Journal of Integer Sequences, Vol. 16 (2013),
Article 13.2.11

On Generalized Pseudostandard Words Over Binary Alphabets

Alexandre Blondin Massé
Laboratoire d'informatique formelle
Université du Québec à Chicoutimi
Chicoutimi, QC, G7H 2B1
Canada
ablondin@uqac.ca

Geneviève Paquin
Département de mathématiques
Cégep de Saint-Jérôme
Saint-Jérôme, QC J7Z 4V2
Canada
gpaquin@cstj.qc.ca

Hugo Tremblay
Laboratoire de combinatoire et d'informatique mathématique
Université du Québec à Montréal
Montréal, QC H3C 3P8
Canada
tremblay.hugo.5@courrier.uqam.ca

Laurent Vuillon
Laboratoire de mathématiques
Université de Savoie
Le-Bourget-du-Lac 73376
France
laurent.vuillon@univ-savoie.fr

Abstract

In this paper, we study generalized pseudostandard words over a two-letter alphabet, which extend the classes of standard Sturmian, standard episturmian and pseudostandard words, allowing different involutory antimorphisms instead of the usual palindromic closure or a fixed involutory antimorphism. We first discuss about *pseudoperiods*, a useful tool for describing words obtained by iterated pseudopalindromic closure. Then, we introduce the concept of *normalized* directive bi-sequence (Θ, w) of a generalized pseudostandard word, that is the one that exactly describes all the pseudopalindromic prefixes of it. We show that a directive bi-sequence is normalized if and only if its set of factors does not intersect a finite set of forbidden ones. Moreover, we provide a construction to normalize any directive bi-sequence. Next, we present an explicit formula, generalizing the one for the standard episturmian words introduced by Justin, that computes recursively the next prefix of a generalized pseudostandard word in term of the previous one. Finally, we focus on generalized pseudostandard words having complexity $2n$, also called *Rote words*. More precisely, we prove that the normalized bi-sequences describing Rote words are completely characterized by their factors of length 2.

1 Introduction

The *Sturmian words* form a well-known class of infinite words over a two-letter alphabet that occurs in many different fields, for instance in astronomy, symbolic dynamics, number theory, discrete geometry, crystallography, and of course, in combinatorics on words (see [13, Chapter 2]). These words have many equivalent characterizations whose usefulness depends on the context. In discrete geometry, they are exactly the words that code the discrete approximations of lines with irrational slopes, using horizontal and diagonal moves. In symbolic dynamics, Sturmian words are obtained from two-intervals exchange transformations. They are also known as the balanced aperiodic infinite words over a two-letter alphabet. A remarkable subclass of the Sturmian words is the class of the so-called *standard Sturmian words*. To each Sturmian word corresponds a standard Sturmian one having the same language, i.e., the same set of factors. Thus, standard Sturmian words are, in a sense, representatives of all Sturmian words having the same language. Geometrically, they correspond to discrete lines starting at the origin. All the words in this subclass can be obtained by a construction called iterated palindromic closure [7]. This operation establishes a bijection between standard Sturmian words and non-eventually constant infinite words over a binary alphabet. It can also be generalized for an alphabet with more than two letters and yields the standard episturmian words.

Another generalization of the standard episturmian words was introduced by de Luca and De Luca in [8], where the authors considered pseudopalindromes instead of palindromes. In the paper, they first define the notion of ϑ-palindrome (called *pseudopalindrome* when ϑ is not mentioned), which is a word fixed by an involutory antimorphism ϑ. Moreover, they introduce the pseudopalindromic closure, which extends the usual palindromic closure to pseudopalindromes. These ideas lead one to naturally define words obtained by *iterated pseudopalindromic closure*. In particular, they consider a generalization of Sturmian and episturmian words: the *pseudostandard words*. Finally, toward the end of their paper, they define an even more general class of infinite words, called *generalized pseudostandard words*, these words being obtained by a directive bi-sequence (Θ, w), where Θ is a sequence of involutory antimorphisms and w is an infinite word. In that case, the type of pseudopalindromic closure changes at each step, applying the nth involutory antimorphism for the nth pseudopalindromic closure, after having added the nth letter of the word w. Different generalizations of standard episturmian words have been introduced and studied (see for instance [4, 5, 8]), but not much is known about generalized pseudostandard words except for the remarkable fact that the famous Thue-Morse word falls within this class of words (see [8]).

In order to study generalized pseudostandard words, it is natural to search for an efficient way to construct it from its directive bi-sequence. In [12], Justin gives a formula that allows one to compute in linear time a prefix of a standard Sturmian (resp., episturmian) word, using its directive sequence, that is the sequence on which the iterated palindromic closure is performed, in order to construct the standard Sturmian (resp., episturmian) word. The main aim of this paper is to extend Justin's formula to generalized pseudostandard words on binary alphabets. In particular, this formula might be useful in the study of generalized pseudostandard word by mean of computer exploration.

The next sections are organized as follows. As usual, we first introduce the definitions and notation used in the next sections. We recall iterated palindromic and pseudopalindromic closure operators as well as the main topic of this paper: the generalized pseudostandard words. Section 3 is devoted to the structure of words having pseudoperiods (i.e. words such that $w[i] = \sigma(w[i+p])$ for some permutation σ and some positive integer p). These results turn out to be very useful for studying generalized pseudostandard words. Next, we introduce in Section 4 the notion of normalized directive bi-sequence. Those normalized bi-sequences are representatives of all directive bi-sequences describing the same word, but having the special additional property that the successive prefixes obtained by pseudopalindromic closure coincide with all pseudopalindromic prefixes of the corresponding generalized pseudostandard word. We first prove the existence of such a normalized bi-sequence and we describe exactly the forbidden factors for a bi-sequence not to be normalized. Then we provide a simple way of constructing a normalized bi-sequence from any directive bi-sequence. In Section 5, we present a generalization of Justin's formula for generalized pseudostandard words whose proof depends strongly on the normalized form. Section 6 is devoted to the particular case of Rote words.

Notice that this paper is an complete and improved version of two conference communications. The content of Sections 4 and 5 was presented in Amiens (France) during the 13th Mons Theoretical Computer Science Days [2] (JM 2010), while the content of Section 6 was the main subject of a communication during the conference in honor of the 20th Anniversary of the Laboratoire de combinatoire et d'informatique mathmatique [3] (LaCIM 2010).

2 Preliminaries

We introduce the definitions and notation in the next sections.

2.1 Words

We first recall notions on words (for more details, see for instance [13]).

An *alphabet* $\mathcal{A}$ is a finite set of symbols called *letters*. A *word over* $\mathcal{A}$ is a sequence of letters from $\mathcal{A}$. The *empty word* ε is the empty sequence. Equipped with the concatenation operation, the set $\mathcal{A}^*$ of *finite words* over $\mathcal{A}$ is a free monoid with neutral element ε and set of generators $\mathcal{A}$, and $\mathcal{A}^+ = \mathcal{A}^* \setminus \varepsilon$. We denote by $\mathcal{A}^\omega$ the set of *(right-) infinite words* over $\mathcal{A}$. The set $\mathcal{A}^\infty$ is defined as the set of finite and infinite words: $\mathcal{A}^\infty = \mathcal{A}^* \cup \mathcal{A}^\omega$. Note that depending on the context, an infinite word is sometimes also called a *sequence*. For sake of clarity, variables denoting infinite words appear in bold.

If, for some words $u, s \in \mathcal{A}^\infty$, $v, p \in \mathcal{A}^*$, $u = pvs$, then v is a *factor* of u, p is a *prefix* of u and s is a *suffix* of u. The *set of factors* of the word u is denoted by $F(u)$. For $u = vw$, with $v \in \mathcal{A}^*$ and $w \in \mathcal{A}^\infty$, $v^{-1}u$ denotes the word w and uw^{-1} denotes the word v. Negative powers are naturally extended by $v^{-n}u = (v^n)^{-1}u$ and $u(w^n)^{-1}$.

As usual, for a finite word u and a positive integer n, the *nth power of* u, denoted by u^n, is the word ε if $n = 0$; otherwise $u^n = u^{n-1}u$. If $u \neq \varepsilon$, u^ω denotes the infinite word obtained by infinitely repeating u. Given a finite or an infinite word u, we denote by $u\texttt{[}i\texttt{]}$ the ith letter of u and by $u\texttt{[}i \dots j\texttt{]}$ the word $u\texttt{[}i\texttt{]}u\texttt{[}i+1\texttt{]} \cdots u\texttt{[}j\texttt{]}$. Given a nonempty finite word $u = u\texttt{[}1\texttt{]}u\texttt{[}2\texttt{]} \cdots u\texttt{[}n\texttt{]}$, the *length* $|u|$ of u is the integer n. One has $|\varepsilon| = 0$. The number of occurrences of the letter a in the word u is denoted by $|u|_a$. If $|u|_a = 0$, then u is called an *a-free word*.

The *reversal* of the finite word $u = u\texttt{[}1\texttt{]}u\texttt{[}2\texttt{]} \cdots u\texttt{[}n\texttt{]}$, also called the *mirror image*, is $R(u)= u\texttt{[}n\texttt{]}u\texttt{[}n-1\texttt{]} \cdots u\texttt{[}1\texttt{]}$ and if $u = R(u)$, then u is called a *palindrome*. The *right-palindromic closure* (*palindromic closure*, for short) of the finite word u, denoted by $u^{(+)}$, is defined by $u^{(+)} = u{\cdot}R(p)$, with $u = ps$ and s is the longest palindromic suffix of u. In other words, it is the shortest palindromic word having u as prefix.

Over a two-letter alphabet $\{0, 1\}$, there is a usual length preserving morphism, the *complementation*, defined by $\overline{0} = 1$ and $\overline{1} = 0$, which extends to words (finite or infinite). For instance, the complement of $u = u\texttt{[}1\texttt{]}u\texttt{[}2\texttt{]} \cdots u\texttt{[}n\texttt{]}$ is the word $\overline{u} = \overline{u\texttt{[}1\texttt{]}}\,\overline{u\texttt{[}2\texttt{]}} \cdots \overline{u\texttt{[}n\texttt{]}}$.

Sturmian words may be defined in many equivalent ways (see Chapter 2 in [13] for more details). For instance, they are the non-ultimately periodic infinite words over a two-letter alphabet that have minimal complexity, that is the number of distinct factors of length n is $(n+1)$, for each positive integer n. They are also the set of non-ultimately periodic binary balanced words. Recall that a word w over $\mathcal{A}$ is *balanced* if for all factors f, f' having same length, and for all letters $a \in \mathcal{A}$, one has $||f|_a - |f'|_a| \leq 1$.

The Sturmian words are also infinite words that describe discrete approximations of irrational slopes (see [13]). More precisely, an infinite word is *Sturmian* if and only if it is equal to one of the two infinite words $s_{\alpha,\rho}, {s'}_{\alpha,\rho} \in \{a, b\}^\omega$, defined by

$$s_{\alpha,\rho}\texttt{[}n\texttt{]} = \begin{cases} a, & \text{if } \lfloor \alpha(n+1) + \rho \rfloor = \lfloor \alpha n + \rho \rfloor; \\ b, & \text{otherwise.} \end{cases}$$

and

$${s'}_{\alpha,\rho}\texttt{[}n\texttt{]} = \begin{cases} a, & \text{if } \lceil \alpha(n+1) + \rho \rceil = \lceil \alpha n + \rho \rceil; \\ b, & \text{otherwise,} \end{cases}$$

where $\alpha, \rho \in \mathbb{R}$, $0 \leq \alpha < 1$ and α irrational.

The parameters ρ and α correspond respectively to the *intercept* and the *slope* of the line approximated by the word s. A Sturmian word is called *standard* (or *characteristic*) if $\rho = \alpha$.

2.2 Pseudopalindromic closure

Given a finite word w, let us denote by $\psi(w)$ the word obtained by iterating palindromic closure over w: $\psi(\varepsilon) = \varepsilon$ and $\psi(wa) = (\psi(w)a)^{(+)}$, for all letters a.

Note that the ψ operator is sometimes denoted by Pal in the works of Justin and Jamet et al (see for instance [11, 12]). By definition of iterated palindromic closure ψ, for any finite word w and letter a, $\psi(w)$ is a prefix of $\psi(wa)$. Thus, one can extend the iterated palindromic closure to any infinite word $\mathbf{w} = (a[n])_{n\geq 1}$ as follows:

$$\psi(\mathbf{w}) = \lim_{n\to\infty} \psi(a[1\dots n]).$$

We say that the word w *directs* the word $\psi(w)$. Also, we know from [7] that ψ gives a bijection between the set of infinite words over $\{a, b\}$ not of the form ua^ω or ub^ω, for some $u \in \{a, b\}^*$, and the set of standard Sturmian words over $\{a, b\}$. The word $\mathbf{w}$ is then called the *directive sequence* of the standard Sturmian word $\psi(\mathbf{w})$. Note that words of the form $\psi(ua^\omega)$ are periodic (see [9]).

The ψ operator is also well-defined over a k-letter alphabet, with $k \geq 3$. In this case, it is known [9] that $\psi(\mathcal{A}^\omega)$ is exactly the set of *standard episturmian words*, a generalization over a k-letter alphabet, $k \geq 3$, of the family of standard Sturmian words (for more details, see [10]).

