku, Finland. karhumak@utu.fi

Article published by EDP Sciences

© EDP Sciences 2011

https://doi.org/10.1051/ita/2011121 Published online by Cambridge University Press

word has a φ^2-repetition as a suffix, then the sequence is ultimately periodic. The quantity φ^2 is optimal in the sense that the claim does not hold true for any real number less than φ^2. In a related paper [6] bounded periods were considered. The authors established strict borderlines for periods of squares distinguishing ultimate periodicity from non-periodicity. See Ph.D. thesis of Lepistö [7] for further results. Powers immediately to the right from each position were recently studied by Saari [11].

In combinatorics of words, abelian analogs of classical problems are often considered, such as abelian complexity, abelian avoidance, abelian powers and their generalizations [1,10]. The purpose of our note is to find abelian analogues of critical factorization theorem. We seek for constraints for abelian powers enforcing a word to be (ultimately) periodic. Our results show that such constraints should be quite strong, but still natural analogies exist. In particular, for every integer k we construct non-periodic words having bounded $2k$-powers centered at every position. We provide bounds for periods in abelian squares enforcing periodicity. In addition, we study abelian powers in Sturmian and Thue-Morse words.

2. Preliminaries

2.1. Critical factorization theorem and its variations

In this section we present necessary definitions and results concerning classical critical factorization theorem.

Given a finite non-empty set Σ (called an alphabet), we denote by Σ^*, Σ^ω and $\Sigma^{\mathbb{Z}}$, the set of finite words, the set of (right) infinite words, and the set of biinfinite words over the alphabet Σ, respectively. For a finite word $u = u_1u_2\dots u_n$ with $n \geq 1$ and $u_i \in \Sigma$, we denote the length n of u by $|u|$. The empty word will be denoted by ε and we set $|\varepsilon| = 0$. A finite word z is a *factor* of a finite or infinite word w if $w = uzv$ for some words u, v. In the special case $u = \varepsilon$ (resp. $v = \varepsilon$), we call z a *prefix* (resp. *suffix*) of w. We say that a biinfinite word w is *periodic*, if there exists a finite word v such that $w = v^{\mathbb{Z}}$. A right infinite word w is *ultimately periodic*, if for some finite words u and v we can write $w = uv^\omega$; w is *purely periodic* (or briefly *periodic*) if $u = \varepsilon$; $|v|$ is a *period* of v.

Given a word w and an integer i such that $w = uv$, $|u| = i$, define z to be the shortest suffix of w_1u which is also a prefix of vw_2 for suitable w_1 and w_2. The word z is the shortest *repetition word* in w centered at position i. If $w_1 = w_2 = \varepsilon$, then at this position we have a "proper" square, otherwise we have a "virtual" square. The *local period* at position i is defined as the length of z. The critical factorization theorem states that the global period of w is the maximum of all local periods. We are interested in consequences of this theorem for infinite words (see *e.g.* [8], pp. 296–297):

Theorem 2.1. *A biinfinite word w is periodic if and only if there exists an integer l such that w has at every position a centered square with period at most l.*

https://doi.org/10.1051/ita/2011121 Published online by Cambridge University Press

Theorem 2.2. *A right infinite word w is periodic if and only if there exists an integer l such that w has at every position a (virtual) centered square with period at most l.*

Theorem 2.3. *A right infinite word w is ultimately periodic if and only if there exists n_0 such that for every $n \geq n_0$ there exists a suffix of $w_1 \dots w_n$ that is also a prefix of $w_{n+1}w_{n+2}\dots$, i.e., for every $n \geq n_0$ there exists a proper central square.*

These theorems show relations between the local and the global regularities, where the local regularity means centered powers, and the global regularity means the (ultimate) periodicity.

In [6,9] by local periods the authors mean powers immediately to the left from each position. Denote by $\mathrm{pref}_k(w)$ a prefix of w of length k. For a rational number $k \geq 1$, we say that a word w is a kth power if there exists a word v such that $w = v^k$, where v^k denotes the word $v'v''$ with $v' = v^{\lfloor k \rfloor}$ and $v'' = \mathrm{pref}_{|v|(k-\lfloor k \rfloor)}v = \mathrm{pref}_{(|w|-\lfloor k \rfloor|v|)}v$. Next we say that w contains a repetition of order $\rho \geq 1$ if it contains as a factor a kth power with $k \geq \rho$. Note that here ρ is allowed to be any real number ≥ 1. We say that w *has a ρ-suffix*, if w contains a repetition of order ρ as its suffix.

