Fine and Wilf's theorem and pseudo-repetitions

Florin Manea[1*], Robert Mercaş[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. The notion of repetition of factors in words is central to considerations on sequences. One of the recent generalizations regarding this concept was introduced by Czeizler et al. (2010) and investigates a restricted version of that notion in the context of DNA computing and bioinformatics. It considers a word to be a pseudo-repetition if it is the iterated concatenation of one of its prefixes and the image of this prefix through an involution. We present here a series of results in the fashion of the Fine and Wilf Theorem in a more general setting where we consider the periods of some word given by a prefix of it and images of that prefix through some arbitrary morphism or antimorphism.

1 Introduction

The notions of repetition and primitivity are two fundamental concepts on words that play an important role in several research areas, e.g., stringology and algebraic coding theory. We call a word a repetition (or power) if there exists a decomposition of it in terms of one of its prefixes. This paper addresses combinatorial questions of a generalization of this concept, namely *pseudo-repetitions in words*. A word w is said to be a pseudo-repetition if it has a decomposition in terms of some prefix t and its image $f(t)$ under some morphism or antimorphism (for short "anti-/morphism") f, more precisely, $w \in t\{t, f(t)\}^+$.

A central combinatorial result regarding repetitions in words is the Fine and Wilf Theorem [1]. 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 greatest common divisor, then the two sequences are equal. Moreover, u and v are both powers of a factor of length equal to the greatest common divisor of their lengths. This theorem addresses the probably most basic natural question one could ask about repetitions in sequences. Therefore, questions in the style of Fine and Wilf are considered

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

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

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

whenever a new form of repetition in words is proposed. Up to now several generalizations of this theorem have been investigated [2–5]. We contribute to that line of research by stating Fine and Wilf style results in a more general setting where not only the words but also their images through some arbitrary functions are considered.

Having as a strong biological motivation the fact that Watson-Crick complementarity can be formalized as an antimorphic involution, and the fact that both a DNA-single stranded molecule and its complementary basically encode the same information, the authors of [5] introduce the notions of *pseudo-repetition* and *pseudo-primitivity*. In particular, a word is a *pseudo-repetition* if it can be expressed as the iterated concatenation between one of its prefixes and its image through a function f. A word is *pseudo-primitive* if it is not a pseudo-repetition. Until now, the considered functions were quite simple, being restricted to cases of anti-/morphic involutions, following the original motivation.

Considering that the notion of repetition is central in the study of combinatorics of words and the plethora of applications that this concept has in many parts of computer science, the study of pseudo-repetitions seems even more attractive, at least from a theoretical point of view. A natural extension of these results is to consider this concept for some more general classes of anti-/morphisms. Although the biological motivation seems appropriate only for the case of involutions, we are interested in identifying factors of words which can be written as the iterated concatenation of a word and its encoding through some arbitrary simple function f. In this setting, pseudo-repetitions can be seen as strings that have an intrinsic repetitive structure, hidden by rewriting some of the factors that define it through some anti-/morphism.

The next section presents some basis for the study of (pseudo-)repetitions and some basic observations that will help us throughout this work. In Section 3 we extend the pseudo-periodicity results obtained for involutory morphisms to arbitrary degree literal morphic functions. The section following it treats the antimorphic case. The results of both sections are shown to be optimal or, in one case, at least as good as the corresponding result for involutions. The paper is concluded with remarks stating the impossibility of deriving similar results for arbitrary anti-/morphisms.

We end this section with an overview of basic concepts used in this paper. For more detailed definitions we refer to [5, 6].

Let V be a finite alphabet. We denote by $\text{alph}(w)$ the alphabet of letters that occur in a word $w \in V^*$ and by ε the *empty word*. The *length* of w is denoted by $|w|$. We say u is a *factor* of w, if $w = xuy$, for some words x, y. Moreover, u is a *prefix* of w, if $x = \varepsilon$ and a *suffix* of it if $y = \varepsilon$. 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 $w[i], \ldots, w[j]$, where $1 \leq i \leq j \leq n$. Moreover, $w = u^{-1}v$, whenever $v = uw$. The powers of w are defined recursively by $w^0 = \varepsilon$, $w^n = ww^{n-1}$ for $n \geq 1$, and $w^\omega = ww\cdots$, 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*.

