Fundamenta Informaticae 135 (2014) 59–72
DOI 10.3233/FI-2014-1111
IOS Press

Reconstructing Words from a σ-palindromic Language*

Srečko Brlek†**, Nadia Lafreniére**

Laboratoire de Combinatoire et d'Informatique Mathématique

Université du Québec à Montréal

C. P. 8888 Succursale "Centre-Ville", Montréal (QC), Canada H3C 3P8

Brlek.Srecko@uqam.ca, nadia.l@lacim.ca

Abstract. We consider words on a finite alphabet Σ and study the structure of its σ-palindromes, i.e. words w satisfying $w = \sigma(\widetilde{w})$ for some involution σ on the alphabet. We provide algorithms for the computation of σ-lacunas in w, that is the positions where the longest σ-palindromic suffix is not uni-occurrent. The σ-palindromic defect is explicitly computed for Sturmian words and the Thue-Morse word. Finally, the problem of reconstructing words from a given fixed set of σ-palindromes is decidable.

Keywords: Generalized palindromes, complexity, σ-palindromic lacunas, σ-palindromic defect.

1. Introduction

The motivation for studying these patterns comes for instance from molecular biology and tiling problems. Indeed, a DNA sequence is a word on the alphabet $\{\mathtt{A}, \mathtt{T}, \mathtt{C}, \mathtt{G}\}$ whose letters code respectively the four nucleotides Adenine (A), Thymine(T), Cytosine (C) and Guanine (G). These nucleotides are arranged in pairs defined by the involution $\sigma : \mathtt{A} \leftrightarrow \mathtt{T}, \mathtt{C} \leftrightarrow \mathtt{G}$. Denoting by $\widetilde{w}$ the mirror image of the word w, the σ-palindromes are words such that $w = \sigma(\widetilde{w})$ like ACCTAGGT. These patterns are known for playing a role in the secondary structure (hair pin) of the DNA, methylation sites, restriction enzymes, and on the Y chromosome. In tiling theory, tiles that tesselate the discrete plane are conveniently encoded on the Freeman alphabet $\{0, 1, 2, 3\}$ corresponding to the canonical elementary unit steps. In this case the

*with the support of NSERC (Canada)

†Address for correspondence: Laboratoire de Combinatoire et d'Informatique Mathématique, Université du Québec à Montréal, C. P. 8888 Succursale "Centre-Ville", Montréal (QC), CANADA H3C 3P8

involution is $\rho : 0 \leftrightarrow 2, 1 \leftrightarrow 3$ and a tile T is described by its contour word. Then, T tiles the plane by translation if and only it can be written as a combination

$$T = u \cdot v \cdot w \cdot \rho(\widetilde{u}) \cdot \rho(\widetilde{v}) \cdot \rho(\widetilde{w})$$

of the 3 ρ-palindromes $u \cdot \rho(\widetilde{u}), v \cdot \rho(\widetilde{v})$, and $w \cdot \rho(\widetilde{w})$ (see [7]). This characterization led to optimal recognition algorithms [18], the exhaustive generation of families of tiles connected to the Fibonacci sequence [12] and showing fractal characteristics [11].

These patterns also generalize palindromic patterns, which have been widely investigated recently. They are closely related to conjugacy and periodicity [30], and to a characterization of Sturmian words as well [19]. Some remarkable properties related to an extension of the Fine and Wilf theorem may be found in [5]. In discrete geometry, they describe local symmetries of discrete figures encoded on the 4-letter Freeman alphabet [15]. Independently, these patterns were extensively studied under the name Watson-Crick palindromes, as they play a fundamental role in the encoding of DNA strands [26, 24, 25].

As factor complexity is one of the many ways of measuring information content, palindromic complexity is a refinement that had many applications in several areas: in physics for the study of Schrödinger operators [2, 6, 23], in number theory [4] and combinatorics on words for being a powerful tool for looking at the local structure of words. It has been also applied to several classes of infinite words, for which the survey of Allouche et al. [3] provides a detailed account. In particular, the palindromic factors completely characterize Sturmian words [28], and provide a connection with the notion of recurrence for the class of smooth words [17, 16]. Droubay, Justin and Pirillo [20] noted that the palindrome complexity $|\operatorname{Pal}(w)|$ of a word w is bounded by $|w| + 1$, and that finite Sturmian (and even episturmian) words realize the upper bound. Moreover they show that the palindrome complexity is computed by a linear algorithm listing the longest palindromic suffixes that are uni-occurrent.

The aim of this article is to give an account of the basic properties of the σ-palindromic language $\operatorname{Pal}_\sigma(w)$ of words w on finite alphabets. In Section 3, we refine the bound of Droubay et al. for the σ-palindromic complexity $|\operatorname{Pal}_\sigma(w)|$ by taking into account the transpositions of σ. Words that realize that bound are no longer full as introduced in [15], and called *saturated*. In the case of infinite words with $\sigma \neq \mathrm{Id}$, we show that for all Sturmian words $|\operatorname{Pal}_\sigma(\boldsymbol{s})|$ is finite. In comparison, for the Thue-Morse word $\boldsymbol{T}$ and its image $\delta(\boldsymbol{T})$ under doubling the letters, $|\operatorname{Pal}_\sigma(\boldsymbol{T})|$ and $|\operatorname{Pal}_\sigma(\delta(\boldsymbol{T}))|$ are infinite. For periodic infinite words $\boldsymbol{w} = w^\omega$, represented conveniently by circular words, a geometric characterization of the finiteness of their σ-palindromic language is provided: $|\operatorname{Pal}_\sigma(w^\omega)|$ is infinite if and only if w is the product of two σ-palindromes, that is, the smallest periodic pattern w is *σ-symmetric*.

The language $\operatorname{Pal}_\sigma(w)$ is computed by scanning w and extracting its longest σ-palindromic suffixes that are uni-occurrent: a *σ-lacuna* is a position where it is not uni-occurrent. The number of σ-lacunas defines the σ-defect $\mathcal{D}_\sigma$, which is computed by a linear algorithm. For infinite words, we deduce that $\mathcal{D}_\sigma(\boldsymbol{s})$ is infinite for Sturmian words, that $\mathcal{D}_\sigma(\delta(\boldsymbol{T}))$ is infinite as well. In the case of periodic infinite words, we prove that the tight bound established in [15] also holds for computing the σ-defect. An optimal algorithm is provided to check if an infinite periodic word is saturated or not.

