Border Correlation of Binary Words

Tero Harju
Dirk Nowotka
Turku Centre for Computer Science, TUCS,
Department of Mathematics, University of Turku

Turku Centre for Computer Science
TUCS Technical Report No 546
August 2003
ISBN 952-12-1205-5
ISSN 1239-1891

Abstract

The border correlation function $\beta\colon A^* \to A^*$, for $A = \{a, b\}$, specifies which conjugates (cyclic shifts) of a given word w of length n are bordered, i.e., $\beta(w) = b_0 b_1 \ldots b_{n-1}$, where $b_i = a$ or b according to whether the i-th cyclic shift $\sigma^i(w)$ of w is unbordered or bordered. Except for some special cases, no binary word w has two consecutive unbordered conjugates ($\sigma^i(w)$ and $\sigma^{i+1}(w)$). We show that this is optimal: in every cyclically overlap-free word every other conjugate is unbordered. We also study the relationship between unbordered conjugates and critical points, as well as, the dynamic system given by iterating the function β. We prove that, for each word w of length n, the sequence $w, \beta(w), \beta^2(w), \ldots$ terminates either in b^n or in the cycle of conjugates of the word $ab^k ab^{k+1}$ for $n = 2k + 3$.

Keywords: combinatorics on words, border correlation, binary words

TUCS Laboratory
Discrete Mathematics for Information Technology

1 Introduction

A word w is said to be *unbordered* (or *self-uncorrelated* [13]), if the only border of w is the word itself, that is, if $w = uv = vu'$ for a nonempty word v, then $v = w$ and, consequently, $u = u' = \varepsilon$, the empty word. Unbordered words and factors of words play a significant role in some proofs concerning combinatorial properties of words. The questions involving periodicity of finite and infinite words are naturally related to the border structure of words see, e.g., [3, 4, 5, 6, 7, 11]. As another example, we mention that the existence of borders in words appear in the study of coding properties of sets of words as well as in unavoidability studies of words; see, e.g., [1, 13].

In this paper we study the border structure of words with respect to conjugation. We shall consider solely binary words. To this end, we fix our alphabet to be $A = \{a, b\}$. Let A^* denote the monoid of all finite words over A including the empty word, denoted by ε. Let σ be the (cyclic) *shift function* of words, $\sigma(cw) = wc$ for all $w \in A^*$ and $c \in A$. The *border correlation function* $\beta\colon A^* \to A^*$ is defined such that $\beta(w)$ specifies which conjugates of w are unbordered: Let $w \in A^*$ be a words of length n. Then $\beta(w) = b_0 b_1 \ldots b_{n-1}$, where

$$b_i = \begin{cases} a & \text{if } \sigma^i(w) \text{ is unbordered,} \\ b & \text{if } \sigma^i(w) \text{ is bordered.} \end{cases}$$

For example, let $w = aabab$. Then

$$\begin{aligned} &\sigma^0(w) = w = aabab,\ \sigma^1(w) = ababa,\ \sigma^2(w) = babaa, \\ &\sigma^3(w) = abaab,\ \sigma^4(w) = baaba, \end{aligned}$$

and hence $\beta(w) = ababb$, since only $\sigma^0(w)$ and $\sigma^2(w)$ are unbordered.

It is rather easy to show (see Lemma 1) that the image $\beta(w)$ of a binary word w cannot have two consecutive a's (except for some trivial words), that is, for no i are both $\sigma^i(w)$ and $\sigma^{i+1}(w)$ unbordered. In Section 2 we show that the bound given by this fact is optimal. Indeed, we prove that in every cyclically overlap-free word every other conjugate (that is, either $\sigma^i(w)$ or $\sigma^{i+1}(w)$ for each i) is unbordered.

There is a close relationship between unbordered conjugates of a word and its critical points, when the latter are defined modulo cyclic shifts. This relation is elaborated on in Section 3.

In Section 4 we shall study the dynamic system given by the border correlation function β. We prove that, for each word w of length n, the sequence $w, \beta(w), \beta^2(w), \ldots$ terminates either in the word b^n or in the cycle of the conjugates of the word $ab^k ab^{k+1}$ for $k = (n-3)/2$.

The border correlation function provides a similarity function among the strings. Related functions of similarity are the *auto-correlation* function of Guibas and Odlyzko [8], and the *border-array* function of Miller, Moore, and Smyth [12].