Example 1. The *infinite Fibonacci word*

$$f = \psi((01)^\omega) = \underline{0}\,\underline{1}0\underline{0}10\underline{1}0010\underline{0}1010010\underline{1}\cdots$$

is a standard Sturmian word directed by the word $(01)^\omega$. Indeed:

$$\begin{array}{rclrcl}
\psi(0) & = & \underline{0} & \psi(010) & = & (\psi(01)0)^{(+)} = 010\underline{0}10 \\
\psi(01) & = & (\psi(0)1)^{(+)} = 0\underline{1}0 & \psi(0101) & = & (\psi(010)1)^{(+)} = 01001\underline{0}10010 \\
& \cdots & & & \cdots &
\end{array}$$

Notice that in the previous example, all the images are palindromes.

Example 2. The infinite word $01201002\cdots$ directs the standard episturmian word

$$\mathbf{w} = \psi(01201002\cdots) = \underline{0}\,\underline{1}0\underline{2}010\underline{0}102010\underline{1}0201001020100\underline{0}\cdots .$$

As usual, we have underlined in the previous examples the letters of w corresponding to the directive words, for sake of clarity.

When generating words by iterated palindromic closure, it is useful to have an efficient way to compute the successive prefixes. In that perspective, it is too costly to compute the longest palindromic suffix every time a letter is added in the directive sequence. The key of this problem is found in a formula introduced by Justin [12].

Proposition 3. [12] *Let $w \in \mathcal{A}^*, a \in \mathcal{A}$. If w is not a-free, then we write $w = v_1 a v_2$ with v_2 a-free and we have*

$$\psi(wa) = \begin{cases} \psi(w)a\psi(w), & \textit{if } w \textit{ is } a\textit{-free;} \\ \psi(w)\psi(v_1)^{-1}\psi(w), & \textit{otherwise.} \end{cases}$$

Algorithm 1 Computation of a word by iterated palindromic closure.

```
function IteratedPalindromicClosure(w)
    Input: Any word w
    Output: The iterated palindromic closure ψ(w) of w
    n ← |w|, u ← ε
    HasOccurred(a) ← false for every letter a
    for i ∈ {1, 2, ..., n} do
        ℓ ← LastLength(w[i])
        LastLength(w[i]) = |u|
        if not HasOccurred(w[i]) then
            HasOccurred(w[i]) ← true
            u ← u · w[i] · u
        else
            s ← |u| − ℓ
            u ← u · Suff_s(u)
        end if
    end for
end function
```

Proposition 3 yields directly Algorithm 1. In the next sections, we generalize it to the so-called *pseudopalindromes* over binary alphabets, where the situation is more complicated, since there are two possible kinds of closure. A few years ago, de Luca and De Luca [8] extended the notion of palindrome to what they call *pseudopalindrome*, using involutory antimorphisms. In order to define it, let us first recall that a map $\vartheta : \mathcal{A}^* \to \mathcal{A}^*$ is called an *antimorphism* of $\mathcal{A}^*$ if for all $u, v \in \mathcal{A}^*$ one has $\vartheta(uv) = \vartheta(v)\vartheta(u)$. Moreover, an antimorphism is *involutory* if $\vartheta^2 = \mathrm{Id}$. Any antimorphism ϑ of $\mathcal{A}^*$ can be constructed as $\vartheta = \sigma \circ R = R \circ \sigma$, with σ a permutation of the alphabet $\mathcal{A}$. We denote by σ_ϑ the permutation associated with the antimorphism ϑ. Over a two-letter alphabet $\{a, b\}$, there are only two involutory permutations of letters $\sigma_R = (a)(b)$ and $\sigma_E = (ab)$, yielding the two antimorphisms R, the reversal antimorphism, and E, defined as $E = \sigma_E \cdot R$. The antimorphism E will be called, as usual, the *exchange antimorphism*.

We can now define the generalization of palindromes given in [8]: a word $w \in \mathcal{A}^*$ is called a *ϑ-palindrome* if it is the fixed point of the involutory antimorphism ϑ of the free monoid $\mathcal{A}^*$: $\vartheta(w) = w$. When the antimorphism ϑ is not mentioned, w is called a *pseudopalindrome*. Notice that the R-palindromes are exactly the usual palindromes.

By analogy to the palindromic closure $^{(+)}$, the *ϑ-palindromic closure* of the finite word u, also called the *pseudopalindromic closure* when the antimorphism is not specified, is defined by $u^{\oplus_\vartheta} = sq\vartheta(s)$, where $u = sq$, with q the longest ϑ-palindromic suffix of u. The ϑ-palindromic closure of u is the shortest ϑ-palindrome having u as prefix.

Example 4. Over the alphabet $\{0, 1\}$, since the longest E-palindromic suffix of $w = 0010$ is 10, $w^{\oplus_E} = 0010 \cdot E(00) = 001011$.

Notice that in the previous example, we have that 001011 is an E-palindrome, also called an *antipalindrome*, since $E(001011) = R(\overline{001011}) = R(110100) = 001011$.

Extending the ψ operator to ϑ-palindromes, the ψ_ϑ operator is naturally defined by $\psi_\vartheta(\varepsilon) = \varepsilon$ and $\psi_\vartheta(wa) = (\psi_\vartheta(w)a)^{\oplus_\vartheta}$, for $w \in \mathcal{A}^*$ and $a \in \mathcal{A}$. Then, for $\mathbf{w} \in \mathcal{A}^\omega$, $\psi_\vartheta(\mathbf{w}) = \lim_{n\to\infty} \psi_\vartheta(w\,[1\ldots n])$. This limit exists since by the definition of ψ_ϑ, for any involutory antimorphism ϑ, $w \in \mathcal{A}^*$ and $a \in \mathcal{A}$, $\psi_\vartheta(w)$ is a prefix of $\psi_\vartheta(wa)$. The infinite word obtained by the ψ_ϑ operator is a *ϑ-standard word*, also called a *pseudostandard word* when the antimorphism is not specified. This new class of words is a general one that includes the standard Sturmian and the standard episturmian ones and was first introduced in [8].

Example 5. Over $\mathcal{A} = \{0, 1\}$:

$$\psi_{\mathrm{E}}(001) = ((\psi_{\mathrm{E}}(0)0)^{\oplus_E}1)^{\oplus_E} = ((01 \cdot 0)^{\oplus_E}1)^{\oplus_E} = (0101 \cdot 1)^{\oplus_E} = 0101100101.$$

2.3 Generalized pseudostandard words

In [8], the authors introduce the pseudostandard words as well as a new class of words, extending the pseudostandard words to a larger class of words: the *generalized pseudostandard words*. It consists mainly in allowing different types of pseudopalindromic closure while constructing the word from its directive sequence.

More formally, let $\mathcal{I}$ be the set of all involutory antimorphisms of $\mathcal{A}^*$, and $\mathcal{I}^\omega$ be the set of infinite sequences over $\mathcal{I}$. Let $\Theta = \vartheta_1\vartheta_2\vartheta_3\ldots \in \mathcal{I}^\omega$ and let ${}^{\oplus_i}$ be the ϑ_i-palindromic closure operator, for all $i \geq 1$. We define inductively an operator ψ_Θ by setting $\psi_\Theta(\varepsilon) = \varepsilon$, and

$$\psi_\Theta(w\,[1\ldots n+1]) = (\psi_\Theta(w\,[1\ldots n])w\,[n+1])^{\oplus_{n+1}}$$

whenever $w\,[i] \in \mathcal{A}$ for $i \geq 1$.

If $\mathbf{w} = \mathbf{w}\,[1]\mathbf{w}\,[2]\cdots\mathbf{w}\,[n]\cdots \in \mathcal{A}^\omega$, $\mathbf{w}\,[i] \in \mathcal{A}$ for $i \geq 1$, then $\psi_\Theta(\mathbf{w}\,[1\ldots i])$ is a prefix of $\psi_\Theta(\mathbf{w}\,[1\ldots i+1])$ for any i, so that the infinite word

$$\psi_\Theta(\mathbf{w}) = \lim_{n\to\infty} \psi_\Theta(\mathbf{w}\,[1\ldots n])$$

is well-defined. We call $\psi_\Theta(\mathbf{w})$ a *generalized pseudostandard word* and the pair (Θ, w) is called the *directive bi-sequence* of $\psi_\Theta(w)$.

When we consider a finite word w and a finite sequence of involutory antimorphisms Θ, both of length $|w|$, we call the word $\psi_\Theta(w)$ a *prefix of a generalized pseudostandard word* and the *length* of its directive bi-sequence is $|w|$.

Notice that one can set $\Theta = R^\omega$ (resp., $\Theta = \vartheta^\omega$) in order to obtain standard episturmian (resp., ϑ-standard) words.

For the remainder of that paper, we fix $\mathcal{A} = \{0, 1\}$ so that the only two involutory antimorphisms are the reversal R and the exchange antimorphism E.

Moreover, we write $\overline{R} = E$ and $\overline{E} = R$.

Example 6. Let $w = 0100$ and $\Theta = RERR$. Then $\psi_\Theta(w) = \underline{0}\underset{\cdot}{1}\underline{0}\,\underline{0}10$.

Note that in the previous example, we have underlined (resp., under-dotted) the letter in the generalized pseudostandard word $\psi_\Theta(w)$ where a R-closure (resp., E-closure) is performed.

A remarkable property of this class of words is that it contains the well-known Thue-Morse word T as shown in [8], whose construction by iterated pseudopalindromic closure coincides with the one consisting of concatenating the successive prefixes with their complement. More precisely:

Theorem 7. [8] *The Thue-Morse word is described by* $\psi_{(ER)^\omega}(01^\omega)$.

Example 8. As an example, let us construct the first prefixes of the Thue-Morse word.

$$
\begin{array}{rclrcl}
\psi_E(0) & = & \d{0}1 & \psi_{ERE}(011) & = & \d{0}1\underline{1}0\d{1}001 \\
\psi_{ER}(01) & = & \d{0}1\underline{1}0 & \psi_{ERER}(0111) & = & \d{0}1\underline{1}0\d{1}001\underline{1}0010110.
\end{array}
$$

3 Pseudoperiodicity

In the conference version of this extended article [2], we introduced the notion of *pseudoperiod*. We also provided a theorem in the spirit of Fine and Wilf for binary alphabets that was generalized to arbitrary alphabets in [1]. We briefly discuss here this concept and provide lemmas (17, 18 and 19) that turn out to be essential in the next section to describe the structure of words obtained by iterated pseudopalindromic closure.

First, we recall the definition of pseudoperiod from [2].

Definition 9. [2] Let w be a finite word and σ be a permutation of the alphabet. We say that the positive integer p is a σ-period of w if for each integer i such that $1 \leq i \leq |w| - p$, we have $w[i] = \sigma(w[i+p])$.

Example 10. Let $\sigma_R = (0)(1)$ and $\sigma_E = (01)$. Then the word 011011 has the σ_R-period 3 and the σ_E-period 5, since $011011 = 011 \cdot \sigma_R(011)$ and $011011 = 01101 \cdot \sigma_E(0)$.

As we can see in the previous example, the σ_R-periods are the usual periods on words. Finally, the reader verifies easily that if p is a σ-period of a word w, then $\mathrm{ord}(\sigma)p$ is a period of w, where $\mathrm{ord}(\sigma)$ is the order of the permutation σ, i.e. the smallest number k such that σ^k is the identity.

It shall be mentioned that the word *pseudoperiod* is also used in [6, 14], but its meaning is not equivalent to Definition 9. More precisely, the authors say from a positive integer p that it is a σ-period of a word w if w may be written as an arbitrary product of elements in $\{u, \sigma(u)\}$, where u is a word of length p. For instance, the number 2 is a σ-period of the word $w = 12323212$ for $\sigma : 1 \mapsto 3, 2 \mapsto 2, 3 \mapsto 1$, since $w = u\sigma(u)\sigma(u)u$, for $u = 12$. On the other hand, Definition 9 imposes an alternance of the word with its complement and applies even if p does not divide $|w|$. Moreover, in the case of a binary alphabet, for every word w, 1 is trivially a σ_E-period of w when considering the definition of pseudoperiod of [6].

It is known that overlapping palindromes yield periodicity (see for instance [7]). These periodic properties may be generalized naturally to overlapping pseudopalindromes. The following lemma shows that pseudoperiodic pseudopalindromes may be extended into longer ones or broken into smaller ones.

Lemma 11. *Let w be a ϑ-palindrome and p be a σ-period of w, where σ is an involutory permutation and $p < |w|$, u is the suffix of w of length p. Moreover, let $i = \lfloor |w|/p \rfloor$. Then*

(i) $w(\sigma(u)u)^j$ *is a* ϑ*-palindrome for all integers* $j \geq -\lfloor i/2 \rfloor$*;*

(ii) $w\sigma(u)(u\sigma(u))^j$ *is a* $(\vartheta \circ \sigma)$*-palindrome for all integers* $j \geq -\lfloor i/2 \rfloor$*.*

Proof. (i) Since p is a σ-period of w, there exist a word x, a nonempty word y and a positive integer k such that $|xy| = p$ and $w \in \{(xy\sigma(xy))^k x, (xy\sigma(xy))^k xy\sigma(x)\}$. We first consider the case $w = (xy\sigma(xy))^k x$. In particular, $k = \lfloor i/2 \rfloor$ and $u = \sigma(y)x$. Moreover, since w is a ϑ-palindrome, we deduce that $x = \vartheta(x)$ and $y = (\vartheta \circ \sigma)(y)$. Let $j \geq -\lfloor i/2 \rfloor$ be an integer. Then $w(\sigma(u)u)^j = (xy\sigma(xy))^{k+j}x$, which is indeed a ϑ-palindrome, since $x = \vartheta(x)$ and $y = (\vartheta \circ \sigma)(y)$. The case $w = (xy\sigma(xy))^k xy\sigma(x)$ is verified similarly, as well as (ii). □

Roughly speaking, Lemma 11 states that if a given ϑ-palindrome is σ-periodic and sufficiently long, then one can construct a sequence of alternating ϑ-palindromes and $(\vartheta \circ \sigma)$-palindromes by erasing the repetitions or by extending the period.