A remarkable result of Mignosi *et al.* [9] connects ρ-suffices with periodicity:

Theorem 2.4. *A right-infinite word w is ultimately periodic if and only if there exists n_0 such that for every $n \geq n_0$ the word $\mathrm{pref}_n(w)$ has a φ^2-suffix, where $\varphi = (1+\sqrt{5})/2$.*

As also shown in [9], Theorem 2.4 is optimal:

Theorem 2.5. *For any real number $\varepsilon > 0$ there exists a natural number $n_0 > 0$ such that for any $n \geq n_0$ the prefix of length n of the infinite Fibonacci word has a $(\varphi^2 - \varepsilon)$-suffix.*

In [6] powers with bounded periods are considered. Let $\rho \geq 1$ be a real number and $p \geq 1$ an integer. An infinite word w is (ρ, p)-*repetitive* if there exists an integer n_0 such that each prefix of w of length at least n_0 ends with a repetition of order ρ of a word of length at most p. It is clear that (ρ, p)-repetitivity implies (ρ, p')-repetitivity for any $p' \geq p$ and (ρ', p)-repetitivity for any $\rho' \leq \rho$. The goal of the paper [6] is to establish connections between (ρ, p)-repetitive and ultimately periodic words. In the paper it is proved, that there exist non-ultimately periodic $(2, 5)$-repetitive words, and all such words are described. Moreover, the pair $(2, 5)$ is optimal in the sense that any (ρ, p)-repetitive word with $\rho > 2$ and $p = 5$ or $\rho = 2$ and $p = 4$ is ultimately periodic.

2.2. Definitions and notation

We continue by fixing some terminology for abelian repetitions. Given a finite word $u \in \Sigma^*$ and $a \in \Sigma$, we let $|u|_a$ denote the number of occurrences of the letter

https://doi.org/10.1051/ita/2011121 Published online by Cambridge University Press

a in u. Two words u and v in Σ^* are *abelian equivalent* if and only if $|u|_a = |v|_a$ for all $a \in \Sigma$. It is easy to see that abelian equivalence is indeed an equivalence relation on Σ^*. An *abelian k-power* is a non-empty word of the form $u = v_1 v_2 \dots v_k$ where the words v_i are pairwise abelian equivalent. In this case we refer to $|v_1|$ as the length of an abelian period of u (or briefly period).

Let w be a biinfinite word over an alphabet Σ. For integer i, by a *position* i in a word w we mean a position between w_i and w_{i+1}. Let k, l be integers. We say that w is (k,l)-*abelian central repetitive* (or briefly (k,l)-ACR), if it has a centered abelian $2k$-power with length of period at most l at every position. Formally, it means that for every i there exists $l' \leq l$ such that $w_{i-l'k+1} \dots w_{i+l'k}$ is an abelian $2k$-power. We say that w is (k,l)-*abelian right* (resp. *left*) *repetitive*, if it has an abelian k-power with length of period at most l immediately to the right (resp. left) from every position, *i.e.*, for every i there exists $l' \leq l$ such that the word $w_{i+1} \dots w_{i+l'k}$ (resp. $w_{i-l'k+1} \dots w_i$) is an abelian k-power. We will refer to (k,l)-abelian right (resp. left) repetitive words as (k,l)-ARR (resp. (k,l)-ALR) for brevity. It is easy to see that a periodic biinfinite word with the period l is (k,l)-ACR, ALR and ARR for every k. Notice also that a (k,l)-ACR word is (k,l)-ARR and ALR as well. It is easy to see that a word having at every position a centered abelian square of fixed length is periodic, and the same holds for left and right squares.

3. Central abelian powers

In this section we study central abelian powers. In particular, contrary to the usual critical factorization theorem, we will prove that there exist infinite non-periodic words with bounded central abelian powers at every position. To build such words, we will need some auxiliary notation.

Let w be a biinfinite word. Let every letter w_i be a centre of an arithmetic progression of one letter of length at least $2k+1$ with difference at most l, *i.e.*, for every i there exists $l' \leq l$ such that $\omega_i = \omega_{i+l'} = \omega_{i-l'} = \omega_{i+2l'} = \omega_{i-2l'} = \ldots = \omega_{i+kl'} = \omega_{i-kl'}$. Then we say that w is (k,l)-*arithmetically centered*. In the further text, we will write that w contains a central arithmetic progression at position i, if w_i, which is a letter immediately to the left from the position i, is a centre of an arithmetic progression.

