Journal of Integer Sequences, Vol. 16 (2013),
Article 13.2.5

Some Properties of Abelian Return Words

M. Rigo, P. Salimov,[1] and E. Vandomme
Department of Mathematics
University of Liège
Grande traverse 12 (B37)
B-4000 Liège
Belgium
M.Rigo@ulg.ac.be E.Vandomme@ulg.ac.be

Abstract

We investigate some properties of abelian return words as recently introduced by Puzynina and Zamboni. In particular, we obtain a characterization of Sturmian words with nonzero intercept in terms of the finiteness of the set of abelian return words to all prefixes. We describe this set of abelian returns for the Fibonacci word but also for the 2-automatic Thue–Morse word. We also investigate the relationship existing between abelian complexity and finiteness of the set of abelian returns to all prefixes.

1 Introduction

Many notions occurring in combinatorics on words have been recently and fruitfully extended to an abelian context. Two words u and v are said to be *abelian equivalent* if u is a permutation of the letters in v and usually, the corresponding concepts are defined up to such an equivalence. This framework gives rise to many challenging questions in combinatorics on words: what kind of information is preserved in the abelian context? To what extent can the classical results be applied? What kind of characterization can we obtain? For instance, consider the classical notion of *factor complexity* $p_{\mathbf{x}}$ which maps an integer $n \geq 0$ to the number of distinct factors of length n occurring in an infinite word $\mathbf{x}$. The well-known theorem of Morse–Hedlund gives a characterization of the ultimately periodic words. See for instance [13]. Sturmian words are defined by the property $p_{\mathbf{x}}(n) = n + 1$ for all $n \geq 0$. The analogue to factor complexity is the *abelian complexity* of $\mathbf{x}$ which maps $n \geq 0$ to the number of distinct abelian classes partitioning the set of factors of length n occurring in $\mathbf{x}$.

[1]The second author is supported by the Russian President's grant no. MK-4075.2012.1, the Russian Foundation for Basic Research grant no. 11-01-00997 and by a University of Liège post-doctoral grant.

This latter notion was already introduced in the 1970's [5]. Some other important questions in combinatorics on words such as avoiding abelian repetitions, were initiated at the same period. See for instance [6]. See also the reference [18] on abelian complexity which contains many relevant bibliographic pointers.

The return word is a classical notion in combinatorics on words and symbolic dynamical systems [11, 12, 20]. For instance, Durand obtained a characterization of primitive substitutive sequences in terms of return words and derived sequences [9]. Let u be a recurrent factor of $\mathbf{x}$, i.e., a factor occurring infinitely many times in $\mathbf{x}$. A *return word* to u is a factor separating two consecutive occurrences of u. In this paper, we consider the abelian analogue of this notion of return word. Such a study has been recently presented by Puzynina and Zamboni during the WORDS 2011 conference. Here we focus on different aspects of abelian returns and we hope that our results can be seen as complementary to those found by Puzynina and Zamboni [17].

The main difference is that we usually consider the set of abelian returns with respect to all the factors of an infinite word $\mathbf{x}$, while Puzynina and Zamboni [17] study the set of abelian returns with respect to each factor taken separately. In particular, their main contribution is a characterization of Sturmian words: *a recurrent infinite word is Sturmian if and only if each of its factors has two or three abelian returns*. Sturmian words have been extensively studied and, in particular, every Sturmian word can be obtained by coding the orbit $(R_\alpha^n(\rho))_{n\geq 0}$ of a point ρ on a circle under a rotation R_α by an irrational angle α when the circle is partitioned in a suitable way into two complementary intervals. The irrational α is called the *slope* of the Sturmian word and the initial point ρ is its *intercept*. See for instance [14, 16]. Many of our results on Sturmian words rely on this definition of Sturmian words.

This paper is organized as follows. In Section 2, we present the main definitions and notation used in this paper. In Section 3, we discuss the relationship with periodicity and we prove that a recurrent word is periodic if and only if its set of abelian returns is finite. We also construct an abelian uniformly recurrent word which is not eventually recurrent. In Section 4, we restrict ourselves to the set $\mathcal{APR}_{\mathbf{x}}$ of abelian returns to all prefixes. In particular, this set is finite for any uniformly recurrent and abelian periodic word. We study the special case of the Thue–Morse word $\mathbf{t}$ [1] and show that the set of abelian returns to all prefixes of $\mathbf{t}$ contains 16 elements. Next, we obtain a characterization of Sturmian words with (non)zero intercept as follows. Let $\mathbf{x}$ be a Sturmian word coding an orbit $(R_\alpha^n(\rho))_{n\geq 0}$. The set $\mathcal{APR}_{\mathbf{x}}$ of abelian returns to the prefixes of $\mathbf{x}$ is finite if and only if $\mathbf{x}$ does not have a null intercept (see Theorem 19). The celebrated Fibonacci word $\mathbf{f}$ can be defined with a slope and an intercept both equal to $1/\tau^2$ where τ is the Golden mean. Therefore our result implies that $\mathcal{APR}_{\mathbf{f}}$ is finite. We show that this set contains exactly 5 elements. Interestingly, our developments can be related to the lengths of the palindromic prefixes of $\mathbf{f}$. See for instance [7, 10]. By contrast the set of abelian returns to all prefixes for the word $\mathtt{0}\mathbf{f}$ is infinite. Then we show that if $\mathbf{x}$ is an abelian recurrent word such that $\mathcal{APR}_{\mathbf{x}}$ is finite, then $\mathbf{x}$ has bounded abelian complexity. In the last section of this paper, we introduce the notion of abelian derived sequence. If a word $\mathbf{x}$ is uniformly recurrent, then $\mathbf{x}$ can be factored in terms of abelian returns to a given prefix of $\mathbf{x}$. This gives rise to a coding that allows one to define a new sequence. Contrary to the non-abelian case and the characterization obtained by Durand, the Thue–Morse word is an example of word having infinitely many derived sequences.

2 Preliminaries

An infinite word $\mathbf{x}$ is said to be *recurrent* if every factor u of $\mathbf{x}$ appears infinitely often in $\mathbf{x}$. An infinite word $\mathbf{x}$ is said to be *uniformly recurrent*, if it is recurrent and the distance between any two consecutive occurrences of any of its factors u is bounded by a constant depending only on u. The language of all the finite factors (resp. prefixes) of an infinite word $\mathbf{x}$ is denoted by $\mathrm{Fac}(\mathbf{x})$ (resp. $\mathrm{Pref}(\mathbf{x})$). Let i, j be such that $i \leq j$. The factor $x_i x_{i+1} \cdots x_j$ of $\mathbf{x} = x_0 x_1 \cdots$ is denoted by $\mathbf{x}[i, j]$. The notation $\mathbf{x}[i, i]$ is shortened to $\mathbf{x}_i$.

2.1 Return words

The classical notion of return word has been used by Durand [9] but was previously introduced by Boshernitzan [4] (see also [8] for the notion of induced transformation in a dynamical context). Let u be a prefix of a uniformly recurrent word $\mathbf{x}$. A nonempty factor w of $\mathbf{x}$ is a *return word* to u, if there exists some $i \geq 0$ such that

- $\mathbf{x}[i, i + |w| - 1] = w$,
- $\mathbf{x}[i, i + |u| - 1] = u = \mathbf{x}[i + |w|, i + |w| + |u| - 1]$,
- $\mathbf{x}[i + j, i + j + |u| - 1] \neq u$ for all $j \in \{1, \ldots, |w| - 1\}$.

We denote by $\mathcal{R}_{\mathbf{x},u}$ the set of return words to u. Since $\mathbf{x}$ is uniformly recurrent, this set is finite because the length of the longest return word to u is bounded by the maximal distance between two successive occurrences of u. If we order the return words to u with respect to their first occurrence in x, then the corresponding map is $\Lambda_{\mathbf{x},u} : \mathcal{R}_{\mathbf{x},u} \to \{\mathtt{1}, \ldots, \#(\mathcal{R}_{\mathbf{x},u})\} =: R_{\mathbf{x},u}$. Since $\mathcal{R}_{\mathbf{x},u}$ is a code [9], i.e., any element in $\mathcal{R}^*_{\mathbf{x},u}$ has a unique factorization as return words to u, $\mathbf{x}$ can be written in a unique way as $\mathbf{x} = m_0 m_1 m_2 \cdots$ and the *derived sequence* $\mathcal{D}_u(\mathbf{x})$ is an infinite word $\Lambda_{\mathbf{x},u}(m_0)\Lambda_{\mathbf{x},u}(m_1)\Lambda_{\mathbf{x},u}(m_2) \cdots$ over $R_{\mathbf{x},u}$.

2.2 Abelian returns

Recently, the notion of return words has been generalized to an abelian framework [17]. In this paper, we will distinguish two cases: abelian return to a prefix and abelian return to a factor. We make such a distinction to be able to define in the first case the abelian derived sequence. Let us start with a few definitions.

Let $A = \{a_1, \ldots, a_k\}$ be a k-letter alphabet. We denote by $|w|_{a_i}$ the number of occurrences of the letter a_i in a word $w \in A^*$. The Parikh mapping $\Psi : A^* \to \mathbb{N}^k$ is defined by $\Psi(w) = (|w|_{a_1}, \ldots, |w|_{a_k})$. Let u, v be two finite words of the same length. We say that u and v are *abelian equivalent* and we write $u \sim_{ab} v$ if $\Psi(u) = \Psi(v)$.