We end this section with some definitions and notations needed in the rest of the paper. We refer to Lothaire's book [11] for more basic and general definitions of combinatorics on words.

We denote the length of a word w by $|w|$. Also, if $w \in A^*$ and $c \in A$, then $|w|_c$ denotes the number of occurrences of letter c in w. For instance, we have for $w = abaab$ that $|w|_a = 3$ and $|w|_b = 2$. A word u is a *factor* of a word w, if $w = w_1 u w_2$ for some words w_1 and w_2. Suppose $w = uv$. Then u is called a *prefix* of w, denoted by $u \le w$, and v is called a *suffix* of w. A nonempty word $u \in A^*$ is a *border* of a word $w \in A^*$, if $w = uv = v'u$ for some suitable nonempty words v and v' in A^*.

We call two words u and v *conjugates*, denoted by $u \sim v$, if $u = \sigma^k(v)$ for some $k \ge 0$. Clearly, $\sim$ is an equivalence relation. Let $[u] = \{v \mid u \sim v\}$ denote the *conjugate class* of u. A word w is *primitive* if it is not a proper power of another word, that is, $w = u^k$ implies $u = w$ and $k = 1$. A word w is called a *Lyndon word* if it is primitive and minimal among all its conjugates with respect to some lexicographic order. In the binary case $A = \{a, b\}$, there are two orders given by $a \lhd b$ and its inverse $b \lhd^{-1} a$. It is well known (see, e.g., Lothaire [11]), that each primitive word w has a unique Lyndon conjugate with respect to a given order. For example, consider $w = abaabb$. Then $aabbab$ and $bbabaa$ are conjugates of w and they are minimal with respect to the order $\lhd$ and $\lhd^{-1}$, respectively. These words are thus Lyndon words.

A word $w \in A^*$ is *overlap-free*, if it does not have overlapping factors, that is, w does not have a factor of the form $axaxa$. Moreover, w is *cyclically overlap-free*, if all its conjugates are overlap-free. The cyclically overlap-free binary words were characterized by Thue [15]; see Section 2.

2 Optimal words for border correlation

Let w be a nonempty word of length n in A^*. If it is not primitive, that is, $w = u^k$ for some u and $k \ge 2$, then it is immediate that all conjugates of w are nonprimitive, and thus bordered. Therefore, $\beta(w) = b^n$ in this case. It is also clear that β is invariant under renaming. That is, if w' is obtained from w by exchanging the letters a and b, then $\beta(w') = \beta(w)$. Therefore β is not injective, and thus not surjective, that is, there are at most 2^{n-1} words of length n that are β-images. In fact, this number is much lower as indicated in Table 1.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	2	4	7	11	18	29	47	76	126	199	316	521	848	1374	2215

Table 1: The number of images $\beta(w)$ for lengths $4 \leq n \leq 16$

The following lemma gives some useful properties of the images $\beta(w)$. By the second case of the lemma, $\beta(w)$ does not contain two adjacent letters a unless w is a conjugate of the special words ab^{n-1} or ba^{n-1}. Notice that $\beta(ab^{n-1}) = aab^{n-2} = \beta(ba^{n-1})$.

Lemma 1. *Let $w \in A^*$ of length n.*

(i) If w is primitive, then $|\beta(w)|_a \geq 2$.

(ii) For each $i = 0, 1, \ldots, n$, $\sigma^i(w)$ or $\sigma^{i+1}(w)$ is bordered, or $w \in [ab^{n-1}]$ or $w \in [ba^{n-1}]$.

(iii) The word w can have at most $\lfloor |w|/2 \rfloor$ unbordered conjugates.

Proof. For (i), we notice, as mentioned in the introduction, that each primitive word w has two Lyndon conjugates. Since Lyndon words are unbordered (see Lothaire [11]), the claim follows.

For (ii), assume that w is not a conjugate of ab^{n-1} nor of ba^{n-1}, and hence, it has at least two occurrences of a and of b. Let $w' = \sigma^i(w)$ be any unbordered conjugate of w. Without loss of generality, we assume that w' begins with a, and, consequently, $w' = ab^k xab^j$, where $j > k$ and the word xa begins with a, since w' is unbordered. (We may have $x = \varepsilon$.) Now, $\sigma(w') = b^k xab^j a$ has a border $b^k a$, and hence, $\sigma^{i+1}(w)$ is bordered, as required.

The claim (iii) is clear from (ii). ☐

In particular, if the length of w is an odd number, then w has two adjacent conjugates that are both bordered.