Finally, a characterization by means of a rational language is given for the language X_P of words whose σ-palindromic factors belong to a fixed and finite set P of σ-palindromes. A finite automaton recognizing X_P is then easy to obtain, and consequently, if there exists a recurrent infinite word having P for σ-palindromic factors, then there exist a periodic one sharing exactly the same σ-palindromic factors.

2. Preliminaries

Given a finite alphabet Σ consisting of *letters*, a *word* $w = w_0w_1w_2\dots w_{n-1}$ is an ordered sequence of letters of Σ. The *length* of w is $|w| = n$ and the unique word of length 0 is denoted by ε. The set of all finite words over Σ is denoted Σ^*, and $\Sigma^\infty = \Sigma^* \cup \Sigma^\omega$ is the set of all finite and infinite words. The set of words of positive length over Σ is noted $\Sigma^+ = \Sigma^* \backslash \varepsilon$.

A *morphism* is a function $\phi : \Sigma_1^* \longrightarrow \Sigma_2^*$ such that $\phi(uv) = \phi(u)\phi(v)$, and is determined by the image of the letters. For later use we denote $\delta : \Sigma^* \longrightarrow \Sigma^*$ the automorphism defined by $\delta(\alpha) = \alpha\alpha$ for each $\alpha \in \Sigma$, which amounts to duplicate each letter of a word.

A *factor* of w is a contiguous subsequence of w. A factor of w occurring at the beginning of w is called a *prefix*, referred to as $\mathrm{Pref}(w)$, and one that is placed at the end is a *suffix* of w. Denote by $\mathcal{L}(w)$ the language of w, i.e. the set of all the factors of w and denote by $\mathcal{L}_n(w)$ the factors of length n in w. The cardinality of this set is denoted by the factor complexity $\mathcal{C}_w(n)$. Two words u and v are said to be *conjugate* if there exist words x and y such that $u = xy$ and $v = yx$.

A *period* of a word w is an integer $m < |w|$ such that $w_i = w_{i+m}$ for any $i < |w| - m$. A factor u on length m of a periodic word w is said to be *primitive* if it is not the power of another word. Let us denote by $|w|_u$ the number of occurrences of u in w. The word w is said to be *recurrent* if $|w|_u$ is infinite and *uniformly recurrent* if the distance between any two consecutive occurrences of u is bounded. For example, periodic words are uniformly recurrent.

The *reversal* of a finite word w is $\widetilde{w} = w_{|w|-1}\dots w_1w_0$. A *palindrome* is a finite word that satisfies $w = \widetilde{w}$. The reversal is an antimorphism, that is, $\widetilde{u \cdot v} = \widetilde{v} \cdot \widetilde{u}$, which commutes with morphisms.

Lemma 2.1. For any morphism $\phi : \Sigma^* \longrightarrow \Sigma^*$, we have $\phi \circ \widetilde{\cdot} = \widetilde{\cdot} \circ \phi$.

Proof:
Let $w \in \Sigma^*$. We must show that $\phi(\widetilde{w}) = \widetilde{\phi(w)}$. We proceed by induction on the length of w. It is clearly true for any letter $\alpha \in \Sigma$. Assume that it is true for all words u such that $|u| < n$. Let $w = u \cdot \alpha$. Then we have $\phi(\widetilde{w}) = \phi(\widetilde{u\alpha}) = \phi(\widetilde{\alpha} \cdot \widetilde{u}) = \phi(\widetilde{\alpha}) \cdot \phi(\widetilde{u}) = \widetilde{\phi(\alpha)} \cdot \widetilde{\phi(u)} = \widetilde{\phi(u)\phi(\alpha)} = \widetilde{\phi(u\alpha)} = \widetilde{\phi(w)}$. □

A word that is the product of two palindromes is said to be *symmetric*[15]. The set of all palindromes of a word u is denoted by $\mathrm{Pal}(u)$ and the function $\mathcal{P}_u(n) = |\,\mathrm{Pal}(u) \cap \mathcal{L}_n(u)|$ is called its *palindromic complexity*.

There is a natural generalization of palindromes. Given an involution σ on Σ, i.e. a permutation of the letters such that $\sigma^2 = \mathrm{Id}$, define $\widetilde{\sigma} = \widetilde{\cdot} \circ \sigma$ which according to Lemma2.1 satisfies

$$\widetilde{\sigma}(w) = \widetilde{\sigma(w)} = \sigma(\widetilde{w}). \tag{1}$$

We write $\widehat{w}$ for $\widetilde{\sigma}(w)$. A *σ-palindrome* is a word w satisfying $\widehat{w} = w$, so that usual palindromes are Id-palindromes. Then $\mathrm{Pal}_\sigma(u)$ is the language of σ-palindrome factors of u, and its *σ-palindromic complexity* is $\mathcal{P}_u^{(\sigma)}(n) = |\,\mathrm{Pal}_\sigma(u) \cap \mathcal{L}_n(u)|$, that is the number of n-length factors of u that are σ-palindromes.

For the rest of the paper, σ is an involution on some finite alphabet Σ.

3. Computation of the σ-palindromic Factors

In order to compute the palindromic language of a finite word w, it is sufficient to compute for each prefix p of w its longest palindromic suffix $\mathrm{LPS}(p)$ which is uni-occurrent, and hence, the cardinality of the palindromic language $\mathrm{Pal}(w)$ is bounded by $|w| + 1$ (see [20]).

For σ-palindromes, the situation is similar. Indeed, consider a nonempty suffix p of w. It suffices to show that there is at most one longest σ-palindromic suffix of p: indeed, assume by contradiction that there exist two σ-palindromic suffixes u and v such that $|u| < |v|$; then $v = xu$ where $x \neq \varepsilon$, and $v = \widehat{v} = \widehat{xu} = u\widehat{x}$, so that u has two occurrences, contradiction. It follows that $\mathrm{Pal}_\sigma(w)$ satisfies the same bound as $\mathrm{Pal}(w)$. However, we can give a more precise bound. Observe that if w_i is the first letter not being fixed by σ, then $\mathrm{L}_\sigma\mathrm{PS}(w[0..i]) = \varepsilon$ which means that there is no nonempty σ-palindromic suffix at position i. Repeating the argument for the subsequent letters which are not fixed by σ one obtains the following more precise bound.

