Multiple Factorizations of Words and Defect Effect

Juhani Karhumäki and Ján Maňuch

Department of Mathematics and
Turku Centre for Computer Science,
University of Turku, Finland
karhumak@cs.utu.fi, manuch@cs.utu.fi

Abstract. We prove that if X is a finite prefix set and w is a non-periodic bi-infinite word possessing 3 disjoint X-factorizations, then the combinatorial rank of X is at most $\mathrm{card}(X) - 2$. This is one of the rare cases when a cumulative defect effect is known to hold. Finally, connections to the critical factorization theorem are discussed.

1 Introduction

Defect theorem is one of the fundamental properties of words. It states that if a set of n words satisfies a nontrivial relation, then these words can be expressed as products of at most $n-1$ words, i.e., any nontrivial relation on words implies a defect effect, cf. [Lo]. Actually, as emphasized in [HK] and [CK], there does not exist just one, but several theorems which formalize the above defect effect, depending on requirements put on these $n-1$ words.

As argued in [HKP] defect theorems can be viewed as a weak dimension property of words. It is weak since a finite set X of words can satisfy several different, or independent as it is formalized in [HKP], relations without forcing a larger defect effect than 1, i.e., a larger defect effect than is forced by a single relation. On the other hand, the compactness property of word equations, established in [AL] and [Gu], guarantees that a finite set of words cannot satisfy infinitely many independent relations.

The above motivates to study the following, in some sense dual, problems. First, how large independent systems of relations on n words can exist that they do not force a defect effect larger than k, i.e., allow a set X of n words of rank at least $n-k$ to satisfy these relations, cf. [KP] and [HKP]. Here the rank can be defined in a number of ways, see [CK], for example as the *combinatorial rank* which is the smallest number of words needed to express all words of X as product of these words. The second problem area asks to find conditions (on relations or sets of words) which imply a cumulative defect effect, i.e., if the set X of n words satisfy k relations, then X is of rank at most $n-k$.

We attack in this note the second problem. We interpret, in a natural way, a relation on words from X as a double X-factorization of some finite or infinite word.

There are only very few results in the direction of our second problem. The graph lemma, Lemma 1 in Section 2, is such an example where the type of relations is restricted. A similar deep result is proved in [Br], extending ideas of [Ka1], [Ka2] and [Ho], where it is shown that if X is a code and has unbounded synchronizing delay in both directions, then the rank of X is at most $\mathrm{card}(X)-2$.

Our starting point is a recent result proved in [KMP1], see also [KMP2], stating that if a non-periodic bi-infinite word possesses two different X-factorizations, then the rank of X is at most $\mathrm{card}(X) - 1$. As emphasized in [KMP1] it is essential to use the notion of the combinatorial rank described above. We ask the following:

Problem 1. Let X be a set of n words and w a non-periodic bi-infinite word. Is it true that if w possesses k disjoint X-factorizations, for $k \leq \mathrm{card}(X)$, then the combinatorial rank of X is at most $\mathrm{card}(X) - k + 1$?

As we said the problem is solved affirmatively in [KMP1] in the case when $k = 2$. We do not have either a counterexample or a proof for larger values of k. However, we are able to prove the following result. If a non-periodic word possesses 3 disjoint X-factorizations, where X is a *prefix* set, then the combinatorial rank of X is at most $\mathrm{card}(X) - 2$. Even this simple case seems to be quite complicated to prove.

Finally, we note that the above problem is connected to the fundamental critical factorization theorem, or more precisely to its application, cf. Chapter 8 in [Lo]. This application studies, for a finite set X of words, the number of disjoint X-interpretations of a given finite word. In particular, it gives a result which is similar to our problem in the case when $k = \mathrm{card}(X)$, cf. Section 4.

2 Preliminaries

In this section we fix our terminology and state 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^+. We are mainly interested in the case when X is a *prefix set*, i.e., no word of X is a prefix of another word of X. We denote the fact that a word u is a prefix of v as $u < v$. The set of all finite, infinite and bi-infinite words over Σ 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 $a_i a_{i+1} \ldots a_{j-1} = \alpha$, i.e., the position i is a starting position of the factor α in w. We say that X-factorization F is *periodic*, if there exist an integer k_0 such that for any integer k, $F(k) = (\alpha, i)$ and $F(k + k_0) = (\beta, j)$ with $\alpha = \beta$. Further, 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 $\alpha = \beta$.

Example 1. Let $X = \{\alpha, \beta, \gamma, \delta\}$, where $\alpha = aa$, $\beta = baab$, $\gamma = baaaab$ and $\delta = aba$. The non-periodic bi-infinite word

$$w_1 = \dots baabaabaaaabaaaab \dots = {}^{\mathbb{N}}(baa)b(aaaab)^{\mathbb{N}}$$

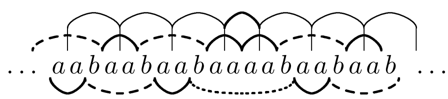
has three different X-factorizations: $F_1 = {}^{\mathbb{N}}(\beta\alpha)(\gamma\alpha\alpha)^{\mathbb{N}}$, $F_2 = {}^{\mathbb{N}}(\alpha\beta)(\alpha\alpha\gamma)^{\mathbb{N}}$ and $F_3 = {}^{\mathbb{N}}\delta(\alpha\delta)^{\mathbb{N}}$, which are pairwise non-shift-equivalent and are depicted as follows:

Clearly, these three factorizations are pairwise disjoint and also non-periodic.

Example 2. Let X be the same set as in the previous example. Take any non-periodic bi-infinite word w_2 in the set $\{aabaab, aabaaaab\}^{\mathbb{Z}}$. Any such bi-infinite word has 3 different X-factorizations: $F_1 \in \{\alpha\beta, \alpha\gamma\}^{\mathbb{Z}}$, $F_2 \in \{\alpha\beta, \alpha\alpha\beta\}^{\mathbb{Z}}$ and $F_3 \in \{\delta\delta, \delta\alpha\delta\}^{\mathbb{Z}}$, assuming that elements in the sets are selected in the same order as for w_2. For example, we will consider a part of w_2 in the form:

$$w_2 = \dots .aabaab.aabaaaab.aabaab. \dots .$$

Corresponding parts of three X-factorizations are depicted as follows:



Again, three X-factorizations are pairwise disjoint, non-shift-equivalent and non-periodic, assuming that the bi-infinite word w_2 is so.