Example 12. Let $\sigma = \sigma_E$, $\vartheta = R$ and $w = 0010111010001011110100$. Then 5 is an σ_E-period of w and w is an R-palindrome. Also, Lemma 11 applies, so that $w(10100)^{-1} = 0010111010001011$ and $w(0101110100)^{-1}(10100)^{-1} = 001011$ are E-palindromes. On the other hand, the words $w(0101110100)^{-1} = 00101110100$ and $w(0101110100)^{-2} = 0$ are R-palindromes.

In [1], the authors generalized Fine and Wilf's Theorem for pseudoperiodic words.

Theorem 13. [1] *Let* p*,* q *be two positive integers and* σ_1*,* σ_2 *two permutations such that* σ_1 *and* σ_2^{-1} *commute. Then any word* w *of length at least* $p + q$ *admitting* p *as a* σ_1*-period and* q *as a* σ_2*-period also admits* $\gcd(p, q)$ *as a* σ*-period, where* $\sigma = \sigma_1^x \sigma_2^{-y}$ *and* x, y *are any integers such that* $\gcd(p, q) = xp - yq$*.*

It is worth mentioning that Theorem 13 holds even if σ_1 and σ_2 are not involutory. The binary case is proved in [2]. It also directly follows from Theorem 13:

Theorem 14. [2] *Let* p*,* q *be two positive integers and* σ_1*,* σ_2 *two permutations on a binary alphabet* $\mathcal{A}$*. Then any word* w *of length at least* $p + q$ *admitting* p *as a* σ_1*-period and* q *as a* σ_2*-period also admits* $\gcd(p, q)$ *as a* σ*-period, where*

$$\sigma = \begin{cases} \sigma_R, & \text{if } \sigma_1 = \sigma_2 = \sigma_R; \\ \sigma_E, & \text{otherwise.} \end{cases}$$

Proof. On binary alphabet, the permutations σ_R and $\sigma_E^{-1} = \sigma_E$ commute. Moreover, any pair of integers verifying $\gcd(p, q) = xp - yq$, also called Bezout coefficients, may be chosen so that one is even and the other is odd. If $\sigma_1 = \sigma_2 = \sigma_R$, then $\sigma = \sigma_R$ and the theorem follows. Otherwise, since x and y have different parity, then $\sigma_1^x \sigma_2^{-y} = \sigma_E = \sigma$, which concludes the proof. □

Example 15. The word $w = 01010101$ admits the σ_R-period 4 and the σ_E-period 3. By Theorem 14, $\gcd(4, 3) = 1$ is also an σ_E-period of w.

Remark 16. The bound $|w| \geq p + q$ in Theorem 14 is tight, as illustrated by the following example. Let p be any positive integer and $q = p + 1$. Moreover, let $w = 0^p1^p$. Then w has the σ_E-periods p and q. On the other hand, $1 = \gcd(p, q)$ is not an σ_E-period of w and $|w| = 2p < 2p + 1 = p + q$. Actually, one may show that there is no word of length $2p + 1$ admitting the σ_E-periods p and q.

We conclude this section with three lemmas about local periodicity in words. The first lemma holds for any alphabet.

Lemma 17. *Let u be a finite word over an arbitrary alphabet, p be a ϑ_1-palindrome and q be a ϑ_2-palindrome for some involutory antimorphisms ϑ_1,ϑ_2, such that $pu = q$. Then q has the $(\sigma_{\vartheta_2} \circ \sigma_{\vartheta_1})$-period $|u|$.*

Proof. Let i be an integer satisfying $1 \leq i \leq |p|$. Since p is a prefix of q and since p is a ϑ_1-palindrome, one has $q\,\texttt{[}i\texttt{]} = \sigma_{\vartheta_1}(q\,\texttt{[}|p| + 1 - i\texttt{]})$. But q is a ϑ_2-palindrome, which implies that

$$\begin{aligned} q\,\texttt{[}i\texttt{]} &= \sigma_{\vartheta_1}(q\,\texttt{[}|p| + 1 - i\texttt{]}) \\ &= (\sigma_{\vartheta_2} \circ \sigma_{\vartheta_1})(q\,\texttt{[}|q| + 1 - |p| - 1 + i\texttt{]}) \\ &= (\sigma_{\vartheta_2} \circ \sigma_{\vartheta_1})(q\,\texttt{[}|q| - |p| + i\texttt{]}) \\ &= (\sigma_{\vartheta_2} \circ \sigma_{\vartheta_1})(q\,\texttt{[}i + |u|\texttt{]}), \end{aligned}$$

and the result follows. ☐

Not any pseudoperiod may exist in a given word.

Lemma 18. *Let w be a finite binary word, p a σ_1-period of w and q a σ_2-period of w, where $\sigma_1, \sigma_2 \in \{\sigma_R, \sigma_E\}$. Assume that $|w| > p > q$ and $p = mq$ for some integer $m \geq 2$. Then one of the two following conditions holds:*

(i) *σ_1 is the identity and m is even;*

(ii) *$\sigma_1 = \sigma_2$ and m is odd.*

Proof. Let i be an integer such that $1 \leq i \leq |w| - p$. Note that σ_1 and σ_2 commute since the alphabet is binary. Therefore,

$$\sigma_R(w\,\texttt{[}i\texttt{]}) = w\,\texttt{[}i\texttt{]} = \sigma_1(w\,\texttt{[}i + p\texttt{]}) = \sigma_1(w\,\texttt{[}i + mq\texttt{]}) = (\sigma_2^m \circ \sigma_1)(w\,\texttt{[}i\texttt{]}),$$

so that $\sigma_R = \sigma_2^m \circ \sigma_1$. As a consequence, if m is even, then $\sigma_1 = \sigma_R$, and if m is odd, then $\sigma_R = \sigma_1 \circ \sigma_2$, which implies $\sigma_1 = \sigma_2$. ☐

It is worth mentioning that Lemma 18 holds for arbitrary alphabets whenever all letters occur in the prefix of length $|w| - p$ of w, i.e. whenever we have $\sigma_R(a) = (\sigma_{\vartheta_2}^m \circ \sigma_{\vartheta_1})(a)$ for all letters a.

The last lemma is a simple extension of Lemma 8.1.3 of [13] giving condition for a local period to propagate to the whole word. As for Lemma 18, it could be extended to arbitrary alphabets provided that all letters occur in some prefix of w, but since this paper is devoted to binary alphabets, for sake of simplicity, we only present this case.

Lemma 19. *Let w be a finite binary word and v be a factor of w. Assume that p is a σ_1-period of w such that $|v| > p$ and q is a σ_2-period of v such that q divides p, where $\sigma_1, \sigma_2 \in \{\sigma_R, \sigma_E\}$. Then q is a σ_2-period of w.*

Proof. First consider the case $q = p$. Then $p = q$ is both a σ_1-period and a σ_2-period of v and $|v| > p$. This means that $\sigma_1 = \sigma_2$ and the lemma follows. For the rest of the proof, we may suppose $q < p$ and $p = qm$ for some positive integer $m \geq 2$.

Let k be an integer, $1 \leq k \leq |w|$, such that $v = w[k]w[k+1]\cdots w[k+|v|-1]$. Let $V = \{k, k+1, \ldots, k+|v|-1\}$ be the set of indices of v in w. Moreover, let i be an integer such that $1 \leq i \leq |w|-q$. Since $|v| > p$, there exists an integer $i' \in V$ such that $i' \equiv i \bmod p$. Therefore, since p is a σ_1-period of w, we have $w[i'] = \sigma_1^{|\ell|}(w[i])$, where ℓ is the integer satisfying $i' - i = p\ell$. Since σ_1 is involutory, we have $w[i] = \sigma_1^{|\ell|}(w[i'])$ as well.

Let $j = i + q$. Using a similar argument as above, we find that there exists at least one integer $j' \in V$ such that $j' \equiv j \bmod p$. In particular, we may choose j' so that $j' \in \{i'+q, i'+q-p\}$. Indeed, we have $i'+q \equiv i'+q-p \equiv i+q \bmod p$ and at least one value among $i'+q$ and $i'+q-p$ must fall in V (since $|v| > p$ and $i' \in V$). Thus, we may write $j' - j = p\ell'$, for some integer $\ell' \in \{\ell-1, \ell\}$, so that $w[j'] = \sigma_1^{|\ell'|}(w[j])$. We distinguish two cases. Assume first that $\ell' = \ell$ so that $j' = i' + q$. Since q is a σ_2-period of v, we have $w[i'] = \sigma_2(w[j'])$. Then $w[j] = (\sigma_1^{|2\ell|} \circ \sigma_2)(w[i]) = \sigma_2(w[i])$. On the other hand, suppose that $\ell' = \ell - 1$ so that $j' = i' + q - p$. Recall that $p = qm$. Then $j' = i' + (1-m)q$ so that $w[i'] = \sigma_2^{|1-m|}(w[j'])$. This implies $w[j] = (\sigma_1^{|2\ell-1|} \circ \sigma_2^{|1-m|})(w[i]) = (\sigma_1 \circ \sigma_2^{|1-m|})(w[i])$. We know from Lemma 18 that either $\sigma_1 = R$ and m is even or $\sigma_1 = \sigma_2$ and m is odd. Both cases imply $w[j] = \sigma_2(w[i])$. Hence, we have shown that $w[i+q] = \sigma_2(w[i])$ for any integer i such that $1 \leq i \leq |v| - q$, i.e. q is a σ_2-period of w. $\square$

4 Normalized form

Words obtained by iterated palindromic closure are the limit of a sequence of palindromes that are prefixes of each other. The idea is the same when considering iterated pseudopalindromic closure. A first trivial and useful observation is the following.

Lemma 20. *Let $\mathbf{u} = \psi(\mathbf{w}) = \psi_{R^\omega}(\mathbf{w})$ be a word on an arbitrary alphabet. The word v is a palindromic prefix of $\mathbf{u}$ if and only if there exists a nonnegative integer n such that $v = \psi_{R^n}(\mathbf{w}[1\ldots n])$.*

Proof. ($\Rightarrow$) By contradiction, assume that such a word v exists. Let n be the integer such that $|\psi_{R^{n-1}}(\mathbf{w}[1\ldots n-1])| < |v| < |\psi_{R^n}(\mathbf{w}[1\ldots n])|$. Then $\psi_{R^n}(\mathbf{w}[1\ldots n])$ is not the shortest palindromic suffix having $\psi_{R^{n-1}}(\mathbf{w}[1\ldots n-1])\mathbf{w}[n]$ as a prefix, contradicting the definition of the palindromic closure. ($\Leftrightarrow$) By definition of palindromic closure, $\psi_{R^n}(\mathbf{w}[1\ldots n])$ is a palindrome for any integer $n \geq 0$. $\square$

Roughly speaking, Lemma 20 states that no palindromic prefix is missed by the iterated palindromic closure. In the following, Lemma 20 is used several times without being referenced. This fact also holds for any ϑ-standard word (see Proposition 4.1, [8]), i.e. no ϑ-palindromic prefix is missed by the iterated ϑ-palindromic closure. However, this is not the case if different pseudopalindromic closures are allowed. For instance, the word $w = \psi_{RERE}(0011) = 00111001100011$ misses the palindrome 00 while the word $u = \psi_{RRE}(011) = 0101$ misses the E-palindrome 01. On the other hand, $w = \psi_{RRERE}(00111)$ and $u = \psi_{RERE}(0101)$, i.e. it is possible to rewrite the directive bi-sequences of w and u so that they do not miss any pseudopalindromic prefixes. As we will see (and prove) in the sequel, it is always possible to rewrite any directive bi-sequence of a generalized pseudostandard word in a “normalized” form.

Definition 21. A finite or infinite directive bi-sequence (Θ, w) is called *normalized* if it verifies the following condition: v is a pseudopalindromic prefix of $\psi_\Theta(w)$ if and only if there exists a non negative integer n such that $v = \psi_{\Theta[1..n]}(w[1..n])$. A pseudopalindromic prefix v that does not satisfy the previous condition is called a *missed pseudopalindrome* and the we say that (Θ, w) *misses* v.

The length of any missed pseudopalindrome is constrained.

Lemma 22. *Let (Θ, w) be a finite directive bi-sequence describing a prefix of a generalized pseudostandard word on any alphabet. If (Θ, w) is normalized and $(\Theta\tau, wx)$ is not, with $x \in \mathcal{A}$ and $\tau \in \{E, R\}$, any missed ϑ-palindromic prefix p is such that $|\psi_\Theta(w)| < |p| < |\psi_{\Theta\tau}(wx)|$.*

Proof. It is obvious that $|p| < |\psi_{\Theta\tau}(wx)|$. If $|p| \le |\psi_\Theta(w)|$, then a contradiction occurs, since (w, Θ) is supposed normalized. □

The next results only apply to binary alphabets. First, the following fact is easily observed.