Proposition 1. Let t be the number of transpositions of σ. For any finite word w, let $k \leq t$ be the number of transpositions of σ such that at least one of the letters appears in w. Then we have

(i) $|\,\mathrm{Pal}_\sigma(w)| \leq |w| + 1$ if w does not contain transposed letters;

(ii) $|\,\mathrm{Pal}_\sigma(w)| \leq |w| + 1 - k$.

Example. The unique nontrivial involution on $\{a, b\}^*$ swaps the letters and is identified by E. It has only one transposition so that for any word $w \in \{a, b\}^*$, $|\,\mathrm{Pal}_E(w)| \leq |w|$. Observe that at position 0 we have $\mathrm{L}_\sigma\mathrm{PS}(w_0) = \varepsilon$. Note also that $\mathrm{Pal}_E(\alpha^k) = \{\varepsilon\}$ for any letter $\alpha \in \Sigma$, and any $k \in \mathbb{N}$.

The longest σ-palindromic suffix of a word w is computed by the following algorithm.

```
Input: Function σ, Word w;
Result: LσPS(w);
Initialization : j := 0,Word v := ε, i := |w|;
while v = ε and j < i do
    if w[j : i]=σ̃(w[j : i]) then
        v := w[j : i];
    else
        j := j + 1;
    end
end
return v.
```

Algorithm 1: Longest σ-Palindromic Suffix

Examples. Taking $Pinzani$ as an example on the alphabet $\Sigma = \{a, i, n, P, z\}$, let σ be defined by the permutation $(5, 2, 3, 4, 1)$ which swaps the letters a and z and leaves the other fixed. Then, $Pinzani$ has the following sequence of longest σ-palindromic suffixes

$$(\varepsilon, P, i, n, \varepsilon, za, nzan, inzani)$$

and therefore $|\operatorname{Pal}_\sigma(Pinzani)| = 7$, realizing the maximal bound according to Proposition 1 (ii). Such words are called *saturated*. By convention the sequence is initialized with ε since the empty word is a factor of every word.

For the word $w = zanziza$, we have the sequence of longest σ-palindromes

$$(\varepsilon, \varepsilon, za, n, anz, i, \varepsilon, za),$$

so that $|\operatorname{Pal}_\sigma(zanziza)| = 5 < 7 + 1 - 1 = 7$, and hence $zanziza$ does not realize the maximal σ-palindromicity.

3.1. Infinite words

For infinite words $\boldsymbol{w} \in \Sigma^\omega$, the computation is not always possible since $\operatorname{Pal}_\sigma(\boldsymbol{w})$ could be either finite or infinite. As one might expect, some properties strongly depend on σ. Sturmian words illustrate this fact perfectly. Recall that a word $\boldsymbol{s}$ is Sturmian if its factor complexity satisfies $\mathcal{C}_{\boldsymbol{s}}(n) = n + 1$ for any $n \in \mathbb{N}$. It is well known that they are saturated with palindromes, since every prefix of a Sturmian word realizes the upper bound given by Theorem 1 (i) (see [20]). This does not hold anymore for the unique non trivial involution $E : a \leftrightarrow b$.

Theorem 2. If $\mathbf{s}$ is Sturmian word then $|\operatorname{Pal}_E(\mathbf{s})| < \infty$.

Proof:
Every Sturmian word $\boldsymbol{s}$ contains either $aa = \widetilde{E}(bb)$ or $bb = \widetilde{E}(aa)$ but not both, and neither is an E-palindrome. Since Sturmian words are uniformly recurrent, the distance between two occurrences of aa (or bb) is bounded, so that the number of E-palindromic factors is necessarily finite. □

Observe also that E-palindromes of Sturmian words are necessarily of the form $(ab)^k$ or $(ba)^l$. The Fibonacci infinite word $\boldsymbol{F}$ is the most studied Sturmian word. Obtained as the fixed point of the morphism $\phi : a \mapsto ab; b \mapsto a$, its first letters are

$$\boldsymbol{F} = abaababaabaababaababaabaababaabaababaababaabaababaababa \ldots.$$

The reader can easily check that $\operatorname{Pal}_E(\boldsymbol{F}) = \{\varepsilon, ab, ba, abab, baba\}$.

Fixed points of morphisms. The Thue-Morse word $\boldsymbol{T}$ provides an example of a non Sturmian word, as exhibited by Morse and Hedlund [31]. Recall that $\boldsymbol{T}$ is defined as the fixed point starting with a of the morphism $\mu : a \mapsto ab; b \mapsto ba$:

$$\boldsymbol{T} = abbabaabbaababbabaababbaabbabaab \cdots.$$

Its factor complexity was established in [13, 29] and its σ-palindromic complexity in [9]. It has been shown in [1] that besides $\boldsymbol{T}$ there is another infinite word sharing the same factor complexity, namely $\delta(\boldsymbol{T})$ where $\delta : \Sigma^* \longrightarrow \Sigma^*$ is the injective morphism that duplicates the letters, that is, $\delta(\alpha) = \alpha\alpha$ for each $\alpha \in \Sigma$. The next lemma shows δ preserves σ-palindromicity.

Lemma 3.1. Let $w \in \Sigma^*$. Then, $w = \widehat{w}$ if and only if $\delta(w) = \widehat{\delta(w)}$.

Proof:
($\Rightarrow$) Let $w = \widehat{w}$. Using Eq. (1), Lemma 2.1 and the fact that σ and δ commute, we have

$$\delta(w) = \delta(\widehat{w}) = (\delta \circ \tilde{\cdot} \circ \sigma)(w) = (\tilde{\cdot} \circ \delta \circ \sigma)(w) = (\tilde{\cdot} \circ \sigma \circ \delta)(w) = \widehat{\delta(w)}.$$

The "only if" part is similar and left to the reader. □

It follows that for every involution σ and for all $w \in \Sigma^* \cup \Sigma^\infty$ we have

$$u \in \mathrm{Pal}_\sigma(w) \quad \text{if and only if} \quad \delta(u) \in \mathrm{Pal}_\sigma(\delta(w)).$$

In the case of the word $\boldsymbol{T}$, we have then a bijection between $\mathrm{Pal}_E(\boldsymbol{T})$ and $\mathrm{Pal}_E(\delta(\boldsymbol{T}))$.

