On Defect Effect of Bi-infinite Words

Juhani Karhumäki
Ján Maňuch
Wojciech Plandowski

Department of Mathematics and
Turku Centre for Computer Science,
University of Turku,
SF-20014 Turku, Finland
karhumak@cs.utu.fi, manuch@cs.utu.fi,
wojtekpl@mimuw.edu.pl

Supported by Academy of Finland under Grant No. 14047.

Turku Centre for Computer Science
TUCS Technical Report No 181
December 1998

ISBN 952-12-0223-8
ISSN 1239-1891

Abstract

We prove the following two variants of the defect theorem. Let X be a finite set of words over a finite alphabet. Then if a nonperiodic bi-infinite word w has two X-factorizations, then the combinatorial rank of X is at most $\mathrm{card}(X) - 1$, i.e. there exists a set F such that $X \subseteq F^+$ with $\mathrm{card}(F) < \mathrm{card}(X)$. Further, if $\mathrm{card}(X) = 2$ and a bi-infinite word possesses two X-factorizations which are not shift-equivalent, then the primitive roots of the words in X are conjugates. Moreover, in the case $\mathrm{card}(F) = \mathrm{card}(X)$, the number of periodic bi-infinite words which have two different X-factorizations is finite and in the two-element case there is at most one such bi-infinite word.

TUCS Research Group
Mathematical Structures of Computer Science

1 Introduction

Defect theorem is one of the fundamental results on words, cf [Lo]. Intuitively it states that if n words satisfy a non-trivial relation, then these words can be expressed as products of at most $n-1$ words. Actually, as discussed in [CK], for example, there does not exist just one defect theorem but several ones depending on restrictions put on the required $n-1$ words.

It is also well-known that the nontrivial relation above can be replaced by a weaker condition, namely by the nontrivial one-way infinite relation. The goal of this note is to look for defect theorems for bi-infinite words. In a strict sense such results do not exist: the set $X = \{ab, ba\}$ of words satisfies a bi-infinite nontrivial relation since $(ab)^{\mathbb{Z}} = (ba)^{\mathbb{Z}}$, but there exists no word ϱ such that $X \subseteq \varrho^+$. However, we are going to prove two results which can be viewed as defect theorems for bi-infinite words.

In terms of factorizations of words defect theorem can be stated as follows: Let $X \subseteq \Sigma^+$ be a finite set of words. If there exists a word $w \in \Sigma^+$ having two different X-factorizations, then the rank of X is at most $\mathrm{card}(X)-1$. Here the rank of X can be defined in different ways, cf again [CK]. For example, it can be defined as a combinatorial rank $r_{\mathrm{c}}(X)$ denoting the smallest number k such that $X \subseteq Y^+$ with $\mathrm{card}(Y) = k$.

To describe our results let w be a bi-infinite word, i.e. an element of $\Sigma^{\mathbb{Z}}$, and X a finite subset of Σ^+. We say that w has an X-factorization if $w \in X^{\mathbb{Z}}$, and that w has two different X-factorizations, if it has two X-factorizations such that they do not match at least in one point of w. We are going to prove the following two results:

(i) If a *nonperiodic* bi-infinite word w has two different X-factorizations, then the combinatorial rank $r_{\mathrm{c}}(X)$ of X is at most $\mathrm{card}(X)-1$. Moreover, if $r_{\mathrm{c}}(X) = \mathrm{card}(X)$, then the number of bi-infinite words with two different X-factorizations is finite.

(ii) Let $\mathrm{card}(X) = r_{\mathrm{c}}(X) = 2$, so that X is a code. If a bi-infinite word w has two different X-factorizations which are not shift-equivalent, then the primitive roots of words in X are conjugates. Moreover, there is at most one bi-infinite word possessing two different X-factorizations.

Note that case *(ii)* is a strict sharpening of case *(i)* for two-element sets. We also want to emphasize that a restriction to nonperiodic words in case *(i)* is necessary, and even more that this theorem requires to consider the combinatorial rank. This seems to be the first result where the defect effect is realized only by the combinatorial rank, and not by the other types of ranks, cf again [CK].

The first part of the result in *(ii)* is related to the main result of [lRlR], and, we believe, deducible from considerations of that paper. However, our proof is self-contained and essentially shorter, and moreover formulated directly to yield a defect-type of theorem.

Our paper is organized as follows.

In Section 2 we fix our terminology and present the auxiliary results needed for our proofs. In Section 3 we prove our general defect theorem for bi-infinite words, i.e. *(i)* above. In Section 4 we prove, as our main result, a defect theorem for binary sets X satisfying a nontrivial bi-infinite relation, i.e. above *(ii)*. To prove this seems to be quite complicated. In Section 5 we prove the second part of *(ii)*, i.e. the uniqueness of the X-ambiguous bi-infinite word in the two-element case. In Section 6 we give a characterization of two-element sets X, which allow an X-ambiguous bi-infinite word. The last section contains conclusions and open problems.

2 Preliminaries

In this section we fix our terminology and recall a few lemmas on combinatorics of words needed for the proofs of our results. For undefined notions we refer to [Lo] or [CK].

Let Σ be a finite alphabet and X a finite subset of Σ^+. The set of all finite, infinite and bi-infinite words over $\Sigma$are denoted by Σ^*, $\Sigma^{\mathbb{N}}$ and $\Sigma^{\mathbb{Z}}$, respectively. Hence, formally a bi-infinite word is a mapping $f_w : \mathbb{Z} \to \Sigma$, usually written as

$$w = \ldots a_{-1}a_0a_1\ldots \qquad \text{with } a_i = f_w(i)\,.$$

An X-factorization of w is any sequence of words from X yielding w as their products. Formally, an X*-factorization* of $w \in \Sigma^{\mathbb{Z}}$ is a mapping $F : \mathbb{Z} \to X \times \mathbb{Z}$ such that for each $k \in \mathbb{Z}$ if $F(k) = (\alpha, i)$ and $F(k+1) = (\beta, j)$, then $a_ia_{i+1}\ldots a_{j-1} = \alpha$, i.e. the position i is a starting position of the factor α in w. We say that two X-factorizations F_1 and F_2 of a bi-infinite word are

- *different*, whenever there is a $k_0 \in \mathbb{Z}$ such that for each $k \in \mathbb{Z}$, $F_1(k_0) \neq F_2(k)$,
- *disjoint*, whenever the starting positions of all factors in F_1 are distinct from the ones in F_2,
- *shift-equivalent*, if there is a k_0 such that whenever $F_1(k) = (\alpha, i)$ and $F_2(k_0 + k) = (\beta, j)$, then $\alpha = \beta$.

An X*-ambiguous* bi-infinite word is a bi-infinite word, which has two different X-factorizations.

Example 1. Let $X = \{a, bab, baab\}$. The word $(baa)^{\mathbb{Z}}$ has two different X-factorizations, namely the ones depicted as

$$\dots\ b\,a\,a\,b\,a\,a\,b\ \dots \quad .$$

They are clearly shift-equivalent. On the other hand the word

$$w = \dots bababaabaab \dots = {}^{\mathbb{N}}(ba)b(aab)^{\mathbb{N}}$$

also has two different X-factorizations, which, however, are not shift-equivalent

$$\dots\ a\,b\,a\,b\,a\,b\,a\,b\,a\,a\,b\,a\,a\,b\,a\,a\,b\,a\,a\ \dots \quad .$$

Clearly, in both of the above cases the two factorizations are disjoint.