Example 2. There are words for which the maximum $\lfloor |w|/2 \rfloor$ is obtained. Every second conjugate of w is unbordered, for instance, in the following cases $w = aabb$ and $w = abaabbaababb$. In these examples, $\beta(w) = (ab)^{|w|}$. However, there is no word of length 10 that has 5 unbordered conjugates (see Theorem 5). Also, e.g., for $w = aabbbab$ of odd length, we have $\beta(w) = ababbab$, and hence, $|\beta(w)|_a = 3 = \lfloor |w|/2 \rfloor$ in this case.

There is a close relationship between overlap-free binary words and the maximum number of unbordered conjugates. Theorems 4 and 5 clarify this relation. Before we prove these theorems, let us recall that the *Thue-Morse* morphism [14, 15] $\tau\colon A^* \to A^*$ is defined by $\tau(a) = ab$ and $\tau(b) = ba$.

The following result is due to Thue [15] (see also [9]).

Lemma 3. *Let $w \in A^*$ be a cyclically overlap-free word.*

(i) Also, $\tau(w)$ is cyclically overlap-free.

(ii) Also, $\tau^{-1}(w)$ is cyclically overlap-free if $w \in \{ab, ba\}^*$.

(iii) Either w or $\sigma(w)$ has a factorization in terms of ab and ba, that is, $w \in \{ab, ba\}^*$.

(iv) For some $u \in \{a, b, aab, abb\}$ and $n \geq 0$, $w \in [\tau^n(u)]$. In particular, $|w| = 2^n$ or $3 \cdot 2^n$ for some $n \geq 0$.

Theorem 4 shows that cyclically overlap-free binary words have a maximum number of unbordered conjugates. In the theorem, "every other conjugate of w is unbordered" means, by Lemma 1(iii), that $\beta(w)$ is a $(ab)^{n/2}$ or $(ba)^{n/2}$.

Theorem 4. *Let $w \in A^*$. Every other conjugate of w is unbordered, if, and only if, w is a cyclically overlap-free word.*

Proof. Let w be a word of length n that contains an overlapping factor, i.e., $w = ucxcxcv$, where $c \in A$ and $u, v, x \in A^*$. Let $i = |ucx|$. Then the conjugates $\sigma^i(w) = cxcvucx$ and $\sigma^{i+1}(w) = xcvucxc$ are both bordered, with borders cx and xc, respectively.

In the other direction, suppose that w is cyclically overlap-free word such that both $\sigma(w)$ and $\sigma^2(w)$ are bordered. Clearly, $|w| \geq 4$. We derive a contradiction which proves the claim. Let u be the shortest border of $\sigma(w)$ and v be the shortest border of $\sigma^2(w)$.

We shall assume that $a \leq w$. The case $b \leq w$ is symmetric, and it can be thus omitted.

Case 1: Assume first that $aa \leq w$. Then $u = a$, and $\sigma(w) \in \{ab, ba\}^*$ by Lemma 3(iii). It follows that $aab \leq w$, and hence $w = aabw_0b$ where $w_0 \in \{ab, ba\}^*$ and the τ-factorization of $\sigma(w)$ is given by $\sigma(w) = (ab)w_0(ba)$. Now, $\sigma^2(w) = bw_0baa$. Note that $v \neq baa$ for the border v of $\sigma^2(w)$, because $w_0 \in \{ab, ba\}^*$. Consequently, $v = bv'baa$ for some $v' \in A^*$. Since $\sigma^2(w) = vzv$ for some nonempty z, and $\sigma(w) \in \{ab, ba\}^*$, w has a conjugate $vvby$ (where $z = by$). This is a contradiction, since v begins with b and so vvb is not overlap-free.

Case 2: Assume that $ab \leq w$. We have now that bb is a suffix of w, since w is unbordered. Therefore again $\sigma(w) \in \{ab, ba\}^*$ which implies that $u = ba$, and also $aba \leq w$, say $w = abaw_0b$. We have $w = abaw_1bb$, since w is unbordered. Moreover, $w = abaw_2abb$, since $\sigma(w) \in \{ab, ba\}^*$. Actually, $w = abaabw_3abb$, since $\tau^{-1}(\sigma(w))$ is cyclically overlap-free by Lemma 3(ii) and thus it is also in $\{ab, ba\}^*$. So, we have the following τ-factorization $\sigma(w) = (ba)(ab)w_3(ab)(ba)$, where $w_3 \in \{ab, ba\}^*$. Now, the shortest border v of $\sigma^2(w)$ is either (2a) $v = aabbab$ or (2b) $v = aabv'abbab$ for some word v'. In Case (2a), we have $\sigma^2(w) = aabbabw_4aabbab$, where $w_4 \neq \varepsilon$ (for, otherwise, $\tau^{-1}(\sigma(w)) \notin \{ab, ba\}^*$). Hence, $\sigma(w) = (ba)(ab)(ba)(bw_4a)(ab)(ba)$ and so $w_4 = aw_5b$, that is,

