Discrete Mathematics and Theoretical Computer Science ....
.... **vol. 26:3 #2 (2024)**

String attractors of Rote sequences

Lubomíra Dvořáková * Veronika Hendrychová†

Czech Technical University in Prague, Czech Republic

revisions 10th Oct. 2023, 4th Mar. 2024, 18th July 2024; accepted 23rd July 2024.

In this paper, we describe minimal string attractors of pseudopalindromic prefixes of standard complementary-symmetric Rote sequences. Such a class of Rote sequences forms a subclass of binary generalized pseudostandard sequences, *i.e.*, of sequences obtained when iterating palindromic and antipalindromic closures. When iterating only palindromic closure, palindromic prefixes of standard Sturmian sequences are obtained and their string attractors are of size two. However, already when iterating only antipalindromic closure, antipalindromic prefixes of binary pseudostandard sequences are obtained and we prove that the minimal string attractors are of size three in this case. We conjecture that the pseudopalindromic prefixes of any binary generalized pseudostandard sequence have a minimal string attractor of size at most four.

Keywords: pseudostandard sequences, generalized pseudostandard sequences, string attractors, Rote sequences, palindromic closure, antipalindromic closure

1 Introduction

In the last years, significant attention has been dedicated to the study of string attractors. Their definition and first results were provided by Kempa and Prezza [10]: a *string attractor* of a finite word $w = w_0 w_1 \cdots w_{n-1}$, where w_i are letters, is a subset Γ of $\{0, 1, \ldots, n-1\}$ such that each non-empty factor of w has an occurrence containing an element of Γ. They are closely related to methods of compression of highly repetitive data, so-called dictionary compressors. On one hand, it was shown by Kempa and Prezza [10] that dictionary compressors can be interpreted as approximation algorithms for the smallest string attractors, due to the measures induced by the compressors being lower bounded by the smallest string attractor size. On the other hand, attractors are able to express bounds for and potentially unify the dictionary compression measures.

However, the general problem of finding the minimum size of a string attractor is NP-complete. It is therefore natural to study the problem in the context of combinatorics on words and to put certain restrictions on the input, which makes the computation tractable.

In this paper we carry on the study of string attractors of important classes of sequences. Attractors of minimum size of factors/prefixes/particular prefixes of the following sequences have been determined so

*Supported by the Ministry of Education, Youth and Sports of the Czech Republic through the project CZ.02.1.01/0.0/0.0/16_019/0000778.

†Supported by Czech Technical University in Prague, through the project SGS23/187/OHK4/3T/14.

ISSN 1365–8050

far: standard Sturmian sequences [13, 6], the Tribonacci sequence [15], episturmian sequences [6], the Thue-Morse sequence [12, 15, 5], the period-doubling sequence [15], the powers of two sequence [15, 11]. Recently, string attractors of fixed points of k-bonacci-like morphisms have been described [8].

When studying string attractors of episturmian sequences [6], we could see usefulness of palindromic closures. It is thus natural to ask as the next question what can be said about string attractors of binary generalized pseudostandard sequences (defined in [3]), *i.e.*, when iterating palindromic and antipalindromic closures. Two subclasses of generalized pseudostandard sequences have been already studied: standard Sturmian sequences and the Thue-Morse sequence. Here, we consider pseudostandard sequences, *i.e.*, binary sequences obtained when iterating solely the antipalindromic closure, and complementary-symmetric Rote sequences. We show that the minimal string attractor of antipalindromic prefixes of pseudostandard sequences is of size three, while the minimal string attractor of pseudopalindromic prefixes of complementary-symmetric Rote sequences is of size two. In the latter case, the description of Rote sequences from [2] plays an important role. Based on computer experiments, we conjecture that the minimum size of string attractors of pseudopalindromic prefixes of binary generalized peudostandard sequences is at most four.

2 Preliminaries

An *alphabet* $\mathcal{A}$ is a finite set of symbols called *letters*. A *word* over $\mathcal{A}$ of *length* n is a string $u = u_0u_1\cdots u_{n-1}$, where $u_i \in \mathcal{A}$ for all $i \in \{0, 1, \ldots, n-1\}$. We let $|u|$ denote the length of u. The set of all finite words over $\mathcal{A}$ together with the operation of concatenation forms a monoid, denoted $\mathcal{A}^*$. Its neutral element is the *empty word* ε and we write $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\varepsilon\}$. If $u = xyz$ for some $x, y, z \in \mathcal{A}^*$, then x is a *prefix* of u, z is a *suffix* of u and y is a *factor* of u.

A *sequence* over $\mathcal{A}$ is an infinite string $\mathbf{u} = u_0u_1u_2\cdots$, where $u_i \in \mathcal{A}$ for all $i \in \mathbb{N}$. We always denote sequences by bold letters.

A sequence $\mathbf{u}$ is *eventually periodic* if $\mathbf{u} = vwww\cdots = v(w)^\omega$ for some $v \in \mathcal{A}^*$ and $w \in \mathcal{A}^+$. If $\mathbf{u}$ is not eventually periodic, then it is *aperiodic*. A *factor* of $\mathbf{u} = u_0u_1u_2\cdots$ is a word y such that $y = u_iu_{i+1}u_{i+2}\cdots u_{j-1}$ for some $i, j \in \mathbb{N}$, $i \leq j$. If $i = j$, then $y = \varepsilon$. In the context of string attractors, the set $\{i, i+1, \ldots, j-1\}$ is called an *occurrence* of the factor y in $\mathbf{u}$. (Usually, only the number i is called an occurrence of y in $\mathbf{u}$.) If $i = 0$, the factor y is a *prefix* of $\mathbf{u}$. A factor w of $\mathbf{u}$ is *left special* if aw, bw are factors of $\mathbf{u}$ for at least two distinct letters $a, b \in \mathcal{A}$. A sequence $\mathbf{u}$ is said to be *closed under reversal* if for each factor $w = w_0w_1\cdots w_{n-1}$, where $w_i \in \mathcal{A}$, $\mathbf{u}$ contains also its mirror image $w_{n-1}\cdots w_1w_0$. A binary sequence $\mathbf{u}$ is called *Sturmian* if $\mathbf{u}$ is closed under reversal and $\mathbf{u}$ contains exactly one left special factor of each length. If moreover each left special factor is a prefix of $\mathbf{u}$, then $\mathbf{u}$ is *standard Sturmian*.

A *string attractor* (or *attractor* for short) of a word $w = w_0w_1\cdots w_{n-1}$, where $w_i \in \mathcal{A}$, is a set $\Gamma \subset \{0, 1, \ldots, n-1\}$ such that every non-empty factor of w has an occurrence in w containing at least one element of Γ. If $i \in \Gamma$ and a word f has an occurrence in w containing i, we say that f *crosses* i and we also say that f *crosses the attractor* Γ. For instance, $\Gamma = \{1, 3\}$ is an attractor of $w = \mathtt{0\underline{1}0\underline{0}10}$ (it corresponds to the underlined positions). The factor $\mathtt{00}$ crosses the position 3 and thus it crosses the attractor. Γ is an attractor of minimum size – minimal attractor for short – since each attractor necessarily contains occurrences of all distinct letters of the word.

3 Palindromic and antipalindromic closures

Throughout the paper, we deal only with binary sequences. Therefore we define the notions of reversal and palindrome over a binary alphabet, too.

Definition 1. The map $R : \{0,1\}^* \to \{0,1\}^*$, called *reversal*, associates with each word its mirror image, *i.e.*, $R(w_0 w_1 \cdots w_{n-1}) = w_{n-1} \cdots w_1 w_0$, where $w_i \in \{0,1\}$ for each $i \in \{0,1,\ldots,n-1\}$. The map $E : \{0,1\}^* \to \{0,1\}^*$, called *exchange antimorphism*, is a composition of reversal and letter exchange, *i.e.*, $E(w_0 w_1 \cdots w_{n-1}) = \overline{w_{n-1}} \cdots \overline{w_1}\, \overline{w_0}$, where $w_i \in \{0,1\}$ and $\overline{0} = 1$ and $\overline{1} = 0$. A word $w \in \{0,1\}^*$ is a *palindrome* if $w = R(w)$ and w is an *antipalindrome* if $w = E(w)$. A word is a *pseudopalindrome* if it is a palindrome or an antipalindrome.