On the other hand, we know from [9] that the E-palindromic complexity of $\boldsymbol{T}$ is

$$\mathcal{P}_{\boldsymbol{T}}^{(E)}(n) = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{if } n \text{ is odd}, \\ 2 & \text{if } n = 2, \\ 4 & \text{if } n \text{ is even and } 2 \cdot 4^k + 2 \le n \le 6 \cdot 4^k, \text{ for } k \ge 0, \\ 2 & \text{if } n \text{ is even and } 6 \cdot 4^k + 2 \le n \le 2 \cdot 4^{k+1}, \text{ for } k \ge 0. \end{cases} \tag{2}$$

Then it follows that

Proposition 3. $|\,\mathrm{Pal}_E(\mathbf{T})| = |\,\mathrm{Pal}_E(\delta(\mathbf{T}))| = \infty$.

Periodic words. Periodic infinite words are a special case of fixed points of uniform morphisms, for which the situation is easier to describe. We have the following result, which is a generalization of Theorem 4 of [15]. Words that are products of two σ-palindromes are said *σ-symmetric*.

Theorem 4. Let σ be an involution on Σ and $w \in \Sigma^*$ be a nonempty word. The following conditions are equivalent:

(i) w is the product of two σ-palindromes;

(ii) $\mathrm{Pal}_\sigma(w^\omega)$ is infinite.

Proof:
(i) $\Longrightarrow$ (ii) : assume that $w = uv$ where u and v are two σ-palindromes such that $uv \neq \varepsilon$. Then, for every integer $n \in \mathbb{N}$, the prefix $(uv)^n u$ of w^ω is a σ-palindrome, and the claim holds.

(ii) $\Longrightarrow$ (i) : since $\mathrm{Pal}_\sigma(w^\omega)$ is infinite, there exists arbitrarily large palindromes in w^ω. Let p in $\mathcal{L}(w^\omega)$ be such that $p > 2|w|$. Observe that $p = xw^k y$, where x and y are respectively a suffix and a prefix of w and $k \geq 1$. Since $p = \widehat{p} = \widehat{y}\widehat{w}^k\widehat{x}$, it follows that $\widehat{w}$ is a factor of ww, so that $ww = u\widehat{w}v$, where u and v are respectively a suffix and a prefix of ww with $|u| + |v| = |w|$. It follows that $w = uv$ and we have $ww = uvuv = u\widehat{w}v$, so that $\widehat{w} = vu$ and $w = \widehat{\widehat{u}\widehat{v}} = uv$. □

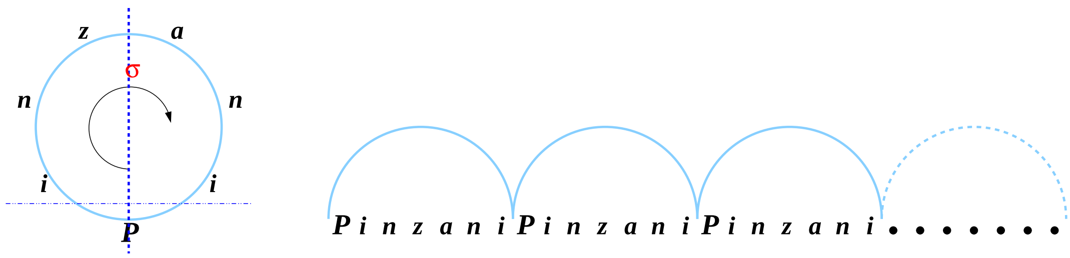

Figure 1. A σ-symmetric word and its periodic expansion.

Example. Since $Pinzani = P \cdot inzani$ which is a product of two σ-palindromes, the periodic word $(Pinzani)^{\omega}$ contains an infinite number of σ-palindromes. Observe also that such a word has a simple geometric representation, when the word is written on a circle as shown in Figure 1.

The vertical line indicates the transposition σ of the letters while the horizontal line indicates the product of the σ-palindromes P and $inzani$. Observe that moving the horizontal line upwards produces conjugates of $Pinzani$, that is

$$P \cdot inzani, \quad iPi \cdot nzan, \quad niPin \cdot za \quad \text{and} \quad aniPinz \cdot \varepsilon.$$

An immediate consequence of Theorem 4 follows (see [15] Theorem 4).

Corollary 3.2. Let σ be an involution on Σ and $w \in \Sigma^*$ be a nonempty word. Then w is the product of two σ-palindromes if and only if every conjugate of w is the product of two σ-palindromes.

This corollary suggests that in order to decide whether w is σ-symmetric or not, it suffices to compute the longest palindromic prefix p and check if the remaining suffix s is a palindrome. This can be achieved in linear time by using for instance an algorithm based on suffix trees [22].

4. σ-palindromic Lacunas

Recall that a *palindromic lacuna* (lacuna for short) is a position i in w where $\mathrm{LPS}(w[0..i])$ is not uni-occurrent [10]. A word w realizing the maximal palindromic complexity is a word without lacunas, and the statistic $\mathcal{D}(w) = |w| + 1 - |\operatorname{Pal}(w)|$ counting the number of its lacunas is called *palindromic defect* (see [15]). Words realizing the maximal palindromic complexity have clearly no lacunas. They were introduced and called *full* in [15]. Later they appeared as *rich* (see [21]) and also *perfect* (see [14]).

When σ is not the identity permutation, Proposition 1 shows that there are necessarily positions where there are no new σ-palindromic suffix: it is a position i where either $\mathrm{L}_\sigma\mathrm{PS}(w[0..i]) = \varepsilon$ or $\mathrm{L}_\sigma\mathrm{PS}(w[0..i]) \neq \varepsilon$ is not uni-occurrent. Both cases are handled in line 5 of Algorithm 2. Call such a position a *σ-lacuna*. Hence, no word appears to be perfect in this case, even though it is *saturated* with σ-palindromes.

Definition 1. Let $\sigma : \Sigma \longrightarrow \Sigma$ be an involution, and $w \in \Sigma^*$. The *σ-defect* of w is defined by

$$\mathcal{D}_\sigma(w) = |w| + 1 - |\operatorname{Pal}_\sigma(w)|. \tag{3}$$

A good way to compute the defect is to count the number of its σ-lacunas with the following algorithm.