Let $\Sigma = \{a_1, \dots, a_n\}$ be an alphabet. Denote by μ_Σ the *generalized Thue-Morse morphism*:

$$\begin{aligned} \mu_\Sigma(a_1) &= a_1 a_2 \dots a_n, \\ \mu_\Sigma(a_2) &= a_2 a_3 \dots a_n a_1, \\ &\dots \\ \mu_\Sigma(a_n) &= a_n a_1 \dots a_{n-1}. \end{aligned}$$

Notice that $\mu = \mu_{\{0,1\}}$ is the usual Thue-Morse morphism, and its fixed point is the Thue-Morse word.

https://doi.org/10.1051/ita/2011121 Published online by Cambridge University Press

Lemma 3.1. *Let w be a (k,l)-arithmetically centered word over an alphabet Σ. Then $\mu_\Sigma(w)$ is $(k,|\Sigma|l)$-ACR word.*

Proof. Consider a position i in $\mu_\Sigma(w)$. If $i = |\Sigma|m$ for an integer m, then we have a central abelian $2k$-power with length of period $|\Sigma|$. The period coincides with a full image of some letter, so in the period all letters from the alphabet are used once. For $i = |\Sigma|m + r$, $0 < r < |\Sigma|$, if w contains at the position m a central arithmetic progression of length $2k+1$ with difference $l' \le l$, then $\mu_\Sigma(w)$ contains at the position i a central abelian $2k$-power with length of period $|\Sigma|l'$. The period includes $l'-1$ full images of letters and two pieces of image of a letter w_m at the beginning and at the end of the period, these two pieces constitute full image of the letter. So in the abelian period every letter is used l' times. □

Theorem 3.2. *For every integer k, there exists a biinfinite non-periodic (k,l)-ACR word with $l = 2(k+1)^2$.*

Proof. Lemma 3.1 implies that it is enough to build a $(k, l/2)$-arithmetically centered word. We build it as follows:

$$w_i = \begin{cases} 1, & \text{if } i \text{ is not divisible by } (k+1), \\ 0, & \text{if } i = j(k+1), j \text{ is not divisible by } (k+1), \\ ?, & \text{if } i \text{ is divisible by } (k+1)^2. \end{cases} \tag{3.1}$$

Here ? means that we can use either 0 or 1.

To prove that this word is $(k, l/2)$-arithmetically centered, we consider separately letters from each group from (3.1). Every $w_i = 1$ with i not divisible by $(k+1)$ is a center of an arithmetic progression with difference $(k+1)$. Every $w_i = 0$ with $i = j(k+1)$, j is not divisible by $k+1$, is a center of an arithmetic progression with difference $l/2$. Every $? = 0$ is a center of an arithmetic progression with difference $(k+1)$, every $? = 1$ is a center of an arithmetic progression with difference 1.

Applying the Thue-Morse morphism to this sequence, we obtain a (k,l)-ACR word. □

There are some remarks to be made about the construction from Theorem 3.2.

Remark 3.3. The construction gives uncountably many non-periodic $(k, 2(k+1)^2)$-ACR words.

Remark 3.4. Note that instead of the Thue-Morse morphism we can also apply the general Thue-Morse morphism for an alphabet Σ, $|\Sigma| > 2$. In this case we obtain a $(k, |\Sigma|(k+1)^2)$-ACR word over Σ.

Remark 3.5. Let w be a (k,l)-ACR word over an alphabet Σ. Consider an alphabet Σ' satisfying $|\Sigma'| < |\Sigma|$, and a surjective map $\varphi : \Sigma \to \Sigma'$. Then the

https://doi.org/10.1051/ita/2011121 Published online by Cambridge University Press

word $\varphi(w)$ is a (k,l)-ACR word over Σ'. This construction allows to construct non-periodic (k,l)-ACR words from non-periodic (k,l)-ACR words over a larger alphabet.

Theorem 3.6. *There exists a biinfinite non-periodic word w, such that for every integer k there exists l such that w is a (k,l)-ACR word.*

Proof. We build the word using the construction from Theorem 3.2 and the Toeplitz construction [3,12]. Recall the definition of Toeplitz words. Let ? be a letter not in Σ. For a word $w \in \Sigma(\Sigma \cup ?)^*$, let

$$T_0(w) = ?^\omega,\ T_{i+1}(w) = F_w(T_i(w)),$$

where $F_w(u)$, defined for any $u \in (\Sigma \cup ?)^\omega$, is the word obtained from w^ω by replacing the sequence of all occurrences of ? by u; in particular, $F_w(u) = w^\omega$ if w contains no ?.