Consider $w \in \{0,1\}^*$. Then w^R is the shortest palindrome having w as prefix and it is called the *palindromic closure* of w. Similarly, w^E is the shortest antipalindrome having w as prefix and it is called the *antipalindromic closure* of w. The *pseudopalindromic closure* is a term covering both palindromic and antipalindromic closure.

Let $w \in \{0,1\}^+$. It follows immediately from the definition that $w^R = vxR(v)$, where $w = vx$ and x is the longest palindromic suffix of w. Similarly, $w^E = vyE(v)$, where y is the longest antipalindromic suffix of w (possibly empty).

Example 2. We have $(000)^R = 000$, $(000)^E = 000111$, $(0001)^R = 0001000$, $(0001)^E = 000111$, $(01101)^R = 0110110$, $(01101)^E = 01101001$.

Definition 3. Let $\Delta = \delta_1\delta_2\cdots$ and $\Theta = \vartheta_1\vartheta_2\cdots$, where $\delta_i \in \{0,1\}$ and $\vartheta_i \in \{E,R\}$ for all $i \in \mathbb{N}, i \geq 1$. The sequence $\mathbf{u}(\Delta,\Theta)$, called *generalized pseudostandard sequence*, is the sequence having prefixes w_n obtained from the recurrence relation

$$w_{n+1} = (w_n\delta_{n+1})^{\vartheta_{n+1}},$$
$$w_0 = \varepsilon.$$

The bi-sequence (Δ,Θ) is called the *directive bi-sequence* of the word $\mathbf{u}(\Delta,\Theta)$.

Example 4. Consider $\mathbf{u} = \mathbf{u}(\Delta,\Theta)$ with $\Delta = 01^\omega$ and $\Theta = (RE)^\omega$, then $\mathbf{u}$ is the Thue-Morse sequence [3]. Here are the first six pseudopalindromic prefixes:

$$\begin{aligned} w_0 &= \varepsilon \\ w_1 &= 0 \\ w_2 &= 01 \\ w_3 &= 0110 \\ w_4 &= 01101001 \\ w_5 &= 0110100110010110. \end{aligned}$$

For some special forms of (Δ,Θ), well-known classes of sequences are obtained:

1. $\mathbf{u}$ is a *standard Sturmian sequence* if $\mathbf{u} = \mathbf{u}(\Delta,\Theta)$ for some Δ containing both letters infinitely many times and $\Theta = R^\omega$.
2. $\mathbf{u}$ is a *pseudostandard sequence* if $\mathbf{u} = \mathbf{u}(\Delta,\Theta)$ for some Δ and $\Theta = E^\omega$.

In the sequel, when examining Rote sequences, we will need the following statement about the form of palindromic prefixes of a standard Sturmian sequence. (i)

(i) The palindromic prefixes of standard Sturmian sequences are also known as central words [1].

Proposition 5 (Proposition 7 [4]). *Let $\mathbf{u}$ be a standard Sturmian sequence and let $(u_n)_{n=0}^{\infty}$ be the sequence of palindromic prefixes of $\mathbf{u}$ ordered by length. If u_n contains both letters, then for some $a \in \{\mathtt{0}, \mathtt{1}\}$*

$$u_n = u_{n-1} a \overline{a} u ,$$

where u is the longest palindromic prefix of u_n followed by $\overline{a}$ (i.e., $u\overline{a}$ is a prefix of u_n).

4 String attractors of Sturmian sequences

String attractors of palindromic prefixes of episturmian sequences were described in [6] (Theorem 7). Let us recall here the statement restricted to the binary alphabet, together with its proof. Similar ideas will be used for pseudostandard sequences.

Theorem 6. *[6] Let v be a non-empty palindromic prefix of a standard Sturmian sequence. For every letter a occurring in v, denote*

$$r_a = \max\{|p| \,:\, p \text{ is a palindrome and } pa \text{ is a prefix of } v\}.$$

Then $\Gamma = \{r_a : a \text{ occurs in } v\}$ is an attractor of v and it is of minimum size.

Proof: To construct a standard Sturmian sequence, we use only palindromic closure in Definition 3 and each palindromic prefix v is equal to w_n for some $n \in \mathbb{N}$. Let us assume that the first letter of Δ is $\mathtt{0}$. We will prove the statement by mathematical induction on n. Let us recall that we index positions from 0, *i.e.*, $v = v_0 v_1 \cdots v_{|v|-1}$.

- For $n = 1$ we have $w_1 = \mathtt{0}$ and its attractor equals $\{0\}$. The longest palindromic prefix of w_1 followed by $\mathtt{0}$ is equal to $w_0 = \varepsilon$ and its length satisfies $|w_0| = 0$.

- For $n \geq 2$ we assume that w_{n-1} has an attractor of the form from the statement. We have $w_n = (w_{n-1}a)^R$ for some $a \in \{\mathtt{0}, \mathtt{1}\}$. The following three situations may occur:

  1. $w_n = w_{n-1}a$: According to the definition of palindromic closure, this happens only for $a = \mathtt{0}$ and $w_{n-1} = \mathtt{0}^{n-1}$. The longest palindromic prefix of $w_n = \mathtt{0}^n$ followed by $\mathtt{0}$ is $w_{n-1} = \mathtt{0}^{n-1}$. The length of w_{n-1} is $n-1$ and, indeed, $\{n-1\}$ is an attractor of w_n.
  2. $w_n = w_{n-1} a w_{n-1}$: By the definition of palindromic closure, this happens only in case when $w_{n-1} = \mathtt{0}^{n-1}$ and $a = \mathtt{1}$. Then $w_n = \mathtt{0}^{n-1}\mathtt{1}\mathtt{0}^{n-1}$ and $\Gamma = \{n-2, n-1\}$ is indeed an attractor of w_n.
  3. $w_n = w_{n-1} a u$ for some $u \neq \varepsilon$ and $u \neq w_{n-1}$: Then w_{n-1} contains both letters. We want to prove that $\Gamma = \{r_{\mathtt{0}}, r_{\mathtt{1}}\}$, as defined in the statement, is an attractor of w_n. Since the longest palindromic prefix of w_n followed by b, where $b \in \{\mathtt{0}, \mathtt{1}\}$, $b \neq a$, is the same as in w_{n-1}, we know by induction assumption that $\{r_b, r'_a\}$ is an attractor of w_{n-1}, where $r'_a = |w_\ell|$ and w_ℓ is the longest palindromic prefix of w_{n-1} followed by a. By the definition of palindromic closure we have

$$w_n = \underbrace{R(u) a w_\ell}_{w_{n-1}} \underline{a} u = R(u) a \underbrace{w_\ell \underline{a} u}_{w_{n-1}} . \tag{1}$$

Then each factor f of w_n either has an occurrence containing the position $|w_{n-1}|$, *i.e.*, corresponding to the second (underlined) a in (1), or f is entirely contained in w_{n-1}. In the latter case, f has an occurrence crossing the attractor of w_{n-1}. Thus, f either crosses the position r_b or f crosses the position $r'_a = |w_\ell|$. In the first case, we are done since f crosses Γ. In the second case, according to (1), the factor f has also an occurrence in w_n containing the position $r_a = |w_{n-1}|$ (corresponding to the underlined a). To sum up, we have proved that each factor of w_n has an occurrence crossing $\{r_a, r_b\} = \{r_{\mathtt{0}}, r_{\mathtt{1}}\} = \Gamma$.