Lemma 23. *The shortest prefixes of a normalized directive bi-sequence containing exactly two different letters are of the form:*

$$(R^{i+1}, a^i\overline{a}) \text{ for } i \ge 2, \text{ or } (R^iE, a^i\overline{a}) \text{ for } i \ge 1 \text{ and } a \in \{0, 1\}.$$

Proof. By direct inspection. One notices in particular that no normalized bi-sequence starts with the antimorphism E. □

Generalized pseudostandard words have periods to different scale. In particular, if a bi-sequence is not normalized, we can extract useful information about the antimorphisms and the letters involved.

Lemma 24. *Let (Θ, w) be a finite normalized directive bi-sequence of length $n \ge 1$ of a prefix of a generalized pseudostandard binary word. Suppose that $(\Theta\tau, wa)$ is not normalized, where $\tau \in \{R, E\}$ and $a \in \mathcal{A}$. Let $u = \psi_\Theta(w)$, $v = \psi_{\Theta\tau}(wa)$ and t be a pseudopalindromic prefix missed by $(\Theta\tau, wa)$. Finally, let $p = |v| - |u|$, $q = |v| - |t|$ and $g = \gcd(p, q)$. Then exactly one of the following conditions hold.*

(i) *g is an σ_E-period of v, $p = 2g$, $q = g$ and $\vartheta_n = \tau$, where ϑ_n is the last antimorphism of Θ;*

(ii) *$(\Theta\tau, wa) = (R^nE, a^{n+1})$;*

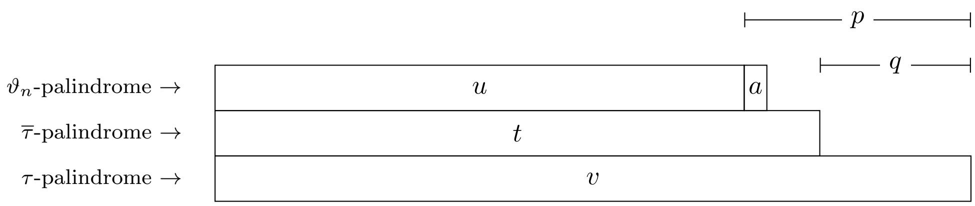

Figure 1: Illustration of the proof of Lemma 24.

Proof. The situation is depicted in Figure 1. Notice that t is a $\overline{\tau}$-palindrome, otherwise v would not be the shortest τ-palindrome having ua as a prefix. Moreover, it follows from Lemma 17 that p is a $(\sigma_{\vartheta_n} \circ \sigma_\tau)$-period of v, since v is a τ-palindrome and u is a ϑ_n-palindrome, where, ϑ_n is the last antimorphism of Θ. Similarly, q is a $(\sigma_\tau \circ \sigma_{\overline{\tau}})$-period (i.e. an σ_E-period) of v.

First, suppose that $|v| \geq p + q$. Then Theorem 14 applies so that v has the σ_E-period $g = \gcd(p, q)$. By Lemma 11, we conclude that the prefix y of v of length $|v| - 2g$ is a τ-palindrome. Since no τ-palindrome occurs between u and v (otherwise v would not be the shortest τ-palindrome having ua as a prefix), we must have $|y| \leq |u|$, i.e. $p = |v| - |u| \leq 2g$. Since $0 < q < p \leq 2g$ and g divides both p and q, we have $p, q \in \{g, 2g\}$. But $q < p$, so that $q = g$ and $p = 2g$. In particular, $u = y$ is a τ-palindrome and we are in case (i).

It remains to consider the case $|v| < p+q$. Notice that, by definition of pseudopalindromic closure, one has $|v| \leq 2|u| + 2$, with $|v| = 2|u| + 2$ only if v is an E-palindrome. We first show that $|v| = 2|u| + 2$. Arguing by contradiction, suppose that $|v| \leq 2|u| + 1$. Then

$$
\begin{aligned}
|v| &< p + q \\
&= |v| - |u| + |v| - |t| \\
&\leq 2|u| + 1 - |t| + |v| - |u| \\
&= |u| + 1 - |t| + |v| \\
&\leq |u| + 1 - |u| - 1 + |v| \qquad (1) \\
&= |v|,
\end{aligned}
$$

which is absurd. Note that Inequality (1) follows from the inequality $|t| \geq |u| + 1$. Hence, $|v| = 2|u| + 2$. This implies that $\tau = E$ and t is a R-palindrome (since $\overline{\tau} = R$).

Next, suppose that $|t| \geq |u| + 2$. Then $q \leq |u|$. Since u is a ϑ_n-palindrome and q is a σ_E-period of u, Lemma 11 implies that the prefix of v of length $|u| + q$ is a $\overline{\sigma_{\vartheta_n}}$-palindrome. Therefore, $\overline{\sigma_{\vartheta_n}} \neq \tau$, otherwise, v would not be of minimum length, which implies $\vartheta_n \neq \tau$ and then $\vartheta_n = R$. As a consequence, $p - q$ is a σ_E-period of t. Moreover, we know that q is a σ_E-period of t as well since it is a period of v. We can apply Theorem 14 to t since $q + (p - q) = p = |u| + 2 \leq |t|$, so that $\gcd(q, p - q) = \gcd(q, p) = g$ is a σ_E-period of t which propagates to v, by Lemma 19. Once again by Lemma 11, we conclude that the prefix of v of length $|u| + 2g$ is a ϑ_n-palindrome, i.e. a τ-palindrome. This implies $|u| + 2g \geq |v| = 2|u| + 2$, so that $2g \geq |u| + 2 = p$. Since g divides both p and q and since $q < p$, this means that $q = g$ and $p = 2g$, which corresponds also to case (i).

Finally, assume that $|t| = |u| + 1$. There are two cases to consider. If $\vartheta_n = R$, then Lemma 17 implies that 1 is a σ_R-period of t, so that $t = a^{n+1}$, $u = a^n$, $v = a^{n+1}\overline{a^{n+1}}$ and

$(\Theta\tau, wa) = (R^n E, a^{n+1})$, which yields case (ii). Otherwise, again by Lemma 17, $\vartheta_n = E$ and 1 is a σ_E-period of t. Since t is an R-palindrome, we must have $t = (a\overline{a})^k a$ for some positive integer k. But every prefix of t is either an R-palindrome or an E-palindrome and (Θ, w) is normalized: Therefore, $|u| = n$ which implies $u = (a\overline{a})^{n/2}$, $t = (a\overline{a})^{n/2}a$ and $v = (a\overline{a})^{n/2+1}$. But $n + 2 = |v| = 2|u| + 2 = 2n + 2$, so that $n = 0$, a contradiction. □

There are forbidden patterns that necessarily lead to non normalized bi-sequences.

Lemma 25. *Let Θ be a sequence of involutory antimorphisms, $\vartheta \in \{R, E\}$, $w \in \mathcal{A}^*$ and $a, b \in \mathcal{A}$, where $\mathcal{A}$ is a binary alphabet. Suppose that $(\Theta\vartheta\overline{\vartheta}, wab)$ is normalized. Also, let $u = \psi_{\Theta\vartheta}(wa)$, $v = \psi_{\Theta\vartheta\overline{\vartheta}}(wab)$, $p = |v| - |u|$ and s be the suffix of length p of v. Then*

(i) *p is the minimum σ_E-period of v;*

(ii) *$\psi_{\Theta\vartheta\overline{\vartheta}\vartheta}(uab\overline{b}) = v\overline{s}$;*

(iii) *$\psi_{\Theta\vartheta\overline{\vartheta}\overline{\vartheta}}(uab\overline{b}) = v\overline{s}s$;*

(iv) *$(\Theta\vartheta\overline{\vartheta}\overline{\vartheta}, uab\overline{b})$ is not normalized, $(\Theta\vartheta\overline{\vartheta}\vartheta\overline{\vartheta}, uab\overline{b}b)$ is normalized and both bi-sequences generate the same word.*

Proof. (i) It follows from Lemma 17 that p is a σ_E-period of v since $\sigma_\vartheta \circ \sigma_{\overline{\vartheta}} = \sigma_E$. It remains to show that p is minimal. By contradiction, assume that there exists a σ_E-period $p' < p$. Let s' be the word of length p' such that us' is a prefix of v. By Lemma 11, us' is a $\overline{\vartheta}$-palindrome and, by construction of v, b is the first letter of s', contradicting the fact that v is the shortest $\overline{\vartheta}$-palindrome having ub as a prefix.

(ii) Since v is a $\overline{\vartheta}$-palindrome having p as a σ_E-period, it follows from Lemma 11 that $v\overline{s}$ is a ϑ-palindrome. Now, assume that $v\overline{s}$ is not the shortest ϑ-palindrome having $v\overline{b}$ as a prefix. Let x be the longest ϑ-palindrome suffix of v. Then $|x| > |v| - p$ and, by Lemma 17, v has the σ_E-period $|v| - |x| < p$, contradicting the minimality of p.

(iii) It suffices to apply the same reasoning as in part (ii).

(iv) Parts (ii) and (iii) implies that $(\Theta\vartheta\overline{\vartheta}\overline{\vartheta}, uab\overline{b})$ is not normalized since it misses the pseudopalindrome $v\overline{s}$. Moreover, $(\Theta\vartheta\overline{\vartheta}\vartheta\overline{\vartheta}, uab\overline{b}b)$ is normalized since it does not verify any of the conditions of Lemma 24. Finally, applying twice part (ii) and once part (iii) shows that both bi-sequence generate the same word. □

Let (Θ, w) be an infinite (resp., a finite) directive bi-sequence of a (resp., prefix of a) generalized pseudostandard word. The concept of factor is naturally extended to bi-sequences. More precisely, $(\vartheta_i \cdots \vartheta_{i+k}, w[i \ldots i+k])$ is called a *factor* of (Θ, w) for any integers $i, k \geq 1$.

We are now ready to describe the forbidden factors and prefixes of normalized bi-sequences.

Proposition 26. *A finite directive bi-sequence s on a binary alphabet is normalized if and only if it does not have a prefix of one of the following forms:*

(i) *$(RR, a\overline{a})$,*

(ii) *$(R^{i-1}E, a^i)$,*

(iii) *$(R^iEE, a^i\overline{a}\,\overline{a})$,*

(iv) *$(\Theta REE, wab\overline{b})$ or $(\Theta ERR, wab\overline{b})$,*

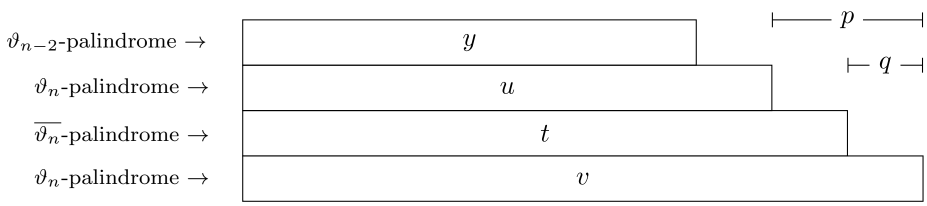

Figure 2: Illustration of Part (⇐) in the proof of Proposition 26. It follows from Lemma 24 that $p = 2g$ and $q = g$, where $g = \gcd(p, q)$. Also, $\vartheta_{n-1} = \vartheta_n$.

where $a, b \in \{0, 1\}$, $i \geq 1$ is an integer and (Θ, w) is a finite directive bi-sequence.

Proof. (⇒) We prove the contrapositive, i.e. we suppose that the prefix of the directive bi-sequence is of one of these four forms and prove that it is not normalized. Clearly, (i) $\psi_{RR}(a\overline{a}) = a\overline{a}a$ misses the E-palindromic prefix $a\overline{a}$, (ii) $\psi_{R^{i-1}E}(a^i) = a^i\overline{a}^i$ misses the palindromic prefix a^i and (iii) $\psi_{R^iEE}(a^i\overline{a}\,\overline{a}) = a^i\overline{a}^{i+1}a^{i+1}\overline{a}^i$ misses the palindromic prefix $a^i\overline{a}^{i+1}a^i$. Case (iv) follows directly from Lemma 25.

(⇐) We prove the contrapositive, i.e. we suppose that s is not normalized. If $|s| \leq 2$, then inspection shows that the only possible non-normalized directive bi-sequences fall in case (i) or (ii). Now, assume that $|s| \geq 3$. Let (Θ, w) be the shortest non-normalized prefix of s and $n = |w|$. Let

$$\begin{aligned} u &= \psi_{\vartheta_1\vartheta_2\cdots\vartheta_{n-1}}(w[1]w[2]\cdots w[n-1]), \\ v &= \psi_\Theta(w) \end{aligned}$$

and t be a $\overline{\vartheta_n}$-palindrome missed by (Θ, w) (see Figure 2). Moreover, let $p = |v| - |u|$, $q = |v| - |t|$ and $g = \gcd(p, q)$. By Lemma 24, either we are in case (ii) or (iv), or (Θ, w) ends with a factor in $\{(RRR, abc), (EEE, abc), (ERR, abc), (REE, abc)\}$, where a, b, c are letters. On the other hand, Lemma 24 implies that $|v| \geq p + q$, g is an σ_E-period of v, $p = 2g$, $q = g$ and u is a ϑ_n-palindrome, i.e. $\vartheta_{n-1} = \vartheta_n$. Let y be the prefix of lenght $n - 2$ of $\psi_\Theta(w)$, that is $y = \psi_{\vartheta_1\vartheta_2\cdots\vartheta_{n-2}}(w[1]w[2]\cdots w[n-2])$.

