

Combinatorics of Pseudo-Repetitions

Florin Manea^{1*}, Robert Mercas^{2**}, Dirk Nowotka^{1***}

¹ Christian-Albrechts-Universität zu Kiel, Institut für Informatik,
D-24098 Kiel, Germany,
{flm,dn}@informatik.uni-kiel.de

² Otto-von-Guericke-Universität Magdeburg, Fakultät für Informatik,
PSF 4120, D-39016 Magdeburg, Germany,
robertmercas@gmail.com

Abstract. Pseudo-repetitions are a generalization of the fundamental notion of repetitions in sequences. We develop the algorithmic foundations for questions on pseudo-repetitions and extend the work by L. Kari et.al. on the combinatorial properties of pseudo-repetitions, who investigated a restricted version of that notion in the context of bioinformatics.

1 Introduction

The notions of repetition and primitivity are fundamental concepts on sequences used in a number of fields, among them being stringology and algebraic coding theory. A word is said to be a repetition (or power) if it can be written as the repeated concatenation of one of its prefixes. In this article, we investigate algorithmic and combinatorial questions of a generalization of that concept, namely *pseudo-repetitions in words*. A word w is said to be a pseudo-repetition if it can be written as the repeated concatenation of one of its prefixes t and its image $f(t)$ under some anti-/morphism f , more precisely, $w \in t\{t, f(t)\}^+$.

Pseudo-repetitions, introduced in a restricted form by L. Kari et.al., lacked so far a developed algorithmic part, something that is usually quite important in the field where this theory originates from — bioinformatics. Our algorithmic results aim to fill this gap. We document results on natural algorithmic questions about finding pseudo-repetitions in a word for a given anti-/morphism and decide pseudo-primitivity of a word for any anti-/morphism thereby improving the results from [1]. Together with these considerations, some fundamental combinatorial properties are established.

A central result for both algorithmic and combinatorial questions regarding sequences is the so called Fine and Wilf theorem [2]. It states in a general context that if one can construct using two different words u and v two different sequences in such a way that one starts with u and the other with v , and they share a common prefix of at least the sum of the lengths of the two words minus their

* The work of Florin Manea is supported by the *DFG* grant 582014.

** The work of Robert Mercas is supported by the *Alexander von Humboldt Foundation*.

*** The work of Dirk Nowotka is supported by the *DFG Heisenberg* grant 590179.

greatest common divisor, then the two sequences are equal and, moreover, u and v are both powers of a factor of length equal to the greatest common divisor of their lengths. Up to now several generalizations of this theorem have been investigated, [3–6]. We contribute to that line of research by stating Fine and Wilf style results for pseudo-repetitions in this paper.

Background and Motivation. The notions of pseudo-repetition and -primitivity were first introduced in [6] by Czeizler, Kari, and Seki. Their motivation originated from the field of computational biology in the fact that the Watson-Crick complement can be formalized as an antimorphic involution, and the fact that both a single-stranded DNA and its complement basically encode the same information. Until now, pseudo-repetitions were considered only in the restricted cases of anti-/morphic involutions, following the original motivation.

A natural extension of these concepts is to consider anti-/morphisms in general, which is done in this paper. Considering that the notion of repetition is central in the study of combinatorics of words and the plethora of applications that this concept has, the study of pseudo-repetitions seems even more attractive, at least from a theoretical point of view. While the biological motivation seems appropriate only for the case when f is an antimorphic involution, one can imagine a series of real-life scenarios where we are interested in identifying factors of words which can be written as the iterated concatenation of a word and its encoding through some simple function f .

Some Basic Concepts. For more detailed definitions we refer to [7, 6].

Let V be a finite alphabet. The *length* of a word $w \in V^*$ is denoted by $|w|$. The *empty word* is denoted by ε . Moreover, we denote by $\text{alph}(w)$ the alphabet of all letters that occur in w . A word u is a *factor* of a word v , if $v = xuy$, for some x, y . We say that u is a *prefix* of v , if $x = \varepsilon$ and a *suffix* of v if $y = \varepsilon$. We denote by $w[i]$ the symbol at position i in w , and by $w[i..j]$ the factor of w starting at position i and ending at position j , consisting of the concatenation of the symbols $w[i], \dots, w[j]$, where $1 \leq i \leq j \leq n$. Moreover, we denote by $w = u^{-1}v$, whenever $v = uw$. The powers of a word w are defined recursively by $w^0 = \varepsilon$, for $n \geq 1$, $w^n = ww^{n-1}$, and $w^\omega = ww \dots$, an infinite concatenation of the word w . If w cannot be expressed as a power of another word, then w is said to be *primitive*.

The following well known result plays an important role in our investigation:

Theorem 1 (Fine and Wilf [2]). *Let u and v be in V^* and $d = \gcd(|u|, |v|)$. If two words $\alpha \in u\{u, v\}^*$ and $\beta \in v\{u, v\}^*$ have a common prefix of length greater than or equal to $|u| + |v| - d$, then u and v are powers of a common word of length d . Moreover, the bound $|u| + |v| - d$ is optimal.*

For some anti-/morphism $f : V^* \rightarrow V^*$ we say that f is *uniform* if there exists a number k with $f(a) \in V^k$, for all $a \in V$; if $k = 1$ then f is called *literal*. If $f(a) = \varepsilon$ for some $a \in V$, then f is called *erasing*, otherwise *non-erasing*. We say that a word w is an *f -repetition*, or, alternatively, an *f -power*, if w is in $t\{t, f(t)\}^+$, for some prefix t of w . If w is not an *f -power*, then w is *f -primitive*.

As an example, we see that $abcaab$ is primitive from the classical point of view (that is, $\mathbf{1}$ -primitive, where $\mathbf{1}$ is the identical morphism) and, moreover, for an f with $f(a) = b$, $f(b) = a$ and $f(c) = c$, the word is also f -primitive. However, when considering the morphism $f(a) = c$, $f(b) = a$ and $f(c) = b$, we note that $abcaab$ is the concatenation of ab , $f(ab) = ca$ and ab , thus, being an f -power. It is worth noting that in [6, 8] the authors were able to prove generalizations of the Fine and Wilf Theorem for f -repetitions, when f is an anti-/morphic involution.