□

Example 7. The most famous standard Sturmian sequence is the *Fibonacci sequence*

$$\mathbf{u} = \mathbf{u}(\Delta, R^\omega)\,,$$

where $\Delta = (\mathtt{01})^\omega$. The first six palindromic prefixes of $\mathbf{u}$ with the positions of the attractor from Theorem 6 underlined read:

$$\begin{aligned} w_0 &= \varepsilon \\ w_1 &= \underline{\mathtt{0}} \\ w_2 &= \underline{\mathtt{01}}\mathtt{0} \\ w_3 &= \mathtt{0}\underline{\mathtt{1}}\mathtt{0}\underline{\mathtt{0}}\mathtt{10} \\ w_4 &= \mathtt{010}\underline{\mathtt{0}}\mathtt{10}\underline{\mathtt{1}}\mathtt{0010} \\ w_5 &= \mathtt{0100101}\underline{\mathtt{1}}\mathtt{00100}\underline{\mathtt{0}}\mathtt{1010010}\,. \end{aligned}$$

5 String attractors of pseudostandard sequences

As a new result we will describe string attractors of antipalindromic prefixes of pseudostandard sequences.

Theorem 8. *Let v be a non-empty antipalindromic prefix of a pseudostandard sequence starting with the letter $\mathtt{0}$. For every letter a occurring in v, denote*

$$e_a = \max\{|q| \,:\, q \text{ is an antipalindrome and } qa \text{ is a prefix of } v\}.$$

If such a prefix does not exist, then set $e_a = e_{\overline{a}}$.
Then $\Gamma = \{e_{\mathtt{0}}, e_{\mathtt{1}}, |v| - e_{\mathtt{1}} - 1\}$ is an attractor of v.
Moreover, if $v = w_n$, $n \geq 2$, from Definition 3, then

- *Γ is of size two if and only if Δ starts with $\mathtt{0}^n$;*
- *Γ is a minimum size attractor if and only if Δ does not start with $\mathtt{01}^{n-1}$.*

Proof: To construct a pseudostandard sequence $\mathbf{u}$, we use only antipalindromic closure in Definition 3 and each antipalindromic prefix v is equal to w_n for some $n \in \mathbb{N}$. We will prove the statement about the form of attractors of w_n by mathematical induction on n. Let us recall that we index positions from 0, *i.e.*, $v = v_0 v_1 \cdots v_{|v|-1}$.

- If $\mathtt{0}^n$ is a prefix of Δ, then we have $w_n = (\mathtt{01})^n$ and $e_{\mathtt{0}} = 2n-2$ and $e_{\mathtt{1}} = e_{\mathtt{0}}$. Then $\Gamma = \{e_{\mathtt{0}}, |w_n| - 1 - e_{\mathtt{0}}\} = \{2n-2, 1\}$ is indeed an attractor of $w_n = \underline{\mathtt{01}}$ for $n = 1$, resp. $w_n = \mathtt{0}\underline{\mathtt{1}}(\mathtt{01})^{n-2}\underline{\mathtt{0}}\mathtt{1}$ for $n \geq 2$ (we underlined the positions of Γ). In particular, this proves the result for all $n \geq 1$ when $\Delta = \mathtt{0}^\omega$.

- Now, assume $\mathtt{0}^k\mathtt{1}$ is a prefix of Δ for some $k \geq 1$. Then $w_{k+1} = (\mathtt{01})^k\mathtt{10}(\mathtt{01})^k$, $e_{\mathtt{0}} = |(\mathtt{01})^{k-1}| = 2k-2$, and $e_{\mathtt{1}} = |(\mathtt{01})^k| = 2k$. It is readily seen that $\Gamma = \{2k-2, 2k, 2k+1\}$ is an attractor of $w_{k+1} = (\mathtt{01})^{k-1}\underline{\mathtt{0}}\mathtt{1}\underline{\mathtt{10}}(\mathtt{01})^k$ (we underlined the positions of Γ). Using the first item, each w_n, for $1 \leq n \leq k$, also has an attractor of the form from the statement.

- For $n > k+1$ we assume that w_{n-1} has an attractor of the form from the statement. The following two situations may occur:

  1. $\delta_n = 1$ and $w_n = (w_{n-1}\mathtt{1})^E = v\mathtt{0}w_\ell\mathtt{1}E(v)$, where $w_\ell \neq \varepsilon$, w_ℓ is the longest antipalindromic prefix followed by $\mathtt{1}$ in w_{n-1}, and v is a prefix of the word w_n. The longest antipalindromic prefix q followed by $\mathtt{0}$ in w_{n-1} and in w_n is the same (q may be empty if $k = 1$). The longest antipalindromic prefix of w_n followed by $\mathtt{1}$ is equal to $w_{n-1} = v\mathtt{0}w_\ell$. We assume that $\{|q|, |w_\ell|, |w_{n-1}| - |w_\ell| - 1\}$ is an attractor of w_{n-1} and we want to show that $\Gamma = \{|q|, |w_{n-1}|, |w_n| - |w_{n-1}| - 1\}$ is then an attractor of w_n. Below, we underline the positions of the attractor of w_{n-1} and the positions of Γ in w_n:

  $$\begin{aligned} w_{n-1} &= q\underline{\mathtt{0}}\cdots = v\underline{\mathtt{0}}w_\ell = w_\ell\underline{\mathtt{1}}E(v)\,; \\ w_n &= q\underline{\mathtt{0}}\cdots = v\underline{\mathtt{0}}w_\ell\underline{\mathtt{1}}E(v)\,. \end{aligned}$$

  Each factor f of w_n either crosses $|w_{n-1}|$, *i.e.*, the underlined $\mathtt{1}$, or f is contained in w_{n-1}. By the form of the attractor of w_{n-1}, the factor f crosses $|q|$ or $|v| = |w_{n-1}| - |w_\ell| - 1 = |w_n| - |w_{n-1}| - 1$ or $|w_\ell|$. In the last case, since $w_{n-1} = w_\ell\mathtt{1}E(v)$ is a suffix of w_n, we can see that f has an occurrence in w_n containing $|w_{n-1}|$.

  2. $\delta_n = \mathtt{0}$ and $w_n = (w_{n-1}\mathtt{0})^E = v\mathtt{1}w_\ell\mathtt{0}E(v)$, where w_ℓ is the longest antipalindromic prefix followed by $\mathtt{0}$ in w_{n-1} (it may be empty), and v is a prefix of the word w_n. The longest antipalindromic prefix of w_n followed by $\mathtt{0}$ is equal to w_{n-1} and the longest antipalindromic prefix q of w_n followed by $\mathtt{1}$ is the same as for w_{n-1}. We assume that $\{|q|, |w_\ell|, |w_{n-1}| - |q| - 1\}$ is an attractor of w_{n-1} and we want to show that $\Gamma = \{|q|, |w_{n-1}|, |w_n| - |q| - 1\}$ is then an attractor of w_n. Below, we underline the positions of the attractor of w_{n-1} and the positions of Γ in w_n:

  $$\begin{aligned} w_{n-1} &= q\underline{\mathtt{1}}\cdots = \cdots\underline{\mathtt{0}}q = w_\ell\underline{\mathtt{0}}E(v)\,; \\ w_n &= q\underline{\mathtt{1}}\cdots = \cdots\underline{\mathtt{0}}q = v\mathtt{1}w_\ell\underline{\mathtt{0}}E(v)\,. \end{aligned}$$

  Each factor f of w_n either crosses $|w_{n-1}|$, *i.e.*, the underlined $\mathtt{0}$ in the last expression for w_n above, or f is contained in w_{n-1}. By the form of the attractor of w_{n-1}, the factor f crosses $|q|$ or $|w_{n-1}| - |q| - 1$ or $|w_\ell|$. In the last two cases, since $w_{n-1} = w_\ell\mathtt{0}E(v)$ is a suffix of w_n, we can see that f has an occurrence in w_n containing $|w_n| - |q| - 1$ or $|w_{n-1}|$.