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 prefix code we conclude that $r_{\mathrm{f}}(X) = 4$.

Example 1 together with results in [KMP1] show that in order to obtain the defect effect for bi-infinite words we have to use the combinatorial rank. It is also necessary to consider non-periodic X-factorizations or non-periodic bi-infinite words:

Example 3. In this example we show that there is a prefix set X without any defect effect and a periodic bi-infinite word with k disjoint X-factorizations.

Let $X = \{\alpha, \beta\}$, where $\alpha = a$ and $\beta = (ba)^{k-1}b$. Clearly, the bi-infinite word $w = (ab)^{\mathbb{Z}}$ has k disjoint X-factorizations of the form $(\alpha\beta)^{\mathbb{Z}}$. They are all shift-equivalent, but different. On the other hand, we have $r_{\mathrm{c}}(X) = 2 = \mathrm{card}(X)$.

Next we recall one crucial result on words that we shall need in our later considerations. For its proof the reader is referred to [CK] or [HK]. We need some terminology. 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 $\mathcal{G}_X$ is equal to X, and the set E_X of edges of $\mathcal{G}_X$ is defined by the condition:

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

We emphasize that the assumption that words in X are non-empty is crucial. Then we have:

Lemma 1 (Graph Lemma). *For each finite set $X \subseteq \Sigma^+$, the combinatorial rank of X is at most the number of connected components of $\mathcal{G}_X$.*

We shall also need the following simple lemma.

Lemma 2. *Consider words $x, y, x', y', v \in \Sigma^+$ and $z_1, z_2, w_1, w_2 \in \Sigma^*$ satisfying equations*

$$\begin{aligned} xz_1 &= vyw_1 \\ x'z_2 &= vy'w_2\,. \end{aligned} \tag{1}$$

If $y = x$, $y' = x'$ or $y = x'$, $y' = x$, i.e., if $\{x, x'\} = \{y, y'\}$, then x and x' are comparable, i.e., one is a prefix of the other.

Proof. Consider, for example, the first case: $y = x$ and $y' = x'$. Without loss of generality we can assume $|x| \leq |x'|$. If $|x'| \leq |v|$, then $x' \leq v$ and also $x \leq v$, which implies that $x \leq x'$, and we are done. Hence assume $|x'| > |v|$, i.e., $v < x'$. Now, if $|x| \leq |v|$, then $x \leq v \leq x'$ and we are again done. Thus, the only case we have to consider is the case $|x'| \geq |x| > |v|$. We can substitute $x = v\bar{x}$, $x' = v\bar{x}'$ for some $\bar{x}, \bar{x}' \in \Sigma^+$. The equations (1) transforms to

$$\bar{x}z_1 = v\bar{x}w_1, \qquad \bar{x}'z_2 = v\bar{x}'w_2\,.$$

We obtained the system of equations of the same type, but with $|\bar{x}| < |x|$ and $|\bar{x}'| < |x'|$. Hence, after a finite number of steps it must happen that the x's obtained, say $\tilde{x}$ and $\tilde{x}'$, are comparable. Clearly, if $\bar{x}, \bar{x}'$ are comparable, then so are x, x'. Inductively, we obtain that x and x' are comparable.

In the second case the proof is the same. □

3 An Example of a Cumulative Defect Effect

In this section we prove our main result.

Theorem 1. *Consider a prefix set $X \subseteq \Sigma^+$. Let w be a bi-infinite word over Σ with 3 different X-factorizations F_0, F_1, F_2. If the word w is non-periodic, then the combinatorial rank of X is at most* $\mathrm{card}(X) - 2$.

Before we start with a proof of the theorem, we define X-differences, triples and minimal triples. Consider a bi-infinite word possessing three disjoint X-factorizations F_0, F_1, F_2. Take an arbitrary occurrence of $\alpha_0 \in X$ in the X-factorization F_0 and find, for $i = 1, 2$, the first end point of an $\alpha_i \in X$ in the X-factorization F_i to the right from the beginning of the α_0, see Figure 1.

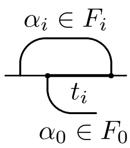


Fig. 1. An illustration of t_i.

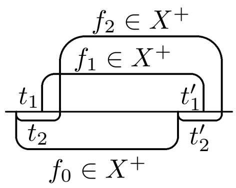


Fig. 2. An illustration of a (t_1, t_2, t_1', t_2')-triple.

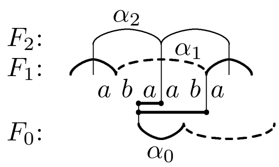
Let us denote the word between the beginning of α_0 and the end of α_i by t_i. We call the pair (t_1, t_2) an X-*difference*, or more precisely, an X-difference with respect to the triple (F_0, F_1, F_2). In what follows we simply call it an X-difference.

Note that t_1 and t_2 are always comparable. Assume that we have an occurrence of X-difference (t_1, t_2) followed by an occurrence of X-difference (t_1', t_2') in w. Figure 2 depicts such a situation, when $|t_1| \leq |t_2|$ and $|t_1'| \leq |t_2'|$. The triple (f_0, f_1, f_2) is called a (t_1, t_2, t_1', t_2')-*triple*. Note that a (t_1, t_2, t_1', t_2')-triple satisfies

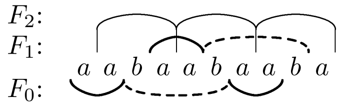
$$f_0 t_1' = t_1 f_1, \qquad f_0 t_2' = t_2 f_2. \tag{2}$$

We say that an X-difference $(\tilde{t}_1, \tilde{t}_2)$ occurs inside a triple (f_0, f_1, f_2), if the first letter of $\tilde{t}_1$ and $\tilde{t}_2$ occurs inside the word f_0. Then we say that a (t_1, t_2, t_1', t_2')-triple is *minimal*, if there is no occurrence of X-difference (t_1, t_2) or (t_1', t_2') inside the triple.