An infinite word $\mathbf{x}$ is *abelian periodic (of period m)*, if it can be factored as $\mathbf{x} = u_1 u_2 u_3 \cdots$ where, for all i, j, the finite words u_i and u_j have the same length m and are abelian equivalent. The smallest m for which such a factorization exists is called the *abelian period* of $\mathbf{x}$. For instance, the Thue–Morse word is an infinite concatenation of `01` and `10` and is thus abelian periodic of period 2.

Let $\mathbf{x}$ be an infinite word. If, for each factor u of $\mathbf{x}$, there exist infinitely many i such that $\mathbf{x}[i, i+|u|-1] \sim_{ab} u$, then $\mathbf{x}$ is said to be *abelian recurrent*.

If $\mathbf{x}$ is abelian recurrent and if, for each factor u of $\mathbf{x}$, the distance between any two consecutive occurrences of factors abelian equivalent to u is bounded by a constant depending only on u, then $\mathbf{x}$ is said to be *abelian uniformly recurrent*.

Remark 1. Note that uniform recurrence implies obviously abelian uniform recurrence. We will show in Proposition 6 that the converse does not hold.

Definition 2. Let u be a prefix of an abelian uniformly recurrent word $\mathbf{x}$. We say that a nonempty factor w of $\mathbf{x}$ is an *abelian return* to u, if there exists some $i \geq 0$ such that

- $\mathbf{x}[i, i+|w|-1] = w$,
- $\mathbf{x}[i, i+|u|-1] \sim_{ab} u \sim_{ab} \mathbf{x}[i+|w|, i+|w|+|u|-1]$,
- $\mathbf{x}[i+j, i+j+|u|-1] \not\sim_{ab} u$, for all $j \in \{1, \ldots, |w|-1\}$.

We denote by $\mathcal{APR}_{\mathbf{x},u}$ the set of abelian returns to the prefix u. Since $\mathbf{x}$ is abelian uniformly recurrent, then the set $\mathcal{APR}_{\mathbf{x},u}$ is finite. We define the set of abelian returns to prefixes as

$$\mathcal{APR}_{\mathbf{x}} := \bigcup_{u \in \mathrm{Pref}(\mathbf{x})} \mathcal{APR}_{\mathbf{x},u}.$$

Observe that if $\mathbf{x}$ is uniformly recurrent, then the length of the longest element in $\mathcal{APR}_{\mathbf{x},u}$ is bounded by the length of the longest element in $\mathcal{R}_{\mathbf{x},u}$.

We will also consider a more general situation where u is not restricted to be a prefix of $\mathbf{x}$. Puzynina and Zamboni [17] called this notion a *semi-abelian return* to the abelian class of u and the number of abelian returns is the number of distinct abelian classes of semi-abelian returns.

Definition 3. If $\mathbf{x}$ is abelian recurrent and if u is a factor of $\mathbf{x}$, we can define as above the notion of abelian return to u. The corresponding set $\mathcal{AR}_{\mathbf{x},u}$ of abelian returns to u is well defined. We define the set of abelian returns as

$$\mathcal{AR}_{\mathbf{x}} := \bigcup_{u \in \mathrm{Fac}(\mathbf{x})} \mathcal{AR}_{\mathbf{x},u}.$$

Remark 4. Let $\mathbf{x}$ be an abelian recurrent word. The set $\mathcal{AR}_{\mathbf{x},u}$ is finite, for each factor u of $\mathbf{x}$, if and only if $\mathbf{x}$ is abelian uniformly recurrent.

3 Finiteness of the set of abelian returns

Puzynina and Zamboni [17] provided a discussion between periodicity and the number of abelian returns. Here we take the finiteness of the set of abelian returns to characterize periodicity.

Theorem 5. *Let $\mathbf{x}$ be a recurrent word. The set $\mathcal{AR}_{\mathbf{x}}$ is finite if and only if $\mathbf{x}$ is periodic.*

Proof. The "if" part is obvious. We prove the "only if" part.

Suppose that $\mathcal{AR}_{\mathbf{x}}$ is finite and that $\mathbf{x}$ is recurrent but not periodic. In this case, for each k, there exists a word u satisfying $|u| > k$ such that $au, bu \in \mathrm{Fac}(\mathbf{x})$ for some letters $a \neq b$. Hence there exist i, j such that $i < j$, $\mathbf{x}[i, i + |u|] = au$ and $\mathbf{x}[j, j + |u|] = bu$. Define $v = \mathbf{x}[i, j - 1]$. Since $\mathbf{x}[i + d, j - 1 + d] \not\sim_{ab} v$ for all $d \in \{1, \ldots, |u|\}$, there is an abelian return to v in $\mathbf{x}$ of length at least k. As we can do the same for arbitrarily large k, the set $\mathcal{AR}_{\mathbf{x}}$ is infinite. $\square$

Obviously, uniform recurrence implies abelian uniform recurrence, but the converse is not true. Recall that an *eventually recurrent word* is an infinite word having a recurrent suffix.

Proposition 6. *There exists an abelian uniformly recurrent word which is not eventually recurrent.*

Proof. Let $\mathbf{t} = t_0 t_1 \cdots = \mathtt{01101001} \cdots$ be the Thue–Morse word and φ_{TM} be the Thue–Morse morphism, $\varphi_{TM}(\mathbf{t}) = \mathbf{t}$. Define the set $I = \{i_0 < i_1 < \ldots\}$ of all positions where an isolated $\mathtt{1}$ occurs. That is, for all n, we have $t_{i_n} = \mathtt{1}$ and $t_{i_n - 1} = t_{i_n + 1} = \mathtt{0}$. Moreover we set $J = \{i_{2^k} \mid k > 0\}$.

Let $\mathbf{y} = y_0 y_1 \cdots$ be the word defined by $y_j = \mathtt{0}$, if $j \in J$, and $y_j = t_j$ otherwise. Define $\mathbf{x} = \varphi_{TM}(\mathbf{y})$.

The word $\mathbf{x}$ coincides with $\mathbf{y}$ almost everywhere, except for positions from the set $2J \cup (2J + 1)$. Hence, each factor of the Thue–Morse word occurs in $\mathbf{x}$ uniformly, i.e., with bounded gaps. At the same time, the factor $\varphi_{TM}(\mathtt{000})$ occurs in $\mathbf{x}$ with strictly growing gaps. Hence $\mathbf{x}$ is not eventually recurrent.

Let us now prove that $\mathbf{x}$ is abelian uniformly recurrent. First we point out a property of the Thue–Morse word: for all $d > 0$ and all $a \in \{\mathtt{0}, \mathtt{1}\}$, there exists k such that $t_k = a \neq t_{k+d}$. This property follows from the well-known fact that the Thue–Morse word does not contain any constant infinite arithmetical subsequence [15].

As $\mathbf{x}$ is abelian periodic (of period 2), the weight (i.e., the sum of digits) of each factor u of $\mathbf{x}$ of odd length is either $\frac{|u|+1}{2}$ or $\frac{|u|-1}{2}$. Note that $y_i = \mathtt{0}$ implies $|\mathbf{x}[2i+1, 2i+|u|]|_{\mathtt{1}} = \frac{|u|+1}{2}$ and $y_i = \mathtt{1}$ implies $|\mathbf{x}[2i + 1, 2i + |u|]|_{\mathtt{1}} = \frac{|u|-1}{2}$. Since $\mathtt{0}$ (resp. $\mathtt{1}$) occurs with bounded gaps in $\mathbf{y}$, gaps between abelian occurrences in $\mathbf{x}$ of a factor of odd length are bounded.

The weight of a factor u of even length of $\mathbf{x}$ can take values $\frac{|u|}{2}$, $\frac{|u|}{2} + 1$ and $\frac{|u|}{2} - 1$. The first case takes place when u occurs at an even position in $\mathbf{x}$, meaning that the gaps between abelian occurrences of u of weight $\frac{|u|}{2}$ in $\mathbf{x}$ are bounded. The last two cases take place if u occurs in $\mathbf{x}$ at an odd position i and if $y_{\frac{i-1}{2}} = \mathtt{1}$ and $y_{\frac{i-1+|u|}{2}} = \mathtt{0}$ or, $y_{\frac{i-1}{2}} = \mathtt{0}$ and $y_{\frac{i-1+|u|}{2}} = \mathtt{1}$. Due to the mentioned property of the Thue–Morse word, there exists k such that $t_k = \mathtt{1} \neq t_{k+\frac{|u|}{2}}$ (resp. $t_k = \mathtt{0} \neq t_{k+\frac{|u|}{2}}$) and since $\mathbf{t}$ is uniformly recurrent, the factor

$\mathbf{t}[k, k+\frac{|u|}{2}]$ occurs infinitely often with bounded gaps in $\mathbf{t}$. Hence abelian occurrences of u in $\mathbf{x}$ appear infinitely often with bounded gaps.

$\square$

4 Finiteness of the set of abelian returns to prefixes

Contrary to the finiteness of $\mathcal{AR}_{\mathbf{x}}$, the finiteness of $\mathcal{APR}_{\mathbf{x}}$ does not imply periodicity nor abelian periodicity of $\mathbf{x}$. Moreover, if $\mathbf{x}$ is uniformly recurrent, it is well-known that

$$\min_{v \in \mathcal{R}_{\mathbf{x},u}} |v| \to \infty, \text{ if } |u| \to \infty,$$

meaning that taking longer prefixes eventually leads to longer return words. Here we show that such a result does not hold for abelian returns to prefixes. Indeed, for the Thue–Morse word the corresponding set $\mathcal{APR}_{\mathbf{t}}$ is finite and can be described precisely. Such a result also holds for the Fibonacci word. In particular, amongst the set of Sturmian words, the finiteness of $\mathcal{APR}_{\mathbf{x}}$ characterizes Sturmian words with nonzero intercept.