Now let us show the attractor's minimality. We could see in the above proof that if $\mathtt{0}^k$ is a prefix of Δ, then the attractor Γ of w_k from the theorem is of minimum size (equal to two).

- As soon as $\mathtt{0}^k\mathtt{1}$ for $k \geq 2$ is a prefix of Δ, then $w_{k+1} = (\mathtt{01})^{k-1}\mathtt{011001}(\mathtt{01})^{k-1}$. It follows from the following observation and Definition 3 that for each $n \geq k+1$, the factor $\mathtt{00}$, resp. $\mathtt{11}$, only occurs as a factor of $\mathtt{011001}$ in w_n. We can observe that the prefix $\mathtt{0}^k\mathtt{10}$ yields

  $$w_{k+2} = (\mathtt{01})^{k-1}\mathtt{011001}(\mathtt{01})^{k-1}\mathtt{011001}(\mathtt{01})^{k-1}\,,$$

and similarly, for the prefix $\mathtt{0}^k\mathtt{11}$, we obtain

$$w_{k+2} = (\mathtt{01})^{k-1}\mathtt{011001}(\mathtt{01})^{k-2}\mathtt{011001}(\mathtt{01})^{k-1}\,.$$

Generally,

$$w_n = (\mathtt{01})^{p_0}\mathtt{011001}(\mathtt{01})^{p_1}\mathtt{011001}\cdots(\mathtt{01})^{p_{j-1}}\mathtt{011001}(\mathtt{01})^{p_j}\,, \tag{2}$$

where $p_i \in \{k-2, k-1\}$ and $p_0 = p_j = k-1$. From this, we can also see that the factors $\mathtt{011001}$ do not overlap anywhere in the word.

An attractor of size two does not exist any more. Let us explain why. All factors of length two, *i.e.*, $\mathtt{00}, \mathtt{01}, \mathtt{10}, \mathtt{11}$, have to cross the attractor. We underline below the possible positions in w_n from (2) for an attractor of size two containing $\mathtt{00}, \mathtt{01}, \mathtt{10}, \mathtt{11}$:

1. $w_n = \cdots(\mathtt{01})^{p_i}\mathtt{01}\underline{\mathtt{1}}\mathtt{0}\underline{\mathtt{0}}\mathtt{1}(\mathtt{01})^{p_{i+1}}\cdots$
2. $w_n = \cdots(\mathtt{01})^{p_i}\mathtt{0}\underline{\mathtt{1}}\mathtt{1}\underline{\mathtt{0}}\mathtt{01}(\mathtt{01})^{p_{i+1}}\cdots$
3. $w_n = \cdots(\mathtt{01})^{p_i}\mathtt{0110}\underline{\mathtt{0}}\mathtt{1}(\mathtt{01})^{p_{i+1}}\cdots(\mathtt{01})^{p_{i+m}}\mathtt{01}\underline{\mathtt{1}}\mathtt{001}(\mathtt{01})^{p_{i+m+1}}\cdots$
4. $w_n = \cdots(\mathtt{01})^{p_i}\mathtt{011}\underline{\mathtt{0}}\mathtt{01}(\mathtt{01})^{p_{i+1}}\cdots(\mathtt{01})^{p_{i+m}}\mathtt{0}\underline{\mathtt{1}}\mathtt{1001}(\mathtt{01})^{p_{i+m+1}}\cdots$
5. $w_n = \cdots(\mathtt{01})^{p_i}\mathtt{01}\underline{\mathtt{1}}\mathtt{001}(\mathtt{01})^{p_{i+1}}\cdots(\mathtt{01})^{p_{i+m}}\mathtt{0110}\underline{\mathtt{0}}\mathtt{1}(\mathtt{01})^{p_{i+m+1}}\cdots$
6. $w_n = \cdots(\mathtt{01})^{p_i}\mathtt{0}\underline{\mathtt{1}}\mathtt{1001}(\mathtt{01})^{p_{i+1}}\cdots(\mathtt{01})^{p_{i+m}}\mathtt{011}\underline{\mathtt{0}}\mathtt{01}(\mathtt{01})^{p_{i+m+1}}\cdots$

where $i, m \in \mathbb{N}, m \geq 1, i+m+1 \leq j$. However, in all the above cases, either $\mathtt{010}$, or $\mathtt{101}$ does not cross the attractor.

- If $\mathtt{01}^k$ for $k \geq 1$ is a prefix of Δ, then $w_{k+1} = \mathtt{01}\underline{\mathtt{1}}\mathtt{0}\underline{\mathtt{0}}\mathtt{1}(\mathtt{1001})^{k-1}$, where we underlined the positions of an attractor of size two. In this case, the attractor Γ from the theorem is not minimal, let us underline its positions: $w_{k+1} = \underline{\mathtt{0}}\mathtt{11}\underline{\mathtt{0}}\mathtt{01}(\mathtt{1001})^{k-2}\underline{\mathtt{1}}\mathtt{001}$ for $k \geq 2$ and $w_{k+1} = \underline{\mathtt{0}}\mathtt{1}\underline{\mathtt{1}}\underline{\mathtt{0}}\mathtt{01}$ for $k = 1$.

- As soon as $\mathtt{01}^k\mathtt{0}$ for $k \geq 1$ is a prefix of Δ, then $w_{k+2} = \mathtt{01}(\mathtt{1001})^k(\mathtt{0110})^k\mathtt{01}$. It follows from Definition 3 that for each $n \geq k+2$, the factor $\mathtt{00}$, resp. $\mathtt{11}$, only occurs as a factor of $\mathtt{011001}$ in w_n. An attractor of size two does not exist – the explanation is analogous as above.

□

Corollary 9. Let $\mathbf{u}$ be a pseudostandard sequence, i.e., $\mathbf{u} = \mathbf{u}(\Delta, E^\omega)$.

- If $\Delta \in \{\mathtt{0}^\omega, \mathtt{1}^\omega, \mathtt{01}^\omega, \mathtt{10}^\omega\}$, then all antipalindromic prefixes of $\mathbf{u}$ have minimal attractors of size two.

- If $\Delta \notin \{\mathtt{0}^\omega, \mathtt{1}^\omega, \mathtt{01}^\omega, \mathtt{10}^\omega\}$, then all sufficiently long antipalindromic prefixes of $\mathbf{u}$ have minimal attractors of size three.

Example 10. Consider $\Delta = \mathtt{01001}\cdots$. The first six prefixes of $\mathbf{u}(\Delta, E^\omega)$ with the positions of attractor underlined read:

$$\begin{array}{rcl}
w_0 & = & \varepsilon \\
w_1 & = & \underline{\mathtt{01}} \\
w_2 & = & \underline{\mathtt{0}}\mathtt{1}\underline{\mathtt{1}}\underline{\mathtt{0}}\mathtt{01} \\
w_3 & = & \mathtt{01}\underline{\mathtt{1}}\mathtt{001}\underline{\mathtt{0}}\mathtt{11}\underline{\mathtt{0}}\mathtt{01} \\
w_4 & = & \mathtt{01}\underline{\mathtt{1}}\mathtt{001011001}\underline{\mathtt{0}}\mathtt{11}\underline{\mathtt{0}}\mathtt{01} \\
w_5 & = & \mathtt{011001011001}\underline{\mathtt{0}}\mathtt{11}\underline{\mathtt{0}}\mathtt{01}\underline{\mathtt{1}}\mathtt{001011001011001}\,.
\end{array}$$

6 String attractors of complementary-symmetric Rote sequences

This section is devoted to the study of attractors of pseudopalindromic prefixes of complementary-symmetric Rote sequences, which form a subclass of generalized pseudostandard sequences. However, besides being generalized pseudostandard sequences, they are also closely related to Sturmian sequences.