Clearly,

$$T(w) = \lim_{i \to \infty} T_i(w) \in \Sigma^\omega$$

is well-defined, and it is referred to as the Toeplitz word determined by the pattern w. Let $p = |w|$ and $q = |w|_?$ be the length of w and the number of ?s in w, respectively. Then $T(w)$ is called a (p,q)-*Toeplitz word*.

We modify this construction as follows. First we build a $(1,l_1)$-ACR word using the construction from Theorem 3.2. Notice that all elements in it except ? are centres of infinite arithmetic progressions. Now we build a $(2,l_2)$-ACR word, using this construction, and put this word on the places of ? of previously built $(1,l_1)$-ACR word. Continuing this line of reasoning to infinity, we obtain a non-periodic binary word, such that for every integer k there exists l such that w is a (k,l)-ACR word. □

Theorems 3.2 and 3.6 actually show differences between usual powers and abelian powers. Contrary to Theorem 2.1, they give the existence of non-periodic words having central abelian powers. To obtain conditions implying periodicity, we have to set quite strong constraints for the powers.

If in Theorem 3.6 we reverse quantifiers "there exists l" and "for every k", then we obtain an opposite result:

Theorem 3.7. *Let w be a biinfinite word. If there exists an integer l such that for every k the word w is a (k,l)-ACR, then w is periodic.*

Proof. Since for every k there exists an abelian centered k-power of length not greater than l, we take $k = l!$. Then for every position i there exists $l_i \leq l$ such that at this position we have a $2l!$-power of length l_i. So, multiplying period by $l!/l_i$, we obtain a $2l_i$-power (and so also 2-power since $1 \leq l_i$) with period $l!$. Therefore, w has a central 2-power of length $l!$ at every position. As mentioned above, a word having at every position a central abelian square with a fixed period is periodic. □

https://doi.org/10.1051/ita/2011121 Published online by Cambridge University Press

The proof of Theorem 3.7 implies that it can be formulated in the following form:

Theorem 3.8. *Let w be a biinfinite non-periodic (k,l)-ACR word. Then $k < l!$.*

Example 3.9. Here we consider a family of non-periodic $(1,8)$-ACR words obtained by the construction described in Theorem 3.2.

First we build a $(1,4)$-arithmetically centered word:

$$w_i = \begin{cases} 1, & \text{if } i \text{ is not divisible by 2,} \\ ?, & \text{if } i = 4j, j \in \mathbb{Z}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

Instead of each ? one can use either 0 or 1. It is easy to see that each letter is a centre of an arithmetic progression of length 1, 2 or 4.

Applying the Thue-Morse morphism

$$\varphi(0) = 01, \varphi(1) = 10$$

to this sequence, we obtain $(1,8)$-ACR word.

Notice that the bound is optimal, *i.e.*, if w is a biinfinite non-periodic $(1,l)$-ACR word, then $l \geq 8$. The nonexistence of non-periodic $(1,l)$-ACR words for $l < 8$ was proved with a technical case study. It was also checked with a computer (and we are grateful for computer experiments to A. Saarela and K. Saari). Notice that here we have a big difference compared to usual squares. For usual squares the central factorization theorem (Thms. 2.1, 2.2) says that bounded central squares imply periodicity. For abelian squares we have to put an additional condition on lengths of abelian squares to enforce periodicity. However, this result for lengths of abelian squares distinguishing periodicity and non-periodicity is similar to the results from [6] on bounded left squares implying periodicity, which we mentioned in Section 2.1.

Remark 3.10. We emphasize that all statements from this paragraph can be reformulated for right infinite words as follows. In these statements instead of "periodicity" we can write "ultimate periodicity", and we suppose that all conditions for abelian repetitions hold starting from some position. For example, the word $w \in \Sigma^\omega$ is (k,l)-ACR, if there exists i_0 such that for every $i \geq i_0$ there is a central $2k$-power with length of period at most l at position i.

4. Right and left abelian powers

In this section we discuss abelian powers occurring immediately to the right (resp., to the left) from each position. Similarly to usual powers, right and left

https://doi.org/10.1051/ita/2011121 Published online by Cambridge University Press

abelian powers give weaker conditions for periodicity than central abelian powers. However, some conditions can be derived to guarantee it.