Lemma 7. *If $\mathbf{x}$ is a uniformly recurrent and abelian periodic word, then the set $\mathcal{APR}_{\mathbf{x}}$ is finite.*

Proof. Let m be the (minimal) abelian period of $\mathbf{x}$. Let us find an upper bound for the length of an abelian return u to a prefix p of $\mathbf{x}$.

Suppose first that $|p| = mk$. In this case, due to abelian periodicity, for all i, we have $\mathbf{x}[mi, m(i+k)-1] \sim_{ab} p$. Hence we get $|u| \leqslant m$.

Suppose now that $|p| = mk + \ell$, where $0 < \ell < m$. Let us denote the word $\mathbf{x}[mk, m(k+1)-1]$ by s. As the word $\mathbf{x}$ is abelian periodic, if there exists i such that the equality $\mathbf{x}[mi, m(i+1)-1] = s$ holds, then $\mathbf{x}[m(i-k), mi+\ell-1] \sim_{ab} p$. Hence, it is sufficient to prove that the set

$$\{i \geq 0 \mid \mathbf{x}[mi, m(i+1)-1] = s\}$$

has bounded gaps.

Let us consider the word $\mathbf{x}'$ on the alphabet of factors of $\mathbf{x}$ of length m, such that $\mathbf{x}'_i = \mathbf{x}[mi, m(i+1)-1]$. It is well-known that the uniform recurrence of $\mathbf{x}$ implies uniform recurrence of $\mathbf{x}'$ (see for instance [19]). Hence, for each letter of $\mathbf{x}'$ there is an upper bound on the gap between two consecutive occurrences of it in $\mathbf{x}'$. Denoting the maximum of such constants by D, we get $|u| \leqslant mD$. $\square$

Remark 8. In Lemma 7, the condition on a word $\mathbf{x}$ to be uniformly recurrent is essential: there exists an abelian periodic word $\mathbf{x}$ which is not uniformly recurrent and such that $\mathcal{APR}_{\mathbf{x},u}$ is infinite for some prefix u of $\mathbf{x}$. Consider the abelian periodic word of period 4 given by $\mathbf{x} = \phi\varphi^{\omega}(\mathtt{0})$ where $\varphi : \mathtt{0} \mapsto \mathtt{010}, \mathtt{1} \mapsto \mathtt{111}$ and $\phi : \mathtt{0} \mapsto \mathtt{01230123}, \mathtt{1} \mapsto \mathtt{0213}$:

$$\mathbf{x} = \mathtt{01230123}\ \ \mathtt{0213}\ \ \mathtt{01230123}\ \ \mathtt{0213}\ \ \mathtt{0213}\ \ \mathtt{0213}\cdots$$

In $\mathbf{x}$ there are unbounded gaps between consecutive abelian occurrences of its prefix $\mathtt{012301}$ that correspond to the occurrences of $\phi(\mathtt{1}^m)$.

Remark 9. In Lemma 7, the condition on abelian periodicity of $\mathbf{x}$ is not necessary to get finiteness of $\mathcal{APR}_{\mathbf{x}}$. We shall give an example below when discussing the case of Sturmian words. Indeed, Sturmian words are not abelian periodic (see Lemma 20) but for instance, the Fibonacci word $\mathbf{f}$ is uniformly recurrent and the corresponding set $\mathcal{APR}_{\mathbf{f}}$ is finite.

Proposition 10. *A word $\mathbf{x}$ is periodic if and only there exists some prefix u such that infinitely many factors of $\mathbf{x}$ are abelian equivalent to u and all the abelian returns in $\mathcal{APR}_{\mathbf{x},u}$ have length* 1.

Proof. If $\mathbf{x} = u^\omega$, then $\mathbf{x}[i, i+|u|-1] \sim_{ab} u$ for all $i \geq 0$. Conversely, if all the abelian returns to some prefix u in $\mathcal{APR}_{\mathbf{x},u}$ have length 1, then $\mathbf{x}[i, i+|u|-1] \sim_{ab} u \sim_{ab} \mathbf{x}[i+1, i+|u|]$ for all $i \geq 0$. There is an abelian return a of length 1 at position i in $\mathbf{x}$ and it also occurs in position $i+|u|$. It follows that $|u|$ is a period of $\mathbf{x}$. □

4.1 Finiteness of $\mathcal{APR}_{\mathbf{t}}$ for the Thue–Morse word

We already know from Lemma 7 that the Thue–Morse word has a finite set of abelian returns to all its prefixes. Here we describe precisely this set.

Lemma 11. *Let $\mathbf{x}$ be a uniformly recurrent word. Let $n \geq 1$ and i, j be such that $i < j$. Assume that $\mathbf{x}[i, i+n-1] \sim_{ab} \mathbf{x}[j, j+n-1]$ and there exists a prefix u of length $j-i$ of $\mathbf{x}$ such that $u \sim_{ab} \mathbf{x}[i, j-1]$. The word $\mathbf{x}[i, i+n-1]$is an occurrence of an abelian return to the prefix u if and only if, for all $\ell \in \{0, \ldots, n-2\}$, $\mathbf{x}[i, i+\ell] \not\sim_{ab} \mathbf{x}[j, j+\ell]$.*

Proof. Since $|u| = j - i$, by assumption we have $\mathbf{x}[i, i+|u|-1] \sim_{ab} u$. Observe first that there exists $\ell \in \{0, \ldots, n-1\}$ such that $\mathbf{x}[i, i+\ell] \sim_{ab} \mathbf{x}[j, j+\ell]$ if and only if $\mathbf{x}[i+\ell+1, i+\ell+|u|] \sim_{ab} u$. In particular, since $\mathbf{x}[i, i+n-1] \sim_{ab} \mathbf{x}[j, j+n-1]$, we get $\mathbf{x}[i+n, i+n+|u|-1] = \mathbf{x}[i+n, j-1+n] \sim_{ab} u$. Moreover, $\ell \in \{0, \ldots, n-2\}$ is such that $\mathbf{x}[i+\ell+1, i+\ell+|u|] \not\sim_{ab} u$ if and only if $\mathbf{x}[i, i+\ell] \not\sim_{ab} \mathbf{x}[j, j+\ell]$. □

Remark 12. From this lemma, we can derive a necessary condition for a word to be an abelian return to a prefix. If a word $w = w_1 \cdots w_n$ of length n is an abelian return to a prefix, then there exists some factor $y = y_1 \cdots y_n$ of $\mathbf{x}$ such that

$$w \sim_{ab} y \text{ and, for all } \ell \in \{1, \ldots, n-1\}, w_1 \cdots w_\ell \not\sim_{ab} y_1 \cdots y_\ell. \tag{1}$$

This condition is not sufficient. For instance, $w = \texttt{001011}$ and $y = \texttt{110010}$ are two factors of length 6 satisfying(1) and occurring in the Thue–Morse word $\mathbf{t}$. But, as shown in the following proposition, w is not an abelian return to any prefix.

Theorem 13. *The set $\mathcal{APR}_{\mathbf{t}}$ of abelian returns to prefixes for the Thue–Morse word $\mathbf{t}$ is*

{0, 1, 01, 10, 001, 011, 100, 110, 0011, 0101, 1010, 1100, 00101, 01011, 10100, 11010}.

Proof. One can check with some computer experiments that the factors given above appear as abelian returns to some prefix of $\mathbf{t}$. Moreover, one can also check that these are the only factors of length $2, \ldots, 5$ in $\mathbf{t}$ satisfying condition (1).

Assume that there exists some abelian return $w = w_1 \cdots w_n = \mathbf{t}[i, i+n-1]$ of length $n \geq 2$ to a prefix of $\mathbf{t}$ occurring at position i. In particular, we may assume that w is an abelian return to the prefix u of length $j - i > 0$ and $y = y_1 \cdots y_n = \mathbf{t}[j, j+n-1]$ satisfies (1). We will show that the length of w is at most 5. Recall that $\mathbf{t}[2k, 2k+1] \in \{\mathtt{01}, \mathtt{10}\}$ for all $k \geq 0$.

Assume first that i, j are even. Since $\mathbf{t}[i, i+1]$ and $\mathbf{t}[j, j+1]$ belong to $\{\mathtt{01}, \mathtt{10}\}$, we conclude that $\mathbf{t}[i, i+1] \sim_{ab} \mathbf{t}[j, j+1]$ and, in that situation, we can only have an abelian return of length at most 2.

Assume now that i is odd and j is even and that $\mathbf{t}_i = \mathtt{0}$ (symmetric cases can be treated in the same way). Our aim is to build the longest possible abelian return. Since $\mathbf{t}_i = \mathtt{0}$ and j is even, we consider $\mathbf{t}[j, j+1] = \mathtt{10}$ because otherwise, $\mathbf{t}[j, j+1] = \mathtt{01}$ and $w_1 = y_1$ (i.e., $\mathbf{t}[i+1, j+1] \sim_{ab} \mathbf{t}[i, j] \sim_{ab} u$ and we get directly an abelian return of length $n = 1$). Now $\mathbf{t}[i, i+2] = \mathtt{001}$ because otherwise, $\mathbf{t}[i, i+2] = \mathtt{010}$ and $w_1 w_2 \sim_{ab} y_1 y_2$. Continuing this way, we have $\mathbf{t}[j, j+3] = \mathtt{1010}$ and $\mathbf{t}[i, i+4] = \mathtt{00101}$. Since $(\mathtt{10})^3$ is not a factor of $\mathbf{t}$, we have $\mathbf{t}[j, j+5] = \mathtt{101001}$ and $\mathbf{t}[i, i+4] \sim_{ab} \mathbf{t}[j, j+4]$. In that situation, we can only have an abelian return of length at most 5.