Definition 11. *Rote sequences* are binary sequences having complexity $2n$ for each $n \geq 1$. A Rote sequence $\mathbf{u}$ is called *complementary-symmetric (CS)* if its language is closed under the letter exchange, *i.e.*, for each factor $v = v_0v_1\cdots v_{n-1}$ of $\mathbf{u}$, the word $\overline{v} = \overline{v_0}\,\overline{v_1}\cdots\overline{v_{n-1}}$ is also a factor of $\mathbf{u}$.

Let $u = u_0u_1\cdots u_{n-1}$ be a binary word on $\{\mathtt{0}, \mathtt{1}\}$ of length at least two. We denote by $S(u)$ the word $v = v_0v_1\cdots v_{n-2}$ defined by

$$v_i = (u_{i+1} + u_i) \bmod 2 \quad \text{for } i = 0, 1, \ldots, n-2.$$

For example, if $u = \mathtt{0011010}$, then $S(u) = \mathtt{010111}$. The definition may be extended to sequences: if $\mathbf{u}$ is a sequence over $\{\mathtt{0}, \mathtt{1}\}$, then $S(\mathbf{u})$ denotes the sequence $\mathbf{v} = v_0v_1v_2\cdots$, where

$$v_i = (u_{i+1} + u_i) \bmod 2 \quad \text{for } i = 0, 1, \ldots$$

CS Rote sequences are connected to Sturmian sequences by a structural theorem.

Theorem 12 (Rote [14])**.** *A binary sequence $\mathbf{u}$ is a CS Rote sequence if and only if the sequence $S(\mathbf{u})$ is a Sturmian sequence.*

We say that a CS Rote sequence $\mathbf{u}$ is *standard* if both $\mathtt{0}\mathbf{u}$ and $\mathtt{1}\mathbf{u}$ are CS Rote sequences. Equivalently, a sequence $\mathbf{u}$ is standard CS Rote if and only if $S(\mathbf{u})$ is standard Sturmian. The relation between pseudopalindromic prefixes of a standard CS Rote sequence $\mathbf{u}$ and palindromic prefixes of a standard Sturmian sequence $S(\mathbf{u})$ is as follows.

Lemma 13 (Lemma 37 [2])**.** *Let $\mathbf{u}$ be a standard CS Rote sequence. Let $u_0 = \varepsilon, u_1, u_2, \ldots$ be the palindromic prefixes of $S(\mathbf{u})$ ordered by length, and $w_0 = \varepsilon, w_1, w_2, \ldots$ be the pseudopalindromic prefixes of $\mathbf{u}$ ordered by length. Then $S(w_{n+1}) = u_n$ for all $n \in \mathbb{N}, n \geq 1$.*

Remark 14. Let us explain that it is not possible to use known attractors of palindromic prefixes of standard Sturmian sequences to obtain attractors of pseudopalindromic prefixes of CS Rote sequences.

Consider the following palindromic prefix $u = \mathtt{010010010}$ of a standard Sturmian sequence. The corresponding standard CS Rote sequence starting with $\mathtt{0}$ has the antipalindromic prefix $w = \mathtt{0011100011}$, *i.e.*, $S(w) = u$. Let us underline the positions of the attractor of u from Theorem 6: $u = \mathtt{0}\underline{\mathtt{1}}\mathtt{0010}\underline{\mathtt{0}}\mathtt{10}$

and also from [13](Theorem 22): $u = \mathtt{010010\underline{01}0}$. Now, the factor $\mathtt{10}$ has a unique occurrence in w, therefore each attractor of w has to contain either the position 4 or 5. However, there is no straightforward way to get such positions from the underlined attractors of u (or their mirror image from Observation 17).

Blondin-Massé at al. [2] showed that standard CS Rote sequences form a subclass of binary generalized pseudostandard sequences. Moreover, they described precisely the form of the corresponding directive bi-sequence.

Theorem 15 ([2]). *A sequence $\mathbf{u}$ is a standard CS Rote sequence if and only if it is aperiodic and $\mathbf{u} = \mathbf{u}(\Delta, \Theta)$ for some directive bi-sequence (Δ, Θ) such that Θ starts with R and no factor of length two of the directive bi-sequence is in the set*

$$\{(ab, EE) : a, b \in \{\mathtt{0}, \mathtt{1}\}\} \cup \{(a\overline{a}, RR) : a \in \{\mathtt{0}, \mathtt{1}\}\} \cup \{(aa, RE) : a \in \{\mathtt{0}, \mathtt{1}\}\} .$$

Moreover, the prefixes w_n from Definition 3 coincide with all pseudopalindromic prefixes of $\mathbf{u}$.

The aperiodicity of a binary generalized pseudostandard sequence may be recognized easily.

Theorem 16 ([7]). *Let (Δ, Θ) be a directive bi-sequence. Then $\mathbf{u} = \mathbf{u}(\Delta, \Theta)$ is aperiodic if and only if there is no bijection $f : \{E, R\} \to \{\mathtt{0}, \mathtt{1}\}$ such that $f(\vartheta_n) = \delta_{n+1}$ for all sufficiently large n.*

In the proof of the main theorem on string attractors of pseudopalindromic prefixes of standard CS Rote sequences, the following statements will be useful.

Observation 17. *If w is a pseudopalindrome with an attractor Γ, the mirror image $\Gamma^R = \{|w| - 1 - \gamma : \gamma \in \Gamma\}$ is an attractor of w, too.*

Lemma 18. *Let $\mathbf{u}$ be a standard CS Rote sequence and let $(w_n)_{n=1}^{\infty}$ be the sequence of all non-empty pseudopalindromic prefixes of $\mathbf{u}$, ordered by length. Then for $n \geq 2$ and $a \in \{\mathtt{0}, \mathtt{1}\}$ such that $w_{n-1}a$ is a prefix of w_n, we have:*

1. *If $w_{n-1} = R(w_{n-1})$ and $w_n = R(w_n)$, then $w_n = w_{n-1}a\overline{w}$, where w is the longest antipalindromic prefix of w_n followed by a.*
2. *If $w_{n-1} = R(w_{n-1})$ and $w_n = E(w_n)$, then $w_n = w_{n-1}a\overline{w}$, where w is the longest palindromic prefix of w_n followed by $\overline{a}$.*
3. *If $w_{n-1} = E(w_{n-1})$ and $w_n = R(w_n)$, then $w_n = w_{n-1}aw$, where w is the longest palindromic prefix of w_n followed by a.*

Proof: Assume without loss of generality that the Rote sequence starts with $\mathtt{0}$. The reader is invited to check the cases, where $S(w_n)$ contains only one letter, *i.e.*, the cases where $w_n = \mathtt{0}^n$ or $w_n = (\mathtt{01})^{\frac{n}{2}}$ for n even or $w_n = (\mathtt{01})^{\frac{n-1}{2}}\mathtt{0}$ for n odd. In the sequel, assume $S(w_n)$ contains both letters. The possible prefixes of (Δ, Θ) are given in Theorem 15.

1. Since $w_n = w_{n-1}aw$ and w_{n-1} as a palindrome has $\mathtt{0}$ as both the first and the last letter, then by Lemma 13, Theorem 12, and Proposition 5, we obtain $S(w_n) = u_n = u_{n-1}a\overline{a}u = S(w_{n-1})a\overline{a}S(w)$, where u_n is the n-th palindromic prefix of the corresponding Sturmian sequence and u is the longest palindromic prefix of u_n followed by $\overline{a}$. Consequently, w_n is equal to $w_{n-1}aw$, where w starts with $\mathtt{1}$ (since $S(aw) = \overline{a}u$), ends with $\mathtt{0}$ (since w_n ends with $\mathtt{0}$) and is a pseudopalindrome by Lemma 13.

Thus $w = E(w)$. Therefore, since w is a suffix of $w_n = R(w_n)$, we get $\overline{w}$ is an antipalindromic prefix of w_n followed by a. Moreover, it is the longest antipalindromic prefix with this property because $S(\overline{w}) = u$, where u is the longest palindromic prefix of $S(w_n)$ followed by $\overline{a}$.