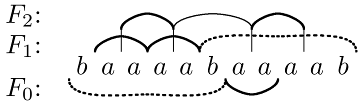
Example 4. Let us illustrate the previous definitions on the bi-infinite word w_1 from Example 1. Take as α_0, for instance, the first α in the factorization F_0 depicted in the figure of Example 1. Then $\alpha_1 = \beta$, $t_1 = aab$ and $\alpha_2 = \delta$, $t_2 = a$:



Hence, we have an occurrence of X-difference (aab, a) marked above with two black lines. Taking as α_0 the first γ in the factorization F_0 we find an occurrence of X-difference (b, ba). We have (aab, a, b, ba)-triple $(\alpha\beta\alpha, \alpha\beta, \delta\delta\delta)$:



It contains inside of it occurrences of X-differences (b, ba) and (aab, a), hence it is not minimal. On other hand the following (b, ba, aab, a)-triple $(\gamma\alpha, \alpha\alpha\gamma, \alpha\delta\alpha)$ is minimal, since it contains inside of it only an occurrence of X-difference $(aaaab, a)$:



Indeed, here it is important to remember that the order of the factorizations is fixed.

Our proof of Theorem 1 is very similar to the proof of Theorem 1 in [KMP2] assuming we have already proved the following two lemmas:

Lemma 3. *Consider a prefix set $X \subseteq \Sigma^+$. Let w be a bi-infinite word over Σ with 3 different X-factorizations F_0, F_1, F_2. If the word w possesses two different minimal (t_1, t_2, t_1, t_2)-triples, then $r_c(X) \leq \mathrm{card}(X) - 2$.*

Lemma 4. *Consider a prefix set $X \subseteq \Sigma^+$. Let w be a bi-infinite word over Σ with 3 different X-factorizations F_0, F_1, F_2. If the word w possesses for some non-empty words t_1, t_2, t_1' and t_2':*

1. *a minimal (t_1, t_2, t_1, t_2)-triple without any occurrence of X-difference (t_1', t_2') inside of it,*
2. *a minimal (t_1, t_2, t_1', t_2')-triple, and*

3. a minimal (t'_1, t'_2, t'_1, t'_2)*-triple,*

then $r_c(X) \leq \mathrm{card}(X) - 2$.

We will prove rather technical Lemma 3 later, while the proof of Lemma 4 is left for the reader since it can be done in a very similar way. Now, let us illustrate the situations these two lemmas deal with on an example.

Example 5. In Example 2 we have exactly the situation considered in Lemma 3. The non-periodic bi-infinite word

$$w_2 \in \{a^2ba^2b, a^2ba^4b\}^{\mathbb{Z}}$$

contains exactly two different minimal (aab, a, aab, a)-triples $(\alpha\beta, \alpha\beta, \delta\delta)$ and $(\alpha\gamma, \alpha\alpha\beta, \delta\alpha\delta)$.

In Example 1 we can find an illustration of the case considered in Lemma 4. The bi-infinite word

$$w_1 = {}^{\mathbb{N}}(ba^2)b(a^4b)^{\mathbb{N}}$$

contains:

1. the minimal (aab, a, aab, a)-triple $(\alpha\beta, \alpha\beta, \delta\delta)$ without any occurrence of X-difference $(aaaab, a)$ inside of it,
2. the minimal $(aab, a, aaaab, a)$-triple $(\alpha\gamma, \alpha\alpha\beta, \delta\alpha\delta)$, and
3. the minimal $(aaaab, a, aaaab, a)$-triple $(\alpha\alpha\gamma, \alpha\alpha\gamma, \alpha\delta\alpha\delta)$.

Proof (of Theorem 1). Since X is a prefix, we can assume that all three X-factorizations are pairwise disjoint. Lemmas 3 and 4 imply that if we assume that the combinatorial rank of X is at least $\mathrm{card}(X) - 1$, then any two minimal (t_1, t_2, t_1, t_2)-triples are equal, and there cannot occur all following three triples in w: a (t_1, t_2, t_1, t_2)-triple without any occurrence of X-difference (t'_1, t'_2) inside, a (t_1, t_2, t'_1, t'_2)-triple and a (t'_1, t'_2, t'_1, t'_2)-triple. Since t_1, t_2 are suffixes of words in X, there are only finitely many different X-differences. By the pigeon hole principle, there exists an X-difference (t'_1, t'_2), which occurs an infinite number of times in the word w. Each two consecutive occurrences define the minimal (t'_1, t'_2, t'_1, t'_2)-triple. If there are infinitely many occurrences to the right and also to the left from an arbitrary point in w, then clearly all three X-factorizations and the bi-infinite word w are periodic, which is a contradiction.

Hence, without loss of generality we assume that there are occurrences of (t'_1, t'_2, t'_1, t'_2)-triple only to the right from a point P. Therefore, by the pigeon hole principle, there must be an X-difference (t_1, t_2) occurring an infinite number of times to the left from the point P in w. Clearly, a (t_1, t_2, t_1, t_2)-triple occurring to the left from the point P in w does not contain any occurrence of X-difference (t'_1, t'_2). Obviously, there is a (t_1, t_2, t'_1, t'_2)-triple in the word w, which is a contradiction. $\square$

Proof (of Lemma 3). Let us consider two different minimal (t_1, t_2, t_1, t_2)-triples (f_0, f_1, f_2) and (f'_0, f'_1, f'_2). Without loss of generality we can assume that $|f_0| \leq |f'_0|$ and $t_1 < t_2$. Note that $t_1 \neq t_2$ since factorizations F_1, F_2 are disjoint.