We say that $f : V^* \to V^*$ is a morphism if $f(xy) = f(x)f(y)$ for any words $x, y \in V^*$. On the other hand, f is an antimorphism if $f(xy) = f(y)f(x)$. Note that, to define an anti-/morphism it is enough to give the definitions of $f(a)$ for all $a \in V$. For some anti-/morphism $f : V^* \to V^*$ 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*. Recall that an anti-/morphism $f : V^* \to V^*$ is an involution when $f^2(a) = a$ for all $a \in V$. Note that, all bijective anti-/morphisms are literal. Furthermore, a bijective morphism is also called *isomorphism*.

A word w is said to be an *f-repetition*, or, an f-power, if $w \in t\{t, f(t)\}^+$ for some prefix t of w. If w is not an f-power, then w is *f-primitive*.

The word $abcaab$ is i-primitive, where i is the identical morphism, and f-primitive for some morphism or antimorphism f with $f(a) = b$, $f(b) = a$ and $f(c) = c$. However, for the morphism $f(a) = c$, $f(b) = a$ and $f(c) = b$ note that $abcaab$ is the concatenation of ab, $f(ab) = ca$ and ab, thus, an f-power in this setting. In [5, 7] the authors investigate generalizations of the Fine and Wilf Theorem for f-repetitions, when f is a morphic or an antimorphic involution.

2 Fine and Wilf's Theorem and pseudo-repetitions

The central concept of our investigation is periodicity with a main role played by the following result:

Theorem 1 (Fine and Wilf [1]). *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.*

Other important parts in this work are played by functions, namely morphisms and antimorphisms of different types.

The study of generalizations of the Fine and Wilf theorem for the case of pseudo-repetitions, that is anti-/morphic involutions, started in [5] and was continued in [7]. The following summarize the existing results for pseudo-repetitions:

Theorem 2. *Let u and v be two words over an alphabet V and $f : V^* \to V^*$ 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 in V^* and $f : V^* \to V^*$ an antimorphic involution.*
1. If $|u| > |v| = 2\gcd(|u|, |v|)$ 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 $u, v \in t\{t, f(t)\}^*$. The bound is optimal.
2. If $|u| > |v| > 2\gcd(|u|, |v|)$ 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 $u, v \in t\{t, f(t)\}^*$.

We start with the simple remark that for a two letter alphabet $\{a, b\}$, the case of bijective anti-/morphisms is quite trivial, since either f is the identity, or f is the involution given by $f(a) = b$ and $f(b) = a$. The results are given by Theorem 1 and its generalizations from [5, 7]. Thus, for the rest of this paper we consider alphabets of three or more letters.

Furthermore, since f is a bijective function from V to V, one can see f as a permutation of V. Thus, there exists a minimum $m > 0$ such that f^m is the identity of V. Generally, this value is denoted by $ord(f)$, called the order of f, and is less than $g(|V|)$, where g is the Landau function[3].

We end this section with a well known lemma and an immediate observation that will help us with the proofs throughout the paper:

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

Lemma 2. *Let $w \in V^*$ be a word and $f : V^* \to V^*$ a bijective anti-/morphism. If $w = f(w)$, then, for any letter $a \in alph(w)$, 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 $1 \leq i \leq n$. □

3 Morphisms and pseudo-repetitions

Using standard techniques similar to the one in [8] one can prove the following first important result:

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

Proof. Note that, if $|u| = |v|$ then $u = v$ and the claim follows trivially. 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 prefix by z and changing appropriately all occurrences from α and β of iterations of f over v with iterations over z, we get