The last case is when i and j are odd. Assume $\mathbf{t}_i = \mathtt{0}$ and $\mathbf{t}_j = \mathtt{1}$. We have $\mathbf{t}_i = \mathtt{0}$ and $\mathbf{t}_{j-1} = \mathtt{0}$ because $\mathbf{t}[j-1, j] = \mathtt{01}$. Moreover, $z = \mathbf{t}[i+1, j-2] \in \{\mathtt{01}, \mathtt{10}\}^*$ and thus $v = \mathbf{t}[i, j-1] = \mathtt{0}z\mathtt{0}$ is a word of even length such that $|v|_\mathtt{0} = 2 + |v|_\mathtt{1}$. Therefore v cannot be abelian equivalent to a prefix u of $\mathbf{t}$. So in such a situation, we cannot have an abelian return to some prefix of $\mathbf{t}$. ☐

Proposition 14. *If a factor of length $n \geq 6$ of the Thue–Morse word satisfies (1), then n is even.*

Proof. Let $w = \mathbf{t}[i, i+n-1]$ and $y = \mathbf{t}[j, j+n-1]$ be factors of $\mathbf{t}$ of length $n \geq 6$ satisfying (1). As $n \geq 6$, i and j are odd. Hence, to satisfy the condition (1), we must have

$$\begin{pmatrix}\mathbf{t}[i, i+n-1]\\ \mathbf{t}[j, j+n-1]\end{pmatrix} \in \begin{pmatrix}\mathtt{0}\\ \mathtt{1}\end{pmatrix}\left\{\begin{pmatrix}\mathtt{01}\\ \mathtt{01}\end{pmatrix}, \begin{pmatrix}\mathtt{10}\\ \mathtt{10}\end{pmatrix}, \begin{pmatrix}\mathtt{01}\\ \mathtt{10}\end{pmatrix}\right\}^* \begin{pmatrix}\mathtt{1}\\ \mathtt{0}\end{pmatrix} \cup \begin{pmatrix}\mathtt{1}\\ \mathtt{0}\end{pmatrix}\left\{\begin{pmatrix}\mathtt{01}\\ \mathtt{01}\end{pmatrix}, \begin{pmatrix}\mathtt{10}\\ \mathtt{10}\end{pmatrix}, \begin{pmatrix}\mathtt{10}\\ \mathtt{01}\end{pmatrix}\right\}^* \begin{pmatrix}\mathtt{0}\\ \mathtt{1}\end{pmatrix}.$$

So n must be even. ☐

For $n = 6, 8, 10, \ldots, 104$, with a computer search, we get the following number of factors of length n satisfying (1): 6, 4, 8, 12, 12, 4, 8, 8, 4, 0, 0, 8, 0, 0, 4, 8, 4, 0, 0, 0, 0, 0, 4, 0, 4, 0, 0, 0, 0, 0, 4, 8, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 8, 0, 0.

4.2 Sturmian words

Sturmian words form one of the most studied classes of infinite words. It can be defined in several ways. For our purpose, the definition in terms of rotations is convenient.

Let C be the one-dimensional torus $\mathbb{R}/\mathbb{Z}$ identified with the interval $[0, 1)$. As usual, we denote by $\{x\}$ the fractional part of x. The *rotation* R_α defined for a real number α is a mapping $C \to C$ that maps an x to $\{x + \alpha\}$.

Theorem 15 (Kronecker [2])**.** *If a number α is irrational, then the set of points $\{R^i_\alpha(\rho) \mid i \in \mathbb{N}\}$ is dense in C for all initial points $\rho \in C$.*

By an *interval* (resp. *half-interval*) of C we mean a set of points that is an image of an interval (resp. half-interval) of $\mathbb{R}$ under operation $\{\cdot\}$. For instance, if $0 \le b < a < 1$, then $[a, 1) \cup [0, b)$ is denoted by $[a, b)$.

Let α be irrational and ρ be real. Without loss of generality we can assume $0 \le \alpha, \rho < 1$. A *Sturmian word* $\mathbf{x} = St(\alpha, \rho)$ (resp. $\mathbf{x} = St'(\alpha, \rho)$) can be defined as

$$x_i = \begin{cases} \mathtt{0}, & \text{if } R_\alpha^i(\rho) \in I_0; \\ \mathtt{1}, & \text{if } R_\alpha^i(\rho) \in I_1, \end{cases} \tag{2}$$

where $I_0 = [0, 1-\alpha)$ and $I_1 = [1-\alpha, 1)$ (resp. $I_0 = (0, 1-\alpha]$ and $I_1 = (1-\alpha, 1]$). See [2, 14].

If $\rho = 0$, then

$$St(\alpha, 0) = \mathtt{0}\mathbf{c}_\alpha \text{ and } St'(\alpha, 0) = \mathtt{1}\mathbf{c}_\alpha$$

and $\mathbf{c}_\alpha$ is said to be the *characteristic Sturmian word of slope* α [14]. If $\mathbf{x} = St(\alpha, 0)$, we say that $\mathbf{x}$ is a Sturmian word *with null intercept*.

Example 16. If $\tau = (1+\sqrt{5})/2$ is the Golden mean, then $St(1/\tau^2, 0) = \mathtt{0}\mathbf{f}$ where $\mathbf{f}$ is the Fibonacci word $\mathtt{0100101001}\cdots$ which is the unique fixed point of the morphism $\varphi : \mathtt{0} \mapsto \mathtt{01}, \mathtt{1} \mapsto \mathtt{0}$. In particular, we have $\mathbf{f} = St(1/\tau^2, 1/\tau^2)$.

Theorem 17. *[14, Theorem 2.1.5] An infinite word $\mathbf{x} \in \{\mathtt{0}, \mathtt{1}\}^\omega$ is Sturmian if and only if it is aperiodic and balanced, i.e., for all $u, v \in Fac(\mathbf{x})$ of same length, we have $||u|_\mathtt{1} - |v|_\mathtt{1}| \le 1$.*

Let $\mathbf{x} = St(\alpha, \rho)$ be a Sturmian word. For a binary word $v = v_0 v_1 \cdots v_m$, we define a half-interval I_v of C as

$$I_v := I_{v_0} \cap R_\alpha^{-1}(I_{v_1}) \cap \cdots \cap R_\alpha^{-m}(I_{v_m}). \tag{3}$$

Hence $\mathbf{x}[i, i+m] = v$ if and only if $R_\alpha^i(\rho) \in I_v$. See [14, Section 2.1.2].

Definition 18. Let $\mathbf{x} = St(\alpha, \rho)$ be a Sturmian word. For each k the number of $\mathtt{1}$'s in a factor of length k in $\mathbf{x}$ can only take the values $\lceil k\alpha \rceil$ or $\lceil k\alpha \rceil - 1$. The corresponding factors will be called respectively *heavy* and *light*. If $\mathbf{x}$ is understood from the context, $H(k)$ (resp. $L(k)$) will denote the set of heavy (resp. light) factors of length k in $\mathbf{x}$. Define

$$I_H(k) := \bigcup_{v \in H(k)} I_v \text{ and } I_L(k) := \bigcup_{v \in L(k)} I_v.$$

So, the word $\mathbf{x}[i, i+k-1]$ is heavy if and only if $R_\alpha^i(\rho) \in I_H(k)$.

Theorem 19. *Let $\mathbf{x}$ be a Sturmian word. The set $\mathcal{APR}_\mathbf{x}$ is finite if and only if $\mathbf{x}$ does not have a null intercept.*

Proof. For the sake of convenience, let $\mathbf{x}$ be defined as $St(\alpha, \rho)$ for half-intervals $I_0 = [0, 1-\alpha)$ and $I_1 = [1-\alpha, 1)$. Let us prove by induction on $k \ge 1$ that

$$I_H(k) = [1 - \{k\alpha\}, 1) \text{ and } I_L(k) = [0, 1 - \{k\alpha\}). \tag{4}$$

It holds true for $k = 1$. Suppose now that the statement holds true for some $k \ge 1$. We consider two cases.

- Assume that $0 \notin R_\alpha^{-k}(I_1)$. Therefore we get $R_\alpha^{-k}(I_1) = [1-\{(k+1)\alpha\}, 1-\{k\alpha\})$ with $1-\{(k+1)\alpha\} < 1-\{k\alpha\}$. By the induction hypothesis, we have $I_H(k) = [1-\{k\alpha\}, 1)$ and consequently,
$$R_\alpha^{-k}(I_1) \cap I_H(k) = \emptyset.$$
This means that all the heavy factors of length k of $\mathbf{x}$ can only be extended with `0` to factors of length $k+1$ of $\mathbf{x}$. In particular, the weights of heavy factors of length k and $k+1$ are the same. At the same time, we have $R_\alpha^{-k}(I_1) \cap I_L(k) = R_\alpha^{-k}(I_1)$, which means that the factors corresponding to elements belonging to this latter set are the light factors of length k that are extended with `1` to heavy factors of length $k+1$. We conclude that
$$I_H(k+1) = I_H(k) \cup R_\alpha^{-k}(I_1) = [1-\{(k+1)\alpha\}, 1)$$
and $I_L(k+1) = I_L(k) \backslash R_\alpha^{-k}(I_1) = [0, 1-\{(k+1)\alpha\})$.