We define the *combinatorial rank* of $X \subseteq \Sigma^+$ by the formula

$$r_{\mathrm{c}}(X) = \min\{\mathrm{card}(Y) \mid X \subseteq Y^+\}\,.$$

For the sake of completeness we remind that

$$r_{\mathrm{c}}(X) \leq r_{\mathrm{f}}(X) \leq \mathrm{card}(X)\,,$$

where $r_{\mathrm{f}}(X)$ denotes the *free rank* (or simply the *rank*) of X defined as the cardinality of the base of the smallest free semigroup containing X, cf [CK].

Example 1 (continued). Clearly, $r_{\mathrm{c}}(X) = 2$, since $X \subseteq \{a, b\}^+$, but for no word ϱ the inclusion $X \subseteq \varrho^+$ holds. On the other hand, since X is a code we conclude that $r_{\mathrm{f}}(X) = 3$.

Example 2. Let $X = \{ab, bc, ca\}$. Then we have $r_{\mathrm{c}}(X) = r_{\mathrm{f}}(X) = 3 = \mathrm{card}(X)$. Note also that the bi-infinite word $(abc)^{\mathbb{Z}}$ has two disjoint, but shift-equivalent, X-factorizations:

$$\dots\ a\ b\ c\ a\ b\ c\ \dots \quad .$$

Next we recall a few basic results on words that we shall need in our later considerations, for their proofs the reader is referred to [Lo] or [CK].

Lemma 1. `(Fine and Wilf)` *Let $u, v \in \Sigma^+$. If the words $u^{\mathbb{N}}$ and $v^{\mathbb{N}}$ have a common prefix of length at least $|u| + |v| - \gcd(|u|, |v|)$, then u and v are powers of a common word.*

Lemma 2. *No primitive word ϱ satisfies a relation $\varrho\varrho = s\varrho p$ with $s \neq 1$ and $p \neq 1$.*

Lemma 3. *If two words u and v satisfy the relation $ut = tv$ for some $u, v, t \in \Sigma^+$, i.e. if they are conjugates, then there exist words p and q such that pq is primitive and*

$$u = (pq)^i, \quad v = (qp)^i \quad \textit{and} \quad t \in p(qp)^* \quad \textit{for some } i \geq 1\,.$$

In Section 5 we shall need also the following result which has been proved in [LyS].

Lemma 4. *Consider nonempty words x, y, z satisfying equation $x^m = y^n z^p$, where $m, n, p \geq 2$. Then all words x, y, z are powers of a common word.*

In order to formulate our fifth, and most crucial lemma, we need some terminology, cf [CK] or [HK]. We associate a finite set $X \subseteq \Sigma^+$ with a graph $\mathcal{G}_X = (V_X, E_X)$, called *the dependency graph* of X, as follows: the set V_X of vertices of $\mathcal{G}_X$ equals to X, and the set E_X of edges of $\mathcal{G}_X$ is defined by the condition

$$(x, y) \in E_X \qquad \text{iff} \qquad xX^{\mathbb{N}} \cap yX^{\mathbb{N}} \neq \emptyset\,.$$

Then we have

Lemma 5. *For each finite set $X \subseteq \Sigma^+$, the combinatorial rank of X is at most the number of connected components of $\mathcal{G}_X$.*

As we shall see, Lemma 5 is particularly suitable for our subsequent considerations. Indeed, in that lemma it is crucial that words in X are nonempty, and that indeed is satisfied in the proofs of our Theorems 1 and 2.

3 The General Case

In this section we prove our first defect theorem for bi-infinite words. Frequently we illustrate our proofs by pictures. In these pictures a horizontal line expresses a bi-infinite word with two X-factorizations F_1, F_2. The sequences of words in the factorization F_1 are depicted above the line by consecutive arcs, similarly the sequences of words in F_2 are depicted by arcs, which are below the line. For example, in Figure 2 we consider the words $f_1, f_1', f_2, f_2' \in X^*$, such that f_1, f_1' are parts of the factorization F_1 and f_2, f_2' are parts of F_2.

Theorem 1. *Consider a set $X = \{\alpha_1, \ldots, \alpha_n\} \subseteq \Sigma^+$. Let w be a bi-infinite word over Σ and F_1, F_2 two different X-factorizations of w. Then the combinatorial rank of X is at most $n - 1$, or both the word w and the X-factorizations F_1, F_2 are periodic. Moreover, if the rank of X is n, then the number of periodic bi-infinite words with two different X-factorizations is finite.*

Proof. If F_1 and F_2 are not disjoint the result follows from Lemma 5. Now we study all words $t \in \Sigma^+$ as depicted in Figure 1 with $x, y \in X$. More precisely, we take the beginning of any $x \in X$ in the lower factorization F_2, i.e. the point A, and we find the closest end of $y \in X$ in the upper factorization F_1 to the right from the point A, i.e. the point B. Now the *X-difference* t is defined as the word between points A and B. Note that there are infinitely many occurrences of X-differences, but only finitely many of different values of t's, since all t's are proper suffices of words in X. By the pigeon hole principle there exists a t such that the X-difference t occurs infinitely many times in the bi-infinite word w.

Consider now two occurrences of an X-difference t. Let the part of the factorization F_1 between the end of the first t and the end of the second t be $f_1 \in X^+$ and the part of F_2 between the beginnings of the t's be $f_2 \in X^+$. We shall call the pair (f_1, f_2) the *t-pair*. Notice that for any t-pair (f_1, f_2) we have $tf_1 = f_2t$. Further, we shall call a t-pair (f_1, f_2) *minimal*, if there is no other occurrence of the X-difference t inside the f_1, f_2.

Let us look at 3 consecutive occurrences of an X-difference t, which has infinitely many occurrences in w, see Figure 2. Clearly, the pairs (f_1, f_2) and (f_1', f_2') are minimal t-pairs.

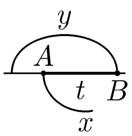

Fig. 1

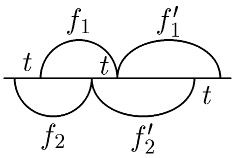

Fig. 2

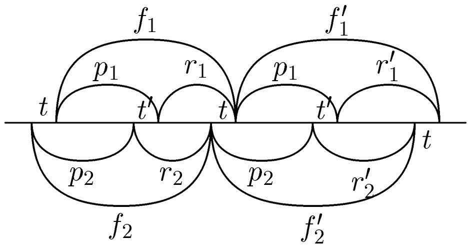

Fig. 3

Assume that we found such 3 occurrences for which at least one of the equalities

$$f_1 = f_1', \qquad f_2 = f_2', \tag{1}$$

does not hold. We prove, that in such a case we have a defect effect. Let $p_i \in X^*$, for $i = 1, 2$, be the longest common prefix of f_i, f_i' over alphabet X and $r_i, r_i' \in X^*$ their suffices, i.e. $f_i = p_i r_i$, $f_i' = p_i r_i'$. At Figure 3 p_2 ends earlier than p_1 in the bi-infinite word w, so the factorizations differ earlier in the pair f_2, f_2'. We shall consider only this case, in the other case the proof is similar.