2 Extensions of the Fine and Wilf Theorem

We say that a function $f : V^* \rightarrow V^*$ is a morphism if $f(xy) = f(x)f(y)$, for any words x and y , over V . Further, f is an antimorphism if $f(xy) = f(y)f(x)$, for any words x and y over V . Note that, when we want to define a morphism or an antimorphism it is enough to give the definitions of $f(a)$, for all $a \in V$. We write anti-/morphism whenever we want to say ‘‘antimorphism or morphism’’. Also, an anti-/morphism $f : V^* \rightarrow V^*$ is an involution when $f(f(a)) = a$ for all $a \in V$.

The following lemma is well known:

Lemma 1. *For a word w , if $w = xwy$ with $x \neq \varepsilon$ and $y \neq \varepsilon$, then x , y and w are powers of the same word t .*

The study of generalizations of the Fine and Wilf theorem for the case of pseudo-repetitions started in [6] and was continued in [8].

Theorem 2. *Let u and v be two words over an alphabet V and $f : V^* \rightarrow V^*$ a morphic involution. If $u\{u, f(u)\}^*$ and $v\{v, f(v)\}^*$ have a common prefix of length greater than or equal to $|u| + |v| - \gcd(|u|, |v|)$, then there exists $t \in V^*$ such that $u, v \in t\{t, f(t)\}^*$. Moreover, the bound $|u| + |v| - \gcd(|u|, |v|)$ is optimal.*

Theorem 3. *Let u and v be two words over an alphabet V and $f : V^* \rightarrow V^*$ an antimorphic involution.*

1. *If $|v| = 2 \gcd(|u|, |v|)$ and $u\{u, f(u)\}^*$ and $v\{v, f(v)\}^*$ have a common prefix of length greater than or equal to $2|u| - \lfloor \gcd(|u|, |v|)/2 \rfloor$, then there exists $t \in V^*$ such that $u, v \in t\{t, f(t)\}^*$.*

2. *If $|v| > 2 \gcd(|u|, |v|)$ and $u\{u, f(u)\}^*$ and $v\{v, f(v)\}^*$ have a common prefix of length greater than or equal to $2|u| + |v| - \gcd(|u|, |v|) - \lfloor \gcd(|u|, |v|)/2 \rfloor$, then there exists $t \in V^*$ such that $u, v \in t\{t, f(t)\}^*$.*

The bulk of this work deals with the case of literal bijective anti-/morphisms; at the end we give a series of results that explain why other types of anti-/morphisms are not interesting in this context. The results presented here generalize the original Fine and Wilf Theorem [2], as well as the corresponding generalizations of this theorem presented by Kari et al. [6, 8].

We start with the simple remark that for a two letter alphabet $\{a, b\}$, the case of bijective literal anti-/morphisms is quite trivial, since either f is the identity, or f is an involution ($f(a) = b$ and $f(b) = a$). The results are given by

Theorem 1 and its generalizations from [6, 8]. Thus, for the rest of this section we consider alphabets of three or more letters.

Note that if f is a bijective function from V to V , then one can see f as a permutation of V . Therefore, there exists a minimum $m > 0$ such that f^m is the identity of V . Generally, this value is denoted by $\text{ord}(f)$, called the order of f , and is less than $g(|V|)$, where g is the Landau function³; clearly, $f^{\text{ord}(f)}(x) = x$, for all $x \in V^*$.

The following observation helps us with proofs throughout the paper:

Lemma 2. *Let w be a word over V and $f : V^* \rightarrow V^*$ a bijective literal anti-/morphism. If $u = f(u)$, then, for any letter $a \in \text{alph}(u)$, we have $f^2(a) = a$.*

Proof. Let us denote $w = a_1 \cdots a_n$ with $a_i \in V$, where $1 \leq i \leq n$. Since $f(w) = w$, then $f^2(w) = f(f(w)) = f(w)$, and it follows that $w = a_1 \cdots a_n = f^2(a_1) \cdots f^2(a_n)$. Thus, $a_i = f^2(a_i)$ for all i with $1 \leq i \leq n$. \square

3 The Fine and Wilf Theorem for morphisms

Using standard techniques similar to the one in [9] one can prove the following first important result:

Theorem 4. *Take $u, v \in V^*$ and $f : V^* \rightarrow V^*$ an isomorphism with $\text{ord}(f) = k + 1$. If a word $\alpha \in u\{u, f(u), \dots, f^k(u), v, f(v), \dots, f^k(v)\}^*$ has a common prefix of length greater than or equal to $|u| + |v| - \gcd(|u|, |v|)$ with a word $\beta \in v\{u, f(u), \dots, f^k(u), v, f(v), \dots, f^k(v)\}^*$, then there exists a $t \in V^*$, such that $u, v \in t\{t, f(t), \dots, f^k(t)\}^*$.*

Proof. In the Fine and Wilf's case the proof follows by induction after the length of $|u| + |v|$. For the base case we consider $|u| = |v|$ (note that this case includes the case when $|u| + |v| = 2$), thus, we get that $u = v$, and the result follows.

Now assume without loss of generality that $|u| > |v|$. Then for some word w we have $u = vw$. Observe that the prefix of length v of $v^{-1}\beta$ is an iteration of $f(v)$. Denoting this iteration by z and changing appropriately all occurrences from α and β of iterations of f over v with iterations over z , one gets

$$v^{-1}\alpha \in w\{w, f(w), \dots, f^k(w), z, f(z), \dots, f^k(z)\}^*$$

and

$$v^{-1}\beta \in z\{w, f(w), \dots, f^k(w), z, f(z), \dots, f^k(z)\}^*.$$

The conclusion follows from a previous step of the induction. \square

³ The Landau function is defined for every natural number n to be the largest order of an element of the symmetric group S_n . Equivalently, $g(n)$ is the largest least common multiple of any partition of n , or the maximum number of times a permutation of n elements can be recursively applied to itself before it returns to its starting sequence. It is known that $\lim_{n \rightarrow \infty} \frac{\ln(g(n))}{\sqrt{n \ln(n)}} = 1$

This generalizes both Fine and Wilf and Kari et al. periodicity results.