2. The proof is similar as before. Since $w_n = w_{n-1}aw$ and w_{n-1} as a palindrome has $\mathtt{0}$ as both the first and the last letter, then by Lemma 13, Theorem 12, and Proposition 5, we obtain $S(w_n) = u_{n-1}a\overline{a}u$, where u is the longest palindromic prefix of u_n followed by $\overline{a}$. Consequently, w_n is equal to $w_{n-1}aw$, where w starts with $\mathtt{1}$ (since $S(aw) = \overline{a}u$), ends with $\mathtt{1}$ (since w_n ends with $\mathtt{1}$) and is a pseudopalindrome by Lemma 13. Therefore, since w is a suffix of $w_n = E(w_n)$, we get $\overline{w}$ is the longest palindromic prefix of w_n followed by $\overline{a}$.

3. Since $w_n = w_{n-1}aw$ and w_{n-1} as an antipalindrome has $\mathtt{1}$ as the last letter, then by Lemma 13, Theorem 12, and Proposition 5, we obtain $S(w_n) = u_{n-1}\overline{a}au$, where u is the longest palindromic prefix of u_n followed by a. Consequently, w_n is equal to $w_{n-1}aw$, where w starts with $\mathtt{0}$, ends with $\mathtt{0}$ and is a pseudopalindrome by Lemma 13. Therefore, since w is a suffix of $w_n = R(w_n)$, we get w is the longest palindromic prefix of w_n followed by a.

□

Let us state the main theorem describing the minimal string attractors of pseudopalindromic prefixes of standard CS Rote sequences.

Theorem 19. *Let* $\mathbf{u}$ *be a standard CS Rote sequence, then the size of the minimal attractor of any pseudopalindromic prefix equals the number of distinct letters contained in the prefix. More precisely, let* $(w_n)_{n=1}^{\infty}$ *be the sequence of all non-empty pseudopalindromic prefixes of* $\mathbf{u}$ *ordered by length and consider and consider* w_n *containing both letters and let* $w_{n-1}a$ *be a prefix of* w_n*, where* $a \in \{\mathtt{0}, \mathtt{1}\}$*. Then the minimal attractor of the pseudopalindromic prefix* w_n *is of the following form:*

1. *If* $w_n = E(w_n)$ *and* w *is the longest antipalindromic prefix of* w_n *followed by* $\overline{a}$*, then*
$$\Gamma = \{|w|, |w_{n-1}|\}$$
is an attractor of w_n*.*

2. *If* $w_n = R(w_n)$*,* $w_{n-1} = E(w_{n-1})$*, and* w *is the longest palindromic prefix of* w_n *followed by* $\overline{a}$*, then*
$$\Gamma = \{|w|, |w_{n-1}|\}$$
is an attractor of w_n*.*

3. *If* $w_n = R(w_n)$*,* $w_{n-1} = R(w_{n-1})$*, and* m *is the minimum index satisfying that* $w_i = R(w_i)$ *for all* $i \in \{m, \ldots, n\}$*, then the attractor of* w_m *from Item 2 is an attractor of* w_n*.*

Proof: First of all, Theorem 15 describes the form of the unique bi-sequence (Δ, Θ) satisfying that the pseudopalindromic prefixes w_n of $\mathbf{u}$ correspond to the prefixes w_n given by Definition 3. It follows that Θ has to start with R. Let us assume without loss of generality that $(\mathtt{0}, R)$ is the first element of (Δ, Θ). If a pseudopalindromic prefix contains one letter, then any position is its attractor. Further on, let us consider pseudopalindromic prefixes w_n containing two distinct letters. Let us proceed by induction on n.

Consider the first pseudopalindromic prefix w_{k+1} containing both letters $\mathtt{0}$ and $\mathtt{1}$. Then by Theorem 15 $(\mathtt{0}^k\mathtt{1}, R^kE)$ is a prefix of (Δ, Θ) and $k \geq 1$. Then $w_{k+1} = \mathtt{0}^k\mathtt{1}^k$ and the longest antipalindromic prefix of w_{k+1} followed by $\mathtt{0}$ is $w_0 = \varepsilon$. Hence $\Gamma = \{0, k\}$ is clearly an attractor of $w_{k+1} = \underline{\mathtt{0}}\mathtt{0}^{k-1}\underline{\mathtt{1}}\mathtt{1}^{k-1}$ (we underlined the positions of Γ).

Assume that for some $n \geq k+1$, we have $w_n = E(w_n)$ and $w_{n-1}a$ is a prefix of w_n for $a \in \{\mathtt{0}, \mathtt{1}\}$ and the claim on the attractor holds, *i.e.*, w_n has the attractor

$$\Gamma_1 \quad = \quad \{|w_i|, |w_{n-1}|\}\,, \tag{3}$$

where w_i is the longest antipalindromic prefix of w_n followed by $\overline{a}$.

Let us assume $\vartheta_n = \vartheta_{n+m+1} = E$ (by Theorems 15 and 16 such an integer m exists and $m \geq 1$), while $\vartheta_\ell = R$ for all $\ell \in \{n+1, \ldots, n+m\}$. We will show that under this assumption, w_{n+1} up to w_{n+m+1} have also the attractors as described in Theorem 19. This will prove the theorem completely.

There are four situations to be considered according to Theorem 15.

1. $m = 1$ and $(\delta_{n-1}\delta_n\delta_{n+1}\delta_{n+2}, \vartheta_{n-1}\vartheta_n\vartheta_{n+1}\vartheta_{n+2}) = (\overline{a}aa\overline{a}, RERE)$;
2. $m \geq 2$ and $(\delta_{n-1}\delta_n\delta_{n+1}\cdots\delta_{n+m}\delta_{n+m+1}, \vartheta_{n-1}\vartheta_n\vartheta_{n+1}\cdots\vartheta_{n+m}\vartheta_{n+m+1}) = (\overline{a}a^{m+1}\overline{a}, RER^mE)$;
3. $m = 1$ and $(\delta_{n-1}\delta_n\delta_{n+1}\delta_{n+2}, \vartheta_{n-1}\vartheta_n\vartheta_{n+1}\vartheta_{n+2}) = (\overline{a}a\overline{a}a, RERE)$;
4. $m \geq 2$ and $(\delta_{n-1}\delta_n\delta_{n+1}\cdots\delta_{n+m}\delta_{n+m+1}, \vartheta_{n-1}\vartheta_n\vartheta_{n+1}\cdots\vartheta_{n+m}\vartheta_{n+m+1}) = (\overline{a}a\overline{a}^m a, RER^mE)$.

We will treat the first two of them. The remaining ones are analogous. In both cases, using Lemma 18, we have

$$w_n = w_{n-1}a\overline{w_j}\,, \tag{4}$$

$$w_n = E(w_n) = w_j\overline{aw_{n-1}}\,, \tag{5}$$

where w_j is the longest palindromic prefix of w_n followed by $\overline{a}$. Since $w_n = (w_{n-1}a)^E$, it follows by (4) and (5) that $w_{n-1} = w_j\overline{a}x$, where x is the longest antipalindromic suffix of w_{n-1} preceded by $\overline{a}$, or equivalently, $\overline{x}$ is the longest antipalindromic prefix of w_{n-1} followed by $\overline{a}$, *i.e.*, $\overline{x} = w_i$, as defined in (3). Hence, $w_{n-1} = w_j\overline{aw_i} = w_i\overline{a}w_j$, where we used palindromicity of w_{n-1} and w_j and antipalindromicity of w_i in the last equality. Therefore, we get the following expressions for w_n, where we underlined the positions of the attractor Γ_1 of w_n from (3) (the first line) and the mirror image attractor ${\Gamma_1}^R$ from Observation 17 (the second line):

$$\begin{aligned} w_n &= w_{n-1}a\overline{w_j} = w_j\overline{aw_i}a\overline{w_j} = w_i\underline{\overline{a}}w_j\underline{a}\overline{w_j}\,,\\ w_n &= w_j\overline{aw_{n-1}} = w_j\underline{\overline{a}}\overline{w_j}\underline{a}w_i = w_j\overline{aw_i}a\overline{w_j}\,. \end{aligned} \tag{6}$$