Now either one of the words r_2, r_2' is empty, or they start with different words $x, x' \in X$. If, for example, r_2 is empty, then $|r_1| < |t|$, which by the definition of X-difference means $r_1 = 1$. But then we have an intermediate occurrence of the X-difference t in the minimal t-pair (f_1', f_2'), which is a contradiction. So assume that r_2, r_2' are both nonempty and their factorizations start with different words $x, x' \in X$, i.e. that $r_2 = xq_2$, $r_2' = x'q_2'$, where $q_2, q_2' \in X^*$. We have 3 equations over $X \cup \{t, t'\}$:

$$t' r_1 = x q_2 t\,, \qquad t' r_1' = x' q_2' t\,, \qquad t f_1 = f_2 t\,,$$

where $x \neq x'$. So, by Lemma 5, we obtain $r_{\rm c}(X \cup \{t, t'\}) \leq n-1$, which implies $r_{\rm c}(X) \leq n-1$. Note, that the last implication need not hold for other ranks.

Assume now that for each 3 occurrences of X-difference t the equalities (1) hold. Take a point in the bi-infinite word w. If we have infinitely many occurrences of X-difference t to the right and also to the left from this point, the bi-infinite word and both X-factorizations are periodic, so we are done. Assume, that there is infinitely many occurrences of t only to the one direction from the point, for example, to the right. By the pigeon hole principle, there must be another X-difference s, which has infinitely many occurrences to the left from the point. We can repeat all considerations for s and either we find a defect effect, or for each 3 occurrences of the X-difference s the equalities (1) hold. If there are infinitely many occurrences of s to the right from the point, we are done. Otherwise, we have a situation depicted in Figure 4, where $e_i, m_i, f_i \in X^+$. In the figure there is a repetition of the t-pair (f_1, f_2) to the right and a repetition of the s-pair (e_1, e_2) to the left. The point is somewhere in the middle of the pair (m_1, m_2), which implies that t-difference does not occur in the s-pair (e_1, e_2).

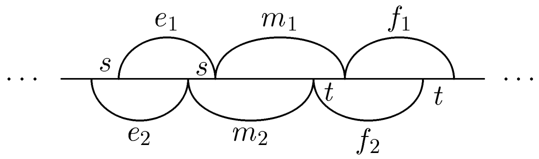

Fig. 4

Now, we can repeat the same consideration for pairs (e_1, e_2) and (m_1, m_2) as we did for two consecutive t-pairs, and we obtain the following system of equations over $X \cup \{s, t, t'\}$:

$$t' r_1^e = x q_2^e s\,, \qquad t' r_1^m = x' q_2^m t\,, \qquad s e_1 = e_2 s\,, \qquad t f_1 = f_2 t\,,$$

where $x \neq x'$. This again implies a defect effect by Lemma 5.

The above also shows that in the case $r_{\rm c}(X) = \mathrm{card}(X)$, for each value of t there is at most one minimal t-pair. Since in this case the bi-infinite word w must be periodic the minimal t-pair specifies the whole bi-infinite word, and so for each value of t there is also at most one X-ambiguous bi-infinite word containing X-difference t. But any t must be a proper suffix of a word from X, hence we have only finitely many bi-infinite words with two different X-factorizations. □

A restriction of Theorem 1 to two-element sets yields the following consequence.

Corollary 1. *Consider set $X = \{\alpha, \beta\} \subseteq \Sigma^+$. Let w be a bi-infinite word over Σ and F_1, F_2 two different X-factorizations of w. Then the words α, β commute or both the word w and the X-factorizations F_1, F_2 are periodic.*

Theorem 1 deserves a few comments. First, the possibility that the two factorizations are both periodic cannot be ruled out. This follows from Example 2. Second, the combinatorial rank cannot be replaced by the free rank, for example. This indeed follows from Example 1. This latter remark is quite interesting since in all previous defect theorems, see [CK], either of our notions of the rank, or even some others, can be used to witness the defect effect.

4 The Two-element Case

In this section we generalize the result of the previous section in the case of two-element sets. First we recall that in a strict sense we cannot have a defect theorem for bi-infinite words even in this simple case.

Example 3. The set $X = \{ab, ba\}$ is of combinatorial rank 2 although the word $(ab)^{\mathbb{Z}}$ has two disjoint, and even non-shift-equivalent, X-factorizations.

As a main result of this paper we, however, show that the above example, and its natural variants, are the only exceptions which may occur. And even in these cases the roots of words in X are conjugates, i.e. they are cyclic permutations of powers of a common word.

Theorem 2. *Consider set $X = \{\alpha, \beta\}$ with $\alpha, \beta \in \Sigma^+$. Let w be a bi-infinite word over Σ and F_1, F_2 two different X-factorizations of w containing together both elements of X. Then one of the following possibilities holds:*

(i) *α and β commute, or*

(ii) *the roots of α and β are conjugates and $F_1 \in \alpha^{\mathbb{Z}}$, $F_2 \in \beta^{\mathbb{Z}}$, or vice versa, or*

(iii) *the two X-factorizations F_1, F_2 are shift-equivalent and there exists an $n \geq 1$ such that $F_1, F_2 \in (\alpha\beta^n)^{\mathbb{Z}}$ and α is primitive or $F_1, F_2 \in (\beta\alpha^n)^{\mathbb{Z}}$ and β is primitive.*

Proof. We can assume that α, β do not commute. Then, by Lemma 5, the factorizations F_1, F_2 must be disjoint. Further, by Corollary 1, the factorizations F_1, F_2 are periodic with the same length of a period. Whenever we find the situation which is shown in Figure 5 then, since the factorizations are periodic, this situation occurs infinitely many times. By the proof of Theorem 1 it is clear that f_1, f_2 are periods of F_1, F_2.

If both α and β are not primitive, we can replace them by powers of their roots $\varrho(\alpha)^k, \varrho(\beta)^m$, where $\varrho(\alpha)^k = \alpha$, $\varrho(\beta)^m = \beta$ and explore the situation over a slightly different set $X = \{\varrho(\alpha), \varrho(\beta)\}$. If we prove that for $\varrho(\alpha), \varrho(\beta)$ the claim holds, then it must hold also for α and β and, moreover, in case *(iii)* we have either $k = 1$ and α is primitive, or $m = 1$ and β is primitive. So it is enough to consider only the case when α and β are primitive.

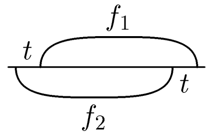

Fig. 5

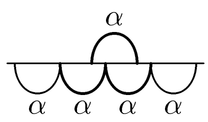

Fig. 6

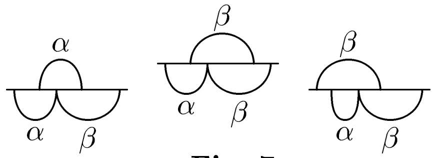

Fig. 7

Without loss of generality we can also assume that $|\alpha| \leq |\beta|$. Now, if F_2 does not contain the factor $\beta\alpha$, then it contains only α's or only β's or there is a point inside F_2 from which to the left there are only α's and to the right only β's. In the last case the factorization F_2 is clearly non-periodic — a contradiction with Corollary 1. Consider now, for example, the case $F_2 = \alpha^{\mathbb{Z}}$. If F_1 contains any α, then we have the situation depicted in Figure 6 which, by Lemma 2, contradicts the primitiveness of α. So we have $F_1 = \beta^{\mathbb{Z}}$, and, by Lemma 1, α and β must be conjugates, which is case *(ii)*.