Notice that $|u| - |y| \leq g$. Otherwise, there would exist a $\overline{\vartheta_n}$-palindromic prefix between y and u, namely the suffix of v of length $|v| - 3g$ by Lemma 11, contradicting the fact that $(\vartheta_1\vartheta_2\cdots\vartheta_{n-1}, w[1]w[2]\cdots w[n-1])$ is normalized. Let $g' = |u| - |y|$. If $g' = g$, then $\vartheta_{n-2} = \overline{\vartheta_n}$ and $c = \overline{b}$, i.e. s ends with $(ERR, ab\overline{b})$ or $(REE, ab\overline{b})$. By Lemma 25, we know that s is not normalized, and this corresponds to case (iv). It remains to consider the case $g' < g$. We show in the next paragraphs that this implies case (iii).

First, we show that $|y| < g$. Proceeding by contradiction, assume that $|y| \geq g$. This implies $|u| \geq g + g'$. Since g is an σ_E-period of v and in particular an σ_E-period of u and, by Lemma 17, g' is a $(\sigma_{\vartheta_{n-2}} \circ \sigma_{\vartheta_n})$-period of u, it follows from Theorem 14 that $g'' = \gcd(g, g')$ is an σ_E-period of u. Moreover, that σ_E-period g'' propagates to the whole word v in virtue of Lemma 19 applies, yielding a contradiction: this would imply that there is a ϑ_n-palindrome between the ϑ_n-palindromes u and v, namely the prefix of v of length $|v| - 2g''$ (by Lemma 11). But $g'' < g$ (since $g' < g$). Hence, the claim $|y| < g$ is proved.

Now, we prove that $|v| = 4g - 2$. Recall that by definition of pseudopalindromic closure, $|v| \leq 2|u| + 2$. Moreover, $|v| = |u| + 2g$, $|y| \leq g - 1$ and $g' = |u| - |y| \leq g - 1$. On the first hand, we have $|v| \leq 2|u| + 2 = 2|v| - 4g + 2$ which implies $|v| \geq 4g - 2$. On the other hand, $|v| = |u| + 2g \leq |y| + g - 1 + 2g \leq g - 1 + g - 1 + 2g = 4g - 2$, thus $|v| = 4g - 2$. In particular, in virtue of the equalities and inequalities $|v| = 4g - 2$, $|v| - |u| = 2g$, $|u| - |y| \leq g - 1$ and $|y| \leq g - 1$, we have $|y| = g - 1$, $|u| = 2g - 2$ and $|v| = 2|u| + 2$. Hence, $\vartheta_n = E$ and $g' = |u| - |y| = g - 1$.

Finally, notice that, since $|v| \geq p + q = 3g$, the prefix z of v of length $|v| - 3g$ is a $\overline{\vartheta_n}$-palindrome, i.e. an R-palindrome, by Lemma 11. But Lemma 17 implies that $|y| - |z| = g - g' = 1$ is an $(\sigma_{\vartheta_{n-2}} \circ \sigma_R)$-period of y, i.e. an $\sigma_{\vartheta_{n-2}}$-period of y. If $\vartheta_{n-2} = R$ and since y has length $g - 1$, contains the letter $\overline{b}$ and $\vartheta_1 = R$, we find $y = \overline{b}^{g-1}$, so that $u = \overline{b}^{g-1} b^{g-1}$ and $v = \overline{b}^{g-1} b^g \overline{b}^g b^{g-1}$, which corresponds to case (iii). It only remains to consider the case $\vartheta_{n-2} = E$. Then $y = (d\overline{d})^{(g-1)/2}$ for some letter d, since 1 is a σ_E-period of y and $|y| = g - 1$. But $\vartheta_n = \vartheta_{n-1} = E$ and $(\vartheta_1\vartheta_2 \cdots \vartheta_{n-1}, w[1]w[2] \cdots w[n-1])$ is normalized, which implies that $w[n-1] \neq d$ (otherwise the palindrome $(d\overline{d})^{(g-1)/2}d$ would be missed). Therefore, $u = (d\overline{d})^{(g-1)/2}(\overline{d}d)^{(g-1)/2}$. Finally, notice that the condition $|v| = 2|u| + 2$ implies that the longest σ_E-palindromic suffix of $uw[n] = \varepsilon$. This is impossible since $d\overline{d}$ is a σ_E-palindromic suffix of $u\overline{d}$ and $\overline{d}\overline{d}(d\overline{d})^{(g-1)/2-1}dd$ is a σ_E-palindromic suffix of ud. $\square$

The following theorem explains how to replace the forbidden factors in order to normalize a directive bi-sequence.

Theorem 27. *Let (Θ, w) be a directive bi-sequence, with Θ a finite or infinite sequence of involutory antimorphisms and w a binary word having same length as Θ. Then there exists a normalized directive bi-sequence (Θ', w') such that $\psi_\Theta(w) = \psi_{\Theta'}(w')$. Moreover, in order to get the normalized directive bi-sequence (Θ', w') from (Θ, w), it is sufficient to replace the prefix (if it is of one of the following forms):*

(i) $(RR, a\overline{a})$ *by* $(RER, a\overline{a}a)$*;*

(ii) $(R^{i-1}E, a^i)$ *by* $(R^iE, a^i\overline{a})$*;*

(iii) $(R^iEE, a^i\overline{a}\,\overline{a})$ *by* $(R^iERE, a^i\overline{a}\,\overline{a}a)$*;*

for $i \geq 1$ and then, to replace from left to right any factor

(iv) $(\vartheta\overline{\vartheta}\overline{\vartheta}, ab\overline{b})$ *by* $(\vartheta\overline{\vartheta}\vartheta\overline{\vartheta}, ab\overline{b}b)$*,*

where $\vartheta \in \{R, E\}$ and $a, b \in \{0, 1\}$.

Proof. Proposition 26 indicates precisely which prefixes and factors cannot occur in a normalized directive bi-sequence, while Lemma 25 tells us how to replace any factor of the form $(\vartheta\overline{\vartheta}\overline{\vartheta}, ab\overline{b})$ in order to normalize the directive bi-sequence. It remains to prove how to correct the non-normalized prefixes of the form (i), (ii) and (iii). By Proposition 26, we know that the prefixes $(RER, a\overline{a}a)$, $(R^iE, a^i\overline{a})$ and $(R^iERE, a^i\overline{a}\,\overline{a}a)$ are normalized, since they are not in the set of forbidden prefixes and factors. In order to conclude the proof, we let the reader verify that $\psi_{RR}(a\overline{a}) = \psi_{RER}(a\overline{a}a)$, $\psi_{R^{i-1}E}(a^i) = \psi_{R^iE}(a^i\overline{a})$ and $\psi_{R^iEE}(a^i\overline{a}\,\overline{a}) = \psi_{R^iERE}(a^i\overline{a}\,\overline{a}a)$. $\square$

Example 28. Let us normalize the directive bi-sequence $d = (RRR, 011)$. Since it has a prefix of the form (i) in Theorem 27, we rewrite d as $d' = (RER \cdot R, 010 \cdot 1)$, with $(RER, 010)$ normalized. The second step is to replace all the factors of the form $(\vartheta\overline{\vartheta\vartheta}, ab\bar{b})$ by $(\vartheta\bar{\vartheta}\vartheta\bar{\vartheta}, ab\bar{b}b)$. There is only one factor of this form: $(ERR, 101)$. We then obtain the new directive bi-sequence $d'' = (RERER, 01010)$, which is normalized. Indeed, one can verify that d'' does not contain any forbidden prefix or factor. Finally, d and d'' direct the same generalized pseudostandard word:

$$\psi_{RRR}(011) = \underline{0}\,\underline{1}0\underline{1}0 \quad \text{and} \quad \psi_{RERER}(01010) = \underline{0}\underset{\cdot}{1}\underline{0}\underset{\cdot}{1}\underline{0}.$$

5 A generalization of Justin's formula

The previous section gives us the main tool to prove Theorem 29, that is the existence of the normalization of a directive bi-sequence: Given a generalized pseudostandard word, it is always possible to find a directive bi-sequence that describes all its pseudopalindromic prefixes.

As pointed out in previous sections, the naive way to compute $\psi_{\Theta R}(wa)$ (resp., $\psi_{\Theta E}(wa)$) when $\psi_\Theta(w)$ is known, is to find the longest palindromic (resp., E-palindromic) suffix p of $\psi_\Theta(w)$ preceded by a. Then $\psi_{\Theta R}(wa) = \psi_\Theta(w)p^{-1}\psi_\Theta(w)$ (resp., $\psi_{\Theta E}(wa) = \psi_\Theta(w)p^{-1}\overline{\psi_\Theta(w)}$). However, this turns out to be very costly since at each step, one must find the longest pseudopalindromic suffix of words that grow more and more in size. We are now ready to state and prove one of the main theorem of this paper, thus providing an efficient way to compute binary generalized pseudostandard words.

Theorem 29. *Let (Θ, w) be a normalized finite directive bi-sequence of length n on a binary alphabet and, for $i = 1, 2, \ldots, n$, let $\psi_i = \psi_{\vartheta_1\vartheta_2\cdots\vartheta_i}(w\,[1 \ldots i])$, $\psi_0 = \varepsilon$ and α_i be the last letter of ψ_i for $1 \leq i \leq n$.*

(i) *If $|w\,[1]w\,[2] \cdots w\,[n-1]|_{w\,[n]} = 0$ or $|\vartheta_1\vartheta_2 \cdots \vartheta_{n-1}|_{\vartheta_n} = 0$, then*

$$\psi_n = \begin{cases} \psi_{n-1}w\,[n]\psi_{n-1}, & \text{if } \vartheta_n = R; \\ \psi_{n-1}\overline{\psi_{n-1}}, & \text{if } \vartheta_n = E \text{ and } \alpha_{n-1} \neq w\,[n]; \\ \psi_{n-1}w\,[n]\overline{w\,[n]}\,\overline{\psi_{n-1}}, & \text{if } \vartheta_n = E \text{ and } \alpha_{n-1} = w\,[n]. \end{cases}$$

(ii) *If one can write $(\Theta, w) = (\Theta'\vartheta_n\Theta'', w'(\vartheta_{n-1} \circ \vartheta_n)(w\,[n])w'')$ such that $i := |w'| = |\Theta'| + 1$ with $|\Theta'|$ maximum, then*

$$\psi_n = \psi_{n-1}(\vartheta_{n-1} \circ \vartheta_n)\left(\psi_i^{-1}\psi_{n-1}\right).$$

(iii) *Otherwise,*

$$\psi_n = \begin{cases} \psi_{n-1}\vartheta_n(\psi_{n-1}), & \text{if } \vartheta_n = R \text{ or } \alpha_{n-1} \neq w\,[n]; \\ \psi_{n-1}w\,[n]\overline{w\,[n]}\,\overline{\psi_{n-1}}, & \text{if } \vartheta_n = E \text{ and } \alpha_{n-1} = w\,[n]. \end{cases}$$

Proof. (i) Assume first that $w = a^{n-1}\overline{a}$ for some $a \in \mathcal{A}$. Then $\Theta \in \{R^n, R^{n-1}E\}$, otherwise (Θ, w) would not be normalized by Lemma 23. The result follows according to the value of Θ_n, w_n and α_n. If $|\vartheta_1\vartheta_2\cdots\vartheta_{n-1}|_{\vartheta_n} = 0$ and $|w[1]w[2]\cdots w[n-1]|_{\overline{w[n]}} \neq 0$, again by Lemma 23, since (Θ, w) is normalized, we know that $\vartheta_1 = R$ so that $\vartheta_n = E$, except if $n = 1$. But since no E-palindrome occurs as a prefix, and then as a suffix of ψ_{n-1}, we deduce that the longest E-palindromic suffix of $\psi_{\vartheta_1\vartheta_2\cdots\vartheta_{n-1}}(w[1]w[2]\cdots w[n-1])w[n]$ is either ε or $\alpha_{n-1}\overline{\alpha_{n-1}}$ and the result follows.

(ii) By hypothesis, there exists a ϑ_n-palindromic prefix ψ_i of ψ_{n-1} followed by the letter $(\vartheta_{n-1} \circ \vartheta_n)(w[n])$. Moreover, since (Θ, w) is normalized, $|\Theta'|$ is maximum and by its construction, ψ_i is exactly the longest ϑ_n-palindromic prefix of ψ_{n-1} followed by the letter $(\vartheta_{n-1}\circ\vartheta_n)(w[n])$. But ψ_{n-1} is a ϑ_{n-1}-palindrome, so that $\vartheta_{n-1}(\psi_i)$ is the longest $(\vartheta_{n-1}\circ\vartheta_n)$-palindromic suffix of ψ_{n-1} preceded by $\vartheta_n(w[n])$. Then $s = \vartheta_n(w[n])\vartheta_{n-1}(\psi_i)w[n]$ is the longest ϑ_n-palindromic suffix of $\psi_{n-1}w[n]$. Therefore,

$$\begin{aligned}
\psi_n &= \psi_{n-1}w[n](s^{-1})\vartheta_n(w[n])\vartheta_n(\psi_{n-1}) \\
&= \psi_{n-1}w[n](\vartheta_n(w[n])\vartheta_{n-1}(\psi_i)w[n])^{-1}\vartheta_n(w[n])\vartheta_n(\psi_{n-1}) && (2)\\
&= \psi_{n-1}w[n]w[n]^{-1}\vartheta_{n-1}(\psi_i)^{-1}\vartheta_n(w[n])^{-1}\vartheta_n(w[n])\vartheta_n(\psi_{n-1}) && (3)\\
&= \psi_{n-1}\vartheta_{n-1}(\psi_i)^{-1}\vartheta_n(\psi_{n-1}) && (4)\\
&= \psi_{n-1}(\vartheta_{n-1}\circ\vartheta_n)(\psi_i^{-1}\psi_{n-1}), && (5)
\end{aligned}$$

which concludes this part. Notice that Equation (3) is obtained from Equation (2), using the fact that for any word $u, v \in \mathcal{A}^*$, $(uv)^{-1} = v^{-1}u^{-1}$, and that Equation (5) follows from Equation (4), since ψ_i is a ϑ_n-palindrome and ψ_{n-1} is a ϑ_{n-1}-palindrome.