1. $m = 1$ and $(\delta_{n-1}\delta_n\delta_{n+1}\delta_{n+2}, \vartheta_{n-1}\vartheta_n\vartheta_{n+1}\vartheta_{n+2}) = (\overline{a}aa\overline{a}, RERE)$:
   - Using the definition of palindromic closure and (4), we obtain

$$w_{n+1} = (w_na)^R = w_{n-1}a\overline{w_j}aw_{n-1} = w_naw_{n-1}\,. \tag{7}$$

We will show that $\Gamma_2 = \{|w_j|, |w_n|\}$ is an attractor of w_{n+1} – see the corresponding positions underlined (we rewrite w_n by (5)):

$$w_{n+1} = w_j\overline{\underline{a}w_{n-1}}\underline{a}w_{n-1}\,. \tag{8}$$

By (7), each factor f of w_{n+1} either crosses $|w_n|$, *i.e.*, the underlined a in (8), or is entirely contained in w_n. In that case, consider the mirror image attractor ${\Gamma_1}^R = \{|w_j|, |w_n|-|w_i|-1\}$ of w_n. Using (6) we can rewrite w_{n+1} as

$$w_{n+1} = \underbrace{w_j\overbrace{\underline{\overline{a}}\overline{w_j}aw_i}^{\overline{w_{n-1}}}}_{w_n}\underline{a}w_{n-1} = \underbrace{w_j\underline{\overline{a}}\overline{w_i}a}_{w_n}\overbrace{\overline{w_j}\underline{a}w_i}^{\overline{w_{n-1}}}\overline{a}w_j\,, \tag{9}$$

where we underlined the positions from Γ_2. From this form, we can see that f contained in w_n either crosses $|w_j|$, *i.e.*, the underlined $\overline{a}$, or f crosses the position $|w_n| - |w_i| - 1$ in (9) and is contained in the word $\overline{w_{n-1}} = \overline{w_j}\underline{a}w_i$, where we underlined the position crossed by f in $\overline{w_{n-1}}$. Observing the right-hand form of w_{n+1} in (9), the factor f has then an occurrence containing the position $|w_n|$ in w_{n+1}, *i.e.*, the underlined a.

- Next, using antipalindromic closure, we obtain

$$w_{n+2} = (w_{n+1}\overline{a})^E = \overbrace{w_{n-1}a\overline{w_n}}^{w_{n+1}}\,\overline{aw_{n-1}}\,. \tag{10}$$

We will show that $\Gamma_3 = \{|w_n|, |w_{n+1}|\}$ is an attractor of w_{n+2}. Each factor f of $w_{n+2} = w_n\underline{a}w_{n-1}\overline{\underline{a}w_{n-1}}$ (we underlined the positions of Γ_3) either crosses $|w_{n+1}|$, *i.e.*, the underlined $\overline{a}$, or is entirely contained in $w_{n+1} = w_naw_{n-1}$. In this case, we can write w_{n+2} as

$$w_{n+2} = \overbrace{\underbrace{w_j\overline{aw_{n-1}}}_{w_n}\underline{a}\underbrace{w_i\overline{a}w_j}_{w_{n-1}}}^{w_{n+1}}\overline{\underline{a}w_{n-1}}\,. \tag{11}$$

By (8), f then either crosses $|w_n|$, *i.e.*, the underlined a, or f is contained in $w_n = w_j\overline{aw_{n-1}}$ and crosses $|w_j|$. However, since w_n forms also a suffix of w_{n+2}, the factor f has an occurrence in w_{n+2} containing the position $|w_{n+1}|$, *i.e.*, the underlined $\overline{a}$.

2. $m \geq 2$ and $(\delta_{n-1}\delta_n\delta_{n+1}\cdots\delta_{n+m}\delta_{n+m+1}, \vartheta_{n-1}\vartheta_n\vartheta_{n+1}\cdots\vartheta_{n+m}\vartheta_{n+m+1}) = (\overline{a}a^{m+1}\overline{a}, RER^mE)$: The proof that $\Gamma_2 = \{|w_j|, |w_n|\}$ is an attractor of w_{n+1} stays the same as above.

- Using the definition of palindromic closure, since w_{n-1} is the longest palindromic prefix of w_{n+1} followed by a, we obtain

$$w_{n+2} = (w_{n+1}a)^R = w_naw_{n-1}a\overline{w_n} = w_naw_{n+1}\,. \tag{12}$$

We will show that $\Gamma_2 = \{|w_j|, |w_n|\}$ is an attractor of w_{n+2}. Indeed, each factor f either crosses $|w_n|$ or by (12) f is entirely contained in w_{n+1}. Since w_{n+1} is a prefix of w_{n+2}, the factor f has an occurrence containing an element of Γ_2.
Similarly, for $k \in \{3, \ldots, m\}$, we have $w_{n+k} = w_naw_{n+k-1}$. The attractor of w_{n+k} is again equal to Γ_2: each factor f either crosses $|w_n|$, or f is entirely contained in the prefix w_{n+k-1} of w_{n+k} and the attractor of w_{n+k-1} is by induction assumption $\Gamma_2 = \{|w_j|, |w_n|\}$.

- Using the definition of the antipalindromic closure, since w_n is the longest antipalindromic prefix followed by a, we have

$$w_{n+m+1} \quad = \quad (w_{n+m}\overline{a})^E = w_{n+m-1}a\overline{w_n a w_{n+m-1}} = w_{n+m}\overline{a w_{n+m-1}}\,.$$

We will show that $\Gamma_3 = \{|w_n|, |w_{n+m}|\}$ is an attractor of w_{n+m+1}. Each factor f of $w_{n+m+1} = w_n\underline{a}w_{n+m-1}\overline{\underline{a}w_{n+m-1}}$ (we underlined the positions of the expected attractor Γ_3) either crosses $|w_{n+m}|$, *i.e.*, the underlined $\overline{a}$, or f is entirely contained in w_{n+m} or in $\overline{w_{n+m-1}}$. The word w_{n+m+1} can be expressed as

$$w_{n+m+1} = \underbrace{\overbrace{w_n\underline{a}w_k}^{w_{n+m-1}}a\overline{w_{n-1}a}w_j}_{w_{n+m}}\underbrace{\overbrace{\overline{\underline{a}w_{n-1}}}^{w_n}aw_j\overline{aw_k}}_{\overline{w_{n+m-1}}}\,, \tag{13}$$

$$w_{n+m+1} = \overbrace{w_{n-1}a\overline{w_j}\underline{a}w_k}^{w_{n+m-1}}a\underbrace{\overbrace{\overline{w_{n-1}a}w_j}^{\overline{w_n}}\overline{\underline{a}w_k a}}_{\overline{w_{n+m-1}}}w_n\,, \tag{14}$$

where $k = n+m-2$ if $m \geq 3$ and $k = n-1$ if $m = 2$.

If f is contained in w_{n+m}, then f crosses the attractor $\Gamma_2 = \{|w_j|, |w_n|\}$ of w_{n+m}. It means that either f crosses $|w_n|$, *i.e.*, the underlined a (see (13)), or f is contained in $w_n = w_j\overline{\underline{a}w_{n-1}}$, where we underlined the position in w_n crossed by f. Then by (13) the factor f has an occurrence in w_{n+m+1} containing $|w_{n+m}|$, *i.e.*, the underlined $\overline{a}$.
If f is contained in $\overline{w_{n+m-1}}$, then f crosses its attractor $\Gamma_2 = \{|w_j|, |w_n|\}$ (taking the positions only in $\overline{w_{n+m-1}}$). By (14) the factor f either crosses $|w_n|$ in $\overline{w_{n+m-1}}$, which means that f crosses $|w_{n+m}|$ in w_{n+m+1}, *i.e.*, the underlined $\overline{a}$, or f is contained in $\overline{w_n} = \overline{w_j}\underline{a}w_{n-1}$ (a prefix of $\overline{w_{n+m-1}}$), where we underlined the position crossed by f. However, then f has an occurrence in w_{n+m+1} containing $|w_n|$, *i.e.*, the underlined a, as can be observed from (14) (since w_{n-1} is a prefix of w_k).