Corollary 1. *Take $u, v \in V^*$ and $f : V^* \rightarrow V^*$ an isomorphism with $\text{ord}(f) = k+1$. If $u\{u, f(u), \dots, f^k(u)\}^*$ and $v\{v, f(v), \dots, f^k(v)\}^*$ have a common prefix of length greater than or equal to $|u| + |v| - \gcd(|u|, |v|)$, then there exists $t \in V^*$, such that $u, v \in t\{t, f(t), \dots, f^k(t)\}^*$. The bound is optimal.*

Next we show that in the case of arbitrary bijective literal morphisms the result of Theorem 4 is optimal also regarding the number of different iterations of the function f that are used in expressing both u and v . The counterexample obtained in this result exploits the algebraic properties of f , as permutation.

Proposition 1. *Let $f : V^* \rightarrow V^*$ be an isomorphism with $\text{ord}(f) = k+1$. There exist $u, v \in V^*$ with $|u| = |v| + \gcd(|u|, |v|)$ and $vf(v)$ a prefix of u^2 , such that u is not part of $t\{f^{i_1}(t), \dots, f^{i_\ell}(t)\}^*$ for any common prefix t of u and v with $v \in t\{t, f(t), \dots, f^k(t)\}^*$, and $\{i_1, \dots, i_\ell\}$ a set strictly included in $\{1, \dots, k\}$.*

Proof. Let us assume that $V = \{a_1, \dots, a_n\}$. As we explained, f is seen as a permutation of V . Assume that f has m disjoint cycles and let $c_i = (j_{i,1}, \dots, j_{i,p_i})$ for $1 \leq i \leq m$ denote these cycles (we assume that the numbers in a cycle are ordered increasingly). Also let x_i be the word obtained by concatenating the letters $a_{i,j}$ of a cycle for $1 \leq j \leq p_i$ and denote $x = x_1 \dots x_m$. Now take

$$u = xf^k(x)f^{k-1}(x) \dots f(x) \text{ and } v = xf^k(x)f^{k-1}(x) \dots f^2(x),$$

where u basically contains all possible iterations of f , while v only k factors. Note that $\gcd(|u|, |v|) = |x|$ and that $|u| = |v| + |x|$. It is straightforward to check that $vf(v)$ is a prefix of length $|u| + |v| - |x|$ of u^2 .

Now we show that there does not exist a word t , such that

$$u \in t\{f^{i_1}(t), \dots, f^{i_\ell}(t)\}^* \text{ and } v \in t\{t, f(t), f^2(t), \dots, f^k(t)\}^*$$

for any set $\{i_1, \dots, i_\ell\}$ strictly included in $\{1, \dots, k\}$.

If such a word t exists, then its length is a divisor of n (as it divides both $|u| = (k+1)n$ and $|v| = kn$). If $|t| = n$ one would not be able to generate all the factors of length n of u using only the factors $f^{i_1}(t), \dots, f^{i_\ell}(t)$, as the order of f is $k+1 > \ell$. If $|t| < n$, then $x = tf^{j_1}(t) \dots f^{j_p}(t)$ for a set of numbers $\{j_1, \dots, j_p\}$ included in $\{i_1, \dots, i_\ell\}$. Let us assume that f is not a cyclic permutation. If t does not contain any symbol of x_m , then these symbols do not appear in $f^\ell(t)$ for any ℓ , thus a contradiction with the fact that $x = tf^{j_1}(t) \dots f^{j_p}(t)$. Hence, t has as suffix a part of x_m and $f^{j_p}(t)$ is included in x_m ; from this we get that t contains only symbols from x_m , another contradiction. It follows that f is a cyclic permutation (thus, of order n) and that all the factors of length n of u begin with a different letter. Therefore, all iterations of f must be used in writing u as the catenation of factors of the form $f^i(t)$. \square

Following the results of Kari et al. a natural questions that comes up is what are good bounds for the case when we consider descriptions given by some prefix of the words and applications of a morphism to that prefix. The rest of this section is dedicated to finding such optimal bounds.

Example 1. Let i be a natural number. Consider the words $u = b^i da^i ca^i e$ and $v = b^i da^i c$, and an isomorphism f with $f(a) = b$, $f(b) = a$, $f(c) = d$, $f(d) = e$ and $f(e) = c$. The words u^2 and $vf(v)^2$ share a prefix of length $|u| + |v| - 1$ and no word t exists, such that $u, v \in t\{t, f(t)\}^*$. \square

Proposition 2. *Take $u, v \in V^*$ such that $|u| > |v| = 2 \gcd(|u|, |v|)$ and $f : V^* \rightarrow V^*$ an isomorphism. If $\alpha \in u\{u, f(u)\}^*$ and $\beta \in v\{v, f(v)\}^*$ have a common prefix of length greater than or equal to $|u| + |v|$, then there exists $t \in V^*$, such that $u, v \in t\{t, f(t)\}^*$. The bound is optimal.*

Proof. Let v_1 be the prefix of length $\gcd(|u|, |v|)$ of v , where $v = v_1 v_2$. It is rather easy to see that $u \in v\{v, f(v)\}^* v_1$ or $u \in v\{v, f(v)\}^* f(v_1)$.

When u ends with v_1 , it follows that v_2 is a prefix of u or $f(u)$, since the first u of α is followed by either u or $f(u)$. In the first case, v_2 is a prefix of v and, thus $v_1 = v_2$. In the second case, we have $v_2 = f(v_1)$. Moreover, looking at what follows v_2 in β , either $f(v_2) = v_1$ or $f(v_2) = f(v_1)$. In both cases, one may take $t = v_1$ and obtain that $u, v \in t\{t, f(t)\}^*$.