Note first, that for every k there exist non-periodic biinfinite $(k, 2(k+1)^2)$-ARR (ALR) words. Moreover, there exists a non-periodic biinfinite word, such that for every k there exists l for which w is a (k,l)-ARR (ALR) word. This follows from Theorems 3.2, 3.6 and an observation that every (k,l)-ACR word is (k,l)-ARR and ALR. So, the results of the previous section imply that l equal to $2(k+1)^2$ is enough for the existence of non-periodic (k,l)-ARR (ALL) words. Actually for right (resp. left) k-powers a smaller bound for l exists:

Theorem 4.1. *There exist biinfinite non-periodic $(k, k+1)$-ARR (ALR) words.*

Proof. We will prove that every w of the form $\{01^{k-1}, 01^k\}^\omega$ is a $(k, k+1)$-ARR (ALR) word.

Denote $a = 01^{k-1}$, $b = 01^k$, then $w \in \{a, b\}^\omega$. To prove this theorem, we will check that every position starts with an abelian k-power of length 1, k or $k+1$ depending on its location in a or b. Consider a position i in an arbitrary occurrence of b. Let after this occurrence of b in w we have a factor consisting of j letters b and $k-j$ letters a.

At the position $i = 0$ we have a right k-power with length of period k. To prove it, first note that the number of occurrences of the letter 0 in the factor of length k^2 starting at position 0 does not depend on j and is equal to k. Any factor of length k of w contains at most one 0, hence we have an abelian right k-power with an abelian period 01^{k-1}.

The position $i = 1$ starts with a right k-power with length of period 1.

For $1 < i < k+1$, $j < i-1$ we have at the position i a right k-power with length of period k. The factor of length $j(k+1)+(k-j)k$ starting at position $k+1$ and consisting of j letters b and $k-j$ letters a contains k letters 0. The suffix of length $k-1$ of this factor is 1^{k-1}, so if we cut this suffix, then we get that the factor v of length $j(k+1)+(k-j)k-(k-1) = k^2-k+1+j$ starting at position $k+1$ contains k letters 0. So the factor of length k^2 starting at the position i is obtained from v by adding 1^{k+1-i} in the beginning and 1^{i-j-2} in the end. So it contains k letters 0. Any factor of length $k+1$ of w contains at least one 0, hence we have an abelian right k-power with abelian period 01^{k-1}.

If $j \geq i-1$, then the position i starts with a right k-power with length of period $k+1$. As above, the factor of length $j(k+1)+(k-j)k$ starting at position $k+1$ and consisting of j letters b and $k-j$ letters a contains k letters 0. The suffix of length $k-1$ of this factor is 1^{k-1}. The factor of length $k(k+1)$ starting at the position i is obtained from v by adding 1^{k+1-i} in the beginning and cutting suffix 1^{j-i+1}, and thus also contains k letters 0. Any factor of length k of w contains at most one 0, hence we have an abelian right k-power with abelian period 01^k.

Consider a position i in an arbitrary occurrence of a. If $i = 0$, then we have an abelian right k-power with length of period $k+1$. For $i > 0$ the proof is the same as for the position $i+1$ inside an occurrence of b. □

We consider separately the minimal case $k = 2$.

https://doi.org/10.1051/ita/2011121 Published online by Cambridge University Press

Example 4.2. We analyze non-periodic biinfinite $(2,2)$- and $(2,3)$-ARR binary words.

Let w be a $(2,2)$-ARR word. If w contains the factor 010, then its suffix is $(10)^\omega$. If w contains the factor 100, then its suffix is of the form $(1100)^*(10)^\omega$ or $(1100)^\omega$. Denote by $\Sigma^{-\omega}$ the set of left infinite words. So, w is contained in the following set of words: $0^{\mathbb{Z}}$, $(01)^{\mathbb{Z}}$, $(1100)^{-\omega}(10)^\omega$, $0^{-\omega}(1100)^\omega$, $(0)^{-\omega}(0011)^*(01)^\omega$, $(0)^{-\omega}(1100)^*(10)^\omega$ (up to renaming 0 and 1). Though some of these words are non-periodic in the sense of our definition, they have a structure similar to ultimately periodic right infinite words. So, we will say that a biinfinite word w is ultimately periodic, if there exist finite words u, v_1, v_2, such that $w = v_1^{-\omega}uv_2^\omega$.