$$\sigma(w) = (ba)(ab)(ba)(ba)w_5(ba)(ab)(ba)\,,$$

and thus $\tau^{-1}(\sigma(w) = babb\tau^{-1}(w_5)bab$, and therefore $babbabb$ is a factor in a conjugate of the preimage $\tau^{-1}(\sigma(w))$ contradicting the overlap-freeness requirement. In Case (2b), we have that $vvay$ occurs in a conjugate of w. This is a contradiction, since v begins with a, and thus vva is an overlapping factor. This completes the proof of the theorem. $\square$

The next theorem shows that words with a maximum number of unbordered conjugates are essentially overlap-free.

Theorem 5. *Let $n \geq 1$. Every word of length $2n$ that has n unbordered conjugates is either cyclically overlap-free or a conjugate of $abbb$ or $aaab$.*

Proof. Note that $\beta(abbb) = aabb$ and $\beta(aaab) = abba$. The claim follows easily now from Lemma 1 and Theorem 4. $\square$

Theorems 4 and 5 show that every word with a maximum number of unbordered conjugates is cyclically overlap-free, except for the conjugates of $abbb$ and $aaab$. By Lemma 3(iv), each such word has length either 2^n or $3 \cdot 2^n$ for some $n \geq 1$.

3 Unbordered Conjugates and Critical Factorizations

In this section we investigate the relation between the border correlation function and critical factorizations. First we introduce the critical points of words.

Let $w = a_0a_1 \ldots a_{n-1} \in A^*$, where $a_i \in A$ for each i. An integer $1 \leq q \leq n$ is a *period* of w, if $a_i = a_{i+q}$ for all $0 \leq i < n - q$. The smallest period of w is denoted by $\partial(w)$. For instance, $\partial(w) = |w|$, if, and only if, w is unbordered. It is easy to see that q, with $1 \le q \le |w|$, is a period of w, if, and only if, there is a word v of length q such that w is a factor of v^n for some $n \ge 1$. Let for example $w = abaababa$. Then the periods of w are $5, 7$, and $8 = |w|$. In this example, $\partial(w) = 5$.

An integer p with $1 \le p < |w|$ is called *point* in w. Intuitively, a point p denotes the place between a_p and a_{p+1} in w above. A nonempty word u is called a *repetition word* at point p if $w = xy$ with $|x| = p$ and there exist x' and y' such that u is a suffix of $x'x$ and a prefix of yy'. For a point p in w, let

$$\partial(w, p) = \min\{|u| \mid u \text{ is a repetition word at } p\}$$

denote the *local period* at point p in w. Let for example $w = abaabab$. Now, for instance, $\partial(w, 2) = 3$, since the shortest repetition word at $p = 3$ is aab. Indeed, $aw = (aab)(aab)ab$. The shortest repetition words of w for the points $p = 1, 2, \ldots, 6$ are, respectively, $ba, aab, aba, babaa, ab$, and ba. We notice that $\partial(w) = 5 = \partial(w, 4)$.

Note, that the repetition word of length $\partial(w, p)$ at point p is necessarily unbordered and $\partial(w, p) \le \partial(w)$. A factorization $w = uv$, with $u, v \ne \varepsilon$ and $|u| = p$, is called *critical*, if $\partial(w, p) = \partial(w)$, and, if this holds, then p is called *critical point*.

We recall the critical factorization theorem next [11] (see also [10]).

Theorem 6. *Every word w, with $|w| \ge 2$, has at least one critical factorization $w = uv$, with $u, v \ne \varepsilon$ and $|u| < \partial(w)$, i.e., $\partial(w, |u|) = \partial(w)$.*

The following lemma is a consequence of the critical factorization theorem. It is proven in [2].