Input: Function σ, Word w;
Result: $\mathcal{D}_\sigma(w)$;
1 **Initialization :** $\mathcal{D} := 0$;
2 **if** $|w| \neq 0$ **then**
3   **for** $i = 0$ **to** $|w| - 1$ **do**
4         $s := \mathrm{L}_\sigma\mathrm{PS}(w[0:i])$;
5   **if** *s is not uni-occurrent in $w[0:i]$* **then**
6             $\mathcal{D} := \mathcal{D} + 1$; /* σ-lacuna at position i */
7   **end**
8   **end**
9 **end**
10 **return** $\mathcal{D}$.

Algorithm 2: σ-Defect

Remark 4.1. Line 4 of the algorithm can be computed in constant time by using means of a linear preprocessing [22]. Line 5 of the algorithm is crucial for ensuring linearity of the algorithm. It amounts to look at occurrences of the factor s in the prefix ending at position i. This is achieved by the classical Boyer-Moore algorithm. Moreover, Line 6 of the algorithm above says that $\mathcal{D}_\sigma$ is an increasing counter, that is, if $w = u\alpha$ where α is a letter then $\mathcal{D}_\sigma(w) - \mathcal{D}_\sigma(u) \leq 1$. Finally, observe that the set of σ-lacunas can be obtained easily with an additional variable.

The following properties of the σ-defect are deduced from the definition.

Lemma 4.2. Let $u, w \in \Sigma^*$ be such that $u \in \mathcal{L}(w)$, and let $\alpha \in \Sigma$. Then we have

(i) $\mathcal{D}_\sigma(w) = \mathcal{D}_\sigma(\widetilde{w})$;

(ii) $\mathcal{D}_\sigma(u) \leq \mathcal{D}_\sigma(u\alpha)$, and $\mathcal{D}_\sigma(u) \leq \mathcal{D}_\sigma(\alpha u)$;

(iii) $\mathcal{D}_\sigma(u) \leq \mathcal{D}_\sigma(w)$;

(iv) if w is saturated then u is saturated.

Proof:
(i) Obviously, $\mathrm{Pal}_\sigma(w) = \mathrm{Pal}_\sigma(\widetilde{w})$. (ii) We have $\mathrm{Pal}_\sigma(u) \subseteq \mathrm{Pal}_\sigma(u\alpha)$, so that $|\mathrm{Pal}_\sigma(u\alpha)| - |\mathrm{Pal}_\sigma(u)| \leq 1 = |u\alpha| - |u|$. Then $|u| - |\mathrm{Pal}_\sigma(u)| \leq |u\alpha| - |\mathrm{Pal}_\sigma(u\alpha)|$, so that $\mathcal{D}_\sigma(u) \leq \mathcal{D}_\sigma(u\alpha)$. For the second part, it suffices to use condition (i). (iii) By induction on the length of u and (ii). (iv) Obvious. □

4.1. The case of infinite words

Although this algorithm only works for finite words, it can be used for infinite words having finite defect or finite σ-palindromic language. First, following [15], observe that the σ-defect of an infinite word is simply defined as the maximal defect of its factors. As a consequence, when the number of σ-lacunas of an infinite word ***w*** is finite, it suffices to compute it for the prefixes of ***w***.

On the other hand, if we know that the σ-palindromic language of an infinite word ***w*** is finite, then necessarily its defect is infinite. This is the case for Sturmian words by Theorem 2.

Theorem 5. Every Sturmian word **s** satisfies $\mathcal{D}_E(\mathbf{s}) = \infty$.

When the σ-palindromic language of a word is infinite, we cannot conclude in general, as shown by the Thue Morse word ***T***. We know that ***T*** has an infinite E-palindromic language (Proposition 3), and that it also has an infinite E-defect. Indeed, let $L(n)$ (resp. $L_E(n)$) be the index where the nth interval of Id-lacunas (resp. E-lacunas) starts and $\ell(n)$ (resp. $\ell_E(n)$) be its length. To emphasize the distinct behaviour of Sturmian sequences we recall from [9](Theorem 2) the following result about ***T***.

Theorem 6. The sequences L, L_E, ℓ and ℓ_E satisfy the following equations :

(i) $L_E(1) = 0$, $L_E(2) = 2$, $L_E(3) = 4$ and $L_E(4) = 12$,

(ii) $\ell_E(n) = 1$ for $n = 1, 2, 3, 4$,

(iii) $L(n) = 2L_E(n+2)$, for $n \geq 1$,

(iv) $\ell(n) = 2\ell_E(n+2)$, for $n \geq 1$,

(v) $L_E(n) = 2L(n-4)$, for $n \geq 5$, and

(vi) $\ell_E(n) = 2\ell(n-4)$, for $n \geq 5$.

Closed formulas for L, L_E, ℓ and ℓ_E are easily obtained :

$$L(n) = \begin{cases} 2^{n+2}, & \text{if } n \text{ is odd,} \\ 2^{n+2} + 2^{n+1}, & \text{if } n \text{ is even.} \end{cases} \quad , \quad L_E(n) = \begin{cases} 2^{n-1}, & \text{if } n \text{ is odd,} \\ 2^{n-1} + 2^{n-2}, & \text{if } n \text{ is even.} \end{cases}$$

$$\ell(n) = \begin{cases} 2^{n}, & \text{if } n \text{ is odd,} \\ 2^{n-1}, & \text{if } n \text{ is even.} \end{cases} \quad , \quad \ell_E(n) = \begin{cases} 2^{n-3}, & \text{if } n \text{ is odd,} \\ 2^{n-4}, & \text{if } n \text{ is even.} \end{cases}$$

Moreover, the first intervals where E-lacunas occur are

$$[0], [2], [4], [12], [16..19], [48..51], [64..79], [192..207], \ldots$$

and those where Id-lacunas occur are

$$[8..9], [24..25], [32..39], [96..103], [128..159], [384..415], \ldots$$

The closed formulas above show also that the lacunas do not intersect [9].

A direct consequence of these computations is

Theorem 7. $\mathcal{D}_E(\boldsymbol{T}) = \mathcal{D}_E(\delta(\boldsymbol{T})) = \infty$.