Let us now analyse the case when u ends with $f(v_1)$. Here, we obtain as above, that $f(v_2)$ is either a prefix of u or of $f(u)$. First, we obtain that $f(v_2) = v_1$, and, looking at what follows the prefix $uf(v_2)$ of β we once more get that $v_2 \in \{v_1, f(v_1)\}$. Similarly, in the second case, $f(v_2) = f(v_1)$, thus, $v_2 = v_1$. In both cases, one may take $t = v_1$ and obtain that $u, v \in t\{t, f(t)\}^*$. The conclusion follows with the optimality derived from Example 1. \square

However, when the length of the shortest word is strictly greater than two times the greatest common divisor of the two words, the result is a bit more complicated. Considering that f is a permutation, and taking into account again the algebraic properties that follow from this, we get the following results.

Proposition 3. *Take $u, v \in V^*$ such that $|u| > |v| > 2 \gcd(|u|, |v|)$, and $f : V^* \rightarrow V^*$ an isomorphism. If $\alpha \in uu\{u, f(u)\}^*$ and $\beta \in v\{v, f(v)\}^*$ have a common prefix of length greater than or equal to $2|u|$, then there exists $t \in V^*$, such that $u, v \in t\{t, f(t)\}^*$. The bound is optimal.*

Proof. Denote by u' the longest prefix of u with $u' \in v\{v, f(v)\}^*$ and by v_1 the prefix of v with $|v_1| = |u| - |u'|$. Obviously, $\gcd(|v_1|, |v|) = d \neq |v|/2$.

Let us assume first that $|v_1| < |v|/2$ and denote $v = v_1 v_2 v_3$, where $|v_2| = |v_1|$.

Consider the case when $\alpha = u' v_1 u \alpha' = u' v_1 v_1 v_2 v_3 u'' \alpha'$, where $\alpha' \in \{u, f(u)\}^*$ and $u = v u''$. Note that u' is a prefix of β , such that $\beta = u' v \beta'$, with $\beta' \in \{v, f(v)\}^*$. The discussion follows now several cases.

If $\beta = u' v v \beta''$, then by Lemma 1 we obtain that both v_1 and v are power of the same word t . Thus, we easily get that $u, v \in t\{t, f(t)\}^*$.

Now take $\beta = u' v f(v) \beta''$. We get that $v_3 = v_1^\ell x$, for some positive number ℓ and $x \in V^*$ a prefix of v_1 such that $|x| < |v_1|$ with x possibly empty. Denoting $v_1 = xy$ we obtain that $yx = f(v_1)$. If u'' starts with v we have the prefix yxv_1 of yxu'' equal to the prefix $f(v_1)f(v_1)$ of β' . Therefore, $f(v_1) = v_1$. It follows that f is the identity on the alphabet of the words u and v , and the conclusion

follows from Theorem 1. If u'' starts with $f(v)$, then $u'' = (f(v_1))^{\ell-1}f(xy)$. But the suffix $f(yx)$ matches either a factor $f(v_1)$ of β or a factor v_1 of β . In the first case we get that f is the identity on the alphabet of u and v , and we conclude by Theorem 1, while in the second case we get that $f^2(v_1) = v_1$, and, thus, f is an involution on the alphabet of u and v , and conclude by Theorem 2.

Next, we analyse the case when $\alpha = u'f(v_1)u\alpha' = u'f(v_1)v_1v_2v_3u''\alpha'$, where $\alpha' \in \{u, f(u)\}^*$ and $u = vu''$. Note that u' is a prefix of β such that $\beta = u'f(v)\beta'$ with $\beta' \in \{v, f(v)\}^*$. Here $f(v_2) = v_1$ and the suffix $f(v_1)$ of the u factor occurring before α' in α matches an $f(v_2)$ or a v_2 factor from β . In the first case we obtain that f is the identity on all letters of u and v , and we conclude by Theorem 1, while in the second case we get that $f^2(v_1) = v_1$ and, thus, f is an involution on the alphabet of u and v , and we conclude by Theorem 2.

We move now to the case when $|v_1| > |v|/2$ and set $v = v_1v_2$ with $|v_2| < |v_1|$.

Assume first that $\alpha = u'v_1u\alpha' = u'v_1v_1v_2u''\alpha'$, where $\alpha' \in \{u, f(u)\}^*$ and $u = vu''$. Note that u' is a prefix of β such that $\beta = u'v\beta'$ with $\beta' \in \{v, f(v)\}^*$. Clearly, v_2 is a prefix of v_1 . If β' starts with v , then by Lemma 1 both v_1 and v are powers of some t , and, therefore, u and v are in $t\{t, f(t)\}^*$. If β' starts with $f(v)$, then $f(v_1)$ has v_2 as a suffix.

If u'' starts with v we obtain that the suffix $f(v_2)$ of the prefix $f(v)$ of β' matches the prefix v_2 of the prefix v of u'' . Thus, f is the identity on the symbols of v_2 . It is easy to see that the symbols of v_1 are those of v_2 and $f(v_2)$, and so, f is the identity also for the symbols of v_1 and, consequently, for the symbols of u and v . The conclusion follows from Theorem 1.

Now, consider the case when u'' starts with $f(v)$. If β' starts with $f(v)f(v)$ we obtain that $f(v_2)$ is a suffix of $f(v_1)$ and, thus, it is equal to v_2 . As in the previous case, this leads to the conclusion that f is the identity on the alphabet of u and v , and the conclusion follows from Theorem 1. If β' starts with $f(v)v$ we obtain that $f(v_2)$ is a suffix of v_1 and, thus, $f^2(v_2)$ is a suffix of $f(v_1)$. Therefore, f is an involution on the alphabet of v_2 and an involution on the alphabet of u and v . The conclusion follows from Theorem 2.

Assume now that $\alpha = u'f(v_1)u\alpha' = u'f(v_1)v_1v_2u''\alpha'$, where $\alpha' \in \{u, f(u)\}^*$ and $u = vu''$. Note that u' is a prefix of β such that $\beta = u'f(v)\beta'$ with $\beta' \in \{v, f(v)\}^*$ and $f(v_2)$ is a prefix of v_1 .

Assume first that β' starts with $f(v)$. If u'' starts with $f(v_1)$, then $f(v_2)$ is a prefix of $f(v_1)$. But $f^2(v_1)$ is a prefix of $f(v_1)$ as well, so f is the identity on v_2 . As in the previous cases, we obtain that f is the identity on all letters of u and v , and with the help of Theorem 1 reach the conclusion.