Lemma 7. *Let $w = uv$ be unbordered and $|u|$ be a critical point. Then vu is unbordered.*

There is no direct relationship between critical points and unbordered conjugates in general, since, for instance, the number of critical points is not invariant under cyclic shifts whereas the border correlation function is; see Remark 12 in the next section. Moreover, if $w = uv$ such that vu is unbordered, then $|u|$ is not a critical point in general.

Example 8. Consider the conjugate class of $w = ababa$

$$[w] = \{ababa,\ babaa,\ abaab,\ baaba,\ aabab\}$$

with 4, 1, 2, 2, and 1 critical points, respectively. However, the word w has exactly two unbordered conjugates $babaa$ and $aabab$.

In general, it is not so that there is a word w' in the conjugate class of some word w such that the critical points of w' mark the unbordered conjugates of w like *babaa* and *aabab* in the above example.

Example 9. Consider the conjugate class of $w = abbabaab$. We have exactly two critical points for every $w' \in [w]$ but four unbordered conjugates in $[w]$.

However, if critical points are considered modulo cyclic shifts, the situation changes. Let w be a word of length n. We call an integer p, with $0 \leq p < n$, an *internal critical point* of w, if $p + n$ is a critical point of www. The following lemma shows that internal critical points are invariant under cyclic shifts.

Lemma 10. *Let w be a word of length n. The point p is internal critical of w, if, and only if, the point $q = p - i \pmod{n}$ is internal critical of $u = \sigma^i(w)$.*

Proof. Clearly, www contains all conjugates of ww. Moreover, it follows from $\sigma(ww) = \sigma(w)\sigma(w)$ that uuu also contains all conjugates of ww. In fact, let $v \in [w]$ such that $v = \sigma^j(w)$, then $vv = \sigma^j(ww)$ and $www = xvvz$ where $|x| = j \pmod{n}$. In particular, $uuu = x'vvz'$, where $|x'| = j - i \pmod{n}$.

Surely, the implication directions of the claim are symmetric to each other. Assume p is an internal critical point of w. Let v be the shortest repetition word at point $p + n$ in www. We have that v is a conjugate of w, since $p + n$ is critical. So, $www = xvvz$ where $|x| = p$. Now, $uuu = x'vvz'$ where $|x'| = p - i \pmod{n}$, and hence, the point $q + n$ is critical, and this proves the claim. $\square$

Theorem 11. *Let w be a primitive word of length n, and let $0 \leq p < n$. Then the following statements are equivalent:*

- *p is an internal critical point of w.*
- *the conjugate $\sigma^p(w)$ is unbordered.*

Proof. Assume p is an internal critical point of w. Then $www = xvvz$ where $|x| = p$ and v is an unbordered factor of length n in ww. Hence, $\sigma^p(w) = v$.

Assume $v = \sigma^p(w)$ is an unbordered conjugate of w. Then $www = xvvz$ with $|x| = p$, and $p + n$ is a critical point of www. Hence, p is an internal critical point of w. $\square$

4 Iterations of the Border Correlation Function

In this section we investigate iterations of the border correlation function. We start by considering the *β-graph* $G_\beta(n)$ for each $n \geq 1$. It is the directed graph with the set $A^n = \{w \mid |w| = n, w \in A^*\}$ as vertices, and with edges determined by the border correlation function β, that is, there is a (directed) edge $u \to v$, if, and only if, $\beta(u) = v$. In order to avoid trivial exceptions, we assume in this section that $n \geq 3$.

Remark 12. It is straightforward to see that $\beta(\sigma(w)) = \sigma(\beta(w))$, that is, the following diagram commutes.

$$\begin{array}{ccc} w & \xrightarrow{\beta} & u \\ {\scriptstyle\sigma}\downarrow & & \downarrow{\scriptstyle\sigma} \\ w' & \xrightarrow{\beta} & u' \end{array}$$

So, the β-graph $G_\beta(n)$ consists of components where each component contains exactly one cycle, since for all members of one conjugate class $[w]$, the images are mapped to the conjugate class $[\beta(w)]$.

In the following we show that any cycle in the graph $G_\beta(n)$ consists of exactly one conjugate class. Moreover, we describe all conjugate classes that form a cycle.

