Multiple Factorizations of Words and Defect Effect

Juhani Karhumäki
Ján Maňuch

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

Turku Centre for Computer Science
TUCS Technical Report No 329
February 2000
ISBN 952-12-0610-1
ISSN 1239-1891

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 $\text{card}(X)-2$. 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.

TUCS Research Group
Mathematical Structures in Computer Science

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 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 [Ka] 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 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 have a quite complicated proof.

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. 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

$$w = \dots a_{-1} a_0 a_1 \dots \qquad \text{with } a_i = f_w(i)\,.$$

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 $a_i a_{i+1} \dots a_{j-1} = \alpha$, 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 $\alpha = \beta$.

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

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

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

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

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

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

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 bi-infinite words:

Example 2. 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 $X = \{a, (ba)^{k-1}b\}$. 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 $u_1, u_2, w_1, w_2 \in \Sigma^*$ satisfying equations*

$$\begin{aligned} xu_1 &= vyw_1 \\ x'u_2 &= vy'w_2 \,. \end{aligned} \tag{1}$$

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 order.

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 done again. 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

$$\bar{x}u_1 = v\bar{x}w_1, \qquad \bar{x}'u_2 = v\bar{x}'w_2 \,.$$

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 disjoint X-factorizations F_0, F_1, F_2. If the word w is non-periodic, then the combinatorial rank of X is at most $|X| - 2$.*

Proof. 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.

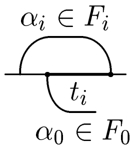

Fig. 1. An illustration of t_i.

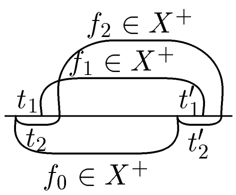

Fig. 2. An illustration of a (t_1, t_2, t_1', t_2')-triple.

Let us denote the word between the beginning of α_0 and the end of α_i by t_i. We call this situation an X-*difference* (t_1, t_2). 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* or (t_1, t_2)-*triple*, if $t_1 = t_1'$ and $t_2 = t_2'$. 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.

We prove that if the word w possesses two different minimal (t_1, t_2)-triples, then we have the defect effect by at least 2. Note that the triples (γ, γ, γ) and $(\alpha\beta, \gamma, \gamma)$, where $\alpha, \beta, \gamma \in X$, are different, even if $\alpha\beta = \gamma$. The similar proof then can be apply in the case when the bi-infinite word w contains:

- a minimal (t_1, t_2)-triple without any occurrence of X-difference (t_1', t_2') inside,
- a minimal (t_1, t_2, t_1', t_2')-triple, and
- a minimal (t_1', t_2')-triple in the bi-infinite word w.

These will imply that if we assume that the combinatorial rank of X is at least $|X| - 1$, then any two minimal (t_1, t_2)-triples are equal and there cannot occur all three triples in w: a (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')-triple. Then: 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')-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')-triple only to the right from a point. Hence, 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 in w. Clearly, a (t_1, t_2)-triple occurring to the left from the point 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.

Let us consider two different minimal (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 $0 \leq a < b \leq 2$,

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

and, for arbitrary $a, b \in \{0, 1, 2\}$,

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

Equations (2) imply

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

The equation (3) represents actually two equations: one with and one without the primed symbols.

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 $f_i' = p_i r_i'$. Note that if $f_i = f_i'$ for any $i = 1, 2, 3$, then by (2) and the choice of (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 $|f_0| \leq |f_0'|$ we have $r_i' \neq \mathit{1}$ 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 3. Note that since the X-factorizations are disjoint, the words u_1, u_2 are nonempty.

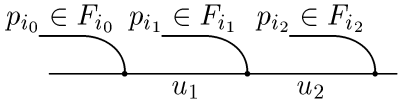

Fig. 3. An illustration of u_i's.

We have

$$t_{i_0} p_{i_0} u_1 = t_{i_1} p_{i_1}, \qquad t_{i_1} p_{i_1} u_2 = t_{i_2} p_{i_2}. \tag{4}$$

Equations (3) and (4) imply

$$\begin{aligned} u_1 r_{i_1}^{(\prime)} s^-_{(i_0,i_1]} &= r_{i_0}^{(\prime)} s^+_{(i_0,i_1]}, \\ u_2 r_{i_2}^{(\prime)} s^-_{(i_1,i_2]} &= r_{i_1}^{(\prime)} s^+_{(i_1,i_2]}. \end{aligned} \tag{5}$$

If $r_{i_j} = 1$, for $j = 0, 1$, then Equation (5) implies $s^{+}_{(i_j, i_{j+1}]} \neq 1$, 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 definitions of the X-difference, we then obtain $r_{i_{j+1}} = 1$, 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 4. But this is impossible by the definition of X-difference, see Figure 1.

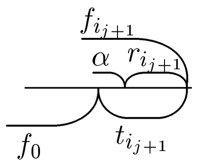

Fig. 4. A hypothetical situation where α would exist.

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)-triple (f_0', f_1', f_2'), which is a contradiction.