(iii) Since the hypothesis of cases (i) and (ii) are not satisfied, we can assume here that ψ_{n-1} does not have a nonempty ϑ_n-palindromic suffix that is preceded by the letter $\vartheta_n(w[n])$ and that $|w[1]w[2]\cdots w[n-1]|_{\overline{w[n]}} \geq 1$. Let us first suppose that $\vartheta_n = R$. Then $w[n] = \alpha_{n-1}$, otherwise we contradict the hypothesis, since it implies that ψ_{n-1} necessarily has a palindromic suffix that is preceded by the letter $\vartheta_n(w[n]) = w[n]$, namely a suffix of the form $w[n]\overline{w[n]}^i$. Thus, $\vartheta_n = R$ implies $w[n] = \alpha_{n-1}$ and consequently, $\psi_n = \psi_{n-1}R(\psi_{n-1})$. Let us now suppose that $\vartheta_n = E$. If $w[n] = \alpha_{n-1}$, one deduces that $\psi_n = \psi_{n-1}w[n]\overline{w[n]}\,\overline{\psi_{n-1}}$, while if $w[n] \neq \alpha_{n-1}$, one obtains $\psi_n = \psi_{n-1}\overline{\psi_{n-1}}$. Combining all those cases yields the statement. $\square$

It is worth mentioning that Proposition 3 is a special case of Theorem 29. Indeed, if $\Theta = R^n$, then (Θ, w) directs a standard Sturmian sequence and we retrieve case (i) if $w = a^{n-1}\overline{a}$, for $a \in \{0, 1\}$ and then, $\psi_n = \psi_{n-1}w[n]\psi_{n-1}$. Otherwise, case (ii) applies and we get $\psi_n = \psi_{n-1}\psi_i^{-1}\psi_{n-1}$.

On the other hand, if $\Theta = E^n$, the directive bi-sequence (E^n, w) cannot be a normalized one, by Lemma 23. In this case, we can use the same idea as in the proof of Theorem 29 in order to get the following generalization of Justin's formula for E-standard words.

Proposition 30. *Let (E^{n+1}, wa) be a directive bi-sequence of a E-standard word, with $w \in \{0,1\}^n$. If w is not a-free, then we write $w = v_1 a v_2$ with v_2 a-free and we have*

$$\psi(wa) = \begin{cases} \psi(w)a\overline{a}E(\psi(w)), & \text{if } w \text{ is } a\text{-free;} \\ \psi(w)\psi(v_1)^{-1}E(\psi(w)), & \text{otherwise.} \end{cases}$$

Notice that the algorithms of normalization and of computation of generalized pseudostandard words have been implemented in Python by the first author and should be included soon in the Sage words library [16].

To conclude this section, let us see how Theorem 29 may be used in order to construct the Thue-Morse word.

Example 31. Theorem 7 tells us that the Thue-Morse word is a generalized pseudostandard word $T = \psi_{(ER)^\omega}(01^\omega)$. In order to apply Theorem 29 to the construction of T, we have to normalize the directive bi-sequence $d = ((ER)^\omega, 01^\omega)$. Using Theorem 27, we get $d' = ((RE)^\omega, 01^\omega)$, which is normalized. Theorem 29 yields the successive prefixes t_i of T, with $t_0 = \varepsilon$:

$$\begin{aligned}
t_1 &= 0; \\
t_2 &= t_1\overline{t_1} = 01, \text{by Theorem 29 (i), second case;} \\
t_3 &= t_2 E(t_0^{-1}t_2) = 01E(01) = 0110, \text{by Theorem 29 (ii);} \\
t_4 &= t_3 E(t_3) = 01101001, \text{by Theorem 29 (iii), first case;} \\
t_5 &= t_4 E(t_4) = 0110100101101001, \text{by Theorem 29 (iii), first case}
\end{aligned}$$

and so on, which corresponds to the usual construction of the Thue-Morse word.

6 Rote words

This section is devoted to the study of Rote words obtained by iterated pseudopalindromic closure. They deserve some attention since they provide a natural characterization of the palindromic prefixes in standard Sturmian words. Moreover, their corresponding normalized bi-sequences are easy to characterize. The so-called *Rote words* or *complementary-symmetric words* are sequences of letters having complexity $2n$ and such that their language is closed under the complementation operator. Let w be a binary word on $\{0, 1\}$. The *difference of* w, denoted by $\Delta(w)$, is the word $v = v_1v_2 \cdots v_{|w|-1}$ defined by

$$v_i = (w_{i+1} - w_i) \bmod 2, \quad \text{for } i = 1, 2, \ldots, |w| - 2.$$

Complementary-symmetric words are connected to Sturmian words by a structural theorem.

Theorem 32 (Rote [15])**.** *An infinite word* $\mathbf{w}$ *is a complementary-symmetric Rote word if and only if the infinite word* $\Delta(\mathbf{w})$ *is a Sturmian word.*

We say that a complementary-symmetric words $\mathbf{r}$ is a *standard Rote word* if both $0\mathbf{r}$ and $1\mathbf{r}$ are complementary-symmetric words. Equivalently, a word $\mathbf{r}$ is standard Rote if and only if $\Delta(\mathbf{r})$ is standard Sturmian.

The aim of this section is to provide an explicit construction of standard Rote words by iterated pseudopalindromic closure. The key idea is to exploit the link with Sturmian words by looking at the palindrome and antipalindrome prefixes of the Rote word. First, we state without proof some elementary properties of the operator Δ.

Lemma 33. *Let* $u, v \in \Sigma^*$*, where* $|u|, |v| \geq 2$*. Then*

(i) $\Delta(u) = \Delta(v)$ *if and only if* $v = u$ *or* $v = \overline{u}$,

(ii) u *is either a palindrome or an antipalindrome if and only if* $\Delta(u)$ *is a palindrome and*

(iii) u *is an antipalindrome if and only if* $\Delta(u)$ *is an odd palindrome with central letter* 1.

Now, we study the palindrome prefixes of standard Sturmian words.
For instance, consider the Fibonacci word:

$$\mathbf{f} = 0100101001001010010 10\cdots$$

Its palindrome prefixes are:

$$\varepsilon, 0, 010, 010010, 01001010010, \ldots$$

We may divide them into three categories

(i) palindromes of even length,

(ii) palindromes of odd length with central letter 0 and

(iii) palindromes of odd length with central letter 1.

Thus, we consider a three-letters alphabet $T = \{\text{A}, \text{E}, \text{O}\}$ and we define a map $\theta : \mathcal{A}^* \to T^*$ by $\theta(\varepsilon) = \text{E}$ and, for $w \in \mathcal{A}^*$ and $\alpha \in \mathcal{A}$,

$$\theta(w\alpha) = \theta(w) \cdot \begin{cases} \text{E}, & \text{if } \psi(w\alpha) \text{ is an even palindrome and;} \\ \text{A}, & \text{if } \psi(w\alpha) \text{ is a palindrome with central letter } 1; \\ \text{O}, & \text{if } \psi(w\alpha) \text{ is a palindrome with central letter } 0. \end{cases}$$

We call $\theta(w)$ the *palindrome type word* of w.

Example 34. One may verify that the Fibonacci word $\mathbf{f}$ satisfies $\theta(\mathbf{f}) = (\text{EOA})^\omega$.

The next proposition establishes an important link between w and $\theta(w)$.

Proposition 35. *We have* $\theta(0) = \text{EO}$ *and* $\theta(1) = \text{EA}$. *Let* $c, d \in \mathcal{A}$, $u \in \mathcal{A}^*$ *and* $\theta(uc) = x\alpha\beta$ *for* $x \in T^*$, $\alpha, \beta \in T$. *Then the two last letters of* $\theta(ucd)$ *are distinct and*

$$\theta(ucd) = \begin{cases} x\alpha\beta\alpha, & \text{if } c = d; \\ x\alpha\beta\gamma, & \text{if } c \neq d. \end{cases} \tag{6}$$

where γ *is the unique letter distinct from* α *and* β.

Proof. The proof is done by induction on $|u|$. Clearly, the two palindrome prefixes of $\psi(0) = 0$ are $\varepsilon, 0$, so that $\theta(0) = \text{EO}$. Similarly, one notices that $\theta(1) = \text{EA}$. Now, there are three cases to consider:

(i) Suppose that $c = d$. Then $\psi(ucd) = \psi(uc)\psi(u)^{-1}\psi(uc)$ by Proposition 3. Therefore, $|\psi(ucd)|$ and $|\psi(u)|$ have same parity. If they are even, then $\theta(ucd)$ ends with $\textsc{e}\beta\textsc{e}$. If they are odd, they share the same central letter, so that $\theta(ucd)$ ends with $\alpha\beta\alpha$, where $\alpha \in \{\textsc{o}, \textsc{a}\}$. In both cases, by the induction hypothesis, the two last letters of $\theta(uc)$ are distinct, and so are the two last letters of $\theta(ucd)$.

(ii) Suppose that $c \neq d$ and $|uc|_d = 0$. Then $\psi(ucd) = \psi(uc)d\psi(uc)$ by Proposition 3. In particular $|\psi(ucd)|$ is odd. Moreover, uc is the power of a letter, which means that the previous palindrome prefixes are all powers of the same letter. Hence, one of the two previous palindrome prefixes is of odd length while the other is of even length, so that $\alpha \neq \beta$. Finally, $|\psi(ucd)|$ is of odd length and has central letter different from all shorter odd palindrome prefixes. Hence, $\gamma \neq \alpha, \beta$, as desired.

(iii) Suppose that $c \neq d$ and $|uc|_d > 0$. Write $u = u_1 dc^k$ with $u_1 \in \mathcal{A}^*$ and k a non negative integer, so that $ucd = u_1 dc^{k+1} d$. Again by Proposition 3, one deduces $\psi(ucd) = \psi(uc)\psi(u_1)^{-1}\psi(uc)$, which means that $\psi(ucd)$ and $\psi(u_1)$ share the same parity and, if odd, the same central letter. Hence, it suffices to show that $\theta(u_1)$ ends with γ.

It follows from (i) that $\theta(uc)$ ends with $(\alpha\beta)^{k/2}$ if k is even and with $\beta(\alpha\beta)^{k/2}$ otherwise. Moreover, by the induction hypothesis $\theta(uc)$ ends with $\gamma(\alpha\beta)^{k/2}$ if k is even, or with $\gamma\beta(\alpha\beta)^{k/2}$ otherwise. In both cases, this implies that $\theta(u_1)$ ends with γ. But $\psi(u_1)$ and $\psi(ucd)$ are of the same type, so that $\theta(u_1)$ ends with γ, as desired. In particular, β and γ are distinct.

The values of $\theta(uc)$ are represented in Figure 3 for short words. □

Example 36. Consider once again the Fibonacci word on $\{0, 1\}$

$$\mathbf{f} = 0100101001001010010\cdots$$

and the Rote word $\mathbf{r}$ starting with 0 such that $\Delta(\mathbf{r}) = \mathbf{f}$

$$\mathbf{r} = 00111001110001100011\cdots$$

By inspection, we may enumerate the palindromic prefixes of $\mathbf{f}$, which are in bijections with the palindromic and antipalindromic prefixes of $\mathbf{r}$, except for the empty word prefix of $\mathbf{r}$

$\mathbf{f}$	$\mathbf{r}$
	ε
ε	0
0	00
010	0011
010010	0011100
01001010010	001110011100
$\cdots$	$\cdots$

One may verify that $\mathbf{r}$ is indeed a generalized pseudostandard word:

$$\mathbf{r} = \underline{0}\,\underline{0}\underset{\cdot}{1}1\underline{1}00\underline{1}1100\underset{\cdot}{0}1100011\cdots$$

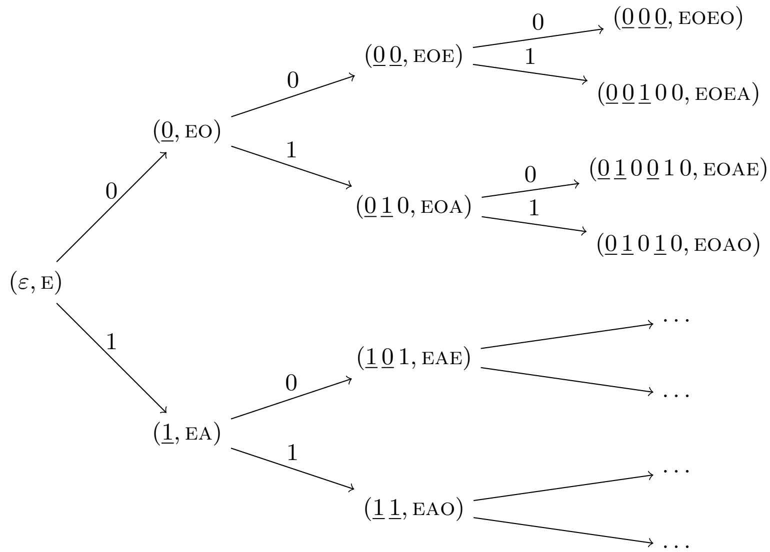

Figure 3: Representation of Proposition 35 by a tree, describing the possible palindrome type sequences for any finite binary word. Each node is a couple $(\psi(w), \theta(w))$, where w is a binary word. For instance, if $w = 010$, then its corresponding node is $(\underline{0}\,\underline{1}\,0\,\underline{0}\,1\,0, \textsc{eoae})$ and one verifies that the palindrome prefixes ε, 0, 010 and 010010 of $\psi(w)$ are indeed of types E, O, A and E.