Let $\kappa\colon A^* \to \mathbb{N}$ where $\kappa(w)$ denotes the minimum k such that ab^ka occurs in any conjugate of w, or w is a conjugate of ab^k, or $w = b^k$. Note, that $k = 0$, if, and only if, a^2 occurs in w or $\sigma(w)$. Let $\mu\colon A^* \to \mathbb{N} \times \mathbb{N}$ be defined such that $\mu(w) = (|w|_a, |w| - \kappa(w))$. Note, that $\mu(w) = \mu(\sigma(w))$. Let $<$ denote the extension of the ordering of natural numbers to the lexicographic order on $\mathbb{N} \times \mathbb{N}$, with other words, $(p, q) < (r, s)$ if $p < r$, or $p = r$ and $q < s$.

Theorem 13. *Let w be a word not in b^* and not in $[ab^k] \cup [ab^kab^{k+1}]$, for all $k \geq 0$. Then $\mu(\beta(w)) < \mu(w)$.*

Proof. Let w be a word of length n that is not in $b^* \cup [ab^{n-1}]$ and not in $[ab^kab^{k+1}]$, for $k = (n-3)/2$. Note, that a occurs at least twice in w. If w is not primitive, then $\beta(w) = b^n$ and, in this case, it is clear that $\mu(\beta(w)) < \mu(w)$. Assume then that w is primitive. Because $\mu(w) = \mu(\sigma(w))$, we can choose any conjugate of w without changing its μ image. Therefore, we can assume that w begins with a and that it is unbordered. For example, we may take the Lyndon word in the conjugate class $[w]$ with respect to the order $a \lhd b$. We have now a unique factorization in the form $w = B_1B_2 \cdots B_r$, where each $B_i = ab^{k_i}$ with $r \geq 2$ and $k_i \geq 0$ for all $1 \leq i \leq r$. Let m be the minimum of all k_i.

Note that $|\beta(w)|_a \leq |w|_a$ by Lemma 1. So, every (occurrence of) letter a in w implies at most one a in $\beta(w)$, since we can get an unbordered conjugate of w only either before or after that occurrence of a, but not in both cases by Lemma 1(ii). If an occurrence of a in w does not imply an a in $\beta(w)$, we say that this occurrence of a is *dropped*.

The claim follows, if $|\beta(w)|_a < |w|_a$, and therefore, we can assume that $|\beta(w)|_a = |w|_a$, that is, no occurrence of a is dropped: for every $i \geq 1$, if the i-th letter in w is an a, then either $\sigma^{i-1}(w)$ or $\sigma^i(w)$ is unbordered. Since w begins with a and is unbordered, we have that $\beta(w) = B'_1 B'_2 \cdots B'_r$, where $B'_i = ab^{k'_i}$ and $k'_i \geq 0$ for all $1 \leq i \leq r$. Note, that the a in B'_i corresponds to the unbordered conjugate of w, if w is factored either before or after the occurrence of a in B_i. We show that $\kappa(w) < \kappa(\beta(w))$ in this case.

Let i be modulo n in the following, and let $j = |B_1 B_2 \cdots B_i|$. If $k_i = k_{i+1}$ then the a in B_{i+1} is dropped, that is, neither $\sigma^j(w)$ nor $\sigma^{j+1}(w)$ is bordered; a contradiction. So, assume that $k_i \neq k_{i+1}$.

Note, that if $k_i > k_{i+1}$ then $\sigma^{j+1}(w)$ is bordered and $\sigma^j(w)$ is unbordered by assumption, and if $k_i < k_{i+1}$ then $\sigma^j(w)$ is bordered and $\sigma^{j+1}(w)$ is unbordered by assumption.

If $k_i > k_{i+1}$ then $k'_i = k_i$, in case $k_{i-1} > k_i$, and $k'_i = k_i - 1$, in case $k_{i-1} < k_i$.

If $k_i < k_{i+1}$ then $k'_i = k_i + 1$.

Now, we have that $|k_i - k'_i| \leq 1$. If $k_i = m$ then $k'_i = k + 1$. However, we get $k'_i = m$, if, and only if, $k_{i-1} = m$ and $k_i = k + 1$ and $k_{i+1} = m$, and $r \geq 4$, since $w \notin [ab^k ab^{k+1}]$ and, by assumption, $|\beta(w)|_a = |w|_a$. Therefore, we also have $k_{i-2} > m$ and $b^{m+1}ab^m ab^{m+1}ab^m a$ occurs in a conjugate of w, and both $\sigma^j(w)$ and $\sigma^{j+1}(w)$ are bordered; a contradiction.