When u'' starts with v , if β' starts with $f(v)v$ we get that v_1 has the suffix v_2 . Thus, $f(v_2) = v_2$ and f is the identity for the alphabet of u and v . The conclusion follows again from Theorem 1. If β' starts with $f(v)f(v)$ we get that u'' starts with either vv or with $vf(v_1)$. In the latter case the conclusion follows as in the case when u'' starts with $f(v_1)$. In the first case, the analysis is restarted, ending up with either a solution as in the case when β' starts with $f(v)v$, or the case when u'' starts with $f(v_1)$, as u ends with $f(v_1)$. Hence, we conclude that this case leads also to what we wanted to prove.

Finally, assume that β' starts with v . If u'' starts with v_1 we obtain that both $f(v_2)$ and v_2 are prefixes of v_1 , so f is the identity on the alphabet of u and v . If u'' starts with $f(v_1)$, then $f(v_1)$ starts with v_2 , so $f^2(v_2) = v_2$. Thus, f is an involution on the alphabet of u and v , and we conclude by Theorem 2.

The optimality of the result is obtained from Example 2. \square

Example 2. Let i be a natural number. Consider the words

$$u = (\text{deadebdec})^i \text{dec} \text{ and } v = (\text{deadebdec})^i,$$

and an isomorphism f with $f(a) = c$, $f(b) = a$, $f(c) = b$, $f(d) = d$ and $f(e) = e$. The words u^2 and $vf(v)^2$ share a common prefix of length $2|u| - 1$ and no word t exists, such that $u, v \in t\{t, f(t)\}^*$. \square

Proposition 4. *Take $u, v \in V^*$ such that $|u| > |v| > 2 \gcd(|u|, |v|)$ and $f : V^* \rightarrow V^*$ an isomorphism. If $\alpha \in uf(u)\{u, f(u)\}^*$ and $\beta \in v\{v, f(v)\}^*$ have a common prefix of length greater than or equal to $2|u| + \gcd(|u|, |v|)$, then there exists $t \in V^*$ such that $u, v \in t\{t, f(t)\}^*$. The bound is optimal.*

Proof. As in the proof of Proposition 3 denote by u' the longest prefix of u with $u' \in v\{v, f(v)\}^*$. Moreover, for some factorization $v = v_1 \cdots v_m$ with $|v_i| = \gcd(|u|, |v|) = d$ for all $1 \leq i \leq m$ denote by $v' = v_1 \cdots v_i$ the prefix of v for which $|v'| = |u| - |u'|$. It is straightforward that $\gcd(|v'|, |v|) = \gcd(|u|, |v|) = d \neq |v|/2$, so $\gcd(i, m) = 1$. Following the same proof of Proposition 3 $u = u'v_1$, where $u' \in v\{v, f(v)\}^*$ and v_1 is a proper prefix of v or $f(v)$ such that $\gcd(|v_1|, |v|) = \gcd(|u|, |v|) = d \neq |v|/2$. Let $v = x_1x_2 \dots x_n$, where $|x_i| = d$. It follows that $v_1 = y_1 \dots y_k$ with $\gcd(k, n) = 1$ and either $y_i = x_i$ for all $1 \leq i \leq k$ or $y_i = f(x_i)$ for all $1 \leq i \leq k$. The rest of the proof is case analysis.

Assume first that $\alpha = u'v_1 \dots v_i f(u)\alpha'$, where $\alpha' \in \{u, f(u)\}^*$, and note that, since u' is a prefix of β , we have a factorization $\beta = u'v\beta'$ with $\beta' \in \{v, f(v)\}^*$. It is rather plain that $v \in \{v_1, f(v_1), \dots, f^k(v_1)\}^*$, where $\text{ord}(f) = k+1$. Indeed, we obtain first that $f(v_1) = v_{i+1}$, and then, we obtain that $f(v_{i+1}) = v_{(2i+1) \bmod m}$ or $f(v_{i+1}) = f(v_{(2i+1) \bmod m})$; this holds as we look at the factor of β' matching the factor $f(v_{i+1})$ from the prefix $f(u)$ of $f(u)\alpha'$ and the choice depends on the starting word of β' , namely v or $f(v)$. So $v_{(2i+1) \bmod m} \in \{f(v_1), f^2(v_1)\}$ and so on: we always look at the factor $f(v_{(\ell i+1) \bmod m})$ from $f(u)$ and see what word of β' matches it. Basically, we get that $v_{(\ell i+1) \bmod m} \in \{v_1, f(v_1), \dots, f^k(v_1)\}$ for all $\ell \geq 1$. Since i and m are coprime, $\{(\ell i+1) \bmod m \mid \ell \in \mathbb{N}\} = \{1, 2, \dots, m\}$ and $v_j \in \{v_1, f(v_1), \dots, f^k(v_1)\}$ for all $1 \leq j \leq m$.

If $i < m/2$, then $f(v_1) = v_{i+1}$ and $f^2(v_1) = f(v_{i+1}) = v_{2i+1}$. For the prefix of α' we may have v_1 or $f(v_1)$, depending if either v or $f(v)$ occur at that position.

When we have $f(v_1)$, this may match a word v_{2i+1} or $f(v_{2i+1})$ from β' . In the first case we obtain that f is the identity on the letters of v_1 and the conclusion follows from Theorem 1, while in the latter $v_1 = v_{2i+1}$, f is an involution on the letters of v_1 and the conclusion follows from Theorem 2.