From now on, we consider only standard Rote word starting with 0. Since the complement of a Rote word is also a Rote word, it is easy to extend the results to standard Rote words starting with 1, but for sake of simplicity, we restrict our study by fixing the first letter as 0. Observations found in Example 36 lead naturally to the following statement:

Lemma 37. *Let* $\mathbf{s}$ *be a standard Sturmian word and* $\mathbf{r}$ *be the standard Rote word starting with* 0 *such that* $\Delta(\mathbf{r}) = \mathbf{s}$. *Let* $p_0 = \varepsilon$, p_1, p_2, $\ldots$ *be the palindrome prefixes of* $\mathbf{s}$, *enumerated with increasing length, and* $q_0 = \varepsilon$, q_1, q_2, $\ldots$ *be the pseudopalindrome prefixes of* $\mathbf{r}$, *enumerated with increasing length as well. Then* $\Delta(q_{i+1}) = p_i$ *for all integers* $i \geq 0$.

Proof. First, notice that since $\mathbf{s}$ has infinitely many palindrome prefixes, then $\mathbf{r}$ has infinitely many pseudopalindrome prefixes, in virtue of Lemma 33(ii). We proceed by contradiction. Let $i \geq 0$ be the smallest integer such that $\Delta(q_{i+1}) \neq p_i$. We know from Lemma 33(ii) that $\Delta(q_{i+1})$ is a palindrome. Hence, it means that $\Delta(q_{i+1}) = p_j$ for some integer $j > i$. This is impossible since the prefix q of $\mathbf{r}$ such that $\Delta(q) = p_i$ is a pseudopalindrome, again by Lemma 33(ii), i.e. there would exist a pseudopalindrome q having length between $|q_i|$ and $|q_{i+1}|$. □

Therefore, standard complementary-symmetric words are generalized pseudostandard words.

Lemma 38. *Let* $\mathbf{s}$ *be a standard Sturmian word directed by some infinite binary sequence* $\mathbf{x}$ *and* $\mathbf{r}$ *be the standard Rote words starting with* 0 *such that* $\Delta(\mathbf{r}) = \mathbf{s}$. *Then there exists an infinite bi-sequence* $(\mathbf{\Theta}, \mathbf{w})$ *such that* $\mathbf{r} = \psi_{\mathbf{\Theta}}(\mathbf{w})$, *i.e.* $\mathbf{r}$ *is a generalized pseudostandard word.*

Proof. As in Lemma 37, let $p_0 = \varepsilon$, p_1, p_2, p_3, … be the palindrome prefixes of $\mathbf{s}$ and $q_0 = \varepsilon$, q_1, q_2, q_3, …, be the list of pseudopalindrome prefixes of $\mathbf{r}$ enumerated with increasing length. For $i = 0, 1, 2, \ldots$, let $w_{i+1} = \mathbf{r}[|q_i| + 1]$, i.e. w_{i+1} is the letter following the pseudopalindrome prefix q_i. Finally, let

$$\vartheta_i = \begin{cases} R, & \text{if } q_i \text{ is an } R\text{-palindrome;} \\ E, & \text{if } q_i \text{ is an } E\text{-palindrome.} \end{cases}$$

We show that $(\mathbf{\Theta}, \mathbf{w}) = (\vartheta_1\vartheta_2\vartheta_3\cdots, w_1w_2w_3\cdots)$ is a directive bi-sequence of $\mathbf{r}$. Let $\psi_i = \psi_{\vartheta_1\vartheta_2\cdots\vartheta_i}(w_1w_2\cdots w_i)$ for $i = 1, 2, \ldots$.

It suffices to prove that $q_{i+1} = \psi_{i+1}$ for $i = 1, 2, \ldots$. Arguing by contradiction, suppose that i is the smallest integer such that $q_i = \psi_i$ but $q_{i+1} \neq \psi_{i+1}$. Clearly, q_{i+1} and ψ_{i+1} are both ϑ_{i+1}-palindromes and share the prefix q_iw_{i+1}. Therefore, $|\psi_{i+1}| < |q_{i+1}|$, otherwise ψ_{i+1} would not be the shortest ϑ_{i+1}-palindrome having $q_iw_{i+1} = \psi_iw_{i+1}$. By Lemma 38, we have $\Delta(q_{i+1}) = p_i$ which implies that $\Delta(q_iw_{i+1}) = p_{i-1}a$, with $a \in \mathcal{A}$, is a prefix of both p_i and $\Delta(\psi_{i+1})$. But $\Delta(\psi_{i+1})$ is a palindrome and $|\Delta(\psi_{i+1})| < |p_i|$, contradicting the fact that p_i is the shortest palindrome having prefix $p_{i-1}a$. □

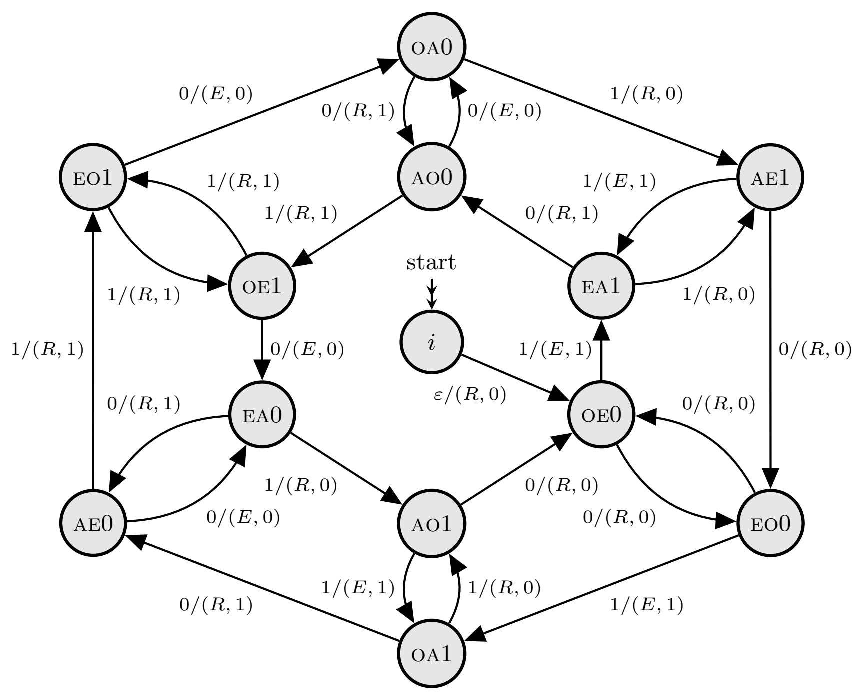

Figure 4: Transducer computing the directive bi-sequence of a Rote standard word $\mathbf{r}$ starting with 0 from the directive sequence of a Sturmian standard word $\mathbf{s}$ such that $\Delta(\mathbf{r}) = \mathbf{s}$.

Consider the transducer $\mathcal{T}$ on $\{0,1\} \times (\{R, E\} \times \{0,1\})$ represented in Figure 4. The next lemma states some observations about $\mathcal{T}$.

Lemma 39. *Let $x = x_1x_2\cdots x_n$ be an input word of length $n \geq 1$ of $\mathcal{T}$ and*

$$(\vartheta_1\vartheta_2\cdots\vartheta_n, y_1y_2\cdots y_n)$$

be its associated output word. Let $q = ab\alpha \neq i, q' = cd\beta$ be the two last states visited when reading x. Then $c = b$, $\beta = x_n$ and $\theta(x)$ ends with cd.

Proof. The proof is done by induction on n. For $n = 1$, we have $q = \text{OE}0$. There are two cases to consider according to the value of x_1. If $x_1 = 0$, then $q' = \text{EO}0$ so that $c = b = \text{E}$, $\beta = x_1 = 0$ and $\theta(x) = \theta(0) = \text{EO}$. On the other hand, if $x_1 = 1$, then $q' = \text{EA}1$ so that $c = b = \text{E}$, $\beta = x_1 = 1$ and $\theta(x) = \theta(1) = \text{EA}$.

Consider now the general case. By inspection of $\mathcal{T}$, one observes that $c = b$ and $\beta = x_i$ for any transition, except the one starting with the initial state i. The fact that $\theta(x)$ ends with cd follows from inspection of $\mathcal{T}$ and Proposition 35. □

We are now ready to show the main theorem of this section. The key idea is to observe that one may derive the directive bi-sequence from the directive sequence of the Sturmian word by looking at the palindrome types (A, E or O) at each step.

Theorem 40. *Let $\mathbf{x}$ be an infinite binary sequence directing some Sturmian word $\mathbf{s}$ and let $\mathbf{r}$ be the Rote word starting with 0 and such that $\Delta(\mathbf{r}) = \mathbf{s}$. Then the output word $(\Theta, \mathbf{y})$ obtained from $\mathbf{x}$ in the transducer $\mathcal{T}$ is a directive bi-sequence of $\mathbf{r}$.*

Proof. We know from Lemma 37 that the palindrome prefixes of $\mathbf{s}$ and the pseudopalindromic prefixes of $\mathbf{r}$ are in 1-to-1 correspondence. It only remains to describe the letters and type of pseudopalindromes involved in the iterated pseudopalindromic closure.

For every positive integer n, let $x = x_1 x_2 \cdots x_n$ and $(\Theta, \mathbf{y}) = (\vartheta_1 \vartheta_2 \cdots \vartheta_n, y_1 y_2 \cdots y_n)$. Moreover, let $q = ab\alpha$ and $q' = cd\beta$ be the two last states visited when reading x. We prove by induction on n that $(\Theta, \mathbf{y})$ is the output word obtained from $\mathcal{T}$ by reading x. This is clear for $n = 1$.

For the general case, we know from Lemma 39 that q' remembers the two last palindrome type encountered in $\psi(x)$, i.e. $\theta(x)$ ends with ab. Moreover, $c = b$ and $\beta = x_n$. Next, observe that since $\psi(x)$ is a palindrome of type $c = b$, we deduce by inspection of $\mathcal{T}$ that $\psi_{\vartheta_1 \vartheta_2 \cdots \vartheta_n}(y_1 y_2 \cdots y_n)$ is a R-palindrome if $c = b \in \{\text{O}, \text{E}\}$ and is a E-palindrome if $c = b = \text{A}$. Finally, one verifies in $\mathcal{T}$ that $y_{n+1} = x_n$ if $c = b \in \{\text{O}, \text{E}\}$ and $y_{n+1} = (x_n + 1) \bmod 2$ if $c = b = \text{A}$.

□

Example 41. Let $x = 001101$. When reading x in $\mathcal{T}$, one visits states i, OE0, EO0, OE0, EA1, AE1, EO0 and OA1 and obtains the output word

$$(\vartheta, y) = (RRRERRE, 0001001).$$

Moreover,

$$\begin{aligned} \psi(x) &= \underline{0}\,\underline{0}\,\underline{1}00\underline{1}00\underline{0}100100\underline{1}000100100 \\ \psi_\vartheta(y) &= \underline{0}\,\underline{0}\,\underline{0}\underset{\cdot}{1}11\underline{0}00\underline{0}111000\underset{\cdot}{1}111000111 \end{aligned}$$

and we indeed have $\Delta(\psi_\vartheta(y)) = \psi(x)$.

As a consequence, we have a complete characterization of standard Rote words obtained from iterated pseudopalindromic closure.

Corollary 42. *Let (Θ, w) be a directive bi-sequence. Then $\psi_\Theta(w)$ is a standard Rote word if and only if no factor of length 2 of (Θ, w) is in the set*

$$D = \{(EE, ab) \mid a, b \in \mathcal{A}\} \cup \{(RR, a\overline{a}) \mid a \in \mathcal{A}\} \cup \{(RE, aa) \mid a \in \mathcal{A}\}.$$

Moreover, (Θ, w) is normalized.

Proof. If $\psi_\Theta(w)$ is a standard Rote word, then it must be obtained from $\mathcal{T}$. But every path of length 2 in $\mathcal{T}$ yields an output word in

$$\begin{array}{cccc} \{(RR, 00), & (RR, 11), & (RE, 01), & (RE, 10), \\ (ER, 00), & (ER, 01), & (ER, 10), & (ER, 11)\}. \end{array}$$

Conversely, if (Θ, w) contains a factor in D, then it cannot be obtained from $\mathcal{T}$. Hence, the first claim is proved. The fact that (Θ, w) is normalized follows directly from Proposition 26 since all forbidden factors for a bi-sequence do not appear in any output word of $\mathcal{T}$. □

7 Concluding remarks

Theorem 29 does not provide the most efficient way of computing the Thue-Morse word. However, it may be very useful for large directive bi-sequences having less regularity. It is also a great tool, combined with the normalization, to compute a generalized pseudostandard word over a two-letter alphabet.