We may assume that F_2 contains the factor $\alpha\beta$. In Figure 7 we can see all possibilities how F_1 covers the border between the above occurrences of α and β. We shall analyze all three cases.

Case 1,2. We can analyze first two cases simultaneously, because when we forget about the relation between lengths of α and β, then they are clearly symmetric. So consider the second case of 3 possibilities drawn in Figure 7. If the word to the right of the β in the factorization F_1 is also β, then β is not primitive which is not the case. Hence, we have the situation shown in Figure 8. Now if $z_{\mathrm{R}} = \alpha$ or $z_{\mathrm{L}} = \alpha$, then $v_1 = v_2$ and we arrive into the situation depicted in Figure 5 with $f_1 = \beta\alpha$ and $f_2 = \alpha\beta$ which is case *(iii)* of our claim. So consider the other case when $z_{\mathrm{R}} = z_{\mathrm{L}} = \beta$. We can continue in this way inductively until sequences of β's exceed α's (on both sides at the same time) or we obtain the situation in Figure 5 with $f_1 = \beta^n\alpha$, $f_2 = \alpha\beta^n$, for some $n \geq 1$, which is again case *(iii)*. The first possibility is shown in Figure 9.

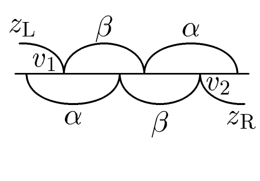

Fig. 8

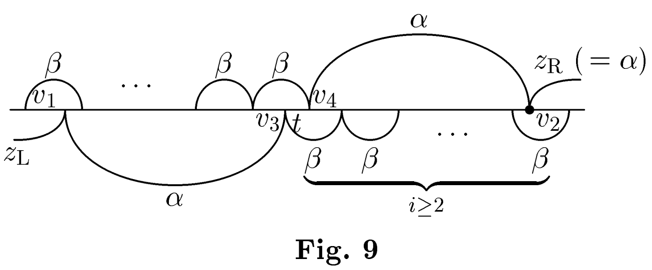

Fig. 9

Now again if $z_{\mathrm{R}} = \beta$, then we have $v_1 = v_2$, and hence we are again in case *(iii)*. So assume that $z_{\mathrm{R}} = \alpha$. We have $\beta = v_3 t = t v_4$, which by Lemma 3 allows us to write $v_3 = (pq)^k$, $v_4 = (qp)^k$, $t = p(qp)^n$, where pq is primitive and $k \geq 1$, $n \geq 0$. We can see that α ends with pq and starts with qp. This means that the word $pqqp$ matches the word $\beta = (pq)^{k+n}p$ around the black point shown in Figure 9. The black point must lie inside β. There are 3 possibilities where the black point inside β can be. In case (1) it occurs inside the first p of β, in case (2) inside the first q and in the last case in the rest of β as it is shown in Figure 10.

At first we analyze some special cases. If the black point matches with the end of the first p in β, then we have according to Figure 9 two equations with unknowns $Y = \{\alpha, p, q\}$:

$$\alpha = v_4 \beta^{i-2} p = q w_1\,, \qquad \alpha = p \beta^{i-2} v_3 = p w_2\,, \qquad \text{where } w_1, w_2 \in Y^*.$$

The dependency graph of this system is then connected which implies that unknowns in Y, and hence also α and β, commute, which is a contradiction.

If the black point matches with the end of any pq in β, then we can write

$$\alpha = v_4 \beta^{i-2} (pq)^l, \quad \alpha e_1 \beta^i \alpha = p(qp)^{k+n-l} e_2 \alpha \beta^{i-1} (pq)^l,$$

where $l \geq 1$, $e_1, e_2 \in X^*$ and the second equation is obtained by taking parts of F_1, F_2 between the black point and the next occurrence (to the right) of the black point, so e_i is a part of the factorization F_i. We can rewrite these two equations as a system of equations with unknowns $Y = \{\alpha, p, q\}$:

$$\alpha = qp w_1\,, \qquad \alpha w_2 = p w_3\,, \qquad \text{where } w_1 \in \{p, q\}^*,\ w_2, w_3 \in Y^*,$$

and, by Lemma 5, we have again a contradiction.

Now we analyze cases (1)–(3) assuming we are not in any special case. In case (1), the qp which follows the black point, lies inside the first pqp in β. But this is a contradiction because then pq cannot be primitive. In case (3), we can use the same argument with the pq which precedes the black point. The situation around the black point in case (2) is shown in Figure 11.

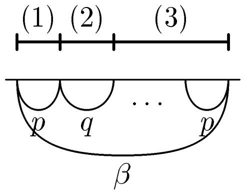

Fig. 10

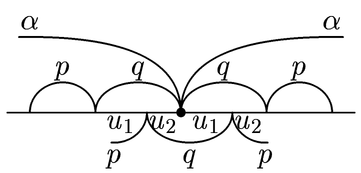

Fig. 11

It follows, by Lemma 2, that q is not primitive, and that there is an s such that $u_1 = s^i$, $u_2 = s^j$ and $q = s^{i+j}$ with $i, j \geq 1$. Now, as above, we have 2 equations with unknowns $Y = \{\alpha, p, s\}$:

$$\alpha = v_4 \beta^{i-2} p u_2 = s^{i+j} w_1,$$
$$s^j w_1 e_1 \alpha = {u_1}^{-i} \alpha e_1 \alpha = (pq)^{k+n-1} p e_2 \alpha \beta^{i-1} p u_2 = p w_2,$$

where $w_1, w_2 \in Y^*$ and $e_1, e_2 \in X^*$ are parts of the X-factorizations.

Case 3. Now we shall analyze the third case of Figure 7. Since β is not primitive there must be α to the right of the β on the upper line. Using the same considerations as in the previous case we come to Figure 12, or we end up in case *(iii)* with $f_1 = \beta\alpha^n$, $f_2 = \alpha^n\beta$ for some $n \geq 1$. In the first case we have $\alpha = v_1v_3 = v_4v_2$, where $|v_1| = |v_2|$ and $|v_3| = |v_4|$. There are again two possibilities.

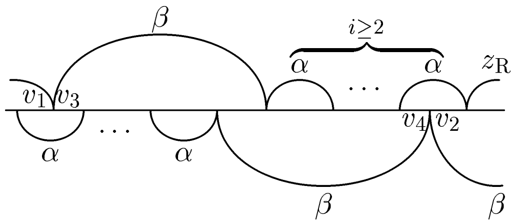

Fig. 12

Assume first that $z_{\mathrm{R}} = \alpha$. We know that β starts with v_3v_1, so, as it is shown in Figure 13, v_3v_1 lies inside $\alpha\alpha = v_1v_3v_1v_3$. This implies that either α is not primitive, or v_3v_1 matches with v_3v_1 in $\alpha\alpha$. In the first case we have a contradiction. In the second case it is obvious from Figure 13 that $v_2 = v_3$, say equal to p, and $v_1 = v_4$, say equal to q, and moreover that $|p| = |q|$. So we have $p\alpha^{i-1}\beta = \beta\alpha^{i-1}q$, which means that $v = p(qp)^{i-1}$ conjugates with $u = q(pq)^{i-1}$. We shall show that this is again a contradiction with the primitiveness of $\alpha = qp$.