In the second case, the word v_1 may match a word v_{2i+1} or a word $f(v_{2i+1})$ from β' . If $v_1 = v_{2i+1}$, then f is an involution on the letters of v_1 and we conclude

by Theorem 2. Otherwise, denote $u = vu''$. If both $f(u'')$ and β' begin with $f(v)$, then the conclusion follows from Lemma 1, since from $f(v_1 \dots v_{m-i})$ and $f(v)$ being powers of some word t' we get that $v_1 \dots v_{m-i}$ and v are powers of a word t with $f(t) = t'$, and the conclusion follows immediately. If $f(u'')$ begins with $f^2(v)$ and β' begins with v we get that $f^2(v_1) = v_{i+1}$, and so f is the identity on the letters of v_1 and the conclusion follows from Theorem 1. Finally, if $f(u'')$ begins with $f^2(v)$ and β' with $f(v)$, or $f(u'')$ begins with $f(v)$ and β' with v we continue the discussion exactly as above but looking at the words that follow in $f(u'')$ and β' , respectively. However, $f(u'')$ has the suffix $f(v_1 \dots v_i)$ and this matches the beginning of a factor $f(v)$ of β' (as $f(v_{2i+1})$ appears at position $|u| + 1$ in $v_{i+1} \dots v_m \beta'$). Thus, $f(v_1) = f(v_{i+1})$, f is the identity on v_1 , and the conclusion follows from Theorem 1.

When $i > m/2$, we have $2i + 1 > m$ and $f^2(v_1) = f(v_{i+1}) \in \{v_{(2i+1) \bmod m}, f(v_{(2i+1) \bmod m})\}$. Both cases are treated analogously to the previously presented ones and the conclusion follows in the same manner.

Finally, assume that $\alpha = u' f(v_1 \dots v_i) f(u) \alpha'$, where $\alpha' \in \{u, f(u)\}^*$ and $u = vu''$. Note that u' is a prefix of β such that $\beta = u' f(v) \beta'$ with $\beta' \in \{v, f(v)\}^*$. As in the previous case, it is rather plain that $v \in \{v_1, f(v_1), \dots, f^k(v_1)\}^*$, where $\text{ord}(f) = k + 1$. Now, we only have to look what is the prefix of β' . If this prefix is $f(v)$ the conclusion follows from Lemma 1. Otherwise, v is a prefix of β' , $f(v_1) = f(v_{i+1})$, and so $v_1 = v_{i+1}$ and $v_{i+1} \in \{f(v_1), f^2(v_1)\}$. Hence, f is either the identity or an involution on v_1 and the conclusion follows.

The optimality of the result is obtained from Example 3. □

Example 3. Let i be a natural number. Consider the words

$$u = (abcabdabe)^i abc \text{ and } v = (abcabdabe)^i$$

and an isomorphism f with $f(a) = a$, $f(b) = b$, $f(c) = d$, $f(d) = e$ and $f(e) = c$. The words $uf(u)ab$ and v^3 share a common prefix of length $2|u| + \gcd(|u|, |v|) - 1$ and no word t exists such that $u, v \in t\{t, f(t)\}^*$. □

4 The Fine and Wilf Theorem for antimorphisms

For a literal bijective antimorphism f and a word t , denote by $f^{-1}(t)$ the unique word x with $f(x) = t$; clearly, $f^{2\text{ord}(f)-1}(t) = f^{-1}(t)$, as $f^{2\text{ord}(f)}(x) = x$, but not necessarily $f^{\text{ord}(f)}(x) = x$, as for some even integer k , $f^{k+1}(x)$ is x mirrored.

First, we note that a result similar to that of Theorem 4 does not hold in this case, even when we allow the size of the common prefix to be arbitrarily large.

Example 4. Let i be a natural number. Consider the words $u = a^i b^i c$ and $v = a^i b^i$, and a literal antimorphism f with $f(a) = e$, $f(b) = d$, $f(c) = c$. Moreover, f can be chosen as involution. The infinite word $w = a^i b^i c (d^i e^i)^\omega$ can be written as $w = uf(v)^\omega = vf(u)f(v)^\omega$ and all three words u, v and w are f -primitive. □

So which are the bounds in the antimorphism case? When $|v| = 2 \gcd(|u|, |v|)$ the following result is not difficult to prove:

Proposition 5. *Take $u, v \in V^*$ with $|u| > |v| = 2 \gcd(|u|, |v|)$ and a bijective literal antimorphism $f : V^* \rightarrow V^*$. If $\alpha \in u\{u, f(u)\}^*$ and $\beta \in v\{v, f(v)\}^*$ have a common prefix of length greater than or equal to $2|u| - \lfloor \gcd(|u|, |v|)/2 \rfloor$, then exists $t \in V^*$ such that $u, v \in t\{t, f(t)\}^*$ or $u, v \in t\{t, f^{-1}(t)\}^*$. The bound is optimal.*

Proof. Since $|v| = 2 \gcd(|u|, |v|)$, it follows that there exists a factorization of $v = v_1 v_2$ and $u = v_1 v_2 \dots v_{2k+1}$ with $|v_1| = |v_2| = \gcd(|u|, |v|)$ and $k \geq 1$.

Assume first that $u \in v\{v, f(v)\}^* v_1$, thus the prefix u is followed by v_2 in α . If uu is a prefix of α , then $v_1 = v_{2k+1}$, $v_2 = v_1$ and $u \in v_1\{v_1, f(v_1)\}^+$. If $uf(u)$ is a prefix of α , then $v_1 = v_{2k+1}$, $v_2 = f(v_{2k+1})$ and, thus, $v_2 = f(v_1)$. When $u \in \{v\}^+ v_1$ we have $u \in v_1\{v_1, f(v_1)\}$. If exists i such that $1 < i \leq k$ and $x_{2i-1} x_{2i} = f(v_2) f(v_1)$, we look at the factors that correspond to $f(x_{2i}) f(x_{2i-1})$ in the occurrence of $f(u)$ of the prefix $uf(u)$ of α that we analyse; note that $2i \leq 2|u| - \lfloor \gcd(|u|, |v|)/2 \rfloor$. We have $f(x_{2i}) f(x_{2i-1}) \in \{v_1 v_2, f(v_2) f(v_1)\}$. In both cases we have that f is an involution and the conclusion follows by Theorem 3.