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

[3] The Landau function is defined for every natural number n as the largest order of an element in 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 $\lim_{n\to\infty} \frac{\ln(g(n))}{\sqrt{n \ln(n)}} = 1$

and

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

and the claim follows by induction. □

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

Corollary 1. *Let $u, v \in V^*$ and $f : V^* \to V^*$ be an isomorphism with $ord(f) = k+1$. If $u\{u, f(u), \ldots, f^k(u)\}^*$ and $v\{v, f(v), \ldots, f^k(v)\}^*$ have a common prefix of length greater than or equal to $|u|+|v|-\gcd(|u|, |v|)$, then there exists $t \in V^*$, such that $u, v \in t\{t, f(t), \ldots, f^k(t)\}^*$. 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 $f : V^* \to V^*$ be an isomorphism with $ord(f) = k+1$. 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 $t\{f^{i_1}(t), \ldots, f^{i_\ell}(t)\}^*$ for any common prefix t of u and v with $v \in t\{t, f(t), \ldots, f^k(t)\}^*$, and $\{i_1, \ldots, i_\ell\}$ a set strictly included in $\{1, \ldots, k\}$.*

Proof. Let us assume that $V = \{a_1, \ldots, a_n\}$. As we explained, f is seen as a permutation of V. Assume that f has m disjoint cycles and let $c_i = (a_{i,1}, \ldots, a_{i,p_i})$ 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 $x = x_1 \ldots x_m$. Now take

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

where u basically contains all possible iterations of f, while v contains 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

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

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

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 $f^{i_1}(t), \ldots, f^{i_\ell}(t)$, as the order of f is $k+1 > \ell$. If $|t| < n$, then $x = tf^{j_1}(t) \ldots f^{j_p}(t)$ for a set of numbers $\{j_1, \ldots, j_p\}$ included in $\{i_1, \ldots, i_\ell\}$. 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 $x = tf^{j_1}(t) \ldots f^{j_p}(t)$. 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 $f^i(t)$. □

Following the results of Kari et al. a natural question 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

$$u = b^i da^i ca^i e \text{ and } v = b^i da^i c,$$

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 $u, v \in t\{t, f(t)\}^*$. □

Proposition 2. *Let $u, v \in V^*$ such that $|u| > |v| = 2\gcd(|u|, |v|)$ and $f : V^* \to V^*$ be 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 $v = v_1v_2$. 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. □

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. *Let $u, v \in V^*$ such that $|u| > |v| > 2\gcd(|u|, |v|)$, and $f : V^* \to V^*$ be 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 $v = v_1v_2v_3$, where $|v_2| = |v_1|$.

Consider the case when $\alpha = u'v_1u\alpha' = u'v_1v_1v_2v_3u''\alpha'$, where $\alpha' \in \{u, f(u)\}^*$ and $u = vu''$. Note that u' is a prefix of β, such that $\beta = u'v\beta'$, with $\beta' \in \{v, f(v)\}^*$. The discussion follows now several cases.

If $\beta = u'vv\beta''$, then since the factors v_1v following u' in α and vv_1 following u' in β match, by Lemma 1 we obtain that v_1 and v are both powers of the same word t. Thus, we easily get that $u, v \in t\{t, f(t)\}^*$.

Now take $\beta = u'vf(v)\beta''$. We have $v_2 = v_1$ and $v_3 = v_1^\ell x$ for some number $\ell \geq 0$ and $x \in V^*$ a possibly empty prefix of v_1 with $|x| < |v_1|$. 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 $f(v)$, this $f(v)$ ends with the suffix $f(yx)$. But this suffix matches either a prefix $f(v_1)$ of β'' or a prefix 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 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 $f^2(v_2)$ 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 have that $v_1v_2 = f(v_2)f(v_1)$ and $v_1v_2 = f(v_2)v_1$. Thus, $v_1 = f(v_1)$ 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. □

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

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

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 $u, v \in t\{t, f(t)\}^*$. □

Using the strategy of the proof of Proposition 3, one gets the following result:

Proposition 4. *Let $u, v \in V^*$ such that $|u| > |v| > 2\gcd(|u|, |v|)$ and $f : V^* \to V^*$ be 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. Denoting again by u' the longest prefix of u with $u' \in v\{v, f(v)\}^*$, for a factorization $v = v_1 \cdots v_m$ with $|v_i| = \gcd(|u|, |v|) = d$ for all $1 \leq i \leq m$ we let $v' = v_1 \cdots v_i$ be the prefix of v for which $|v'| = |u| - |u'|$. It is straightforward that $\gcd(|v'|, |v|) = \gcd(|u|, |v|) = d \neq |v|/2$, so $\gcd(i, m) = 1$.