Case (ii). Let $r_{i_0}^{(\prime)} \in X^+$ start with $x^{(\prime)} \in X$ and $r_{i_1}^{(\prime)} \in X^+$ with $y^{(\prime)} \in X$. By Lemma 2 the pair $\{x, x'\}$ is different from the pair $\{y, y'\}$.

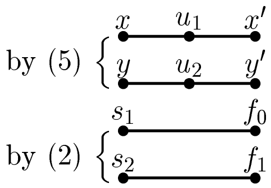

Fig. 5. A part of the dependency graph for the case *(iii)*.

The dependency graph contains at least 6 distinct edges, as shown in Figure 5, $|X| + 4$ vertices and, most importantly, at most $|X| - 2$ components. Hence, we can bound the combinatorial rank of X using Graph Lemma:

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

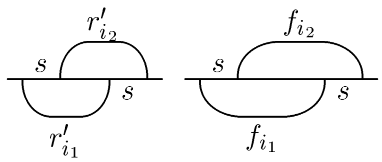

Fig. 6. An illustration of Equation (6).

Case (iii). By Equations (5) and (3), we can conclude that

$$s_{(i_1,i_2]}r'_{i_2} = r'_{i_1}s_{(i_1,i_2]}\,, \qquad s_{(i_1,i_2]}f_{i_2} = f_{i_1}s_{(i_1,i_2]}\,. \tag{6}$$

Let us denote $s = s_{(i_1,i_2]}$. Figure 6 illustrates these parts of factorizations F_{i_1} and F_{i_2} corresponding to Equations (6). Using Equation (5) we obtain

$$u_1u_2s^-_{(i_0,i_2]} = r_{i_0}s^+_{(i_0,i_2]}, \qquad u_1u_2r'_{i_2}s^-_{(i_0,i_2]} = r'_{i_0}s^+_{(i_0,i_2]}. \tag{7}$$

Assume that $r^{(\prime)}_{i_0}$ starts with $x^{(\prime)} \in X$. If $|x| < |u_1u_2|$ or $|x'| < |u_1u_2|$, then x and x' are comparable, a contradiction to the prefix property of X. Thus, we have $u_1u_2 < x$, which implies that $s^-_{(i_0,i_2]} \neq 1$, and also $i_1 < i_2 < i_0$.

By Equation (3), we derive $s^-_{(i_0,i_2]} < f_{i_2}$. If f_{i_2} is comparable with r'_{i_2}, then we can also derive $r'_{i_2} > s^-_{(i_0,i_2]}$. This and Equations (7) imply that x and x' are comparable. Hence, we may assume that f_{i_2} and r'_{i_2} are not comparable.

Let us analyze Equations (7). Let $\bar{p}_j$ be the maximal common prefix of r'_{i_j} and f_{i_j} over the alphabet X and let $r'_{i_j} = \bar{p}_j\bar{r}_{r,j}$ and $f_{i_j} = \bar{p}_j\bar{r}_{f,j}$, for some $\bar{p}_j, \bar{r}_{r,j}, \bar{r}_{f,j} \in X^*$. Note that $\bar{r}_{r,j}$ and $\bar{r}_{f,j}$ are nonempty, otherwise f_{i_2} and r'_{i_2} are comparable. There are two possibilities:

(a) $\bar{p}_1$ ends later than $\bar{p}_2$, i.e., $\bar{p}_1 = s\bar{p}_2u$ for some $u \in \Sigma^+$;

(b) $\bar{p}_2$ ends later than $\bar{p}_1$, i.e., $\bar{p}_1u = s\bar{p}_2$ for some $u \in \Sigma^+$.

Case (a). We obtain

$$u\bar{r}_{f,1}s = \bar{r}_{f,2}, \qquad u\bar{r}_{r,1}s = \bar{r}_{r,2}.$$

Assume that $\bar{r}_{f,2}$ and $\bar{r}_{r,2}$ start with different symbols y and y'. 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 (7) and the fact $i_2 < i_0$ we derive

$$\begin{aligned}
&\underline{u_1u_2\bar{p}_2}\bar{r}_{r,2}s_{(i_2,i_0]} = r'_{i_0},\\
&\underline{u_1u_2\bar{p}_2}\bar{r}_{f,2}s_{(i_2,i_0]} = u_1u_2f_{i_2}s_{(i_2,i_0]} \overset{(3)}{=} u_1u_2s_{(i_2,i_0]}f_{i_0} \overset{(7)}{=} r_{i_0}f_{i_0}\,.
\end{aligned}$$

Setting in Lemma 2, $v = u_1u_2\bar{p}_2$ we obtain $\{y, y'\} \neq \{x, x'\}$.

Case (b). The proof is almost the same as in the previous case. □

In fact we have proved a slightly stronger version of Theorem 1.

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 $|X| - 2$.*

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

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

$$xwy = x_1 \ldots x_n,$$

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

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.*

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

It was even conjectured by Schützenberger that the exact value in Proposition 1 is $\mathrm{card}(X) - 1$. If this is 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 $k = \mathrm{card}(X)$.

Corollary 1. *Let X be a set of n nonempty 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:

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

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

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

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. □ □

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.

[Ka] 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.

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**