Denote $t_0 = \mathit{1}$ and let $s_1, s_2 \in \Sigma^+$ be such words that $t_1 = s_1$, $t_2 = t_1 s_2$. We define, for $0 \le a \le b \le 2$,

$$s_{(a,b]} = s_{a+1} \dots s_b .$$

Notice that for $a = b$, $s_{(a,b]} = 1$; for $0 \le a \le 2$, $t_a = s_{(0,a]}$; and for $a \le b \le c$, $s_{(a,c]} = s_{(a,b]} s_{(b,c]}$. Next, we define $\pm$-notation: for arbitrary $a, b \in \{0, 1, 2\}$ let

$$s^+_{(a,b]} = \begin{cases} s_{(a,b]}, & \text{if } a < b, \\ \mathit{1}, & \text{otherwise,} \end{cases} \qquad s^-_{(a,b]} = \begin{cases} s_{(b,a]}, & \text{if } b < a, \\ \mathit{1}, & \text{otherwise.} \end{cases}$$

Equations (2) imply

$$f_a^{(\prime)} s_{(a,b]} = s_{(a,b]} f_b^{(\prime)}, \text{ for } a < b . \tag{3}$$

Equation (3), for each $a < b$, represents actually two equations: one with and one without the primed symbols, hence the notation $f^{(\prime)}$. Note that for $a = 0$ we have only another transcription of Equations (2) and for $a = 1, b = 2$ we have

$$f_1 s_2 = s_2 f_2, \qquad f_1' s_2 = s_2 f_2' .$$

Using our $\pm$-notation we can restate Equation (3) for any $0 \le a, b \le 2$:

$$s^-_{(a,b]} f_a^{(\prime)} s^+_{(a,b]} = s^+_{(a,b]} f_b^{(\prime)} s^-_{(a,b]} . \tag{4}$$

Indeed, it is easy to check that for $a < b$ and $b < a$ we get exactly Equation (3) and for $a = b$ tautology $f_a^{(\prime)} = f_b^{(\prime)}$.

Let $p_i \in X^*$ be a common prefix of f_i, f_i' over the alphabet X and let $r_i, r_i' \in X^*$ be words such that $f_i = p_i r_i$ and $f_i' = p_i r_i'$ (see Figure 3).

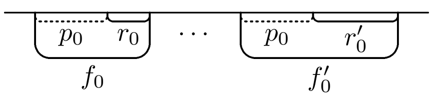


Fig. 3. An illustration of the definition of p_i, r_i, r_i' for $i = 0$.

Note that if $f_i = f_i'$ for any $i = 1, 2, 3$, then by (2) and the choice of (t_1, t_2, t_1, t_2)-triples, $f_i = f_i'$ for all $i = 1, 2, 3$. This is impossible since X is a code and the triples are different. Thus, since $|f_0| \le |f_0'|$ we have $r_i' \ne \mathit{1}$ for all i.

Let i_0, i_1, i_2 be the order of the ends of p_i's in the bi-infinite word w, as depicted in Figure 4. Note that since the X-factorizations are disjoint, the words u_1, u_2 are non-empty. Hence, we have that $|t_{i_0} p_{i_0}| < |t_{i_1} p_{i_1}| < |t_{i_2} p_{i_2}|$, where, we remind, $t_0 = 1$.

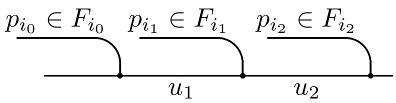


Fig. 4. An illustration of u_i's.

Example 5 (continued). Let us change the indices of factorizations F_1 and F_2, so that the condition $t_1 < t_2$ is satisfied. Hence, we will consider (a, aab, a, aab)-triples $(\alpha\beta, \delta\delta, \alpha\beta)$ and $(\alpha\gamma, \delta\alpha\delta, \alpha\alpha\beta)$ with $s_1 = a$ and $s_2 = ab$. Then

$$\begin{aligned} p_0 &= \alpha, & r_0 &= \beta, & r'_0 &= \gamma, \\ p_1 &= \delta, & r_1 &= \delta, & r'_1 &= \alpha\delta, \\ p_2 &= \alpha, & r_2 &= \beta, & r'_2 &= \alpha\beta\,. \end{aligned}$$

Since $|p_0| < |t_1p_1| < |t_2p_2|$, the order of p_i's is $i_0 = 0$, $i_1 = 1$ and $i_2 = 2$, and $u_1 = ba$ and $u_2 = a$.

Proof (continued). We have

$$t_{i_0}p_{i_0}u_1 = t_{i_1}p_{i_1}\,, \qquad t_{i_1}p_{i_1}u_2 = t_{i_2}p_{i_2}\,, \qquad t_{i_0}p_{i_0}u_1u_2 = t_{i_2}p_{i_2}\,. \tag{5}$$

Taking the first equation and multiplying both sides by $r_{i_1}s^-_{(i_0,i_1]}$ we obtain

$$s_{(0,i_0]}p_{i_0}u_1r_{i_1}s^-_{(i_0,i_1]} = t_{i_0}p_{i_0}u_1r_{i_1}s^-_{(i_0,i_1]} \overset{(5)}{=} t_{i_1}p_{i_1}r_{i_1}s^-_{(i_0,i_1]} = s_{(0,i_1]}f_{i_1}s^-_{(i_0,i_1]}\,.$$

This is equivalent to

$$s^-_{(i_0,i_1]}p_{i_0}u_1r_{i_1}s^-_{(i_0,i_1]} = s^+_{(i_0,i_1]}f_{i_1}s^-_{(i_0,i_1]} \overset{(4)}{=} s^-_{(i_0,i_1]}f_{i_0}s^+_{(i_0,i_1]} = s^-_{(i_0,i_1]}p_{i_0}r_{i_0}s^+_{(i_0,i_1]}\,,$$

hence $u_1r_{i_1}s^-_{(i_0,i_1]} = r_{i_0}s^+_{(i_0,i_1]}$. In the similar way we obtain

$$u_1r^{(\prime)}_{i_1}s^-_{(i_0,i_1]} = r^{(\prime)}_{i_0}s^+_{(i_0,i_1]}, \tag{6}$$

$$u_2r^{(\prime)}_{i_2}s^-_{(i_1,i_2]} = r^{(\prime)}_{i_1}s^+_{(i_1,i_2]}, \tag{7}$$

$$u_1u_2r^{(\prime)}_{i_2}s^-_{(i_0,i_2]} = r^{(\prime)}_{i_0}s^+_{(i_0,i_2]}. \tag{8}$$

If $r_{i_j} = \mathit{1}$, for $j = 0, 1$, then Equations (6) and (7) imply $s^+_{(i_j,i_{j+1}]} \neq \mathit{1}$, and hence also $i_j < i_{j+1}$ and $|r_{i_{j+1}}| < |s_{(i_j,i_{j+1}]}| \leq |t_{i_{j+1}}|$. By the definition of the X-difference, we then obtain $r_{i_{j+1}} = \mathit{1}$, otherwise there is an $\alpha \in X$ in the X-factorization $F_{i_{j+1}}$, which ends before the end of $f_{i_{j+1}}$ and after the end of f_0, as illustrated in Figure 5. But this is impossible by the definition of X-difference, see Figure 1.