So, $k'_\ell > m$, for all $1 \leq \ell \leq r$, if $|\beta(w)|_a = |w|_a$, and therefore we have $\mu(\beta(w)) < \mu(w)$. □

Lemma 14. *Let $w \in [ab^k ab^{k+1}]$ with $k \geq 0$. Then*

$$[ab^k ab^{k+1}] = \{\beta^i(w) \mid 0 \leq i < |w|\} .$$

Proof. We have that $w = b^r ab^s ab^t$, where either $r + t = k$ and $s = k + 1$, or $r + t = k + 1$ and $s = k$. Now $\beta(w) = b^{r+1}ab^{s-1}ab^t = \sigma^s(w)$ in the former case and $\beta(w) = b^r ab^{s+1}ab^{t-1} = \sigma^{s+1}(w)$ in the latter case. That is, $\beta(w) = \sigma^{k+1}(w)$, and the claim follows, since $2k + 3$ and $k + 1$ are relatively prime. □

We are now ready to show that iterations of β on any binary word result in a word of a certain shape.

Theorem 15. *For every word w, there exists an $i \geq 0$ such that $\beta^i(w) \in b^*$ or $\beta^i(w) \in [ab^k ab^{k+1}]$.*

Proof. Let w be a word of length n. Note, that $\beta(w) = b^n$, if w is not primitive. Assume thus that w is primitive. Note, that if $\mu(w) \neq \mu(u)$ then $[w] \neq [u]$, and that $\beta(w) \notin [ab^{n-1}]$, since w has at least two unbordered conjugates. If $w \in [ab^{n-1}]$ then $\beta(w) \in [aab^{n-2}]$. If $w \in [ab^k ab^{k+1}]$ then $\beta(w) \in [ab^k ab^{k+1}]$ by Lemma 14.

Suppose, w is different from b^n and w is not in $[ab^{n-1}] \cup [ab^k ab^{k+1}]$ for $k = (n-3)/2$. Since the values of μ strictly decrease after an application of β, by Theorem 13, we conclude that there exists an $i \geq 1$ such that $\beta^i(w) = b^n$ or $\beta^i(w) \in [ab^k ab^{k+1}]$. ☐

Consider then the graph $G_\beta^\sim(n)$, which consists of the conjugate classes $[w]$, for $|w| = n$, as its vertices and there is an edge $[u] \to [v]$ if $\beta(u) = v$. By the above results, this graph is well defined, and it consists of trees when disregarding reflexive loops $[u] \to [u]$. (See Figure 1 for the graph $G_\beta^\sim(7)$.)

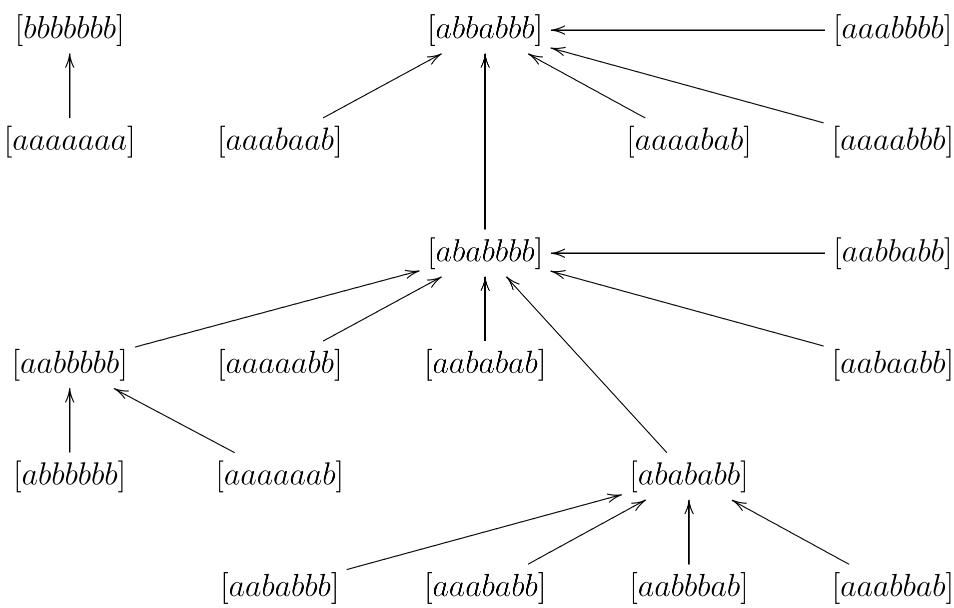

Figure 1: The graph $G_\beta^\sim(7)$. We have omitted the loops of the vertices $[b^7]$ and $[abbabbb]$.