- Assume now that $0 \in R_\alpha^{-k}(I_1)$, i.e., $1-\{(k+1)\alpha\} > 1-\{k\alpha\}$. In this case, using again the induction hypothesis, $R_\alpha^{-k}(I_1) \cap I_H(k) = [1-\{(k+1)\alpha\}, 1)$ is nonempty. This interval corresponds to the heavy factors of length k having an extension with `1` making them the only heavy factors of length $k+1$ in $\mathbf{x}$.

Now we are ready to prove the main part of the statement. First of all, let us prove that, if $\mathbf{x}$ has a null intercept, then $\mathcal{APR}_\mathbf{x}$ is infinite. Let $k \geq 1$ and p be the prefix of length k of $\mathbf{x}$. As $\rho = 0$, we have $0 \in I_p$. Since the interval I_p corresponds exactly to one word p which is either light or heavy, we have $I_p \subseteq I_L(k)$ or $I_p \subseteq I_H(k)$. As $0 \in I_p$, we conclude that $I_p \subseteq I_L(k)$ using (4). In other words, we have just shown that each prefix of $\mathbf{x}$ is a light factor.

Now we show that, for all n, there exists a length ℓ such that gaps between consecutive occurrences of light factors of length ℓ in $\mathbf{x}$ can be larger than n. Let $n \geq 1$. Define the set of points
$$S_n := \{R_\alpha^i(0) \mid 0 \leqslant i \leqslant n\}$$
and denote by d the minimal length of intervals having endpoints in S_n. Due to Kronecker's theorem, we can find some ℓ such that $|I_L(\ell)| < d$ and it follows that $I_L(\ell) \cap S_n = \{0\}$. With our definitions, it means that the light prefix of $\mathbf{x}$ of length ℓ is followed by at least n heavy consecutive factors of length ℓ. Since this can be done for any n, the set $\mathcal{APR}_\mathbf{x}$ is infinite.

Let us prove that, if $\mathbf{x}$ does not have a null intercept, then $\mathcal{APR}_\mathbf{x}$ is finite. The main difference with the previous situation is that a prefix can now be a heavy or a light factor depending on its length: $\rho \geq 1-\{k\alpha\}$ if and only the prefix of length k is heavy. We will show that there exists a constant c such that, for all prefixes p of $\mathbf{x}$, the gap between consecutive occurrences of factors abelian equivalent to p is bounded by c.

Let $n \geq 1$. Consider as before the set S_n and order its elements $0 = s_0 < s_1 < \cdots < s_n$. Denote by $D(n)$ the maximal length of the intervals $[s_0, s_1)$, $\ldots$, $[s_{n-1}, s_n)$, $[s_n, s_0)$ whose endpoints are consecutive points in S_n. Due to Kronecker's theorem, there exists some c such that $2D(c) < \min\{\rho, 1-\rho\}$.

Suppose that the prefix of length k of $\mathbf{x}$ is a light word. Then we have $\rho \in I_L(k)$ and, consequently, $|I_L(k)| > \rho$. Assume that there is a light factor of length k occurring at position i in $\mathbf{x}$, i.e., $R^i_\alpha(\rho) \in I_L(k)$. We consider two cases. If $R^i_\alpha(\rho) \geq D(c)$, there exists $j \in \{1, \ldots, c\}$ such that $R^j_\alpha(0) = s_c$ and $\theta = 1 - R^j_\alpha(0) \in (0, D(c)]$. Hence the point $R^j_\alpha(R^i_\alpha(\rho)) = R^{i+j}_\alpha(\rho) = R^i_\alpha(\rho) - \theta$ belongs to $I_L(k)$, i.e., the factor of length k at position $i + j$ in $\mathbf{x}$ is light again. If $R^i_\alpha(\rho) < D(c)$, there exists $j \in \{1, \ldots, c\}$ such that $R^j_\alpha(0) = s_1 \leq D(c) < \rho/2$. Hence the point $R^j_\alpha(R^i_\alpha(\rho)) = R^{i+j}_\alpha(\rho) = R^i_\alpha(\rho) + s_1$ is less than ρ and belongs to $I_L(k)$.

A similar proof can be done for the case of a heavy prefix of length k. Assume that $\rho \in I_H(k)$ and that, for some $i \geq 0$, $R^i_\alpha(\rho) \in I_H(k)$. If $R^i_\alpha(\rho) < 1-D(c)$, then $R^i_\alpha(\rho)+s_1 < 1$. If $R^i_\alpha(\rho) \geq 1-D(c)$, then $R^i_\alpha(\rho)-(1-s_c) \geq 1-2D(c) > \rho$. We can derive the same conclusion as above.

Hence the number c is an upper bound on the length of abelian returns to any prefix and therefore $\mathcal{APR}_\mathbf{x}$ is finite. $\square$

Lemma 20. *No Sturmian word is abelian periodic.*

Proof. Proceed by contradiction and assume that $\mathbf{x} = St(\alpha, \rho)$ is abelian periodic of period m with α irrational. Then all factors of the kind $\mathbf{x}[tm, (t+1)m-1]$, $t \in \mathbb{N}$, are abelian equivalent, i.e., have the same weight. Assume that, for all t, $R^{tm}_\alpha(\rho) = R^t_{m\alpha}(\rho) \in I_L(m)$. But since α is irrational, $m\alpha$ is also irrational and thanks to Kronecker's theorem, $\{R^t_{m\alpha}(\rho) \mid t \geq 0\}$ is dense in C contradicting the fact that $\{R^t_{m\alpha}(\rho) \mid t \geq 0\} \cap I_H(m)$ should be empty. $\square$

4.3 Finiteness of $\mathcal{APR}_\mathbf{f}$ for the Fibonacci word

From Theorem 19, since the Fibonacci word $\mathbf{f}$ is given by $St(1/\tau^2, 1/\tau^2)$, we already know that $\mathcal{APR}_\mathbf{f}$ is finite. Here we exhibit exactly the elements of this set in Theorem 22. As a first attempt, (1) gives a necessary condition that allows one to exclude some words as abelian returns. This condition will not be used in the proof of Theorem 22 but, interestingly, our developments can be related to the lengths of the palindromic prefixes of $\mathbf{f}$, [7, 10].

Proposition 21. *For the Fibonacci word, there exist exactly two factors of length n satisfying* (1) *if n is a Fibonacci number. Otherwise, no factor of length n satisfies such a condition.*

Proof. Consider two factors x, y of length n satisfying (1) and occurring in

$$\mathbf{f} = \mathtt{010010100100} \cdots .$$

Assume that x starts with $\mathtt{0}$. Then to fulfill (1), y starts with $\mathtt{1}$. Since $\mathbf{f}$ is Sturmian, for any two words of the same length x' and y' which are prefixes of x and y respectively, we have $||x'|_\mathtt{1} - |y'|_\mathtt{1}| \leq 1$. Therefore, we deduce that x and y must be of the form $x = \mathtt{0}u\mathtt{1}$ and $y = \mathtt{1}u\mathtt{0}$ for some $u \in \{\mathtt{0}, \mathtt{1}\}^*$. This means that u is a bispecial factor of the Fibonacci word.

Recall that the left special factors in $\mathbf{f}$ are its prefixes and its right special factors are the mirror images of its prefixes [3, Prop. 4.10.3]. So bispecial factors of $\mathbf{f}$ are its palindromic prefixes. If $(\ell_i)_{i\geq 1}$ denotes the increasing sequence of all lengths of palindromic prefixes in $\mathbf{f}$, it is well-known that $(\ell_i)_{i\geq 1} = (0, 1, 3, 6, 11, \ldots)$ is given by $\ell_i = F_{i+1} - 2$ where F_i is the ith Fibonacci number. See [7, Thm. 5] and [10]. Hence n must be a Fibonacci number.

Conversely, for any bispecial factor u of $\mathbf{f}$, it is easy to show that either $\mathtt{0}u\mathtt{0}$ or $\mathtt{1}u\mathtt{1}$ is not a factor occurring in $\mathbf{f}$ (see for instance [14, p. 47]). Therefore, amongst the four words

$0u0$, $1u1$, $0u1$ and $1u0$, the last two must occur in $\mathbf{f}$ and we get exactly two factors of length $|u|+2$ satisfying (1). Indeed, assume that $0u0$ does not occur in $\mathbf{f}$. Then for u to be left (resp. right) special, $0u1$ (resp. $1u0$) must occur in $\mathbf{f}$. ☐

The reader may notice that the computations carried out in the proofs of the next two results could also be adapted to other Sturmian words.

Theorem 22. *The set $\mathcal{APR}_{\mathbf{f}}$ of abelian returns to prefixes for the Fibonacci word $\mathbf{f}$ contains exactly the words* 0, 1, 01, 10, 001.