We have already analyzed this situation. Since the word u is a conjugate with v, a factor of the word uu must be equal to the word v. The word u starts with qp and ends with pq, so the middle point of the word $pqqp$ lies inside the word $v = p(qp)^{i-1}$. There are again 3 possibilities (see Figure 10) and two special cases. Since $|p| = |q|$ in the special cases we have $p = q$, so that $\alpha = v_1v_3 = qp = p^2$ proving that α is not primitive, a contradiction. In cases (1) and (3) we also have a contradiction with the primitiveness of α as we already proved. In case (2) we have $u_1 = s^i$, $u_2 = s^j$, $q = s^{i+j}$ (see Figure 11), and since $|p| = |q|$ we also have $p = u_2u_1 = s^{i+j} = q$, which is again a contradiction.

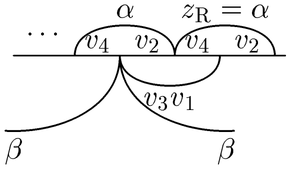

Fig. 13

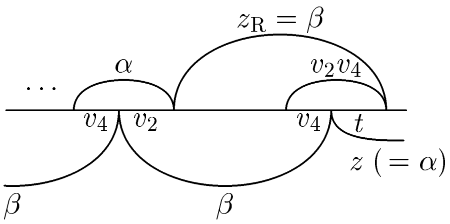

Fig. 14

So it remains the case $z_{\mathrm{R}} = \beta$. The situation is drawn in Figure 14. Since β is not primitive z must be α. It is obvious that $|t| = |v_2| = |v_1|$, which implies $t = v_1$. We know that β ends with v_2v_4, and hence there is v_2v_4 at the end of the last upper β in Figure 14, and v_4 at the end of the last lower β. But since $|v_2v_4| = |v_4t|$ we have the equation $v_2v_4 = v_4t = v_4v_1$. According to Figure 12 and Figure 14 we have the following system of equations with unknowns $Y = \{\beta, v_1, v_2, v_3, v_4\}$:

$$\begin{aligned} v_2v_4 &= v_4v_1\,, & v_3(v_1v_3)^{i-1}\beta &= \beta(v_4v_2)^{i-1}v_4\,,\\ (\alpha =)\ v_1v_3 &= v_4v_2\,, & v_2\beta &= \beta v_1\,. \end{aligned}$$

The dependency graph associated with this system is connected, and hence all unknowns commute, in particular α commutes with β. This completes the proof of the theorem. □

Theorem 2 deserves a few comments. The number of different X-factorizations of the bi-infinite word w having an X-factorization is very different in cases *(i)–(iii)*. In case *(i)* there exist nondenumerably many such X-factorizations, in case *(ii)* there are finitely many different X-factorizations, and if we consider shift-equivalent X-factorization as the same, then there are exactly two of them. Finally, in case *(iii)* there are also finitely many different X-factorizations, which are all shift-equivalent. This actually means that in case *(iii)* no bi-infinite word can be expressed in two different ways as a product of words from X. Hence, indeed, Theorem 2 shows a defect effect of a two-element set for bi-infinite factorizations.

In Theorem 2 we showed that if the words of X do not commute and are not conjugates, then only the case *(iii)* is possible. But if they do not commute and are conjugates Theorem 2 allows either case *(ii)* or *(iii)*. Now we shall prove that only case *(ii)* is possible. According to the last part of the proof of Theorem 2, we can formulate the following lemma.

Lemma 6. *If pq is primitive, then $p(qp)^n$ and $q(pq)^n$ are not conjugates for any $n \geq 1$.*

This yields easily

Corollary 2. *If α and β are different conjugates, then $\alpha\beta$ must be primitive.*

Proof. Assume the contrary that $\alpha\beta$ is not primitive, so we have $\alpha\beta = t^i$, where t is primitive and $i \geq 2$. Now if i is even, then immediately $\alpha = \beta$, which is a contradiction. For odd $i = 2n + 1$ we have $\alpha = t^n p$, $\beta = qt^n$, where $t = pq$. But α and β are conjugates and so, by Lemma 6, we have a contradiction. □

In fact Corollary 2 is a special case of the claim in [LeS] which states under the additional assumption that α, β are primitive, that $\alpha\beta^m$ is primitive for all natural numbers m. The proof is not difficult, but we need only this special case to prove the next result.

Corollary 3. *Consider set $X = \{\alpha, \beta\}$ with $\alpha, \beta \in \Sigma^+$. Let w be a bi-infinite word over Σ and F_1, F_2 two different X-factorizations of w containing together both elements of X. If the roots of α, β are non-commuting conjugates, then $F_1 \in \alpha^{\mathbb{Z}}$, $F_2 \in \beta^{\mathbb{Z}}$, or vice versa.*

Proof. Again as in the proof of Theorem 2, we can assume that α, β are primitive. We have to show that the one of the X-factorizations F_1, F_2 consists only of α's and the other only of $\beta's$. So assume the contrary that the lower factorization contains both α and β. Without lost of generality we have the situation shown in Figure 15.

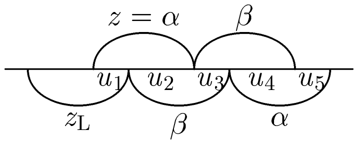

Fig. 15

Since β is primitive we have $z = \alpha$, by Lemma 2. We can write $\beta = u_2 u_3 = u_3 u_4$, and so, by Lemma 3, we have

$$u_2 = (pq)^i, \quad u_4 = (qp)^i, \; i \geq 1, \quad u_3 = p(qp)^n, \; n \geq 0,$$

where pq is primitive. If $z_{\mathrm{L}} = \alpha$, then, by Lemma 2, $\alpha\beta$ is not primitive, which contradicts Corollary 2. Hence assume $z_{\mathrm{L}} = \beta$, which implies $u_1 = u_3$. Similarly $u_5 = u_3$ and we have $\alpha = u_1 u_2 = p(qp)^n (pq)^i = u_4 u_5 = (qp)^i p(qp)^n$. By Lemma 5, then p and q commute, which is again a contradiction. $\square$

5 The Uniqueness of the Bi-infinite Word

In Section 3 we proved that if the rank of the set X equals to $\mathrm{card}(X)$, then the number of X-ambiguous bi-infinite words is finite. In this section we shall prove that in the two-element case, for each set X, there is at most one X-ambiguous bi-infinite word. This holds also in the case when $r_{\mathrm{c}}(X) = 1$, since then both elements of X are powers of a common word t and the only possible bi-infinite word is $t^{\mathbb{Z}}$. The situation is also trivial in the case when roots of elements of $X = \{\alpha, \beta\}$ are conjugates, by Corollary 3 the only possible bi-infinite word is $w = \alpha^{\mathbb{Z}} = \beta^{\mathbb{Z}}$. So we need to consider only the case when the roots of α and β are not conjugates.

In this case, by Theorem 2, we know that an X-ambiguous bi-infinite word must be of the form $(\alpha\beta^n)^{\mathbb{Z}}$ or $(\alpha^n\beta)^{\mathbb{Z}}$. Moreover, since w has two X-factorizations, the word $\alpha\beta^n$ or the word $\alpha^n\beta$ cannot be primitive, by Lemma 2.