Periodic words. We consider now the computation of the σ-defect of periodic words. In a previous paper, Brlek *et al.* [15] established that the Id-defect of periodic words for which the smallest period is symmetric is bounded. We could expect to extend this theorem to any involution σ and indeed we have:

Theorem 8. Let $w = uv$, with $u,v \in \mathrm{Pal}_\sigma(\Sigma^*)$, be a primitive σ-symmetric word, and $\boldsymbol{w} = w^\omega$. Then $\mathcal{D}_\sigma(\boldsymbol{w}) = \mathcal{D}_\sigma(x)$ where x is a prefix of $\boldsymbol{w}$ of length $|uv| + \lfloor|\frac{|u|-|v|}{3}|\rfloor$.

To prove the theorem, we need the following lemma taken from Proposition 1.3.4 in [27] :

Lemma 4.3. Assume that there exist $x, z \in \Sigma^+$, $y \in \Sigma^*$ such that $xy = yz$. Then there exist $u, v \in \Sigma^*$ and an integer k such that

$$x = uv,\ z = vu,\ y = u(vu)^k.$$

Proof:
[Theorem 8] We need to show that for any prefix p of $\boldsymbol{w}$ of length greater than $|uv| + \lfloor|\frac{|u|-|v|}{3}|\rfloor$, the longest σ-palindromic suffix of p occurs only once in p. Recall that u and v being σ-palindromes, they are of the form

$$u = x\alpha\widehat{x}, v = y\beta\widehat{y}$$

with $x, y, \alpha, \beta \in \Sigma^*$, α and β being either the empty word or a letter that is fixed by the involution σ. We thus divide the proof into three cases according to the length of p :

- $|p| > |uvuv|$. Since $|p| > 2|w|$, there exists $m \in \mathbb{N}$ such that:

$$p = \begin{cases} x\alpha\widehat{x}y\beta\widehat{y}(x\alpha\widehat{x}y\beta\widehat{y})^m z, & \text{with } z \in \mathrm{Pref}(x), \quad (4)\\ x(\alpha\widehat{x}y\beta\widehat{y}x)^m\alpha, & \quad (5)\\ x(\alpha\widehat{x}y\beta\widehat{y}x)^m\alpha z, & \text{with } z \in \mathrm{Pref}(\widehat{x}), \quad (6)\\ x\alpha\widehat{x}y\beta\widehat{y}x(\alpha\widehat{x}y\beta\widehat{y}x)^m\alpha\widehat{x}z, & \text{with } z \in \mathrm{Pref}(y), \quad (7)\\ x\alpha\widehat{x}y(\beta\widehat{y}x\alpha\widehat{x}y)^m\beta, & \quad (8)\\ x\alpha\widehat{x}y(\beta\widehat{y}x\alpha\widehat{x}y)^m\beta z, & \text{with } z \in \mathrm{Pref}(\widehat{y}). \quad (9) \end{cases}$$

The longest σ-palindromic suffix of p is, respectively :

$$\mathrm{L}_\sigma\mathrm{PS}(p) = \begin{cases} \widehat{z}y\beta\widehat{y}(x\alpha\widehat{x}y\beta\widehat{y})^m z, & (10)\\ (\alpha\widehat{x}y\beta\widehat{y}x)^m\alpha, & (11)\\ \widehat{z}(\alpha\widehat{x}y\beta\widehat{y}x)^m\alpha z, & (12)\\ \widehat{z}x(\alpha\widehat{x}y\beta\widehat{y}x)^m\alpha\widehat{x}z\ \ (\beta\widehat{y}x\alpha\widehat{x}y)^m\beta, & (13)\\ \widehat{z}(\beta\widehat{y}x\alpha\widehat{x}y)^m\beta z. & (14) \end{cases}$$

Thus, for all $|p| > 2|w|$, $p = s\,\mathrm{L}_\sigma\mathrm{PS}(p)$ and $|s| < |\,\mathrm{L}_\sigma\mathrm{PS}(p)|$. This implies that the longest σ-palindromic suffix is uni-occurrent. Otherwise, by Lemma 4.3, two occurrences of $\mathrm{L}_\sigma\mathrm{PS}(p)$ overlap and contradict the choice of $\mathrm{L}_\sigma\mathrm{PS}(p)$.

- $|uvu| < |p| \le |uvuv|$. In this case p has one of the following forms :

$$p = \begin{cases} x\alpha\widehat{x}y\beta\widehat{y}x\alpha\widehat{x}z, & \text{with } z \in \mathrm{Pref}(y), \quad (15)\\ x\alpha\widehat{x}y\beta\widehat{y}x\alpha\widehat{x}y\beta, & \quad (16)\\ x\alpha\widehat{x}y\beta\widehat{y}x\alpha\widehat{x}y\beta z, & \text{with } z \in \mathrm{Pref}(\widehat{y}). \quad (17) \end{cases}$$

The longest σ-palindromic prefix of p is, respectively,

$$\mathrm{L}_\sigma\mathrm{PS}(p) = \begin{cases} \widehat{z}x\alpha\widehat{x}z, & (18)\\ \beta\widehat{y}x\alpha\widehat{x}y\beta, & (19)\\ \widehat{z}\beta\widehat{y}x\alpha\widehat{x}y\beta z. & (20)\end{cases}$$

We then write $p = s\,\mathrm{L}_\sigma\mathrm{PS}(p)$. If the situation is as (19) or (20), then $|\,\mathrm{L}_\sigma\mathrm{PS}(p)| > |s|$ and it follows from the first part of proof that $\mathrm{L}_\sigma\mathrm{PS}(p)$ occurs only once. Suppose now that $\mathrm{L}_\sigma\mathrm{PS}(p)$ occurs at least twice in p. Then, u overlaps itself. By Lemma 4.3, there exist two σ-palindromes a and b and an integer k such that $u = (ab)^k a$. Thus, baz is a prefix of v and $\widehat{baz} = \widehat{z}ab$ is a suffix of v. It is then clear that $\widehat{z}ab(ab)^k az$ is a σ-palindromic suffix of p and thus contradicts the choice of $\mathrm{L}_\sigma\mathrm{PS}(p)$.