Proof. Using the same notation as in Theorem 19, for $c=7$, we have $D(7)\approx 0.145898$ which is such that $2D(7)<\min\{1/\tau^2, 1-1/\tau^2\}\approx 0.381$. Hence, all abelian returns to prefixes of the Fibonacci word have length at most 7. Actually, this value can be reduced:

Lemma 23. *There is no abelian return of length greater than 3 to prefixes in the Fibonacci word.*

Proof. With the notation of the proof of Theorem 19, we set $\rho=\alpha=1/\tau^2\approx 0.381$. Let i be a natural number. Define the four points $\rho_{i,t}=R_\alpha^{i+t}(\rho)$ for $t=0,1,2,3$.

Recall that, for all $k\geq 1$, the unit circle $[0,1)$ is split into two half-intervals $I_H(k)=[1-\{k\alpha\},1)$ and $I_L(k)=[0,1-\{k\alpha\})$ such that two factors $\mathbf{f}[i,i+k-1]$ and $\mathbf{f}[j,j+k-1]$ are abelian equivalent if and only if the points $R_\alpha^i(\rho)$ and $R_\alpha^j(\rho)$ belong to the same interval $I_H(k)$ or $I_L(k)$.

Let I be any of the two intervals $I_H(k)$ or $I_L(k)$. What we are going to prove is that if ρ and $\rho_{i,0}$ belong to I, then either $\rho_{i,2}$ or $\rho_{i,3}$ belongs to I. In other words, if $\mathbf{f}[i,i+k-1]\sim_{ab}\mathbf{f}[0,k-1]$, then either $\mathbf{f}[i+2,i+k+1]$ or $\mathbf{f}[i+3,i+k+2]$ is abelian equivalent to $\mathbf{f}[i,i+k-1]$ which gives the upper bound on the length of any abelian return to a prefix of $\mathbf{f}$.

Note, that $\rho_{i,0}=R_\delta(\rho_{i,2})$, $\rho_{i,2}=R_{-\delta}(\rho_{i,0})$ and $\rho_{i,3}=R_{\alpha-\delta}(\rho_{i,0})$, where $\delta=1-2\alpha\approx 0.2361$. Assume that the factor of length k starting in position i is abelian equivalent to the prefix of $\mathbf{f}$ of length k, i.e., ρ and $\rho_{i,0}$ are both light or heavy words. We consider two cases. Suppose first that $\rho,\rho_{i,0}\in I_L(k)$. In this case, we have $[0,\rho]\subseteq I_L(k)$.

- If $\rho\leq\rho_{i,0}<1$, then $[0,\rho_{i,0}]\subseteq I_L(k)$ and $0<\rho_{i,2}=R_{-\delta}(\rho_{i,0})<\rho_{i,0}$. Thus $\rho_{i,2}$ belongs also to $I_L(k)$.

- If $\rho>\rho_{i,0}>0$, either $\rho_{i,0}\geq\delta$ and then $\rho_{i,2}=R_{-\delta}(\rho_{i,0})\in[0,\rho)$ meaning that $\rho_{i,2}\in I_L(k)$ or, $0<\rho_{i,0}<\delta$, i.e., $-\delta<\rho_{i,2}-1<0$ and then $0<\alpha-\delta<\rho_{i,3}=R_\alpha(\rho_{i,2})<\rho$ meaning that $\rho_{i,3}\in I_L(k)$.

Suppose now that $\rho,\rho_{i,0}\in I_H(k)$. In this case, as $\rho\in I_H(k)$, we have $[\rho,1)\subseteq I_H(k)$.

- If $\rho>\rho_{i,0}>0$, then $[\rho_{i,0},1)\subseteq I_H(k)$ and $\rho_{i,3}=R_{\alpha-\delta}(\rho_{i,0})$ belongs to $I_H(k)$.

- If $\rho\leq\rho_{i,0}<1$, either $\rho_{i,0}\geq\rho+\delta$ and then $\rho_{i,2}=R_{-\delta}(\rho_{i,0})\in I_H(k)$ or, $\rho\leq\rho_{i,0}<\rho+\delta$, i.e., $\rho-\delta\leq\rho_{i,2}<\rho$ and then $\rho<\rho-\delta+\alpha\leq\rho_{i,3}=R_\alpha(\rho_{i,2})<\rho+\alpha<1$ meaning that $\rho_{i,3}\in I_H(k)$.

☐

The factors of length at most 3 occurring in $\mathbf{f}$ are ε, 0, 1, 00, 01, 10, 001, 010, 100 and 101. Clearly, 00, 010 and 101 do not satisfy (1) and cannot be abelian returns. To conclude the proof, we just have to show that 100 is also forbidden.

Lemma 24. *The set $\mathcal{APR}_{\mathbf{f}}$ of abelian returns to prefixes for the Fibonacci word $\mathbf{f}$ does not contain* 100.

Proof. We continue with notation of Lemma 23. Suppose that $100 \in \mathcal{APR}_{\mathbf{f}}$. There exists a prefix p of $\mathbf{f}$ of length k and a position $i \geq 0$ such that

1. $\mathbf{f}[i, i+2] = 100$,
2. $\mathbf{f}[i, i+k-1] \sim_{ab} p$, i.e., ρ and $\rho_{i,0}$ belong to the same interval $I \in \{I_L(k), I_H(k)\}$,
3. for $t = 1, 2$, $\mathbf{f}[i+t, i+t+k-1] \not\sim_{ab} p$, i.e., $\rho_{i,1}$ and $\rho_{i,2}$ do not belong to I,
4. $\mathbf{f}[i+3, i+2+k] \sim_{ab} p$.

To get a contradiction, let us prove that either $\rho_{i,1}$ or $\rho_{i,2}$ belongs to I. Since $\mathbf{f}_i = 1$, $\rho_{i,0}$ belongs to $I_1 = [1-\alpha, 1)$. If $I = I_L(k)$, then we have $\rho_{i,1} \in [0, \rho) \subseteq I_L(k)$. If $I = I_H(k)$, then we have $\rho_{i,2} \in [\rho, \rho + \alpha) \subseteq I_H(k)$. $\square$

That concludes the proof of Theorem 22. $\square$

4.4 Link with abelian complexity

The *abelian complexity* of an infinite word $\mathbf{x}$ is the function $p^{ab}_{\mathbf{x}} : \mathbb{N} \to \mathbb{N}$ that maps $n \geq 0$ to the number of distinct abelian equivalence classes of factors of length n in $\mathbf{x}$. Let $C > 0$. Recall that an infinite word $\mathbf{x} \in A^\omega$ is *C-balanced*, if for all $u, v \in \mathrm{Fac}(\mathbf{x})$ such that $|u| = |v|$, we have $||u|_a - |v|_a| \leq C$ for all $a \in A$.

Lemma 25. *[18] An infinite word has bounded abelian complexity if and only if it is C-balanced for some $C > 0$.*

Proposition 26. *If $\mathbf{x}$ is an abelian recurrent word such that $\mathcal{APR}_{\mathbf{x}}$ is finite, then $\mathbf{x}$ has bounded abelian complexity.*

Proof. Suppose $\mathbf{x}$ satisfies the assumptions of the proposition but that $\mathbf{x}$ has unbounded abelian complexity. From the previous lemma, we deduce that there exists a symbol a such that the maximum of differences $|u|_a - |v|_a$ for factors u, v in $\mathbf{x}$ having equal length can be arbitrarily large.

Let $\delta > 0$. There exist $u, v \in \mathrm{Fac}(\mathbf{x})$ of equal length n such that $|u|_a - |v|_a \geq \delta$. Let $p = x_0 x_1 \ldots x_{n-1}$ be the prefix of length n of $\mathbf{x}$. Without loss of generality, we may assume that

$$||u|_a - |p|_a| \geqslant \frac{\delta}{2}.$$

Indeed, if $||u|_a - |p|_a| < \delta/2$ and $||v|_a - |p|_a| < \delta/2$, then one would deduce that $||u|_a - |v|_a| < \delta$.

As $\mathbf{x}$ is abelian recurrent, factors abelian equivalent to p (resp. to u) occur infinitely often in $\mathbf{x}$. Therefore there exist $i < j < k$ such that

1. $\mathbf{x}[i, i+n-1] \sim_{ab} p$, $\mathbf{x}[k, k+n-1] \sim_{ab} p$,

2. for all t such that $i < t < k$, we have $\mathbf{x}[t, t+n-1] \not\sim_{ab} p$,

3. $\mathbf{x}[j, j+n-1] \sim_{ab} u$.

This just means that we can consider two consecutive factors abelian equivalent to p separated by a factor abelian equivalent to u. Note that, for all t,

$$||x[t+c, t+n-1+c]|_a - |x[t, t+n-1]|_a| \leqslant c, \quad \forall c \leq n.$$

Hence, $j - i \geq \delta/2$ and $k - j \geqslant \delta/2$. Therefore we get $k - i \geq \delta$ which means that the abelian return $\mathbf{x}[i, k-1]$ to the prefix p has length at least δ. As δ can be chosen arbitrarily large, the set $\mathcal{APR}_{\mathbf{x}}$ is infinite and that is a contradiction. □

Note that any Sturmian word $\mathbf{x}$ satisfies $p^{ab}_{\mathbf{x}}(n) = 2$ for all $n \geq 1$: there are exactly two kinds of factors of length n, the light ones and the heavy ones. But thanks to Theorem 19, if $\mathbf{x}$ is a Sturmian word with null intercept, then $\mathcal{APR}_{\mathbf{x}}$ is infinite. In other words, bounded abelian complexity does not imply the finiteness of $\mathcal{APR}_{\mathbf{x}}$.