The proof of Theorem 4.1 implies that $\{01, 011\}^{\mathbb{Z}}$ give a family of non-periodic biinfinite $(2,3)$-ARR words. Notice that this family gives uncountably many words.

So, in the case of right abelian squares the bound distinguishing periodic and non-periodic words is the following: for $l = 1$ there exist only periodic words, namely, words of the form $a^{\mathbb{Z}}$, where $a \in \Sigma$. For $l = 2$ there exist only periodic and ultimately periodic words, for $l \geq 3$ there exist non-periodic words. Remark that we proved the optimality of the bound $l = k + 1$ for the existence of non ultimately periodic (k,l)-ARR words for $k \leq 6$ with a technical case study.

5. Examples

In this section we consider the examples of the Sturmian words and the Thue-Morse word. These words have quite strong properties concerning the balance between number of 0's and 1's in factors. Therefore, they are good candidates for non-periodic words with abelian repetitive properties.

Example 5.1 (Sturmian words)**.** Let s be a Sturmian word. Then

- s is not (k,l)-ACR for any pair (k,l);
- for every k there exists l such that s is (k,l)-ARR.

The nonexistence of bounded centered abelian squares follows from the fact that a Sturmian word contains factors of the form $u01u$, where u is a special palindromic factor, and the lengths of bispecial palindromic factors are unbounded (see *e.g.* [8], Chap. 2).

The existence of bounded right abelian powers follows from the following theorem proved by Richomme *et al.*:

Theorem 5.2 (Thm. 6.1. in [10])**.** *For every Sturmian word s and every positive integer k, there exist two integers l_1 and l_2 such that each position in s begins in an abelian k-power $u_1u_2\dots u_k$ with abelian period l_1 or l_2, that is $|u_i| \in \{l_1, l_2\}$. In particular, every Sturmian word begins in an abelian k-power for all positive integers k.*

https://doi.org/10.1051/ita/2011121 Published online by Cambridge University Press

Example 5.3 (the Thue-Morse word)**.** For the Thue-Morse word t we will prove that

- t is $(1, 14)$-ACR and not (k, l)-ACR for every $k > 1$ or $l < 14$;
- t is $(2, 5)$-ARR and not (k, l)-ARR for every $k > 2$ or $l < 5$[1].

These facts are proved using the definition of the Thue-Morse word via binary expansion and Lemma 3.1. Recall that the Thue-Morse word can be defined in the following way: $t_i = 1$, if the number of ones in the binary expansion of i is odd, and $t_i = 0$ otherwise. In this definition the starting index is 0.

First, we prove that the Thue-Morse word has bounded centered abelian squares at every position. At every odd position (between blocks) t has a centered abelian square with length of period 2. At even positions (inside blocks) centered abelian squares with odd periods do not exist. Odd periods are impossible, since a word of odd length to the right from an even position contains several full blocks and a letter a immediately to the right from the position, and a word to the left from this even position contains several full blocks and a letter $\bar{a}$ immediately to the left from the position. So these words cannot be abelian equivalent. It is easy to see that t contains an abelian square at a position $2i$ with period $2l$ if and only if t contains at the position i an arithmetic progression of length 3 with difference l. So, it is enough to prove that every position i is the centre of an arithmetic progression of length 3 with bounded difference. To prove it, we will consider several cases. Denote the binary expansion of i by $f(i)$. We will consider different forms of $f(i)$ and in each case we will find a number m, such that adding and subtracting $f(m)$ from $f(i)$ does not change the parity of 1's. So the elements t_{i-m}, t_i, t_{i+m} form an arithmetic progression of one letter.

If $f(i)$ ends with 10^{2j}, $j \geq 1$, then we take $f(m) = 11$. Then $f(i) + f(m)$ ends with $10^{2j-2}11$, $f(i) - f(m)$ ends with $01^{2j-2}01$, the beginning is the same as $f(i)$.

If $f(i)$ ends with 10^{2j+1}, or $10^{2j}10$, or $01^{2j}01$, or 01^{2j+1}, $j \geq 1$, then we take $f(m) = 110$.

If $f(i)$ ends with 01^{2j}, or $10^{2j}1$, or $01^{2j}0$, $j \geq 1$, then we take $f(m) = 11$.

If $f(i)$ ends with $10^{2j+1}1$, or $01^{2j+1}0$, $j \geq 1$, then we take $f(m) = 111$.