As we stated in the previous section, if α and β are conjugates, then the words $\alpha\beta^n$ and $\alpha^n\beta$ are primitive for all n. Now, we shall show a similar result for α, β being non-conjugates, i.e. we shall show that at most one word of words $\alpha\beta^n$, $\alpha^m\beta$ is not primitive. By this result, we have that also in the last case there is at most one X-ambiguous bi-infinite word. We need two lemmas.

Lemma 7. *Let α, β be primitive and not conjugates. Then it is not possible that $\alpha\beta^n$ and $\alpha\beta^m$ are both non-primitive for $n \neq m$.*

Proof. Assume the contrary that both $\alpha\beta^n$, $\alpha\beta^m$ are non-primitive with $m < n$. For $m = 0$ the claim is obvious, so we can assume $m \geq 1$, and so $n \geq 2$. We can write

$$\left.\begin{aligned} \alpha\beta^n &= s^i \\ \alpha\beta^m &= t^j \end{aligned}\right\} \implies \quad s^i = t^j\beta^{n-m}, \tag{2}$$

where s, t are primitive and $i, j \geq 2$. Now if $n - m \geq 2$, then by Lemma 4 s, t and β are powers of a common word, and so are α and β, which is a contradiction. So we can assume $m = n-1$, and thus equation (2) simplifies to $s^i = t^j\beta$.

Now if $|s| \leq (n-1)|\beta|$, then $|\beta| + |s| \leq |\beta^n|$, so that, by the equation $\alpha\beta^n = s^i$, words $\beta^{-\mathbb{N}}$ and $s^{-\mathbb{N}}$ have a common suffix of a length at least $|\beta| + |s|$. Applying Lemma 1 we conclude that words s and β are powers of a common word, which again yields to a contradiction. So we have

$$|s| > (n-1)|\beta| \geq |\beta|\,, \text{ and similarly} \tag{3}$$
$$|t| > (m-1)|\beta| = (n-2)|\beta|\,. \tag{4}$$

Inequality (3) together with equation (2) implies $|t^j| = i|s| - |\beta| > (i-1)|s|$. So, if $i \geq 3$ we have $|t^j| > 2|s|$, and since $j \geq 2$ also that $|t^j| > |s| + |t|$. Then equation (2) and Lemma 1 implies that s and t commutes which leads to a contradiction. Hence we can assume $i = 2$.

If $|t| + |\beta| \leq |s|$, then using equation (2) we derive inequality $|t^j| \geq (i-1)|s| + |t|$ and, by Lemma 1, we have a contradiction again. Hence we may assume that

$$|t| + |\beta| > |s|\,. \tag{5}$$

Now consider the case $n \geq 3$. We have

$$2|t| \geq \Big(1 + \frac{1}{n-2}\Big)|t| \overset{(4)}{>} |t| + |\beta| \overset{(5)}{>} |s| \overset{(2)}{=} j|t| + |\beta| - |s| \overset{(5)}{>} (j-1)|t|\,, \tag{6}$$

which implies that $j = 2$ and also that $|t| \geq |\beta|$ by the two first inequalities. The second inequality and the equation $\alpha\beta^n = t^2$ imply that $t = x\beta$ for some $x \neq 1$. Thus equation (2) yields to $s^2 = x\beta x\beta\beta$. Since $|x\beta| < |s|$ and $\frac{3}{2}|\beta| < |s|$, we can write $s = x\beta y = z\beta_2\beta$ for some $y, z \neq 1$, where

$|y| = |\beta_2| = \frac{|\beta|}{2}$, $\beta = \beta_1\beta_2$ and $|x| = |z|$. We can divide this equation into two parts: $x = z$ and $\beta y = \beta_2\beta$, where the second one, by Lemma 2, contradicts the primitiveness of β.

The last case we have to analyze is $n = 2$. Now if $|t| \geq |\beta|$, then, by (6), we have $2|t| \geq |t| + |\beta| > (j-1)|t|$ and $j = 2$, which is again the previous case. So consider the case $|t| < |\beta| < |s|$, where the second inequality comes from (3). By the equations $\alpha\beta = t^j$ and $\alpha\beta^2 = s^2$ we can write $\beta = xt$ and $s = yt$ for some $x, y \neq 1$. Hence equation (2) leads to $ytyt = t^j\beta$. Therefore by (2) and (3) we obtain $|yt| = |s| = |t^j| + |\beta| - |s| < |t^j|$, so that we can write $ytz = t^j$, $z \neq 1$. Now either t is not primitive by Lemma 2, or t matches with some t in t^j, but then we have $y = t^k$, and hence also $s = t^{k+1}$, so that words t, s are powers of a common word. In both cases we come to a contradiction. □

Lemma 8. *Let α, β be primitive and not conjugates. Then it is not possible that $\alpha\beta^n$ and $\alpha^m\beta$ are both non-primitive for any $(n, m) \neq (1, 1)$.*

Proof. Cases $m = 0$ and $n = 0$ are trivial. The case $m = 1$ is a special case of Lemma 7. In the case $n = 1$ we can exchange α and β and take reverses of words, and we are again in the case $m = 1$. We shall use this reasoning again later, so let us call it *the reverse argument.* Consider $n, m \geq 2$ and assume the contrary that $\alpha\beta^n = s^i$, $\alpha^m\beta = t^j$, where $i, j \geq 2$ and s, t are primitive. Using the same argument as in the proof of the previous lemma we have

$$|s| > (n-1)|\beta| \geq |\beta|\,, \qquad |t| > (m-1)|\alpha| \geq |\alpha|\,. \tag{7}$$

Hence

$$\begin{aligned} |\alpha| &= i|s| - n|\beta| > (in - i - n)|\beta| \\ |\beta| &= j|t| - m|\alpha| > (jm - j - m)|\alpha|\,, \end{aligned} \tag{8}$$

which implies that

$$\big[(i-1)(n-1) - 1\big] \cdot \big[(j-1)(m-1) - 1\big] < 1\,.$$

So we have either $i = n = 2$, or $j = m = 2$. Now by the reverse argument the first case is equivalent to the second one, so it is enough to consider only the case $j = m = 2$. If $|t| < |\beta|$, then, by (7), we obtain

$$|\alpha| \overset{(7)}{<} |t| < |\beta| \overset{(7)}{<} |s|\,.$$

Together with (8) we have $(i-1)(n-1) - 1 < 1$, which implies that also $i = n = 2$. Now we can apply the reverse argument and the inequality $|s| > |\alpha|$ transforms to the inequality $|t| > |\beta|$. So, without loss of generality, we can assume that $|t| > |\beta|$. We have the situation depicted in Figure 16, where $|u_1| = |u_2| = \frac{1}{2}|\beta|$.