5 Discussion

We have investigated the border correlation function β of binary words. The shape of β images for words with a minimal and maximal number of unbordered conjugates has been clarified. Nevertheless, the set $\beta(A^*)$ has not been completely described. We conjecture that Lemma 1 and the Theorems 4 and 5 in Section 2 describe the range of β. Let $\mathcal{M} = \mathbb{N} \setminus \{2^n, 3 \cdot 2^n \mid n \geq 0\}$.

Conjecture 16.

$$\beta(A^*) = b^*aab^* \cup ab^*a \cup \{w \mid |w|_a \geq 2,\ a^2 \textit{ not in } ww\} \setminus \{(ab)^k, (ba)^k \mid k \in \mathcal{M}\}\,.$$

This conjecture has been checked by a computer program for all words up to length 30.

Apart from the border correlation function β one could investigate an extension $\beta'\colon A^* \to \mathbb{N}^*$ of that function such that a word w of length n is mapped to $m_0 m_1 \cdots m_{n-1}$ where m_i is the length of the shortest border of $\sigma^i(w)$ for all $0 \leq i < n$. We just notice here that β' is injective, since, if $u = wau'$ and $v = wbv'$, then clearly the shortest borders of the $|w|$-th conjugates $au'w$ and $bv'w$ are different, because one of them is equal to 1, and the other is not.

References

[1] J. Berstel and D. Perrin. *Theory of codes*, volume 117 of *Pure and Applied Mathematics*. Academic Press Inc., Orlando, FL, 1985.

[2] D. Breslauer, T. Jiang, and Z. Jiang. Rotations of periodic strings and short superstrings. *J. Algorithms*, 24(2), 1997.

[3] C. Choffrut and J. Karhumäki. Combinatorics of words. In A. Salomaa and G. Rozenberg, editors, *Handbook of Formal Languages*, volume 1, pages 329–438. Springer-Verlag, Berlin, 1997.

[4] W.-F. Chuan. Unbordered factors of the characteristic sequences of irrational numbers. *Theoret. Comput. Sci.*, 205(2):337–344, 1998.

[5] J. C. Costa. Biinfinite words with maximal recurrent unbordered factors. *Theoret. Comput. Sci.*, 290(3):2053–2061, 2003.

[6] J.-P. Duval. Relationship between the period of a finite word and the length of its unbordered segments. *Discrete Math.*, 40(1):31–44, 1982.

[7] A. Ehrenfeucht and D. M. Silberger. Periodicity and unbordered segments of words. *Discrete Math.*, 26(2):101–109, 1979.

[8] L. J. Guibas and A. Odlyzko. String overlaps, pattern matching, and nontransitive games. *J. Combin. Theory Ser. A*, 30(2):183–203, 1981.

[9] T. Harju. On cyclically overlap-free words in binary alphabets. In G. Rozenberg and A. Salomaa, editors, *The Book of L*, pages 125–130. Springer, Berlin, 1985.

[10] T. Harju and D. Nowotka. Density of critical factorizations. *Theor. Inform. Appl.*, 36(3):315–327, 2002.

[11] M. Lothaire. *Algebraic Combinatorics on Words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, United Kingdom, 2002.

[12] D. Miller, D. Moore, and W. F. Smyth. Counting distinct strings. *Algorithmica*, 23(1):1–13, 1999.

[13] H. Morita, A. J. van Wijngaarden, and A. J. Han Vinck. On the construction of maximal prefix-synchronized codes. *IEEE Trans. Inform. Theory*, 42:2158–2166, 1996.

[14] M. Morse. Recurrent geodesics on a surface of negative curvature. *Trans. Amer. Math. Soc.*, 22(1):84–100, 1921.

[15] A. Thue. Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Det Kongelige Norske Videnskabersselskabs Skrifter, I Mat.-nat. Kl. Christiania*, 1:1–67, 1912.

Turku Centre for Computer Science
Lemminkäisenkatu 14
FIN-20520 Turku
Finland

http://www.tucs.fi

University of Turku
- **Department of Information Technology**
- **Department of Mathematics**

Åbo Akademi University
- **Department of Computer Science**
- **Institute for Advanced Management Systems Research**

Turku School of Economics and Business Administration
- **Institute of Information Systems Science**