If $f(i)$ ends with $10^{2j+1}10$, or $01^{2j+1}01$, $j \geq 0$, then we take $f(m) = 101$.

It is not difficult to check that all possible endings of $f(i)$ are considered.

The maximal m we used is 7, $f(7) = 111$. It is easy to see that smaller m is not possible in $f(i)$ ending with $10^{2j+1}1$, $j \geq 1$. So the Thue-Morse word is $(1, 14)$-ACR word, and it is not $(1, l)$-ACR for $l < 14$.

To prove the nonexistence of bounded abelian centered 4-powers, we first notice that at positions inside blocks periods of centered abelian 4-powers should be even (since it holds already for squares). So in $\mu^{-1}(t)$ there should be centered arithmetic progressions of length 5 with bounded difference at every position, and hence right arithmetic progressions with bounded difference of length 3. We will prove that the existence of right bounded abelian progressions of length 3 is

[1]Remark that the right abelian powers for the Thue-Morse word were also studied independently and simultaneously in [4].

https://doi.org/10.1051/ita/2011121 Published online by Cambridge University Press

impossible. Namely, we will prove that the positions of the form $i = 2^j - 1$ start with arithmetic progressions of length 3 only with growing differences. Let m be the difference of an arithmetic progression of length 3 starting in i. Notice that $f(i) = 1^j$, $f(2m) = f(m)0$. We can take i as large as needed, while m is bounded. So for i large enough the number of 1's in $1^j + f(m)$ and $1^j + f(m)0$ differs by 1 and hence $t_{i+m} \neq t_{i+2m}$.

Now we proceed to the proof of the nonexistence of bounded abelian right cubes. Suppose that there exists l_0 such that t is $(3, l_0)$-ARR. Similarly to even periods of centered squares inside blocks, we get for right cubes that in the positions inside blocks the periods cannot be odd. So, if at a position $2i$ the Thue-Morse word contains an abelian cube of length $2l \leq l_0$, then $\mu^{-1}(t) = t$ at the position i starts with an arithmetic progression of length 4 with difference $l \leq l_0/2$, which as we have just seen is impossible at some positions.

It remains to prove the existence of bounded right abelian squares. Between blocks we have right abelian powers of any length with period 2. We will prove that at positions inside blocks we have bounded squares with odd periods. Considering $\mu^{-1}(t) = t$, one can see that the existence of abelian square at position $2i$ with odd period l is equivalent to the condition $t_i = 1 - t_{i+l}$. We will prove that in every position we can find such $l \leq 5$. Suppose the converse, *i.e.*, that there exists a position i in which for all odd $l' \leq 5$ we do not have $t_i = 1 - t_{i+l'}$, which means $t_i = t_{i+1} = t_{i+3} = t_{i+5}$. Then a block begins in $i + 1$, i is odd, and hence in the preimage we have $t_{\frac{i+1}{2}} = t_{\frac{i+1}{2}+1} = t_{\frac{i+1}{2}+2}$. This is impossible since the Thue-Morse word is cube-free (see *e.g.* [8], p. 113). The optimality of $l = 5$ is trivially achieved, for example, at position 2.

We would like to emphasize that the Sturmian and the Thue-Morse words show very different properties concerning abelian powers. While Sturmian words have bounded right k-powers for every k, they do not have bounded centered powers (even squares!). The Thue-Morse word has bounded centered abelian squares and right cubes, and does not have bounded centered and right abelian 4-powers.

6. Conclusion and open questions

In this paper we studied abelian versions of critical factorization theorem. These results lie in a streamline of research on connections between local and global regularities. We consider such problems in infinite words, where the global regularity is specified as a periodicity, and a local regularity as an abelian central, right or left repetitivity. We considered centered abelian powers and powers to the right and to the left from each position. First, we proved that for every k there exist non-periodic $(k, 2(k+1)^2)$-ACR words (Thm. 3.2). These results show big differences compared to usual powers (Thms. 2.1, 2.2), where bounded centered squares imply periodicity. Secondly, we established the existence of a word having for every k uniformly bounded centered abelian k-powers at every position, the bound depends on k (Thm. 3.6). On the other hand, we established the nonexistence of a word having for any k centered abelian powers bounded by a fixed l (Thm. 3.7).

https://doi.org/10.1051/ita/2011121 Published online by Cambridge University Press

Thirdly, we found a construction for $(k, k+1)$-ARR (ALR) words (Thm. 4.1). Besides that, we studied the abelian repetitive properties of Sturmian and the Thue-Morse words.