The proof consists of several case analysis just as that of Proposition 3.

First we consider $\alpha = u'v_1 \ldots v_i f(u)\alpha'$, where $\alpha' \in \{u, f(u)\}^*$, and thus $\beta = u'v\beta'$ with $\beta' \in \{v, f(v)\}^*$, and analyze what happens when $i < m/2$ (we have in this case $f(v_1) = v_{i+1}$ and $f^2(v_1) = f(v_{i+1}) = v_{2i+1}$), and then what happens when $i > m/2$ (now we have $2i + 1 > m$ and $f^2(v_1) = f(v_{i+1}) \in \{v_{(2i+1) \bmod m}, f(v_{(2i+1) \bmod m})\}$).

Finally, we consider the case when $\alpha = u'f(v_1 \ldots v_i)f(u)\alpha'$, where $\alpha' \in \{u, f(u)\}^*$ and $u = vu''$.

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 Pseudo-repetitions for antimorphisms

For a bijective antimorphism f and a word t, denote by $f^{-1}(t)$ the unique word x with $f(x) = t$. Clearly, $f^{2ord(f)-1}(t) = f^{-1}(t)$, as $f^{2ord(f)}(x) = x$, but not necessarily $f^{ord(f)}(x) = x$, 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 common prefixes of arbitrarily large length.

Example 4. Let i be a natural number. Consider the words

$$u = a^i b^i c \text{ and } v = a^i b^i,$$

and a bijective 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. *Let $u, v \in V^*$ with $|u| > |v| = 2\gcd(|u|, |v|)$ and $f : V^* \to V^*$ be a bijective antimorphism. 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 there 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 $|v| = 2\gcd(|u|, |v|)$, there exist factorizations $v = v_1 v_2$ and $u = v_1 v_2 \dots v_{2k+1}$, where $k \geq 1$ and $|v_i| = \gcd(|u|, |v|)$ for $1 \leq i \leq 2k+1$.

Assume first that $u \in v\{v, f(v)\}^* v_1$, thus the prefix of length $|u|$ of β is followed by v_2. 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 $v_{2i-1}v_{2i} = f(v_2)f(v_1)$, we look at the factor that corresponds to $f(v_{2i})f(v_{2i-1})$ 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 $f(v_{2i})f(v_{2i-1}) \in \{v_1 v_2, f(v_2)f(v_1)\}$. 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 of length $|u|$ of β is followed by $f(v_1)$. 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 there exists i such that $1 < i \leq k$ and $v_{2i-1}v_{2i} = f(v_2)f(v_1)$ we look at the factor that corresponds to $f(v_{2i})f(v_{2i-1})$ in the occurrence of $f(u)$ of the prefix $uf(u)$ of α that we analyse; note that $2i \leq 2|u| - \lfloor \gcd(|u|, |v|)/2 \lfloor$. 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 $2|u| - \lfloor \gcd(|u|, |v|)/2 \lfloor$ follows from the optimality result in Theorem 3. □

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 a 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 antimorphism $f : V^* \to V^*$ 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 $a_1, a_2, \ldots, a_k$ with $k \geq 3$, $f(a_i) = a_{i+1}$ for $1 \leq i \leq k-1$ and $f(a_k) = a_1$. Consider the words

$$u = a_1 a_k a_{k-1} \ldots a_3 a_2 a_1 a_2 \ldots a_{k-1} a_k a_1 a_k a_{k-1} \ldots a_3 a_2$$

and

$$v = a_1 a_k a_{k-1} \ldots a_3 a_2 a_1 a_2 \ldots a_{k-1} a_k.$$

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 $x = a_1 a_k a_{k-1} \ldots a_3 a_2$, 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 antimorphism $f : V^* \to V^*$, 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,

$$u = a_1 a_k a_{k-1} \ldots a_3 a_2 a_3 a_4 \ldots a_k a_1 a_2 a_k a_{k-1} \ldots a_3 a_2$$

and

$$v = a_1 a_k a_{k-1} \ldots a_3 a_2 a_3 a_4 \ldots a_k a_1 a_2.$$

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 $u, v \in t\{t, f(t)\}^*$. □

