Abelian bordered factors and periodicity
Extended abstract

Émilie Charlier[*], Tero Harju[†], Svetlana Puzynina[‡] and Luca Q. Zamboni[§]

June 2014

Abstract

A finite word is *bordered*, if it has a non-empty proper prefix which is equal to its suffix, and *unbordered* otherwise. Ehrenfeucht and Silberger [3] proved that an infinite word is (purely) periodic if and only if it contains only finitely many unbordered factors. We are interested in an abelian modification of this fact; namely, we have the following question: Let w be an infinite word such that all sufficiently long factors are abelian bordered. Is w (abelian) periodic? We also consider a *weakly abelian* modification of this question, when only the frequencies of letters are taken into account. Besides that, we answer a question from [1] concerning abelian central factorization theorem.

1 Preliminaries

In this section we give some basics on words following terminology from [4] and introduce our notions.

Given a finite non-empty set Σ (called the alphabet), we denote by Σ^* and Σ^ω, respectively, the set of finite words and the set of (right) infinite words over the alphabet Σ. Given a finite word $u = u_1 u_2 \cdots u_n$ 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 $|\varepsilon| = 0$. Given the words w, x, y, z such that $w = xyz$, x is called a *prefix*, y is a *factor* and z a *suffix* of w. The word x is a *proper prefix* if $0 < |x| < |w|$. The factor of w starting at position i and ending at position j is denoted by $w[i, j] = w_i w_{i+1} \ldots w_j$. An infinite word w is *ultimately periodic*, if for some finite words u and v it holds $w = uv^\omega = uvv \cdots$; w is *purely periodic* (or briefly *periodic*) if $u = \varepsilon$. An infinite word is *aperiodic* if it is not ultimately periodic.

Given a finite word $u = u_1 u_2 \cdots u_n$ with $n \geq 1$ and $u_i \in \Sigma$, for each $a \in \Sigma$, we let $|u|_a$ denote the number of occurrences of the letter a in u. Two words u and v in Σ^* are *abelian equivalent*, denoted $u \sim_{ab} v$, if and only if $|u|_a = |v|_a$ for all $a \in \Sigma$. It is easy to see that abelian equivalence is an equivalence relation on Σ^*. An infinite word w is called *abelian (ultimately) periodic*, if $w = v_0 v_1 \cdots$, where $v_k \in \Sigma^*$ for $k \geq 0$, and $v_i \sim_{ab} v_j$ for all integers $i, j \geq 1$.

For a finite non-empty word $w \in \Sigma^*$, we define *frequency* $\rho_a(w)$ of a letter $a \in \Sigma$ in w as $\rho_a(w) = \frac{|w|_a}{|w|}$.

[*]Université de Liège, Belgium, echarlier@ulg.ac.be, supported in part by the Academy of Finland under FiDiPro grant

[†]University of Turku, Finland, harju@utu.fi

[‡]Corresponding author, University of Turku, Finland, and Sobolev Institute of Mathematics, Russia, svepuz@utu.fi, supported in part by the Academy of Finland under grant 251371 and by Russian Foundation of Basic Research (grants 12-01-00448 and 12-01-00089)

[§]Université de Lyon 1, France zamboni@math.univ-lyon1.fr, supported in part by the Academy of Finland under FiDiPro grant

Definition 1. *An infinite word w is called* weakly abelian (ultimately) periodic*, if $w = v_0 v_1 \dots$, where $v_i \in \Sigma^*$, $\rho_a(v_i) = \rho_a(v_j)$ for all $a \in \Sigma$ and all integers $i, j \geq 1$.*

In other words, a word is weakly abelian periodic if it can be factorized into words of possibly different lengths with the same frequencies of letters. For more on weak abelian periodicity see [2]. In what follows we usually omit the word "ultimately". An infinite word w is called *bounded weakly abelian periodic*, if it is weakly abelian periodic with bounded lengths of blocks, i. e., there exists C such that for every i we have $|v_i| \leq C$.

We make use of the following geometric interpretation of weak abelian periodicity. We translate an infinite word $w = w_1 w_2 \dots \in \Sigma^\omega$ to a graph visiting points of the lattice $\mathbb{Z}^{|\Sigma|}$ by interpreting letters of w as drawing instructions. In the binary case, we associate 0 with a move by vector $\mathbf{v}_0 = (1, -1)$, and 1 with a move $\mathbf{v}_1 = (1, 1)$. We start at the origin $(x_0, y_0) = (0, 0)$. At step n, we are at a point (x_{n-1}, y_{n-1}) and we move by a vector corresponding to the letter w_n, so that we come to a point $(x_n, y_n) = (x_{n-1}, y_{n-1}) + v_{w_n}$, and the two points (x_{n-1}, y_{n-1}) and (x_n, y_n) are connected with a line segment. We let g_w denote the corresponding graph. So, for any word w, its graph is a piecewise linear function with linear segments connecting integer points (see an example on Figure 1). A weakly abelian periodic word w has a graph with infinitely many integer points on some line with a rational slope. We remark that instead of the vectors $(1, -1)$ and $(1, 1)$, one can use any other pair of noncollinear vectors $\mathbf{v}_0$ and $\mathbf{v}_1$. For a k-letter alphabet one can consider a similar graph in $\mathbb{Z}^k$.