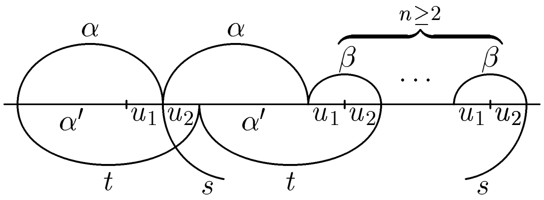

Fig. 16

We can see that $u_2\alpha' = \alpha' u_1$, and thus Lemma 3 gives us

$$\left.\begin{aligned} u_2 &= (pq)^k \\ u_1 &= (qp)^k \\ \alpha' &= p(qp)^l \end{aligned}\right\} \Longrightarrow \begin{cases} \alpha = p(qp)^{k+l} \\ \beta = (qp)^k (pq)^k \end{cases},$$

where $k \geq 1$, $l \geq 0$ and pq is primitive. Now considerring the last occurrence of s in Figure 16 we can, by (7), write $s = s'\beta = s'(qp)^k(pq)^k$. We also have

$$|s| = |\alpha| + n|\beta| - (i-1)|s| \leq |\alpha| + n|\beta| - |s| \overset{(7)}{<} |\alpha| + |\beta|,$$

which yields

$$s'(qp)^k(pq)^k r = sr = \alpha\beta = \underbrace{p(qp)^{2k+l}}_{w}(pq)^k,$$

for some $r \neq 1$. The first occurrence of qp in s after s' must match with qp in w, otherwise qp is not primitive. But then, since $r \neq 1$, the first occurrence of pq in s after $s'(qp)^k$ matches with some qp in w, so we have $pq = qp$, which is again a contradiction with the primitiveness of pq. ☐

Finally, we can state the result of this section.

Theorem 3. *Consider set $X = \{\alpha, \beta\}$ with $\alpha, \beta \in \Sigma^+$. There is at most one X-ambiguous bi-infinite word over Σ.*

6 The Existence of the Bi-infinite Word

We consider again only the two-element case in this section. In the previous section we proved that there is at most one X-ambiguous bi-infinite word. It is natural to ask when such a word exists. It is easy to see that there are sets X for which there is no X-ambiguous bi-infinite word. For example, take set $X = \Sigma = \{a, b\}$. We shall now characterize under which condition there exists an X-ambiguous bi-infinite word.

In case *(i)* of Theorem 2, when words of X are powers of a common word t, the bi-infinite word $t^{\mathbb{Z}}$ has infinitely many X-factorizations. In particular, in the case there is always an X-ambiguous bi-infinite word. In case *(ii)*, when roots of words in $X = \{\alpha, \beta\}$ are conjugates, the bi-infinite word $\alpha^{\mathbb{Z}} = \beta^{\mathbb{Z}}$ has exactly two different X-factorizations, so it is X-ambiguous.

Consider now the last case *(iii)*, and the set $X = \{\alpha, \beta\}$. By Theorem 2, an X-ambiguous bi-infinite word is of the form $(\alpha\beta^n)^{\mathbb{Z}}$ where $\alpha\beta^n$ is not primitive, i.e. there is an $i \geq 2$ such that the equation

$$\alpha\beta^n = s^i \tag{9}$$

has a solution. Conversely, if for some $n \geq 1$, $i \geq 2$ the equation (9) has a solution, then clearly the bi-infinite word $(\alpha\beta^n)^{\mathbb{Z}}$ has exactly i shift-equivalent, but different X-factorizations. We formalize this as a lemma.

Lemma 9. *Let $X = \{\alpha, \beta\} \subseteq \Sigma^+$ be a set of two non-commuting words such that their roots are not conjugates. Then there is an X-ambiguous bi-infinite word if and only if one of the equations $\alpha\beta^n = s^i$, with $n \geq 1$, $i \geq 2$, has a solution.*

We shall also try to give a characterization of the solutions of the equation (9). We need the following lemma.

Lemma 10. *The only non-commuting solutions of the equation*

$$u_1u_2 = u_3(u_2u_3)^m, \quad m \geq 1 \tag{10}$$

are

$$\begin{aligned} u_3 &= qp\,, \\ u_2 &= p(qp)^k, \\ u_1 &= u_3(u_2u_3)^{m-1}pq\,, \end{aligned} \tag{11}$$

where $p, q \in \Sigma^+$, $k \geq 0$.

Proof. It is easy to check that (11) is really a solution of equation (10). Now we shall prove that if equation (10) has a solution, then it is of the form (11). We proceed by induction.

Consider first the case $m = 1$. We have the equation $u_1u_2 = u_3u_2u_3$. It is obvious that $|u_1| > |u_3|$, so we can write $u_1 = u_3t$. The equation transforms into $tu_2 = u_2u_3$, which has, by Lemma 3, the only solutions $t = pq$, $u_3 = qp$ and $u_2 = p(qp)^k$, $k \geq 0$. This implies that $u_1 = qppq$, so we have a solution of the form (11) for $m = 1$.

Consider now equation (10) with $m \geq 2$. Again we have $|u_1| > |u_3|$, so we can substitute $u_1 = u_3t$ and equation (10) becomes $tu_2 = u_2u_3(u_2u_3)^{m-1}$. By Lemma 3, we have $t = uv$, $u_3(u_2u_3)^{m-1} = vu$, $u_2 = u(vu)^l$. If $l \geq$

1, then $|vu| = |u_3(u_2u_3)^{m-1}| \geq 2|u| + |v| + |u_3|$. This implies that $u = u_3 = 1$, which leads to a commuting solution. Hence, consider the case $l = 0$. We have $u_2 = u$, $u_1 = u_3u_2v$ and $vu_2 = u_3(u_2u_3)^{m-1}$. Now we can apply induction hypothesis on the last equation and we obtain that all non-commuting solutions are of the form

$$u_3 = qp\,, \quad u_2 = p(qp)^k, \quad v = u_3(u_2u_3)^{m-2}pq\,, \quad k \geq 0\,,$$

which implies $u_1 = u_3u_2v = u_3(u_2u_3)^{m-1}pq$. We obtained exactly solution (11), which completes the proof. □

The following lemma gives us the characterization of solutions of equation (9) and hence also of sets X allowing an X-ambiguous bi-infinite word in case *(iii)*.

Lemma 11. *Assume that α and β do not commute. All solutions of equation* (9) *satisfying $n \geq 1$, $i \geq 2$ are*

$$\begin{aligned} \beta &= p(qp)^j, \\ s &= qp\beta^{n-1}, \\ \alpha &= s^{i-1}\beta^{-1}pq\,, \end{aligned} \tag{12}$$

where $p, q \in \Sigma^+$, $j \geq 0$ and $j < i$ if $n = 1$.

Proof. It is easy to check that (12) is a solution of equation (9). For the converse implication we analyze 3 cases.
Case 1. Assume that $|s| > |\beta^n|$. Then we have $\alpha = s^{i-1}q$ and $s = q\beta^n$ for some $q \neq 1$. This is solution (12) for $j = 0$, $p = \beta$.
Case 2. Assume that $|s| < |\beta^n|$ and $n = 1$. The situation is depicted in Figure 17.

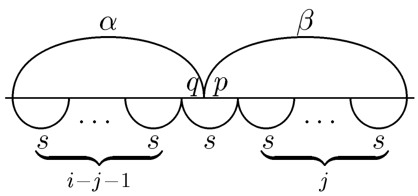

Fig. 17

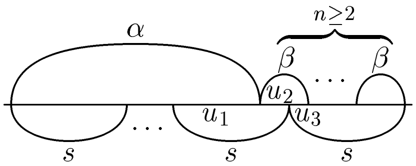

Fig. 18

Directly from the figure we can write

$$\beta = p(qp)^j, \quad s = qp\,, \quad \alpha = q(pq)^{i-j-1},$$

where $p, q \neq 1$ and $j < i$. Since

$$s^{i-1}\beta^{-1}pq = (qp)^{i-1}\left[p(qp)^j\right]^{-1}pq = (qp)^{i-j-1}q = \alpha\,,$$

we have solution (12) for $n = 1$.