Now assume that $u \in v\{v, f(v)\}^* f(v_2)$, that is the prefix u is followed by $f(v_1)$ in α . If uu is a prefix of α , then $f(v_2) = v_{2k+1}$ and $f(v_1) = v_1$. Looking at the prefix of length $|v|$ of the second occurrence of u in α we obtain that $v_2 = f(v_2)$ or $v_2 = v_1$. In the first case, f is an involution and we conclude by Theorem 3, while in the second case we have $u \in v_1\{v_1, f(v_1)\}^+$. If $uf(u)$ is a prefix of α , then $f(v_2) = v_{2k+1}$ and $f(v_1) = f(v_{2k+1})$, and, thus, $f(v_2) = v_1$. As above, if $u \in \{v\}^+ f(v_2)$, then $u \in v_1\{v_1, f^{-1}(v_1)\}$. If exists i such that $1 < i \leq k$ and $x_{2i-1} x_{2i} = f(v_2) f(v_1)$ we look at the factors that correspond to $f(x_{2i}) f(x_{2i-1})$ in the occurrence of $f(u)$ of the prefix $uf(u)$ of α that we analyse; note that $2i \leq 2|u| - \lfloor \gcd(|u|, |v|)/2 \rfloor$. The conclusion follows as above.

In conclusion, there always exists a prefix t of u such that $u, v \in t\{t, f(t)\}^*$ or $u, v \in t\{t, f^{-1}(t)\}^*$. The optimality of the bound $2|u| - \lfloor \gcd(|u|, |v|)/2 \rfloor$ follows from the optimality result in Theorem 3. \square

In fact, the following example shows that there are words u and v as in the statement of the previous proposition for which there exists an unique t such that $u, v \in t\{t, f(t)\}^*$ (or, alternatively, $u, v \in t\{t, f^{-1}(t)\}^*$).

Example 5. This example shows that for any bijective literal antimorphism $f : V^* \rightarrow V^*$ which is not an involution there exist two words $u, v \in V^*$ such that $|u| > |v| = 2 \gcd(|u|, |v|)$ and the words $\alpha \in u\{u, f(u)\}^*$ and $\beta \in v\{v, f(v)\}^*$ having a common prefix of length greater than or equal to $2|u| + \gcd(|u|, |v|)$ such that there exists a unique prefix x of v such that $u, v \in x\{x, f^{-1}(x)\}$ and there exists no prefix t of v such that $u, v \in t\{t, f(t)\}^*$.

Since f is not an involution, f has at least one cycle of length greater than or equal to 3; denote the elements of this cycle with x_1, x_2, \dots, x_k with $k \geq 3$, $f(x_i) = x_{i+1}$ for $1 \leq i \leq k-1$ and $f(x_k) = x_1$. Consider the words

$$u = x_1 x_k x_{k-1} \dots x_3 x_2 x_1 x_2 \dots x_{k-1} x_k x_1 x_k x_{k-1} \dots x_3 x_2$$

and

$$v = x_1 x_k x_{k-1} \dots x_3 x_2 x_1 x_2 \dots x_{k-1} x_k.$$

The words $uf(u)$ and $vf(v)^2$ are equal, but no word t exists such that u and v are both in $t\{t, f(t)\}^*$. Clearly, an infinite iteration of $uf(u) = vf(v)^2$ still has two different factorizations: one as a word from $u\{u, f(u)\}^*$ and one from $v\{v, f(v)\}^*$, respectively. Also, $u = xf^{-1}(x)x$ and $v = xf^{-1}(x)$, for $x = x_1x_kx_{k-1}\dots x_3x_2$, and there is no other prefix t of u and v such that $u, v \in t\{t, f^{-1}(t)\}^*$.

Similar examples can be devised to show that, for any bijective literal antimorphism $f : V^* \rightarrow V^*$, there exist two words $u, v \in V^*$ with $|u| > |v| = 2 \gcd(|u|, |v|)$ and the words $\alpha \in u\{u, f(u)\}^*$ and $\beta \in v\{v, f(v)\}^*$ having a common prefix of length greater than or equal to $2|u| + \gcd(|u|, |v|)$ such that there exists a unique prefix x of v such that $u, v \in x\{x, f(x)\}^*$ and there exists no prefix t of v such that $u, v \in t\{t, f^{-1}(t)\}^*$. Just take, in the above setting,

$$u = x_1x_kx_{k-1}\dots x_3x_2x_3x_4\dots x_kx_1x_2x_kx_{k-1}\dots x_3x_2$$

and

$$v = x_1x_kx_{k-1}\dots x_3x_2x_3x_4\dots x_kx_1x_2.$$

If f is an involution, then we have $f^{-1}(x) = f(x)$ for any word x . Assume that f is over an alphabet including $\{a, b\}$, with $f(a) \notin \{a, b\}$. Let i be a prime number, and consider the words $u = (ab)^i f((ab)^i)(ab)^i$ and $v = (ab)^i f((ab)^i)$. As in the previous cases, $uf(u) = v(f(v))^2$ and $u, v \in x\{x, f(x)\}^*$ for $x = (ab)^i$, but there is no other prefix t of u and v such that $u, v \in t\{t, f(t)\}^*$. \square

The following result represents a variation of Lemma 1. The proof is done identifying factors that give equalities as in Lemma 2 and conclude that the antimorphism is an involution.

Lemma 3. *For a word w and a bijective literal antimorphism f defined on the alphabet of w , if w or $f(w)$ are proper factors of $\{w, f(w)\}^2$, such that not all three factors are equal, it is the case that f is an involution.*

Proof. Assume first that $w = w_1 \dots w_n$ is a proper factor of $wf(w)$, where n is the length of w . It follows that for some j with $1 < j \leq n$ we have that $w_j \dots w_n = f(w_n) \dots f(w_j)$ and by Lemma 2 we get that for the alphabet of this factor, f is an involution. Looking now at the equality $w_1 \dots w_{j-1} = w_{n-j+1} \dots w_n$, one can easily prove that the alphabet of this factor is the same as the one of $w_j \dots w_n$, and, therefore, f is an involution for all letters in w .

If w is a proper factor of $f(w)w$, then $w_1 \dots w_j = f(w_j) \dots f(w_1)$ and, again by Lemma 2, for the alphabet of this factor f is an involution. The equality $w_{j+1} \dots w_n = w_1 \dots w_{n-j}$ shows that f is an involution for w .

If w is a proper factor of $f(w)f(w)$, then $w_1 \dots w_j = f(w_j) \dots f(w_1)$ and $w_{j+1} \dots w_n = f(w_n) \dots f(w_{j+1})$. Again by Lemma 2, we conclude that f is an involution for w .