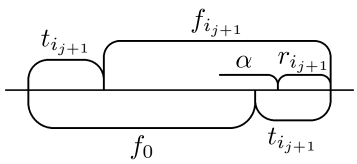


Fig. 5. A hypothetical situation where α would exist.

We have three possibilities:

(i) $r_{i_0} = r_{i_1} = r_{i_2} = 1$, $i_0 < i_1 < i_2$;
(ii) $r_{i_0} \neq 1$, $r_{i_1} \neq 1$.
(iii) $r_{i_0} \neq 1$, $r_{i_1} = r_{i_2} = 1$, $i_1 < i_2$;

Case (i). In this case there is a middle occurrence of the X-difference (t_1, t_2) in the minimal (t_1, t_2, t_1, t_2)-triple (f'_0, f'_1, f'_2), which is a contradiction.

Case (ii). Let $x, x', y, y' \in X$ be the first letters of $r_{i_0}, r'_{i_0}, r_{i_1}, r'_{i_1}$, respectively. Hence, $r^{(\prime)}_{i_0} = x^{(\prime)} \bar{r}^{(\prime)}_{i_0}$ and $r^{(\prime)}_{i_1} = y^{(\prime)} \bar{r}^{(\prime)}_{i_1}$ with some $\bar{r}_{i_0}, \bar{r}'_{i_0}, \bar{r}_{i_1}, \bar{r}'_{i_1} \in X^*$. Clearly, since p_{i_0} and p_{i_1} are maximal common prefixes, $x \neq x'$ and $y \neq y'$. Using this notation Equations (6) transform to

$$\begin{aligned} u_1 y \, \overbrace{\bar{r}_{i_1} s^-_{(i_0,i_1]}}^{w_1} &= x \, \overbrace{\bar{r}_{i_0} s^+_{(i_0,i_1]}}^{z_1} \quad , \\ u_1 y' \, \underbrace{\bar{r}'_{i_1} s^-_{(i_0,i_1]}}_{w_2} &= x' \, \underbrace{\bar{r}'_{i_0} s^+_{(i_0,i_1]}}_{z_2} \quad . \end{aligned}$$

Hence, by Lemma 2, the pair $\{x, x'\}$ is different from the pair $\{y, y'\}$.

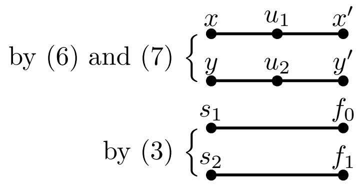


Fig. 6. A part of the dependency graph for the case *(ii)*.

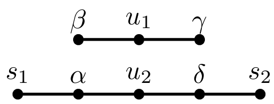
The dependency graph contains at least 6 distinct edges, as shown in Figure 6, $\mathrm{card}(X) + 4$ vertices and, most importantly, at most $\mathrm{card}(X) - 2$ components. Hence, we can bound the combinatorial rank of X using Graph Lemma:

$$r_{\mathrm{c}}(X) \leq r_{\mathrm{c}}(X \cup \{u_1, u_2, s_1, s_2\}) \leq \mathrm{card}(X) - 2\,.$$

Example 7 (continued). We have $x = \beta$, $x' = \gamma$, $y = \delta$ and $y' = \alpha$, so $\{x, x'\} \neq \{y, y'\}$. By Equations (3), (6) and (7) the set X satisfies the following 6 equalities:

$$\begin{aligned} \alpha\beta s_1 &= aa.baab.a &&= a.aba.aba &&= s_1\delta\delta, \\ \delta\delta s_2 &= aba.aba.ab &&= ab.aa.baab &&= s_2\alpha\beta, \\ u_1\delta &= ba.aba &&= baab.a &&= \beta s_1, \\ u_1\alpha\delta &= ba.aa.aba &&= baaaab.a &&= \gamma s_1, \\ u_2\beta &= a.baab &&= aba.ab &&= \delta s_2, \\ u_2\alpha\beta &= a.aa.baab &&= aa.aba.ab &&= \alpha\delta s_2. \end{aligned}$$

Hence, the dependency graph



has 2 components, so by Graph Lemma, $r_{\mathrm{c}}(X) \leq 2$.

Proof (continued). Case (iii). Let us recall Equations (8), where we set $r_{i_2} = 1$:

$$u_1 u_2 s^-_{(i_0,i_2]} = r_{i_0} s^+_{(i_0,i_2]}, \qquad u_1 u_2 r'_{i_2} s^-_{(i_0,i_2]} = r'_{i_0} s^+_{(i_0,i_2]}. \tag{9}$$

Assume that $r^{(\prime)}_{i_0}$ starts with $x^{(\prime)} \in X$, where again x must be different from x'. Note that x and x' are connected in the dependency graph through u_1. If $|x| \leq |u_1 u_2|$, then x and x' are comparable, a contradiction to the prefix property of X. Thus, we have $u_1 u_2 < x$, which implies that $s^-_{(i_0,i_2]} \neq 1$, and also $i_1 < i_2 < i_0$. Therefore, $i_0 = 2$, $i_1 = 0$ and $i_2 = 1$. Equations (9) simplify to

$$u_1 u_2 s_2 = r_2, \qquad u_1 u_2 r'_1 s_2 = r'_2, \tag{10}$$

where r_2 and r'_2 start with x and x', respectively.

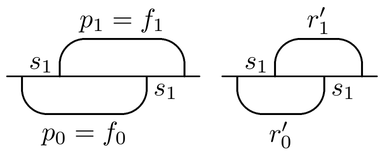


Fig. 7. An illustration of Equations (11) and (12).

Since $r_0 = r_1 = 1$, Equation (3) for $a = 0$, $b = 1$ implies

$$s_1 p_1 = p_0 s_1 . \tag{11}$$

Again since $r_0 = r_1 = 1$, Equation (7) without primes gives $u_2 = s_1$. Hence, Equation (7) with primes simplifies to

$$s_1 r'_1 = r'_0 s_1 . \tag{12}$$

Figure 7 illustrates the parts of factorizations F_{i_1} and F_{i_2} corresponding to Equations (11) and (12).