Case 3. Finally assume that $|s| < |\beta^n|$ and $n \geq 2$. Since we are looking for non-commuting solutions, necessarily $|s| > |\beta^{n-1}|$ (see the proof of Lemma 7). Hence, we have a situation shown in Figure 18. According to this figure we can write $\beta = u_2 u_3$, $\alpha = s^{i-2} u_1$ and $u_1 u_2 = s = u_3 \beta^{n-1} = u_3 (u_2 u_3)^{n-1}$, which is equation (10). Now, Lemma 10 implies

$$\begin{aligned}
&\beta = u_2 u_3 = p(qp)^{k+1} = p(qp)^j \text{ for } j = k+1, \\
&s = u_1 u_2 = u_3 (u_2 u_3)^{n-2} pqp(qp)^k = qp\beta^{n-2}\beta = qp\beta^{n-1} \text{ and} \\
&\alpha = s^{i-2} u_1 = s^{i-2} u_3 (u_2 u_3)^{n-2} pq = s^{i-2} qp\beta^{n-2}\beta\beta^{-1} pq = s^{i-1}\beta^{-1} pq\,.
\end{aligned}$$

This is exactly solution (12). $\square$

The following theorem summarizes the previous results.

Theorem 4. *Consider set $X \subseteq \Sigma^+$ with $|X| = 2$. There exists an X-ambiguous bi-infinite word if and only if one of the following conditions is satisfied:*

(i) X is not a code,

(ii) $X = \{u^n, v^m\}$, where u and v are conjugates and $n, m \geq 1$,

(iii) $X = \{\alpha, \beta\}$, where α, β satisfy

$$\beta = p(qp)^j, \quad s = qp\beta^{n-1}, \quad \alpha = s^{i-1}\beta^{-1} pq\,,$$

for $p, q, s \in \Sigma^+$, $n \geq 1$, $i \geq 2$, $j \geq 0$ and $j < i$ if $n = 1$.

Theorem 4 gives us a characterization of the two-element sets possessing an X-ambiguous bi-infinite word. This characterization does not help us, if we want to decide if there is an X-ambiguous bi-infinite word for the certain set X, but we can use this theorem to generate all such sets.

Example 4. Let us choose in (12) $p = a$, $q = b$, $n = 2$, $i = 2$ and $j = 2$. We have

$$\beta = ababa\,, \quad s = baababa\,, \quad \alpha = baab\,.$$

The bi-infinite word $(\alpha\beta^2)^{\mathbb{Z}}$ has two different X-factorizations:

$$\ldots\,\overbrace{baab}^{\alpha}\overbrace{ababa}^{\beta}\overbrace{ababa}^{\beta}\ldots$$
$$\underbrace{\ldots ba}_{\beta}\underbrace{ababa}_{\beta}\underbrace{baab}_{\alpha}\underbrace{aba\ldots}_{\beta}\;.$$

7 Conclusions and Open Problems

Our Theorem 2 is closely related to the main result of [lRlR], where it is characterized when a finite word can have two disjoint X-interpretations for a binary set X. Our result could be concluded, with some effort, from the considerations in this paper. However, our proof is simpler, due to the use of the graph lemma (Lemma 5), and moreover directly formulated to obtain a defect type of theorems.

We pose an open problem asking whether Theorem 2 can be extended to arbitrary sets.

Open problem 1. Let $X \subseteq \Sigma^+$ be a finite set such that $r_{\mathrm{c}}(X) = \mathrm{card}(X)$. Does there exist a bi-infinite word w having two X-factorizations F_1 and F_2 satisfying:

(i) both F_1 and F_2 contain all elements of X, and

(ii) F_1 and F_2 are not shift-equivalent.

Observe here that even in the case of a two-element set X without the assumption that all elements of X occur in both factorizations the answer to this problem is trivially negative.

The answer is also negative if the condition $r_c(X) = \mathrm{card}(X)$ is replaced by a weaker one involving the free rank: $r_{\mathrm{f}}(X) = \mathrm{card}(X)$. This is verified by Example 1.

Another open problem asks whether Corollary 3 can be generalized for an arbitrary finite X.

Open problem 2. Let $X \subseteq \Sigma^+$ be a finite set satisfying $r_{\mathrm{c}}(X) = \mathrm{card}(X)$. Suppose that primitive roots of all elements of X are conjugates and that a bi-infinite word w has at least two different X-factorizations. Are all X-factorizations of w of the form $\alpha^{\mathbb{Z}}$, where $\alpha \in X$?

Example 5. The answer to the above question is negative if we omit the assumption $r_{\mathrm{c}}(X) = \mathrm{card}(X)$. Indeed, let $X = \{\alpha_1, \alpha_2, \alpha_3\}$, where $\alpha_1 = baa$, $\alpha_2 = aba$, $\alpha_3 = aab$. Then clearly $\alpha_1 \sim \alpha_2 \sim \alpha_3$ and the word $(abaaab)^{\mathbb{Z}}$ has two different, and even non-shift-equivalent, X-factorizations: $(\alpha_1\alpha_2)^{\mathbb{Z}}$ and $(\alpha_2\alpha_3)^{\mathbb{Z}}$.

In Section 5 we proved that in the two-element case there is at most one bi-infinite word possessing two different X-factorizations. In Section 3 we showed that if $r_{\mathrm{c}}(X) = \mathrm{card}(X)$, then there are finitely many such bi-infinite words, or more precisely, at most $\mathrm{s}(X) - \mathrm{card}(X)$ such words, where $\mathrm{s}(X)$ is the sum of lengths of words in X. We conclude with our last open problem.

Open problem 3. Let $X \subseteq \Sigma^+$ be a finite set satisfying $r_c(X) = \text{card}(X)$. Find a good upper bound for the number of bi-infinite words possessing two different X-factorizations. Does there exist an upper bound which depends only on card(X)?

Acknowledgment

The authors are grateful to Dr Tero Harju for useful discussions.

References

[CK] Choffrut, C., Karhumäki, J., Combinatorics of words, in: G. Rozenberg and A. Salomaa (eds), Handbook of Formal Languages, Vol. I, Springer, 329–438, 1997.

[lRlR] Le Rest, E., Le Rest, M. Sur la combinatoire des codes a deux mots, Theoretical Computer Science 41, 61–80, 1985.

[Lo] Lothaire, M., Combinatorics on words, Addison-Wesley, 1983.

[HK] Harju, T., Karhumäki, J., On the defect theorem and simplifiability, Semigroup Forum 33, 199–217, 1986.

[LeS] Lentin, A., Schützenberger, M. P., A combinatorial problem in the theory of free monoids, in: R.C. Bose and T.W. Dowling (eds), Combinatorial Mathematics and its Applications, Univ. North Carolina Press, 128–144, 1969.

[LyS] Lyndon, R.C., Schützenberger, M. P., The equation $a^m = b^n c^p$ in a free group, Michigan Mathematical Journal 9, 289–298, 1962.

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

http://www.tucs.abo.fi

University of Turku
- **Department of Mathematical Sciences**

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**