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 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 $w = a_1 \cdots a_n$ 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 $a_j \cdots a_n = f(a_n) \cdots f(a_j)$ and by Lemma 2 we get that for the alphabet of this factor, f is an involution. Looking now at the equality $a_1 \cdots a_{j-1} = a_{n-j+1} \cdots a_n$, one can easily prove that the alphabet of this factor is the same as the one of $a_j \cdots a_n$, and, therefore, f is an involution for all letters in w.

If w is a proper factor of $f(w)w$, then $a_1 \cdots a_j = f(a_j) \cdots f(a_1)$ and, again by Lemma 2, for the alphabet of this factor f is an involution. The equality $a_{j+1} \cdots a_n = a_1 \cdots a_{n-j}$ shows that f is an involution for w.

If w is a proper factor of $f(w)f(w)$, then $a_1 \cdots a_j = f(a_j) \cdots f(a_1)$ and $a_{j+1} \cdots a_n = f(a_n) \cdots f(a_{j+1})$. 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 $f(a_n) \cdots f(a_j) = a_j \cdots a_n$, and by Lemma 2 for the alphabet of this factor f is an involution. From the equality $f(a_{j-1}) \cdots f(a_1) = f(a_n) \cdots f(a_{n-j+1})$, one can prove that the alphabet of this factor is the same as that of $a_j \cdots a_n$, and so f is an involution for all letters in w.

If $f(w)$ is a proper factor of $f(w)w$, then $f(a_j) \cdots f(a_1) = a_1 \cdots a_j$ and by Lemma 2 for the alphabet of this factor f is an involution. From the equality $f(a_n) \cdots f(a_{j+1}) = f(a_{n-j}) \cdots f(a_1)$, one concludes again that f is an involution for the entire alphabet of w.

Finally, take $f(w)$ a proper factor of ww. Since $f(a_n) \cdots f(a_j) = a_j \cdots a_n$ and $f(a_{j-1}) \cdots f(a_1) = a_1 \cdots a_{j-1}$, by Lemma 2 we conclude that f is an involution for $\text{alph}(w)$. □

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. *Let $u, v \in V^*$ be such that $|u| > |v| > 2\gcd(|u|, |v|)$ and let $f : V^* \to V^*$ be a bijective antimorphism. 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. Let $d = \gcd(|u|, |v|)$. 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. Further, it is worth noting that, when α has $uf(u)$ as a prefix, the suffix $f(v)$ of $f(u)$ is a proper factor of a word from $\{v, f(v)\}^2$. 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 $k' = 2$, a contradiction.

Now, in both cases above, since $|v| \geq 3d$, we get that there exist the words $x, y, z \in \{v, f(v)\}$ such that x is a proper factor of yz. If not all three factors x, y, and z are equal, we conclude by Lemma 3. Otherwise, we have $x = y = z = v$, and it follows by Lemma 1 that v is a power, or $x = y = z = f(v)$. In the last case, we get by Lemma 1 that $f(v) = t^j$, for some word t with $|t| \mid |u|$ and positive natural number j. Hence, we have $v = f^{-1}(t^j) = (t')^j$. As $|t'| = |t| \mid |u|$, we get that $u \in t'\{t', f(t')\}^*$. □

5 Conclusion

We end this work with some concluding remarks.

First note that the result of Proposition 6 matches the one existing for antimorphic involutions, see Theorem 3. Thus, an optimality of this bound derives an optimality for the antimorphic involution bound, or vice-versa.

The following three examples show that results similar to the ones presented here cannot be derived 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 that includes $\{a, b\}$ and some natural number $i \geq 1$.