- $|uv| + \lfloor|\frac{|u|-|v|}{3}|\rfloor < |p| \leq |uvu|$. In this case $p = uvs$ and s is prefix of u of length greater than $\lfloor|\frac{|u|-|v|}{3}|\rfloor$. Let $r = \mathrm{L}_\sigma\mathrm{PS}(p)$. Then,

$$\begin{aligned}|r| &\geq |s| + |v| + |\widehat{s}|\\ &= 2|s| + |v|.\end{aligned}$$

Assume that there is another occurrence of r. By Lemma 4.3 and because r is the longest σ-palindromic suffix, the two occurrences of r do not overlap:

$$\begin{aligned}|r| &\leq |uvs| - |v| - 2|s|\\ &= |u| - |s|.\end{aligned}$$

Thus we have the following situation :

$$\begin{aligned}|u| - |s| &\geq 2|s| + |v|\\ \Rightarrow 3|s| &\leq |\,|u| - |v|\,|\\ \Rightarrow |s| &\leq \lfloor|\frac{|u|-|v|}{3}|\rfloor,\end{aligned}$$

contradicting the hypothesis on the length of s. □

For instance, $\mathcal{D}_\sigma((Pinzani)^\omega) = \mathcal{D}_\sigma(P \cdot inzani \cdot P) = 1$. Note that this result provides another algorithm for deciding whether a word is σ-symmetric or not. Moreover, if we are lucky enough, that is when $|u|$ and $|v|$ are close, we can do better (see [15]).

Corollary 9. ([15])
Let $w = uv$ be a primitive word such that $u, v \in \mathrm{Pal}_\sigma(\Sigma^*)$. Then the following properties hold:

(i) if $|v| \leq |u| \leq |v| + 2$. Then $\mathcal{D}_\sigma((uv)^\omega) = \mathcal{D}_\sigma(uv)$;

(ii) for some conjugate w' of w we have $\mathcal{D}_\sigma(w^\omega) = \mathcal{D}_\sigma(w')$.

Proof:
(i) Obvious. (ii) According to Figure 1 one can choose a conjugate $w' = u'v'$ with $|v'| \leq |u'| \leq |v'| + 2$. Then Theorem 8 applies and we have $\mathcal{D}_\sigma(w^\omega) = \mathcal{D}_\sigma((w')^\omega) = \mathcal{D}_\sigma(w')$. □

5. Words with a Fixed σ-palindromic Language

Let $P \subset \mathrm{Pal}_\sigma(\Sigma^*)$ be a fixed and finite set of σ-palindromes. Since each σ-palindrome p contains its own σ-palindromic factors, we assume that P is *factorially closed* with respect to σ-palindromes, that is, for each $p \in P$,

$$q \in \mathcal{L}(p) \quad \text{and} \quad q \in \mathrm{Pal}_\sigma(\Sigma^*) \implies q \in P.$$

We consider in the first place the problem of constructing words w whose σ-palindromic language is included in P. Define the set Q to be the set of minimal elements of $\mathrm{Pal}_\sigma(\Sigma^*) \setminus P$, where the minimality is taken with respect to the *factorial* partial order: $u \leq v$ iff u is a factor of v.

Theorem 10. The maximal language whose σ-palindromes are contained in P is rational and is given by $X_P = \Sigma^* \setminus \Sigma^* Q \Sigma^*$.

The proof is the same as in [15] and is omitted.

Of course, the language X_P may be finite in the case where $\sigma = \mathrm{Id}$. This no longer true for $\sigma \neq \mathrm{Id}$. On the two-letter alphabet $\Sigma = \{a, b\}$ the unique involution without fixed point is the exchange of letters, so that X_P necessarily contains a^k and b^k for arbitrary k since these words have ε for unique σ-palindrome. This construction can be carried out for any involution σ without fixed points as well: assume for instance that $\Sigma = \{a, b, c, d\}$ then $\alpha^* \subseteq X_P$ for each letter $\alpha \in \{a, b, c, d\}$.

Example. Consider the set $P = \{bbaa, ba, \varepsilon\}$. Then $Q = \{bbbaaa, ab\}$ and the solution is therefore $X_P = (\varepsilon \cup b \cup bb) \cdot a^* \cup b^* \cdot (\varepsilon \cup a \cup aa)$. Observe that the language $L \subseteq X_P$ such that $\mathrm{Pal}_\sigma(\boldsymbol{w}) = P$ is infinite and, moreover, the infinite word $\boldsymbol{w} = bba^\omega$ has exactly P for σ-palindromic language.

Since the langage X_P in the theorem is rational, it is recognizable by a finite trim automaton $\mathbf{A}$, which necessarily contains circuits since X_P is infinite. An immediate consequence of the Pumping Lemma is the existence of infinite words whose language of σ-palindromes is included in P. The proof of the next proposition is left to the reader.

Proposition 11. Let P be a finite set of σ-palindromes factorially closed. The following reconstruction problems are decidable:

(i) there exists an infinite word $\boldsymbol{w}$ such that $\mathrm{Pal}_\sigma(\boldsymbol{w}) \subseteq P$;

(ii) there exists an infinite periodic word $\boldsymbol{w}$ such that $\mathrm{Pal}_\sigma(\boldsymbol{w}) \subseteq P$;

(iii) there exists an infinite word $\boldsymbol{w}$ such that $\mathrm{Pal}_\sigma(\boldsymbol{w}) = P$.

Another immediate consequence is

Corollary 12. There exists an infinite periodic word $\boldsymbol{w}$ such that $\mathrm{Pal}_\sigma(\boldsymbol{w}) = P$ if and only if there exists an infinite recurrent word $\boldsymbol{u}$ such that $\mathrm{Pal}_\sigma(\boldsymbol{u}) = P$.

6. Concluding Remarks

The saturation property strongly depends on the involution σ as shown by $w = Pinzani$ in the examples above. Indeed, with $\sigma = \text{Id}$, the defect is $\mathcal{D}(w) = 2$ and therefore w is not saturated. Moreover, in that case $Pinzani$ is not a product of palindromes and therefore does not produce an infinite periodic word w^ω having infinitely many palindromic factors. On the other hand, with the involution σ swapping a and z, it is saturated and also a product of σ-palindromes so that the infinite periodic word w^ω has infinitely many palindromes. This raises the problem of finding for a given word w the involutions satisfying some properties like saturation and/or symmetry, and in turn the corresponding reconstruction problems.

Another investigation concerns the reconstruction of words from a fixed σ-palindromic length sequence in the spirit of [8].

The dissertation of the second author will address some of the problems mentioned above and will be completed and available online in 2015.

Acknowledgements We are grateful to the anonymous reviewers for their careful reading and suggestions that improved the final presentation.