Let us analyze Equations (11) and (12). For $j = 0, 1$, let $\tilde{p}_j$ be the maximal common prefix of $p_j = f_j$ and r'_j over the alphabet X and let $p_j = \tilde{p}_j \tilde{r}_{p_j}$ and $r'_j = \tilde{p}_j \tilde{r}_{r'_j}$, for some $\tilde{p}_j, \tilde{r}_{p_j}, \tilde{r}_{r'_j} \in X^*$. There are two possibilities:

(a) $\tilde{p}_0$ ends later than $\tilde{p}_1$, i.e., $\tilde{p}_0 = s_1 \tilde{p}_1 u$ for some $u \in \Sigma^+$;
(b) $\tilde{p}_1$ ends later than $\tilde{p}_0$, i.e., $\tilde{p}_0 u = s_1 \tilde{p}_1$ for some $u \in \Sigma^+$.

Since the factorizations F_0 and F_1 are disjoint, u must be non-empty.

Case (a). Equations (11) and (12) imply

$$u\tilde{r}_{p_0}s_1 = \tilde{r}_{p_1}, \qquad u\tilde{r}_{r'_0}s_1 = \tilde{r}_{r'_1}.$$

Note that both, $\tilde{r}_{p_1}$ and $\tilde{r}_{r'_1}$, are non-empty. Hence, we can assume that $\tilde{r}_{p_1}$ and $\tilde{r}_{r'_1}$ start with different symbols y and y' and so, y and y' are connected in the dependency graph through u. It is enough to show that the pair $\{y, y'\}$ is different from the pair $\{x, x'\}$, since after that the end of the proof is essentially the same as the one in the case *(ii)*. Using Equations (3) and (10) we derive

$$\begin{aligned}
\overbrace{u_1u_2\tilde{p}_1}^{v}\,\tilde{r}_{r'_1}s_2 &\overset{(10)}{=} && && r'_2,\\
\underbrace{u_1u_2\tilde{p}_1}_{v}\,\tilde{r}_{p_1}s_2 &= u_1u_2f_1s_2 &&\overset{(3)}{=} u_1u_2s_2f_2 &&\overset{(10)}{=} r_2f_2\,.
\end{aligned}$$

Setting in Lemma 2, $v = u_1u_2\tilde{p}_1$ we obtain $\{y, y'\} \neq \{x, x'\}$.

Example 8. Consider again set $X = \{\alpha, \beta, \gamma, \delta\}$ from Example 1. Recall that $\alpha = aa$, $\beta = baab$, $\gamma = baaaab$ and $\delta = aba$. We will take any non-periodic word w_3 in the set $\{aabaab, aabaabaa\}^{\mathbb{Z}}$. It has 3 different X-factorizations in the sets: $\{\alpha\beta, \alpha\beta\alpha\}^{\mathbb{Z}}$, $\{\delta\delta, \delta\delta\alpha\}^{\mathbb{Z}}$ and $\{\alpha\beta, \alpha\gamma\}^{\mathbb{Z}}$. Note that this example is equivalent to Example 2, we have only changed order of X-factorizations.

The pieces of all 3 factorizations of w_3 can be illustrated as follows:

We have two different (a, aab, a, aab)-triple $(\alpha\beta, \delta\delta, \alpha\beta)$ and $(\alpha\beta\alpha, \delta\delta\alpha, \alpha\gamma)$, with $s_1 = a$, $s_2 = ab$ and

$$\begin{aligned}
p_0 &= \alpha\beta, & r_0 &= 1, & r'_0 &= \alpha,\\
p_1 &= \delta\delta, & r_1 &= 1, & r'_1 &= \alpha,\\
p_2 &= \alpha, & r_2 &= \beta, & r'_2 &= \gamma\,.
\end{aligned}$$

The order of p_i's is $i_0 = 2$, $i_1 = 0$ and $i_2 = 1$ with $u_1 = b$ and $u_2 = a = s_1$. This is case *(iii)*. It is easy to check that the set X satisfies Equations (11) and (12):

$$s_1\delta\delta = \alpha\beta s_1, \qquad s_1\alpha = \alpha s_1\,.$$

Next, we have

$$\begin{aligned}
\tilde{p}_0 &= \alpha, & \tilde{r}_{p_0} &= \beta, & \tilde{r}_{r'_0} &= 1,\\
\tilde{p}_1 &= 1, & \tilde{r}_{p_1} &= \delta\delta, & \tilde{r}_{r'_1} &= \alpha\,.
\end{aligned}$$

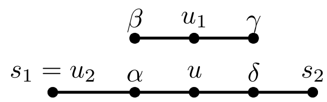
Hence, $|s_1\tilde{p}_1| < |\tilde{p}_0|$, i.e., $\tilde{p}_0$ ends in w_3 later than $\tilde{p}_1$, so we are in case *(a)* with $u = a$. Equations (11) and (12) imply

$$u\beta s_1 = \delta\delta, \qquad us_1 = \alpha,$$

hence, α and δ are connected through u in the dependency graph of the set X. On other hand Equations (10)

$$u_1u_2s_2 = \beta, \qquad u_1u_2\alpha s_2 = \gamma$$

imply that β and γ are connected through u_1. Hence again, the dependency graph has 2 components:



Proof (continued). Case (b). The proof is similar as in the previous case. Equations (11) and (12) imply

$$\tilde{r}_{p_0}s_1 = u\tilde{r}_{p_1}, \qquad \tilde{r}_{r'_0}s_1 = u\tilde{r}_{r'_1}. \tag{13}$$

Equation (3) implies that s_2 is a prefix of $f_1 = p_1 = \tilde{p}_1\tilde{r}_{p_1}$. We will show that x and x' are comparable, and hence X is not a prefix.

First, assume that $s_2 < \tilde{p}_1$. Since, by (10), $x < r_2 = u_1u_2s_2$, it is enough to show that x' and $u_1u_2s_2$ are comparable. Again by Equation (10) we have $x' < r'_2 = u_1u_2r'_1s_2 = u_1u_2\tilde{p}_1\tilde{r}_{r'_1}s_2$. And since $s_2 < \tilde{p}_1$, x' and $u_1u_2s_2$ are comparable.