Example 6. Consider the words

$$u = b^i a^i b^i a^{2i} b^{3i} \text{ and } v = b^i a^i b^i a^{2i} b^i,$$

and a function f with $f(a) = \varepsilon$ and $f(b) = b$. Then $w = (uf(u)^2)^\omega = (vf(v)^4)^\omega$ and one can check that there is no t with $|t| \leq |v|$ such that $u, v \in t\{t, f(t)\}^*$. □

Example 7. Consider the words

$$u = a^i b^i a^{2i} \text{ and } v = a^i b^i a^i,$$

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

Finally we consider strictly increasing anti/morphisms.

Example 8. Consider the words

$$u = (ab)^{2i-1} a \text{ and } v = a,$$

and a function f with $f(a) = bab$ and $f(b) = aba$. Then $w = (uf(u))^\omega = (vf(v))^\omega$, but u is not part of $\{v, f(v)\}^*$. □

References

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

6 Appendix

Proof of Proposition 4: 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 $1 \le i \le m$, we denote by $v' = v_1 \cdots v_k$ the prefix of v, for which $|v'| = |u| - |u'|$. It is straightforward that $\gcd(|v'|, |v|) = \gcd(|u|, |v|) = d \neq |v|/2$. Thus, $\gcd(k, m) = 1$. As in the proof of Proposition 3 we have that $u = u'v'_1$, where $u' \in v\{v, f(v)\}^*$ and v'_1 is a proper prefix of v or $f(v)$ such that $\gcd(|v'_1|, |v|) = \gcd(|u|, |v|) = d \neq |v|/2$. It follows that $v'_1 = y_1 \dots y_k$, with $\gcd(k, n) = 1$ and either $y_i = v_i$ for all $1 \le i \le k$ or $y_i = f(v_i)$ for all $1 \le i \le k$. The rest of the proof consists in analysing a series of cases.

Assume first that $\alpha = u'v_1 \dots v_i f(u)\alpha'$, where $\alpha' \in \{u, f(u)\}^*$. 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 $ord(f) = k+1$. Indeed, we obtain first that $f(v_1) = v_{i+1}$, and then, we obtain that $f(v_{i+1}) = v_{(2i+1) \mod m}$ or $f(v_{i+1}) = f(v_{(2i+1) \mod m})$; 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 $v_{(2i+1) \mod m} \in \{f(v_1), f^2(v_1)\}$, and so on: we always look at the factor $f(v_{(\ell i+1) \mod m})$ from $f(u)$ and see what word of β' matches it. Basically, we get that $v_{(\ell i+1) \mod m} \in \{v_1, f(v_1), \dots, f^k(v_1)\}$, for all $\ell \ge 1$. Since i and m are coprime, it follows that $\{(\ell i + 1) \mod m \mid \ell \in \mathbb{N}\} = \{1, 2, \dots, m\}$, thus, $v_j \in \{v_1, f(v_1), \dots, f^k(v_1)\}$ for all $1 \le j \le m$. Further we analyse two subcases.

Let us first look at the case when $i < m/2$. It follows that $f(v_1) = v_{i+1}$ and $f^2(v_1) = f(v_{i+1}) = v_{2i+1}$. Looking at 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 we obtain that $v_1 = v_{2i+1}$, and so 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'. In the first case, we get that f is an involution on the letters of v_1, and the conclusion follows from Theorem 2. In the second case, a more careful analysis is needed. Let us 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, $f(v_1) = f(v_{i+1})$, and we obtain that f is the identity on v_1, and the conclusion follows from Theorem 1.

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

Finally, assume that $\alpha = u' f(v_1 \ldots v_i) f(u) \alpha'$, 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 $v \in \{v_1, f(v_1), \ldots, f^k(v_1)\}^*$, where $ord(f) = k + 1$. 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'. In this case we obtain that $f(v_1) = f(v_{i+1})$, 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. Therefore, the conclusion follows in this case, as well.

The conclusion follows, since the optimality is obtained from Example 3. □