5 Abelian derived sequences

We refer the reader to definitions and notation introduced in Section 2.1. As was studied by Durand [9] for classical return words, we introduce the notion of abelian derived sequence which is the factorization of an infinite word with respect to its abelian returns to prefixes in their order of occurrence. The next result allows us to define such a sequence.

Lemma 27. *Let u be a prefix of a uniformly recurrent word $\mathbf{x}$. The word $\mathbf{x}$ has a factorization as a sequence $m_0m_1m_2\cdots$ of elements in $\mathcal{APR}_{\mathbf{x},u}$ computed as follows. Consider the sequence of indices $(i_n)_{n\geq 0}$ such that, for all $j \geq 0$, $\mathbf{x}[i_j, i_j + |u| - 1] \sim_{ab} u$ and, for all $i \notin \{i_n \mid n \geq 0\}$, we have $\mathbf{x}[i, i + |u| - 1] \not\sim_{ab} u$. Set $m_n := \mathbf{x}[i_n, i_{n+1} - 1]$.*

As shown in Example 29, the factorization of $\mathbf{x}$ with elements in $\mathcal{APR}_{\mathbf{x},u}$ is not necessarily unique.

Definition 28. We define a map $\mu_{\mathbf{x},u} : \mathcal{APR}_{\mathbf{x},u} \to \{1, \ldots, \#(\mathcal{APR}_{\mathbf{x},u})\} =: A_{\mathbf{x},u}$ analogous to $\Lambda_{\mathbf{x},u}$. The *abelian derived sequence* $\mathcal{E}_u(\mathbf{x})$ is the corresponding infinite word

$$\mu_{\mathbf{x},u}(m_0)\mu_{\mathbf{x},u}(m_1)\mu_{\mathbf{x},u}(m_2)\cdots$$

over $A_{\mathbf{x},u}$ where the sequence $m_0m_1m_2\cdots \in \mathcal{APR}^{\omega}_{\mathbf{x},u}$ is the one computed in the previous lemma. The inverse map $\mu^{-1}_{\mathbf{x},u}$ defines a morphism $\theta_{\mathbf{x},u}$ from $A^*_{\mathbf{x},u}$ to $\mathcal{APR}^*_{\mathbf{x},u}$

Observe that $\mathcal{E}_u(\mathbf{x})$ is uniformly recurrent. Indeed, if $a_1 \cdots a_n$ is a factor occurring in $\mathcal{E}_u(\mathbf{x})$, it comes from a factor $m_1 \cdots m_n \in \mathcal{APR}^*_{\mathbf{x},u}$ such that $m_1 \cdots m_n v$ occurs in $\mathbf{x}$ for some $v \sim_{ab} u$ and $\mu_{\mathbf{x},u}(m_1) \cdots \mu_{\mathbf{x},u}(m_n) = a_1 \cdots a_n$. Since $\mathbf{x}$ is uniformly recurrent, $m_1 \cdots m_n v$ occurs infinitely often with bounded gaps in $\mathbf{x}$.

Example 29. Consider the Thue–Morse word where the first few symbols are

$$\mathbf{t} = \texttt{0110100110010110100101100110100110010110011010010 1}\cdots .$$

Take the prefix $u = \texttt{011}$. We get the derived sequence over $R_{\mathbf{t},u} = \{\texttt{1}, \ldots, \texttt{4}\}$

$$\mathcal{D}_u(\mathbf{t}) = \texttt{123412431234312412341243124123431234124312343124 12}\cdots$$

where the set of return words to u in order of occurrence in $\mathbf{t}$ is given by

$$\mathcal{R}_{\mathbf{t},u} = \{\texttt{011010},\ \texttt{011001},\ \texttt{01101001},\ \texttt{0110}\}.$$

The abelian derived sequence over $A_{\mathbf{t},u} = \{\texttt{1}, \ldots, \texttt{6}\}$ is

$$\mathcal{E}_u(\mathbf{t}) = \texttt{123142125216123142161252123142125216125212314216 12}\cdots$$

where the set of abelian returns to u in order of occurrence in $\mathbf{t}$ is given by

$$\mathcal{APR}_{\mathbf{t},u} = \{\texttt{0},\ \texttt{1},\ \texttt{1010},\ \texttt{1100},\ \texttt{10100},\ \texttt{110}\}.$$

Note that, since $\texttt{0}, \texttt{1} \in \mathcal{APR}_{\mathbf{t},u}$, there are infinitely many factorizations of $\mathbf{t}$ in terms of elements belonging to $\mathcal{APR}_{\mathbf{t},u}$.

Proposition 30. *Let u be a prefix of a uniformly recurrent word $\mathbf{x}$. There exists a morphism h_u from $R_{\mathbf{x},u}$ to $A^*_{\mathbf{x},u}$ such that $h_u(\mathcal{D}_u(\mathbf{x})) = \mathcal{E}_u(\mathbf{x})$.*

Proof. Each return word m occurring in $\mathbf{x}$ is followed by u. Consider the procedure of Lemma 27 applied to mu. It will define the image by h_u of $\Lambda_{\mathbf{x},u}(m)$. Indeed, one has to take into account a factor u appended to m because some suffix of m and a prefix of u can give a word $v \sim_{ab} u$ leading to some abelian return in the decomposition of m. More precisely, we consider all the occurrences $0 = i_1 < \cdots < i_t = |m|$ of factors abelian equivalent to u in $w = mu$. Then

$$h_u(\Lambda_{\mathbf{x},u}(m)) := \mu_{\mathbf{x},u}(w[i_1, i_2 - 1]) \cdots \mu_{\mathbf{x},u}(w[i_{t-1}, i_t - 1]).$$

$\square$

Example 31 (Example 29 continued)**.** There exists a morphism h_u from $R_{\mathbf{t},u}$ to $A^*_{\mathbf{t},u}$ such that $h_u(\mathcal{D}_u(\mathbf{t})) = \mathcal{E}_u(\mathbf{t})$. Take

$$h_u(\texttt{1}) = \texttt{123},\ h_u(\texttt{2}) = \texttt{142},\ h_u(\texttt{3}) = \texttt{1252},\ h_u(\texttt{4}) = \texttt{16}.$$

Let us explain how to get $h_u(\texttt{2})$. We have the following factorization where the vertical bars indicate the occurrence of a factor abelian equivalent to u:

$$\Lambda^{-1}_{\mathbf{t},u}(\texttt{2})\, u = (|\texttt{0}|\texttt{1100}|\texttt{1})\, \texttt{011}.$$

Definition 32. A map $h : A^\omega \to B^\omega$ is a *t-block morphism*, if there exists some map $f : A^t \to B^*$ such that, for all $\mathbf{w} \in A^\omega$,

$$h(\mathbf{w}) = f(\mathbf{w}[0, t-1])f(\mathbf{w}[1, t])f(\mathbf{w}[2, t+1]) \cdots .$$

By abuse of notation, the second map f will also be denoted by h.

Proposition 33. *Let u be a prefix of a uniformly recurrent word $\mathbf{x}$. Let v be a prefix of $\mathbf{y} = \mathcal{E}_u(\mathbf{x})$. There exist $t \le |u| - 1$ and a t-block morphism $h_{u,v} : (A_{\mathbf{y},v})^t \to A^*_{\mathbf{x},u}$ such that*

$$h_{u,v}(\mathcal{E}_v(\mathcal{E}_u(\mathbf{x}))) = \mathcal{E}_u(\mathbf{x}).$$

Proof. Note that any element $\theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a))$ with $a \in A_{\mathbf{y},v}$ is a concatenation of abelian returns to u. Now consider a factor $a_0a_1 \cdots a_{t-1}$ occurring in $\mathcal{E}_v(\mathcal{E}_u(\mathbf{x}))$. We have to determine the unique factorization of $\theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_0))$ with abelian returns to u given by Lemma 27. This one is completely determined when one knows the $|u|-1$ symbols occurring next. Without that extra knowledge we cannot uniquely determine the factorization for the last $|u|-1$ symbols possibly occurring in $\theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_0))$. This is the reason to consider the suffix $a_1 \cdots a_{t-1}$ in such a way that $\theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_1)) \cdots \theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_{t-1}))$ has length at least $|u| - 1$. One takes t large enough to ensure this property for any initial symbol $a_0 \in A_{\mathbf{y},v}$. More precisely, consider all the occurrences $0 = i_1 < \cdots < i_s$ of factors abelian equivalent to u in $w = \theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_0)) \cdots \theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_{t-1}))$. Let r be the largest integer such that $i_r < |\theta_{\mathbf{x},u}(\theta_{\mathbf{y},v}(a_0))|$. Then

$$h_{u,v}(a_0 \cdots a_{t-1}) := \mu_{\mathbf{x},u}(w[i_1, i_2 - 1]) \cdots \mu_{\mathbf{x},u}(w[i_r, i_{r+1} - 1]).$$

Note that the above definition is only meaningful if $a_0 \cdots a_{t-1}$ is a factor of $\mathcal{E}_v(\mathcal{E}_u(\mathbf{x}))$. Since this is the only relevant situation, in any other case, the image of $h_{u,v}$ is set to ε. □

Observe that if we iterate the process, since the composition of a t-block morphism and an s-block morphism is an (st)-block morphism, then there exists an r-block morphism h such that $h(\mathcal{E}_{u_k}(\cdots(\mathcal{E}_{u_2}(\mathcal{E}_{u_1}(\mathbf{x}))) \cdots)) = \mathcal{E}_{u_1}(\mathbf{x})$ where prefixes $u_1, \ldots, u_k$ are considered accordingly.