Second, we will assume that $|\tilde{p}_1| < |s_2|$, i.e., $s_2 = \tilde{p}_1\omega$ for some $\omega \in \Sigma^+$. By the definition of X-difference, we have $|\tilde{p}_0\tilde{r}_{p_0}| = |p_0| = |f_0| = |f_2| > |s_1s_2|$. On the other hand, $|\tilde{p}_0| < |\tilde{p}_0u| = |s_1\tilde{p}_1| < |s_1s_2|$. Therefore, the word $\tilde{r}_{p_0}$ must be non-empty. If also $\tilde{r}_{r'_0} \neq 1$, then we can proceed as in case *(a)* choosing y and y' to be the starting symbols of $\tilde{r}_{p_0}$ and $\tilde{r}_{r'_0}$ over the alphabet X.

Unfortunately, it can also happen that $\tilde{r}_{r'_0} = 1$. Let us consider this case. We have $r'_0 = \tilde{p}_0$. The word $\tilde{p}_1$ in r'_1 on Figure 7 should end after the beginning of s_1. By the definition of X-difference, this is only possible if $r'_1 = \tilde{p}_1$ and $\tilde{r}_{r'_1} = 1$. The second of Equations (13) implies $u = s_1$, so we have

$$s_1\tilde{p}_1 = \tilde{p}_0s_1, \text{ by (12)}, \qquad s_1\tilde{r}_{p_1} = \tilde{r}_{p_0}s_1, \text{ by (13)}.$$

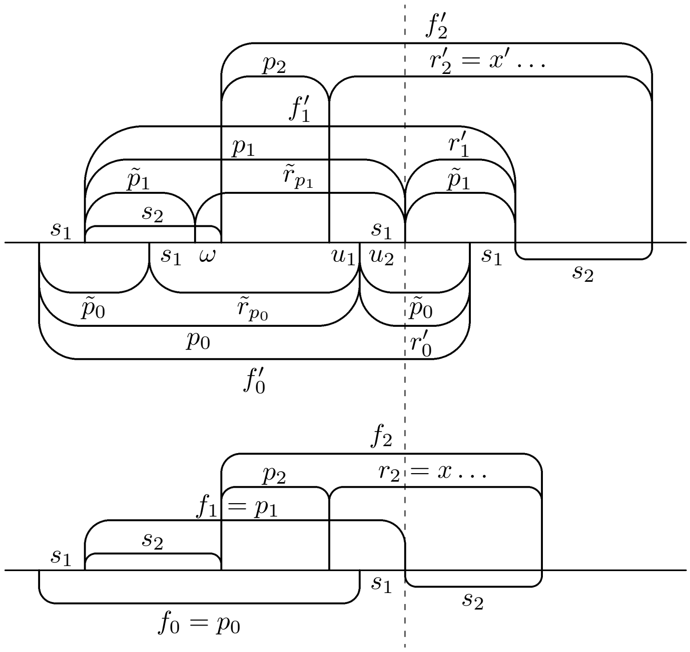


Fig. 8. An illustration of triples (f_0, f_1, f_2) (down) and (f_0', f_1', f_2') (up) in case *(iii.b)* when $\tilde{r}_{r_0'}$ is the empty word. Note that triples are same to the left from the dashed line, but they can differ to the right from it.

Thus, we can perform the same kind of analysis as we did for Equations (11) and (12). Then either $\tilde{r}_{p_0} < \tilde{p}_0$, or $\tilde{p}_0 < \tilde{r}_{p_0}$, or the words $\tilde{p}_0$ and $\tilde{r}_{p_0}$ are not comparable. In the first case, we will show that x and x' are comparable, as we wanted. In the second case, we obtain again equations of type (11) and (12), and we can continue inductively. Since the words in the new equations are shorter, we have to arrive to one of the other two cases after a final number of steps. In the third case, we can take the starting symbols of $\tilde{p}_0$ and $\tilde{r}_{p_0}$ over the alphabet X for the values of y and y' and proceed as in case *(a)*.

Hence, consider the first case $\tilde{r}_{p_0} < \tilde{p}_0$. Multiplying this inequality by s_1 and using Equations (13) we obtain $\tilde{r}_{p_1} < \tilde{p}_1$. Since $\tilde{p}_1\omega = s_2 < f_1 = \tilde{p}_1\tilde{r}_{p_1}$, we have $\omega < \tilde{r}_{p_1} < \tilde{p}_1$. Hence:

$$u_1u_2s_2 = u_1u_2\tilde{p}_1\omega < u_1u_2\tilde{p}_1\tilde{p}_1 < u_1u_2\tilde{p}_1\tilde{p}_1\omega = u_1u_2\tilde{p}_1s_2 = u_1u_2r_1's_2\,,$$

which together with Equation (10) gives that x' and $u_1u_2s_2$ are comparable. This is, as we have seen before, a contradiction. □

Note that in the proof of Theorem 1 we have used the fact that at least 1 of the X-factorizations is non-periodic, and not the fact that the bi-infinite word w is non-periodic. Hence, we have proved the following theorem:

Theorem 2. *Consider a prefix set $X \subseteq \Sigma^+$. Let w be a bi-infinite word over Σ with 3 disjoint X-factorizations F_0, F_1, F_2. If at least one of three X-factorizations is non-periodic, then the combinatorial rank of X is at most* $\mathrm{card}(X)-2$.

Nevertheless, under assumption that X is a prefix, this theorem is equivalent to Theorem 1. The following example shows that in Theorem 2, but not in Theorem 1, we have to put some assumptions on the set X, for example, that it is a code.

Example 9. Let $X = \{\alpha, \beta, \gamma\}$, where $\alpha = ababa$, $\beta = b$ and $\gamma = ababab$. Then the periodic bi-infinite word $w = (ab)^{\mathbb{Z}}$ has three disjoint X-factorizations of the form $\{\alpha\beta, \gamma\}^{\mathbb{Z}}$. We can choose them to be non-periodic and not shift-equivalent. The combinatorial rank of the set X is 2, so in this case the defect effect is only by 1.