Algorithm 2 takes as input any directive bi-sequence of length n for a two-letter alphabet, normalized or not, and computes, in linear time with respect to the length of the directive bi-sequence, a prefix of the corresponding generalized pseudostandard word, using both the normalization and the generalization of Justin's formula (Theorem 29) for generalized pseudostandard words.

The main results of this paper are the generalization of Justin's formula for generalized pseudostandard words (Theorem 29) over a two-letter alphabet and the characterization of the directive bi-sequences of standard Rote words. They should be of considerable help for future investigations about generalized pseudostandard words. It shall be noted that some stated lemmas could be generalized for words over an alphabet with three or more letters, but it remains an interesting open problem to generalize all of them.

Another topic of interest would be to compute the maximum complexity of generalized pseudostandard words. Since they include Sturmian words, which have complexity $n + 1$, Rote words of complexity $2n$ and the Thue-Morse word, whose complexity oscillates around $3n$, it seems reasonable to conjecture that it is at most linear. Indeed, empirical observation suggests that, over binary alphabets, no word has a complexity greater than $4n$:

Conjecture 43. Let $\mathbf{w}$ be a generalized pseudostandard words over a binary alphabet. Then there exists some integer n_0 such that $f_{\mathbf{w}}(n) \leq 4n$, for all $n \geq n_0$.

Interestingly, the bi-sequence yielding the highest complexity we have been able to provide, so far, is constructed only from forbidden factors as detailed in Proposition 26. For instance, in Sage, we may compute the number of factors of each length between 1 and 300 for the prefix of length 5 000 of the word $\psi_{(REEEEEERRRE)^\omega}((0001101011)^\omega)$:

Algorithm 2 Computation of a finite generalized pseudostandard word.

function ITERATEDPSEUDOPALINDROMICCLOSURE(Θ, w)

  if $(RR, a\overline{a})$ is prefix of (Θ, w) **then**

    $\Theta \leftarrow RER(RR)^{-1}\Theta, w \leftarrow a\overline{a}a(a\overline{a})^{-1}w$

  else if $(R^{i-1}E, a^i)$ is prefix of (Θ, w), for some integer $i \geq 1$ **then**

    $\Theta \leftarrow R\Theta, w \leftarrow a^i\overline{a}a^{-i}w$

  else if $(R^iEE, a^i\overline{a}\,\overline{a})$ is prefix of (Θ, w) for some integer $i \geq 1$ **then**

    $\Theta \leftarrow R^iERE(R^iEE)^{-1}\Theta, w \leftarrow a^i\overline{a}\,\overline{a}a(a^i\overline{a}\,\overline{a})^{-1}w$

  end if

  $\psi_0 \leftarrow \varepsilon$

  for $i \in \{1, 2, \ldots, n\}$ **do**

    if $i \leq n-2$ and $(\vartheta_i\vartheta_{i+1}\vartheta_{i+2}, w\,[\,i \ldots i+2\,]) \in \{(ERR, ab\bar{b}), (REE, ab\bar{b})\}$, $a, b \in \mathcal{A}$ **then**

      $\Theta \leftarrow \vartheta_1 \cdots \vartheta_{i+1}\overline{\vartheta}_{i+1}\vartheta_{i+2}\vartheta_{i+3} \cdots \vartheta_n$, $w \leftarrow w\,[\,1 \ldots i+2\,]\overline{w\,[\,i+2\,]}w\,[\,i+3 \ldots n\,]$

    end if

    if $|w\,[\,1 \ldots i-1\,]|_{w\,[\,i\,]} = 0$ or $|\vartheta_1 \cdots \vartheta_{i-1}|_{\vartheta_i} = 0$ **then**

      if $\vartheta_i = R$ **then** $\psi_i \leftarrow \psi_{i-1}w\,[\,i\,]\psi_{i-1}$

      else if $\vartheta_i = E$ and $\vartheta_{i-1}(w\,[\,1\,]) \neq w\,[\,i\,]$ **then** $\psi_i \leftarrow \psi_{i-1}\overline{\psi}_{i-1}$

      else $\psi_i \leftarrow \psi_{i-1}w\,[\,i\,]\overline{w\,[\,i\,]}\,\overline{\psi_{i-1}}$

      end if

    else if $\exists j$ such that $(\vartheta_1 \cdots \vartheta_i, w\,[\,1 \ldots i\,]) = (\vartheta_1 \cdots \vartheta_j\vartheta_i\vartheta_{j+2}\vartheta_{j+3} \cdots \vartheta_i, w\,[\,1 \ldots j+1\,]\vartheta_{i-1} \circ \vartheta_i(w\,[\,i\,])w\,[\,j+3 \ldots i\,])$ **then** $\psi_i \leftarrow \psi_{i-1}(\vartheta_{i-1} \circ \vartheta_i)(\psi_j^{-1}\psi_{i-1})$

    else

      if $\vartheta_i = R$ or $\vartheta_{i-1}(w\,[\,1\,]) \neq w\,[\,i\,]$ **then** $\psi_i \leftarrow \psi_{i-1}\vartheta_i(\psi_{i-1})$

      else $\psi_i \leftarrow \psi_{i-1}w\,[\,i\,](\vartheta_{i-1} \circ \vartheta_i)(w\,[\,i\,]\psi_{i-1})$

      end if

    end if

  end for

  return ψ_n

end function

```
sage: A = Words([0,1])
sage: R = WordMorphism({0:0,1:1}, codomain=A)
sage: E = WordMorphism({0:1,1:0}, codomain=A)
sage: T = [R,E,E,E,E,E,R,R,R,E] + [R,E,E,E]
sage: w = A([0,0,0,1,1,0,1,0,1,1] + [0,0,0,1])
sage: u = iterated_right_palindromic_closure(w,T)
sage: v = u[:5000]
sage: [round(float(v.number_of_factors(i)/(4*i)),3) for i in range(1,300)]
[0.5, 0.5, 0.5, 0.5, 0.5, 0.583, 0.643, 0.688, 0.722, 0.8, 0.864, 0.917, 0.962,
1.0, 1.033, 1.063, 1.088, 1.111, 1.132, 1.15, 1.167, 1.182, 1.196, 1.208, 1.22,
1.212, 1.204, 1.196, 1.19, 1.183, 1.177, 1.172, 1.167, 1.162, 1.157, 1.153,
1.149, 1.145, 1.141, 1.137, 1.134, 1.131, 1.128, 1.125, 1.122, 1.12, 1.117,
1.115, 1.112, 1.11, 1.108, 1.106, 1.104, 1.102, 1.1, 1.098, 1.096, 1.095,
1.093, 1.092, 1.09, 1.089, 1.087, 1.086, 1.085, 1.083, 1.082, 1.081, 1.08,
1.079, 1.077, 1.076, 1.075, 1.074, 1.073, 1.072, 1.071, 1.071, 1.07, 1.069,
1.068, 1.067, 1.066, 1.065, 1.065, 1.064, 1.063, 1.063, 1.062, 1.061, 1.06,
1.06, 1.059, 1.059, 1.058, 1.057, 1.057, 1.056, 1.056, 1.055, 1.054, 1.054,
1.053, 1.053, 1.052, 1.052, 1.051, 1.051, 1.05, 1.05, 1.05, 1.049, 1.049,
1.048, 1.048, 1.047, 1.047, 1.047, 1.046, 1.046, 1.045, 1.045, 1.045, 1.044,
1.044, 1.044, 1.043, 1.043, 1.043, 1.042, 1.042, 1.042, 1.041, 1.041, 1.041,
1.04, 1.04, 1.04, 1.04, 1.039, 1.039, 1.039, 1.038, 1.038, 1.038, 1.038, 1.037,
1.037, 1.037, 1.037, 1.036, 1.036, 1.036, 1.036, 1.035, 1.035, 1.035, 1.035,
1.035, 1.034, 1.034, 1.034, 1.034, 1.034, 1.033, 1.033, 1.033, 1.033, 1.033,
1.032, 1.032, 1.032, 1.032, 1.032, 1.031, 1.031, 1.031, 1.031, 1.031, 1.031,
1.03, 1.03, 1.03, 1.03, 1.03, 1.03, 1.029, 1.029, 1.029, 1.029, 1.029, 1.029,
1.028, 1.028, 1.028, 1.028, 1.028, 1.028, 1.028, 1.027, 1.027, 1.027, 1.027,
1.027, 1.027, 1.027, 1.027, 1.026, 1.026, 1.026, 1.024, 1.021, 1.019, 1.016,
```

```
1.014, 1.012, 1.009, 1.007, 1.005, 1.002, 1.0, 0.998, 0.996, 0.993, 0.991,
0.989, 0.987, 0.985, 0.983, 0.98, 0.978, 0.976, 0.974, 0.972, 0.97, 0.968,
0.966, 0.964, 0.962, 0.96, 0.959, 0.957, 0.955, 0.953, 0.951, 0.949, 0.947,
0.946, 0.944, 0.942, 0.94, 0.938, 0.937, 0.935, 0.933, 0.932, 0.93, 0.928,
0.927, 0.925, 0.923, 0.922, 0.92, 0.919, 0.917, 0.915, 0.914, 0.912, 0.911,
0.909, 0.908, 0.906, 0.905, 0.903, 0.902, 0.9, 0.899, 0.897, 0.896, 0.895,
0.893, 0.892, 0.89, 0.889, 0.888, 0.886, 0.885, 0.884, 0.882, 0.881, 0.88,
0.878, 0.877, 0.876, 0.875, 0.873, 0.872, 0.871, 0.87]
```

Observe that the $f(n)$ may be greater than $4n$ up to $n = 200$, but it then decreases beyond $4n$, suggesting Conjecture 43.

8 Acknowledgement

The authors wish to thank the referees for a number of comments and suggestions that have led to improvements in this paper.

References

[1] A. Blondin Massé, S. Gaboury, and S. Hallé, Pseudoperiodic words, in *DLT'2012 - 16th International Conference on Developments in Language Theory (Taipei, Taiwan) August 14–17*, Lect. Notes in Comp. Sci., Vol. 7410, Springer, 2012, pp. 308–331.

[2] A. Blondin Massé, G. Paquin and L. Vuillon, A Fine and Wilf's theorem for pseudoperiods and Justin's formula for generalized pseudostandard words, in *JM'2010 - 13th Mons Theoretical Computer Science Days*, 2010. Available at http://tinyurl.com/bbqxtou.

[3] A. Blondin Massé, and L. Vuillon, Generalized pseudostandard words having complexity $2n$, in *LaCIM'2010 - 20th Anniversary of the Laboratoire de combinatoire et d'informatique mathématique*, 2010.

[4] M. Bucci, A. de Luca, A. De Luca, and L. Q. Zamboni, On different generalizations of episturmian words, *Theor. Comput. Sci.*, **393** (2008) 23–36.

[5] M. Bucci, A. de Luca, A. De Luca, and L. Q. Zamboni, On some problems related to palindromic closure, *RAIRO-Theor. Inf. and Appl.*, **42** (2008), 679–700.

[6] E. Czeizler, E. Czeizler, L. Kari, and S. Seki, An extension of the Lyndon Schützenberger result to pseudoperiodic words, in *DLT '09: Proceedings of the 13th International Conference on Developments in Language Theory*, Springer-Verlag, 2009, pp. 183–194.

[7] A. de Luca, Sturmian words: structure, combinatorics, and their arithmetics, *Theor. Comput. Sci.* **183** (1997), 45–82.

[8] A. de Luca, and A. De Luca, Pseudopalindrome closure operators in free monoids, *Theor. Comput. Sci.* **362** (2006), 45–82.

[9] X. Droubay, J. Justin, and G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, *Theor. Comput. Sci.*, **255** (2001), 539–553.

[10] A. Glen, and J. Justin, Episturmian words: a survey, *RAIRO Theor. Info. Appl.*, **43** (2009) 402–433.

[11] D. Jamet, G. Paquin, G. Richomme and L. Vuillon, On the fixed points of the iterated pseudopalindromic closure, in *WORDS'09 - 7th International Conference on Words*, 2010, Paper 122.

[12] J. Justin, Episturmian morphisms and a Galois theorem on continued fractions, *RAIRO Theor. Info. Appl.*, **39** (2005), 207–215.

[13] M. Lothaire, *Algebraic Combinatorics on Words*, Vol. 90, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2002.

[14] F. Manea, R. Mercaş and D. Nowotka, Fine and Wilf's theorem and pseudo-repetitions, in B. Rovan, V. Sassone, and P. Widmayer, eds., *MFCS 2012: Mathematical Foundations of Computer Science*, Lecture Notes in Computer Science, Vol. 7464, Springer-Verlag, 2012, pp. 668–680.

[15] G. Rote, Sequences with subword complexity $2n$, *J. Number Theory*, **46** (1994), 196–213.

[16] W. A. Stein et al., Sage Mathematics Software (Version 4.3.3), The Sage Development Team (2010), `http://www.sagemath.org`.

2010 *Mathematics Subject Classification*: Primary 68R15; Secondary 05A05, 37E05.
Keywords: palindrome, pseudopalindrome, iterated pseudopalindromic closure, generalized pseudostandard word, normalization, directive bi-sequence, Justin's formula, Fine and Wilf theorem.

Received July 4 2012; revised version received January 1 2013. Published in *Journal of Integer Sequences*, March 2 2013.

Return to Journal of Integer Sequences home page.