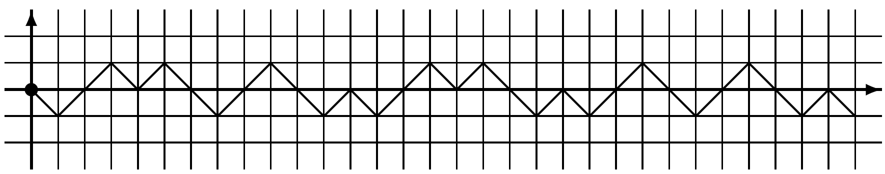

Figure 1: The graph of the Thue-Morse word with $\mathbf{v}_0 = (1, -1)$, $\mathbf{v}_1 = (1, 1)$.

A finite word u is *bordered*, if there exists a non-empty word z which is a proper prefix and a suffix of u. A finite word u is *abelian bordered*, if it has a non-empty prefix which is abelian equivalent to its suffix. Abelian bordered words has been recently considered in [5]. A finite word u is *weakly abelian bordered*, if it has a non-empty proper prefix and a suffix with the same frequencies of letters.

The following theorem states a well known connection between periodicity and unbordered factors:

Theorem 1. [3, 6] *An infinite word w is periodic if and only if there exists a constant C such that every factor v of w with $|v| \geq C$ is bordered.*

In the next section we discuss similar connections in the abelian and weakly abelian setting. We provide the results without proofs, sometimes giving only a rough sketch of the proofs.

2 Main results

First we show that finitely many abelian unbordered factors is not a sufficient condition for the periodicity:

Proposition 1. *There exists an infinite aperiodic word w and a constant C such that every factor v of w with $|v| \geq C$ is abelian bordered.*

For example, any aperiodic word $w \in \{010100110011, 0101001100110011\}^\omega$ satisfies the condition with $C = 15$. The proof can be done by a fairly straightforward case study.

On the other hand, it is not hard to find an example showing that abelian periodicity does not imply that a word has finitely many abelian unbordered factors:

Example 1. Consider the Thue-Morse word $t = 0110100110010110\cdots$ defined as a fixed point of the morphism $0 \to 01$, $1 \to 10$. Clearly, Thue-Morse word is abelian periodic with period 2. Although it has infinitely many abelian unbordered factors. It can be proved that it has factors of the form $0p1$, where p is a palindrome, and these factors are clearly abelian unbordered.

We do not know whether the converse is true: Let w be an infinite word with finitely many abelian unbordered factors; does it follow that w is abelian periodic (see Question 1)? However, we are able to answer a similar question in a weak abelian setting. We proceed with the following theorem relating weak abelian borders and weak abelian periodicity, giving one way analog of the characterization of periodicity in terms of unbordered factors (Theorem 1):

Theorem 2. *Let w be an infinite binary word. If there exists a constant C such that every factor v of w with $|v| \geq C$ is weakly abelian bordered, then w is bounded weakly abelian periodic. Moreover, its graph lies between two rational lines and has points on each of these two lines with bounded gaps.*

The proof of the theorem relies heavily on the graph representation of the word. It consists of several lemmas restricting the form of the word. Namely, we first prove that if a factor $w_i \cdots w_j$ of w is such that the graph of the word between i and j lies above or below the line segment connecting the points $(i, g_w(i))$ and $(j, g_w(j))$, then $w_i \cdots w_j$ is weakly abelian unbordered:

Lemma 1. *Let w be an infinite binary word, i, j be integers, $i < j$. If $g_w(k) > \frac{g_w(i)j - g_w(j)i}{j-i}k + \frac{g_w(j)-g_w(i)}{j-i}$ for each $i < k < j$, then the factor $w[i..j]$ is weakly abelian unbordered.*

We also note that a symmetric assertion holds for the case when the graph lies below this line: simply invert the inequality in the statement of the lemma.

A word $w \in \Sigma^\omega$ is called *K-balanced*, if for each letter a and two factors u, v of w such that $|u| = |v|$ the inequality $||u|_a - |v|_a| \leq K$ holds. We then show that the word satisfying the conditions of Theorem 2 is K-balanced for some constant K, and that the frequencies of letters must be rational::

Lemma 2. *Let w be as in Theorem 2. Then w is K-balanced for some K.*

Lemma 3. *Let w be as in Theorem 2. Then the frequencies of letters are rational.*

Finally, we show that the graph of the word lies between two lines with the same rational slope and has points on each of these two lines with bounded gaps.