A variant of our problem related to Theorem 2 is as follows:

Problem 2. Let X be a code and w a bi-infinite word. Is it true that if, for $k \leq \mathrm{card}(X)$, w possesses k disjoint X-factorizations, such that at least one of them is non-periodic, then the combinatorial rank of X is at most $\mathrm{card}(X) - k + 1$?

There are two natural directions to generalize Theorems 1 and 2: either relax the prefix condition or assume more than 3 factorizations. We believe that the ideas of our above proof can be used to prove at least some such extensions.

4 A Connection to the Critical Factorization Theorem

In this section we recall an application of the critical factorization theorem, cf. Chapter 8 in [Lo], and look at how it is connected to Problem 1 in the cases $k = \mathrm{card}(X)$ and $k = \mathrm{card}(X) + 1$. First, we need a few definitions.

Let $w = a_1 a_2 \dots a_n$ be a word of length n over the alphabet Σ. A positive integer p is a *period* of w if, for any $i \in \{1, \dots, n-p\}$, $a_{i+p} = a_i$. The minimal period of w is called *the period* of w, denoted as $\mathrm{p}(w)$.

Let X be a set of non-empty words over the alphabet Σ. An *X-interpretation* of a word $w \in \Sigma^+$ is a sequence $x, x_1, \dots, x_n, y$ of words such that

$$xwy = x_1 \dots x_n,$$

where, for any $i \in \{1, \dots, n\}$, $x_i \in X$, x is a proper prefix of x_1 and y is a proper suffix of x_n. Two X-factorizations $x, x_1, \dots, x_n, y$ and $x', x'_1, \dots, x'_m, y'$ of w are *disjoint*, if for each $i \leq n$ and $j \leq m$, we have $x^{-1}x_1 \dots x_i \neq {x'}^{-1}x'_1 \dots x'_j$.

The application of the critical factorization theorem states, cf. [Lo]:

Proposition 1. *Let $w \in \Sigma^+$ and $X \subseteq \Sigma^+$ be a finite set satisfying* $\mathrm{p}(x) < \mathrm{p}(w)$ *for all $x \in X$. Then w has at most* $\mathrm{card}(X)$ *disjoint X-interpretations.*

Already in [Lo] it was noted that the bound in the proposition is close to the optimal: for each $n \geq 2$, words of the form $w \in (a^{2n-2}b)^+$ have exactly n disjoint X-interpretations for $X = \{a^n, b, aba, \ldots, a^{n-1}ba^{n-1}\}$.

It was conjectured by Schützenberger that the exact value in Proposition 1 is $\mathrm{card}(X) - 1$. If this would be true, it would imply that a non-periodic bi-infinite word can possess at most $\mathrm{card}(X) - 1$ disjoint X-factorizations, which is also an immediate consequence of an affirmative answer to our Problem 1 in the case $k = \mathrm{card}(X)$. But in [Ma] there was given an example showing that the bound in Proposition 1 is optimal.

Corollary 1. *Let X be a set of non-empty words and w a non-periodic bi-infinite word. Then w can possess at most* $\mathrm{card}(X)$ *disjoint X-factorizations.*

Proof. Let $w_i \in \Sigma$, $i \in \mathbb{Z}$ be letters of the bi-infinite word w:

$$w = \ldots w_{-2}w_{-1}w_0w_1w_2 \ldots .$$

We define the sequence $\{u_i\}_{i\geq 0}$ of finite words as follows

$$u_i = w_{-i} \ldots w_{-1}w_0w_1 \ldots w_i .$$

Clearly, $\mathrm{p}(u_{i+1}) \geq \mathrm{p}(u_i)$. Hence, the sequence $\{\mathrm{p}(u_i)\}_{i\geq 0}$ is a non-decreasing sequence. Assume that it is upper bounded, i.e., there are positive integers j, p such that, for all $i \geq j$, $\mathrm{p}(u_i) = p$. Then the bi-infinite word w is periodic with a period p, which is a contradiction. Thus, there exists a positive integer j such that $\mathrm{p}(u_j)$ is greater than the periods of words in X. Assume that w possesses $\mathrm{card}(X) + 1$ disjoint X-factorizations. We can construct, in a natural way, $\mathrm{card}(X) + 1$ disjoint X-interpretations of the word u_j. But this together with Proposition 1 yields to a contradiction. $\square$

The above motivates us to formulate a stronger version of our Problem 1:

Problem 3. Let $w \in \Sigma^+$ and $X \subseteq \Sigma^+$ be a finite set satisfying $\mathrm{p}(x) < \mathrm{p}(w)$ for all $x \in X$. Is it true that w has at most $\mathrm{card}(X) + 1 - r_{\mathrm{c}}(X)$ disjoint X-interpretations?

References

[AL] Albert, M. H., Lawrence, J., *A proof of Ehrenfeucht's conjecture*, Theoretical Computer Science 41, no. 1, 121-123, 1985.

[Br] Bruyère, V., *Codes*, Chapter 7, in: M. Lothaire (ed), Algebraic combinatorics on words (to appear).

[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.

[Gu] Guba, V.S., *Equivalence of infinite systems of equations in free groups and semigroups to finite subsystems*, Mat. Zametki 40, no. 3, 321–324, 1986.

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

[HKP] Harju, T., Karhumäki, J., Plandowski, W., *Independent systems of equations*, Chapter 14, in: M. Lothaire (ed), Algebraic combinatorics on words (to appear).

[Ho] Honkala, J., *A defect property of codes with unbounded delays*, Discrete Applied Mathematics 21, 261–264, 1988.

[Ka1] Karhumäki, J., *On three-element codes*, Theoretical Computer Science 40, 3–11, 1985.

[Ka2] Karhumäki, J., *A property of three-element codes*, Theoretical Computer Science 41, 215–222, 1985.

[KMP1] Karhumäki, J., Maňuch, J., Plandowski, W., *On defect effect of bi-infinite words*, in: Proceedings MFCS'98, 23rd International Symposium, LNCS 1450, Springer, 674–682, 1998.

[KMP2] Karhumäki, J., Maňuch, J., Plandowski, W., *A defect theorem for bi-infinite words*, submitted to Theoretical Computer Science.

[KP] Karhumäki, J., Plandowski, W., *On the size of independent systems of equations in semigroups*, Theoretical Computer Science 168, 105–119, 1996.

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

[Ma] Maňuch, J., personal communication.