References

[1] Aberkane, A., Brlek, S.: Suites de même complexité que celle de Thue-Morse, *Actes des Journées Montoises d'Informatique Théorique*, 2002.

[2] Allouche, J.-P.: Schrödinger operators with Rudin-Shapiro potentials are not palindromic, *Journal of Mathematical Physics*, **38**, April 1997, 1843–1848.

[3] Allouche, J.-P., Baake, M., Cassaigne, J., Damanik, D.: Palindrome complexity, *Theoretical Computer Science*, **292**(1), 2003, 9–31.

[4] Allouche, J.-P., Shallit, J.: Sums of Digits, Overlaps, and Palindromes, *Discrete Mathematics & Theoretical Computer Science*, **4**(1), 2000, 1–10.

[5] Anne, V., Zamboni, L., Zorca, I.: Palindromes and pseudo-palindromes in episturmian and pseudo-palindromic infinite words, *Proc. Words, 5th International Conference On Words, September 13 - 17, 2005, Montréal (QC) Canada* (S. Brlek, C. Reutenauer, Eds.), 36, Publications du LaCIM, 2005.

[6] Baake, M.: A Note on Palindromicity, *Letters in Mathematical Physics*, **49**(3), 1999, 217–227.

[7] Beauquier, D., Nivat, M.: On Translating one Polyomino to Tile the Plane, *Discrete Computational Geometry*, **6**, 1991, 575–592.

[8] Blondin Massé, A., Brlek, S., Frosini, A., Labbé, S., Rinaldi, S.: Reconstructing words from a fixed palindromic length sequence, *5th IFIP International Conference On Theoretical Computer Science - TCS 2008, IFIP 20th World Computer Congress, TC 1, Foundations of Computer Science, September 7-10, 2008, Milano, Italy* (G. Ausiello, J. Karhumäki, G. Mauri, C.-H. L. Ong, Eds.), 273, Springer, 2008.

[9] Blondin Massé, A., Brlek, S., Garon, A., Labbé, S.: Combinatorial properties of f-palindromes in the Thue-Morse sequence, *Pure Mathematics and Applications*, **19**(2-3), 2008, 39–52.

[10] Blondin Massé, A., Brlek, S., Labbé, S.: Palindromic lacunas of the Thue-Morse word, *Proc. GASCom 2008 (16-20 June 2008, Bibbiena, Arezzo-Italia)*, 2008.

[11] Blondin Massé, A., Brlek, S., Labbé, S., Mendès France, M.: Complexity of the Fibonacci snowflake, *Fractals*, **20**(03n04), 2012, 57–60.

[12] Blondin Massé, A., Garon, A., Labbé, S.: Combinatorial properties of double square tiles, *Theoretical Computer Science*, **502**, 2013, 98–117.

[13] Brlek, S.: Enumeration of factors in the Thue-Morse word, *Discrete Applied Mathematics*, **24**(1-3), 1989, 83–96.

[14] Brlek, S.: TBA, *Conf. in honour of C.Reutenauer on the occasion of his 60th birthday (1-5 July 2013, Cetraro, Italia)*, 2013.

[15] Brlek, S., Hamel, S., Nivat, M., Reutenauer, C.: On The Palindromic Complexity Of Infinite Words, *International Journal on Foundation of Computer Science*, **15**(2), 2004, 293–306.

[16] Brlek, S., Jamet, D., Paquin, G.: Smooth words on 2-letter alphabets having same parity, *Theoretical Computer Science*, **393**(1-3), 2008, 166–181.

[17] Brlek, S., Ladouceur, A.: A note on differentiable palindromes, *Theoretical Computer Science*, **302**(1-3), 2003, 167–178.

[18] Brlek, S., Provençal, X.: An Optimal Algorithm for Detecting Pseudo-squares, *Discrete Geometry for Computer Imagery, 13th International Conference, DGCI 2006, Szeged, Hungary, October 25-27, 2006, Proceedings* (A. Kuba, L. G. Nyúl, K. Palágyi, Eds.), 4245, Springer, 2006.

[19] de Luca, A., De Luca, A.: Pseudopalindrome closure operators in free monoids, *Theoretical Computer Science*, **362**(1-3), 2006, 282–300.

[20] Droubay, X., Justin, J., Pirillo, G.: Episturmian words and some constructions of de Luca and Rauzy, *Theoretical Computer Science*, **255**(1-2), 2001, 539–553.

[21] Glen, A., Justin, J., Widmer, S., Zamboni, L. Q.: Palindromic richness, *European Journal of Combinatorics*, **30**(2), 2009, 510–531.

[22] Gusfield, D.: Finding all maximal palindromes in linear time, in: *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*, Cambridge University Press, 1997, 197–199.

[23] Hof, A., Knill, O., Simon, B.: Singular continuous spectrum for palindromic Schrödinger operators, *Communications in Mathematical Physics*, **174**(1), 1995, 149–159.

[24] Hussini, S., Kari, L., Konstantinidis, S.: Coding properties of DNA languages, *Theoretical Computer Science*, **290**(3), 2003, 1557–1579.

[25] Kari, L., Kitto, R., Thierrin, G.: Codes, Involutions, and DNA Encodings, *Formal and Natural Computing* (W. Brauer, H. Ehrig, J. Karhumäki, A. Salomaa, Eds.), LNCS 2300, Springer, 2002.

[26] Kari, L., Mahalingam, K.: Watson-Crick palindromes in DNA computing, *Natural Computing*, **9**(2), 2010, 297–316.

[27] Lothaire, M.: *Combinatorics on Words*, Cambridge University Press, Cambridge, 1997.

[28] de Luca, A.: Sturmian Words: Structure, Combinatorics, and Their Arithmetics, *Theoretical Computer Science*, **183**(1), 1997, 45–82.

[29] de Luca, A., Varricchio, S.: Some Combinatorial Properties of the Thue-Morse Sequence and a Problem in Semigroups, *Theoretical Computer Science*, **63**(3), 1989, 333–348.

[30] Massé, A. B., Brlek, S., Garon, A., Labbé, S.: Equations on palindromes and circular words, *Theor. Comput. Sci.*, **412**(27), 2011, 2922–2930.

[31] Morse, M., Hedlund, G. A.: Symbolic Dynamics, *American Journal of Mathematics*, **60**, 1938, 815–866.