There remain some open questions about the abelian versions of the critical factorization theorem. Below we summarize some of those.

Open problem 6.1. For any k, find l (depending on k) such that there exists a non-periodic (k, l)-ACR (ARR) word, and all $(k, l-1)$-ACR (ARR) words are periodic.

For abelian squares the answer is given by Examples 3.9 and 4.2. We have that for any k there exists an l such that a non-periodic (k, l)-ACR (ARR) word exists (see Thms. 3.2, 4.1). The general straightforward way to obtain such l is as follows: for every $i = 2k, 4k, 6k, \ldots$ we build the directed graph whose vertices are all words of length i having an abelian $2k$-power in the middle. There is an arc from a vertex u to a vertex v, if v is obtained from u by deleting first letter and adding one letter in the end of the word. Actually, this graph is a subgraph of De Bruijn graph of order i. Each path in this graph corresponds to a word. If the graph does not contain two intersecting cycles, then there exist only periodic words for this length and we increase i by $2k$. If it does, then there exists a non-periodic $(k, i/2k)$-centered word and for the smallest such i we have $l = i/2k$. Remark that this graph construction implies that if for some length we have non-periodic words, then we have uncountably many words, since we have free choice of a cycle at infinitely many points. But of course this way is ineffective, since the number of all words of length i is exponential. Notice that we do not know examples which give smaller l than constructions from Theorems 3.2, 4.1.

Open problem 6.2. Let k be an integer. Does there exist a non-periodic word w and integers l_1, l_2, such that w contains a central abelian $2k$-power of length l_1 or l_2 at every position?

Notice that the construction from Theorem 3.2 gives an example of such a word using three periods. One period trivially implies periodicity. So two periods is an open question. Note also that for right (left) powers the answer to the question is given by Theorem 5.2, which implies that for every k Sturmian words have abelian k-powers with one of two periods at every position.

Open problem 6.3. Critical factorization theorem for abelian periods.

We will say that a biinfinite word w is *abelian periodic* if there exists a finite word v such that w can be written as a concatenation $w = \ldots v_i v_{i+1} \ldots$, where $v_i \approx^{ab} v$ for every i. The problem is to find connections between abelian periodicity and *ACR*, *ARR*, *ALR* properties, *i.e.*, find abelian analogues of the critical factorization theorem where not only powers, but also the period is considered in the abelian sense.

https://doi.org/10.1051/ita/2011121 Published online by Cambridge University Press

References

[1] S.V. Avgustinovich and A.E. Frid, Words avoiding abelian inclusions. *J. Autom. Lang. Comb.* **7** (2002) 3–9.

[2] Y. Césari and M. Vincent, Une caractérisation des mots périodiques. *C.R. Acad. Sci. Paris, Ser. A* **286** (1978) 1175–1177.

[3] J. Cassaigne and J. Karhumäki, Toeplitz words, generalized periodicity and periodically iterated morphisms. *Eur. J. Comb.* **18** (1997) 497–510.

[4] J. Cassaigne, G. Richomme, K. Saari and L.Q. Zamboni, Avoiding Abelian powers in binary words with bounded Abelian complexity. *Int. J. Found. Comput. Sci.* **22** (2011) 905–920.

[5] J.-P. Duval, Périodes et répetitions des mots du monoide libre. *Theoret. Comput. Sci.* **9** (1979) 17–26.

[6] J. Karhumäki, A. Lepistö and W. Plandowski, Locally periodic versus globally periodic infinite words. *J. Comb. Th. (A)* **100** (2002) 250–264.

[7] A. Lepistö, *On Relations between Local and Global Periodicity*. Ph.D. thesis (2002).

[8] M. Lothaire, *Algebraic combinatorics on words*. Cambridge University Press (2002).

[9] F. Mignosi, A. Restivo and S. Salemi, Periodicity and the golden ratio. *Theoret. Comput. Sci.* **204** (1998) 153–167.

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

[11] K. Saari, Everywhere α-repetitive sequences and Sturmian words. *Eur. J. Comb.* **31** (2010) 177–192.

[12] O. Toeplitz, Beispiele zur theorie der fastperiodischen Funktionen. *Math. Ann.* **98** (1928) 281–295.

Communicated by G. Richomme.
Received November 2, 2010. Accepted August 2, 2011.

https://doi.org/10.1051/ita/2011121 Published online by Cambridge University Press