□

Example 20. Let us consider a standard CS Rote sequence with the bi-sequence starting with $(\texttt{0011001}, RRERERE)$, *i.e.*, corresponding to the situation of $m = 1$ treated in the proof of Theorem 19. The attractors' positions given in Theorem 19 for prefixes w_n containing both letters are underlined.

$$
\begin{aligned}
w_1 &= \mathtt{0}\\
w_2 &= \mathtt{00}\\
w_3 &= \overbrace{\underline{\mathtt{0}}\mathtt{0}}^{w_2}\,\underline{\mathtt{1}}\mathtt{1}\\
w_4 &= \underbrace{\overbrace{\mathtt{0}}^{w_1}\,\underline{\mathtt{0}}\mathtt{11}}_{w_3}\underline{\mathtt{1}}\mathtt{00}\\
w_5 &= \underbrace{\overbrace{\mathtt{0011}}^{w_3}\underline{\mathtt{1}}\mathtt{00}}_{w_4}\underline{\mathtt{0}}\mathtt{11}\\
w_6 &= \underbrace{\overbrace{\mathtt{00}}^{w_2}\,\underline{\mathtt{1}}\mathtt{1100011}}_{w_5}\underline{\mathtt{0}}\mathtt{0011100}\\
w_7 &= \underbrace{\overbrace{\mathtt{0011100011}}^{w_5}\underline{\mathtt{0}}\mathtt{0011100}}_{w_6}\underline{\mathtt{1}}\mathtt{1100011}
\end{aligned}
$$

For illustration of the case $m \geq 2$ from the proof of Theorem 19, let us consider a bi-sequence starting with $(\mathtt{001100001}, RRERERRRE)$. The steps for $n \leq 6$ are identical with the previous example.

$$
\begin{aligned}
w_1 &= \mathtt{0}\\
w_2 &= \mathtt{00}\\
w_3 &= \underline{\mathtt{0}}\mathtt{0}\underline{\mathtt{1}}\mathtt{1}\\
w_4 &= \mathtt{0}\underline{\mathtt{0}}\mathtt{11}\underline{\mathtt{1}}\mathtt{00}\\
w_5 &= \mathtt{0011}\underline{\mathtt{1}}\mathtt{00}\underline{\mathtt{0}}\mathtt{11}\\
w_6 &= \underbrace{\overbrace{\mathtt{00}}^{w_2}\,\underline{\mathtt{1}}\mathtt{1100011}}_{w_5}\underline{\mathtt{0}}\mathtt{0011100}\\
w_7 &= \underbrace{\overbrace{\mathtt{00}}^{w_2}\,\underline{\mathtt{1}}\mathtt{1100011}}_{w_5}\underline{\mathtt{0}}\mathtt{001110001100011100}\\
w_8 &= \underbrace{\overbrace{\mathtt{00}}^{w_2}\,\underline{\mathtt{1}}\mathtt{1100011}}_{w_5}\underline{\mathtt{0}}\mathtt{00111000110001110001100011100}\\
w_9 &= \underbrace{\overbrace{\mathtt{0011100011}}^{w_5}\underline{\mathtt{0}}\mathtt{00111000110001110001100011100}}_{w_8}\underline{\mathtt{1}}\mathtt{11000111001110001110011100011}
\end{aligned}
$$

7 Open problems

It still remains an open problem to find string attractors of pseudopalindromic prefixes w_n of binary generalized pseudostandard sequences. Recall partial steps done in this paper and also those done previously. For standard Sturmian sequences, the minimal attractors of factors containing two letters are of size two. For pseudostandard sequences we have described attractors of size three for antipalindromic prefixes and we have shown that, up to an exceptional case, they are minimal. For pseudopalindromic prefixes of standard CS Rote sequences we have found attractors of size two. In both previous cases, we studied only pseudopalindromic prefixes, neither prefixes nor factors in general. Even if we keep restricting our focus to pseudopalindromic prefixes of binary generalized pseudostandard sequences, the attractor may be larger. For instance, the minimal attractors of pseudopalindromic prefixes of length at least 8 of the Thue-Morse sequence are of size four [12]. Based on computer experiments, we conjecture that the size of minimal string attractors of pseudopalindromic prefixes of binary generalized pseudostandard sequences is of size at most four. An even more demanding open problem is to study attractors of d-ary generalized pseudostandard sequences for $d > 2$ (defined in [3] and studied in [9]).

8 Acknowledgements

We would like to thank Karel Břinda for his great help with the programming part. We address our cordial thanks also to Francesco Dolce and Edita Pelantová for their careful reading of our manuscript. Last, but not least, we express our thanks to two anonymous referees whose thorough reports helped improve the paper.

References

[1] J. Berstel and P. Séébold. Sturmian words. In M. Lothaire, editor, *Algebraic Combinatorics on Words*, volume 90 of *Encyclopedia of Mathematics and Its Applications*, pages 45–110. Cambridge University Press, 2002.

[2] A. Blondin-Massé, G. Paquin, H. Tremblay, and L. Vuillon. On generalized pseudostandard words over binary alphabet. *J. Integer Seq.*, 16:Article 13.2.11, 2013.

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

[4] A. de Luca and F. Mignosi. Some combinatorial properties of Sturmian words. *Theoretical Computer Science*, 136(2):361–385, 1994.

[5] F. Dolce. String attractors for factors of the Thue-Morse word. In A. Frid and R. Mercaş, editors, *Combinatorics on Words*, pages 117–129, Cham, 2023. Springer Nature Switzerland.

[6] L. Dvořáková. String attractors of episturmian sequences. *Theoretical Computer Science*, 986:114341, 2024.

[7] L. Dvořáková and J. Florian. On periodicity of generalized pseudostandard words. *Electronic Journal of Combinatorics*, 23:P1.2, 2016.

[8] F. Gheeraert, G. Romana, and M. Stipulanti. String attractors of fixed points of k-bonacci-like morphisms. In A. Frid and R. Mercaş, editors, *Combinatorics on Words*, pages 192–205, Cham, 2023. Springer Nature Switzerland.

[9] T. Jajcayová, E. Pelantová, and Š. Starosta. Palindromic closures using multiple antimorphisms. *Theoretical Computer Science*, 533:37–45, 2014.

[10] D. Kempa and N. Prezza. At the roots of dictionary compression: String attractors. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, page 827–840, New York, NY, USA, 2018. Association for Computing Machinery.

[11] T. Kociumaka, G. Navarro, and N. Prezza. Towards a definitive measure of repetitiveness. In *LATIN 2020: Theoretical Informatics: 14th Latin American Symposium, São Paulo, Brazil, January 5-8, 2021, Proceedings*, page 207–219, Berlin, Heidelberg, 2021. Springer-Verlag.

[12] K. Kutsukake, T. Matsumoto, Y. Nakashima, S. Inenaga, H. Bannai, and M. Takeda. On repetitiveness measures of Thue-Morse words, 2020.

[13] S. Mantaci, A. Restivo, G. Romana, G. Rosone, and M. Sciortino. A combinatorial view on string attractors. *Theoretical Computer Science*, 850:236–248, 2021.

[14] G. Rote. Sequences with subword complexity $2n$. *Journal of Number Theory*, 46:196–213, 1993.

[15] L. Schaeffer and J. Shallit. String attractors for automatic sequences, 2020.