Assume that $f(w)$ is a proper factor of $wf(w)$. It follows that for some j with $1 < j \leq n$ we have that $f(w_n) \dots f(w_j) = w_j \dots w_n$, and by Lemma 2 for the alphabet of this factor f is an involution. From the equality $f(w_{j-1}) \dots f(w_1) = f(w_n) \dots f(w_{n-j+1})$, one can prove that the alphabet of this factor is the same as that of $w_j \dots w_n$, and so f is an involution for all letters in w .

If $f(w)$ is a proper factor of $f(w)w$, then $f(w_j) \cdots f(w_1) = w_1 \cdots w_j$ and by Lemma 2 for the alphabet of this factor f is an involution. From the equality $f(w_n) \cdots f(w_{j+1}) = f(w_{n-j}) \cdots f(w_1)$, one concludes again that f is an involution for the entire alphabet of w .

Finally, take $f(w)$ a proper factor of ww . Since $f(w_n) \cdots f(w_j) = w_j \cdots w_n$ and $f(w_{j-1}) \cdots f(w_1) = w_1 \cdots w_{j-1}$, by Lemma 2 we conclude that f is an involution for $\text{alph}(w)$. \square

The case of $|v| \geq 3 \gcd(|u|, |v|)$ is proved by looking at the alignment of the prefix v , or, respectively, suffix $f(v)$, of the second factor of length $|u|$ of α with the corresponding factors from β .

Proposition 6. *Take $u, v \in V^*$ such that $|u| > |v| > 2 \gcd(|u|, |v|)$ and a bijective literal antimorphism $f : V^* \rightarrow V^*$. If $\alpha \in u\{u, f(u)\}^*$ and $\beta \in v\{v, f(v)\}^*$ have a common prefix of length greater than or equal to $2|u| + |v| - \gcd(|u|, |v|) - \lfloor \gcd(|u|, |v|)/2 \rfloor$, then there exists $t \in V^*$, such that $u, v \in t\{t, f(t)\}^*$.*

Proof. The proof of this is based on the key remark that the prefix u in α is followed by either v , the prefix of u , or $f(u)$, which has $f(v)$ as a suffix. In both cases, since $|v| \geq 3 \gcd(|u|, |v|)$, we get that either v or $f(v)$ are proper factors of some word in $\{v, f(v)\}^2$. If not all factors are equal, we conclude by Lemma 3.

It is important to note here that, when α has $uf(u)$ as a prefix, the suffix $f(v)$ of $f(u)$ is a proper factor of $\{v, f(v)\}^2$. This is true since otherwise we have that for some coprime integers k, k' with $|u| = kd$ and $|v| = k'd \geq 3d$ there exists an integer h such that $2kd = hk'd$. Thus, from $2k = hk'$ and the fact that k and k' are coprime, we get that $k = h$ and $k' = 2$, which is a contradiction.

In the other case, denoting $u \in v\{v, f(v)\}^*v'$ or $u \in v\{v, f(v)\}^*f(v'')$ for some appropriate factorization $v = v'v''$ and using Lemma 1, we have that v' and v'' are powers of the same word. The conclusion easily follows since, then, also u is an f -power of the same word.

Remark that for $f(v)$ a proper factor of $f(v)f(v)$, we have $f(v) = t^j$ for some integer j with $1 < t \leq k + 1$. In this case $v = f^{k-j+1}(t)$ and $f(v) = f^{k-j+2}(t)$.

It is important to note that the bound we propose in this theorem matches the one existing for antimorphic involutions, see Theorem 3. \square

5 Conclusion

We conclude this work with three examples showing that results similar to the ones presented here cannot be devised for more general anti-/morphisms. In all the following examples f can be considered both as a morphism and as an antimorphism, over an alphabet including $\{a, b\}$, and $i \geq 1$ a natural number.

Example 6. 1. Let $u = b^i a^i b^i a^{2i} b^{3i}$, $v = b^i a^i b^i a^{2i} b^i$, $f(a) = \varepsilon$ and $f(b) = b$. Then, $w = (uf(u)^2)^\omega = (vf(v)^4)^\omega$; there is no $|t| \leq |v|$ such that $u, v \in t\{t, f(t)\}^*$.

2. Let $u = a^i b^i a^{2i}$ and $v = a^i b^i a^i$, and $f(a) = f(b) = a$. We have $w = (uf(u)^2)^\omega = (vf(v)^3)^\omega$ and there is no t with $|t| \leq |v|$ such that w is in $t\{t, f(t)\}^*$.

3. Let $u = (ab)^{2i-1}a$ and $v = a$, $f(a) = bab$ and $f(b) = aba$. Then $w = (uf(u))^\omega = (vf(v))^\omega$, but u is not part of $\{v, f(v)\}^*$.

References

1. Chiniforooshan, E., Kari, L., Xu, Z.: Pseudopower avoidance. *Fundamenta Informaticae* **114**(1) (2012) 55–72
2. Fine, N.J., Wilf, H.S.: Uniqueness theorem for periodic functions. *Proceedings of the American Mathematical Society* **16** (1965) 109–114
3. Berstel, J., Boasson, L.: Partial words and a theorem of Fine and Wilf. *Theoretical Computer Science* **218** (1999) 135–141
4. Constantinescu, S., Ilie, L.: Generalised Fine and Wilf’s theorem for arbitrary number of periods. *Theoretical Computer Science* **339**(1) (2005) 49–60
5. Constantinescu, S., Ilie, L.: Fine and Wilf’s theorem for abelian periods. *Bulletin of EATCS* **89** (2006) 167–170
6. Czeizler, E., Kari, L., Seki, S.: On a special class of primitive words. *Theoretical Computer Science* **411** (2010) 617–630
7. Lothaire, M.: *Combinatorics on Words*. Cambridge University Press (1997)
8. Kari, L., Seki, S.: An improved bound for an extension of Fine and Wilf’s theorem, and its optimality. *Fundamenta Informaticae* **191** (2010) 215–236
9. Shallit, J.: <http://www.cs.uwaterloo.ca/~shallit/Talks/wilf3.pdf>