Except for the case $\rho_0 = \rho_1 = 1/2$ (equal frequencies), we do not know whether the converse of Theorem 2 is true (see Question 2).

3 A non-abelian periodic word with bounded abelian square at each position

In [1], the following open question was proposed: Let w be an infinite word and C be an integer such that each position in w is a centre of an abelian square of length at most C. Is w abelian periodic? We answer this question negatively by providing an example (actually, a family of examples).

Consider a family of infinite words of the following form:

$$(0001010101110001110000(111000)^*111010101)^\omega$$

A straightforward case study shows that words of this form have abelian square of length at most 12 at each position. It is not hard to see that this family contains abelian aperiodic words.

4 Conclusions and open questions

We conclude with the summary of our results and propose two open problems.

Question 1. *Let w be an infinite word and C a constant such that every factor v of w with $|v| \geq C$ is abelian bordered. Does it follow that w is abelian periodic?*

Recall that we showed that there exist examples of words satisfying these conditions which are not periodic (Proposition 1), but all examples we have are abelian periodic.

The next question asks whether the converse of Theorem 2 is true:

Question 2. *Let w be an infinite binary bounded weakly abelian periodic word such that its graph lies between two rational lines and has points on each of these two lines with bounded gaps. Does it follow that it has only finitely many weakly abelian unbordered factors?*

We remark that in the case of frequencies of letters equal to 1/2 the answer to this question is positive. In this case it could be easily seen from the graph of the word. In fact, one can prove that all sufficiently long factors have weak abelian borders with frequencies 1/2. But we do not know whether it holds for other frequencies.

Our main results and open questions are summarized in Figure 1. We remind that we mostly focus on the binary case (more precise statements are in the text).

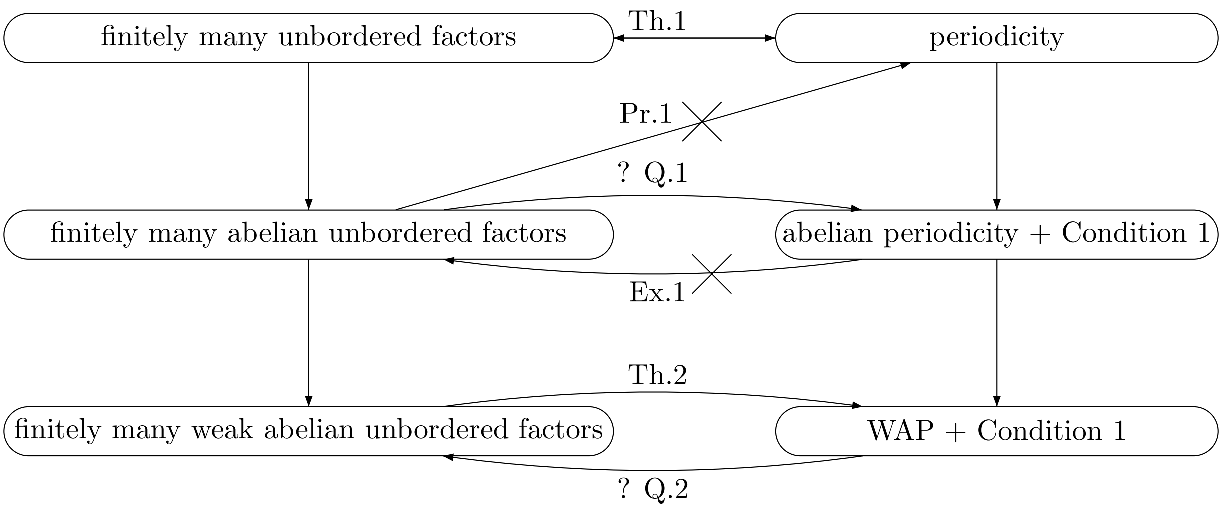

Figure 2: Results and open questions. Condition 1: the graph of the word lies between two rational lines and has points on each of these two lines with bounded gaps.

References

[1] S. V. Avgustinovich, J. Karhumäki, S. Puzynina: On abelian versions of critical factorization theorem. RAIRO - Theor. Inf. and Applic. 46(1): 3–15 (2012).

[2] S. V. Avgustinovich, S. Puzynina: Weak Abelian Periodicity of Infinite Words. LNCS 7913 (CSR 2013): 258–270.

[3] A. Ehrenfeucht and D. Silberger, Periodicity and unbordered segments of words, Discrete Math. **26** (1979), 101–109.

[4] M. Lothaire. Algebraic combinatorics on words. Cambridge University Press, 2002.

[5] N. Rampersad, M. Rigo, P. Salimov: On the Number of Abelian Bordered Words. LNCS 7907 (DLT 2013): 420–432.

[6] K. Saari, Lyndon words and Fibonacci numbers, J. Combin. Theory Ser. A **121** (2013), 34–44.