Example 34 (Example 29 continued)**.** We can iterate the process of computing the abelian derived sequence, for instance by taking each time the corresponding prefix of length 3:

$$\mathcal{E}_{\mathtt{123}}(\mathcal{E}_{\mathtt{011}}(\mathbf{t})) = \mathtt{121314151213151412131415141213151213141512131514 12} \cdots ,$$

$$\mathcal{E}_{\mathtt{121}}(\mathcal{E}_{\mathtt{123}}(\mathcal{E}_{\mathtt{011}}(\mathbf{t}))) = \mathtt{12341243123431241234124312412343123412431234312412} \cdots ,$$

$$\mathcal{E}_{\mathtt{123}}(\mathcal{E}_{\mathtt{121}}(\mathcal{E}_{\mathtt{123}}(\mathcal{E}_{\mathtt{011}}(\mathbf{t})))) = \mathtt{12341432123432141234143214123432123414321234321412} \cdots .$$

Let us illustrate the previous result. Take again $u = \mathtt{011}$, $\mathbf{y} = \mathcal{E}_u(\mathbf{t})$, $v = \mathtt{123}$. We have

$$\mathcal{APR}_{\mathbf{y},v} = \{\mathtt{1},\ \mathtt{23142125216},\ \mathtt{23142161252},\ \mathtt{231421252161252},\ \mathtt{2314216}\}.$$

Observe that $\theta_{\mathbf{t},u}(\theta_{\mathbf{y},v}(\mathtt{1})) = \theta_{\mathbf{t},u}(\mathtt{1}) = \mathtt{0}$ and, for all $a \in \{\mathtt{2}, \ldots, \mathtt{5}\}$, $\theta_{\mathbf{y},v}(a)$ has a prefix $\mathtt{23}$, so $\theta_{\mathbf{t},u}(\theta_{\mathbf{y},v}(a))$ has prefix $\mathtt{110} \sim_{ab} u$. Let us assume that $h_{u,v}$ is a 3-block morphism. We define $h_{u,v}(\mathtt{1}ab) = \mathtt{1}$, for all $a, b \in A_{\mathbf{y},v}$ and $a \neq \mathtt{1}$. We get $h_{u,v}(\mathtt{213}) = \mathtt{23142125216}$ because, if vertical bars denote occurrences of a factor abelian equivalent to u, we get the following factorization:

$$\theta_{\mathbf{t},u}(\theta_{\mathbf{y},v}(\mathtt{2}))\,\theta_{\mathbf{t},u}(\theta_{\mathbf{y},v}(\mathtt{13})) =$$

$$(|\mathtt{1}|\mathtt{1010}|\mathtt{0}|\mathtt{1100}|\mathtt{1}|\mathtt{0}|\mathtt{1}|\mathtt{10100}|\mathtt{1}|\mathtt{0}|\mathtt{110})\,|\mathtt{0110100110010110011010001}.$$

Proposition 35. *Let σ be a primitive substitution and u be a prefix of its fixed point $\mathbf{x} = \sigma(\mathbf{x}) \in A^\omega$. There exists a 2-block morphism $\sigma_u : A^*_{\mathbf{x},u} \to A^*_{\mathbf{x},u}$ such that $\mathcal{E}_u(\mathbf{x})$ is fixed point of σ_u and*

$$\theta_{\mathbf{x},u}(\sigma_u(\mathcal{E}_u(\mathbf{x}))) = \sigma(\theta_{\mathbf{x},u}(\mathcal{E}_u(\mathbf{x}))).$$

Proof. We may replace σ by a convenient power of σ in such a way that, for all $a \in A$, $\sigma(a)$ contains an occurrence of a factor abelian equivalent to u. For all $a, b \in A_{\mathbf{x},u}$, consider all the occurrences $i_1 < \cdots < i_t$ of a factor abelian equivalent to u occurring in $w = \sigma(\theta_{\mathbf{x},u}(ab))$. With our choice of σ, at least one of these i_j belongs to $[0, |\sigma(\theta_{\mathbf{x},u}(a))|-1]$ (resp. $[|\sigma(\theta_{\mathbf{x},u}(a))|, |w|-1]$). Let r be the largest integer such that $i_r < |\sigma(\theta_{\mathbf{x},u}(a))|$. We define

$$\sigma_u(ab) = \mu_{\mathbf{x},u}(w[i_1, i_2-1]) \cdots \mu_{\mathbf{x},u}(w[i_r, i_{r+1}-1]).$$

□

Corollary 36. *Let σ be a primitive substitution and u be a prefix of its fixed point $\mathbf{x} = \sigma(\mathbf{x}) \in A^\omega$. The sequence $\mathcal{E}_u(\mathbf{x})$ is primitive substitutive, i.e., there exists a primitive morphism $\tau_u : B \to B^*$ and a coding $\phi : B \to A_{\mathbf{x},u}$ such that $\mathcal{E}_u(\mathbf{x}) = \phi(\tau_u^\omega(b))$ for some $b \in B$.*

Proof. We may replace σ by a convenient power of σ in such a way that, for all $a \in A$, $\sigma(a)$ contains occurrences of two factors abelian equivalent to u. Consider the alphabet

$$B = \{(a, b) \mid a, b \in A_{\mathbf{x},u} \wedge ab \text{ is a factor of } \mathcal{E}_u(\mathbf{x})\}.$$

For all $a, b \in A_{\mathbf{x},u}$ such that $(a, b) \in B$, consider all the occurrences $i_1 < \cdots < i_t$ of a factor abelian equivalent to u occurring in $w = \theta_{\mathbf{x},u}(ab)$. Let r be the smallest integer such that $i_r \geq |\theta_{\mathbf{x},u}(a)|$. Note that $r \geq 3$. We define

$$\tau_u((a, b)) = (\mu_{\mathbf{x},u}(w[i_1, i_2-1]), \mu_{\mathbf{x},u}(w[i_2, i_3-1])) \cdots (\mu_{\mathbf{x},u}(w[i_{r-1}, i_r-1]), \mu_{\mathbf{x},u}(w[i_r, i_{r+1}-1])).$$

Let e_0e_1 be the prefix of length 2 of $\mathcal{E}_u(\mathbf{x})$. We have

$$\mathcal{E}_u(\mathbf{x}) = \phi(\tau_u^\omega((e_0, e_1)))$$

where $\phi : B \to A_{\mathbf{x},u}$ is the coding that maps $(a, b) \in B$ to a.

Observe that, for all $(a, b) \in B$, $|\tau_u^n((a, b))| \geq 2^n$. Let us show that τ_u is primitive. Since $\mathcal{E}_u(\mathbf{x})$ is uniformly recurrent, there exists K such that any factor of length K of $\mathcal{E}_u(\mathbf{x})$ contains all elements in $\{cd \mid (c, d) \in B\}$. Therefore any factor of length K of $\tau_u^\omega((e_0, e_1))$ contains all the elements of B. Take N such that $2^N \geq K$. Then, for all $(a, b), (c, d) \in B$, $\tau_u^N((a, b))$ contains (c, d) which means that τ_u is primitive. □

Example 37 (Example 29 continued)**.** Take again $u = \texttt{011}$ and the morphism $\sigma : \texttt{0} \mapsto \texttt{01101001}, \texttt{1} \mapsto \texttt{10010110}$ generating $\mathbf{t}$. We have

$$\sigma(\theta_{\mathbf{t},u}(\texttt{12})) = (|\texttt{0}|\texttt{1}|\texttt{1010}|\texttt{0}|\texttt{1})\texttt{100}|\texttt{1}|\texttt{0}|\texttt{110} \text{ and } \sigma_u(\texttt{12}) = \texttt{12314}$$

$$\sigma(\theta_{\mathbf{t},u}(\texttt{23})) = (\texttt{100}|\texttt{1}|\texttt{0}|\texttt{1}|\texttt{10})\texttt{100}|\texttt{1}|\texttt{0}|\texttt{110}\cdots \text{ and } \sigma_u(\texttt{23}) = \texttt{2125}$$

$$\sigma(\theta_{\mathbf{t},u}(\texttt{31})) = (\texttt{100}|\texttt{1}|\texttt{0}|\texttt{110}|\texttt{0}|\texttt{1}|\texttt{1010}|\texttt{0}|\texttt{1100}|\texttt{1}|\texttt{0}|\texttt{110}|\texttt{0}|\texttt{1}|\texttt{10100}|\texttt{1})|\texttt{0}|\texttt{1}|\texttt{101001}$$
$$\text{and } \sigma_u(\texttt{21}) = \texttt{216123142161252}.$$

Using the above corollary, we get

$$\tau_u(1,2) = (1,2)(2,3)(3,1)(1,4)(4,2),\ \tau_u(2,3) = (2,1)(1,2)(2,5)(5,2), \ldots.$$

5.1 Abelian derivatives of the Thue–Morse word

Proposition 38. *For the Thue–Morse word* $\mathbf{t}$*, the set* $\{\mathcal{E}_u(\mathbf{t}) \mid u \in \mathrm{Pref}(\mathbf{t})\}$ *is infinite.*

Proof. It is sufficient to show that the set $\{\mathcal{E}_u(\mathbf{t}) \mid u \in \mathrm{Pref}(\mathbf{t}) : |u| \equiv 1 \pmod 2\}$ is infinite. Proceed by contradiction and suppose that the set $\{\mathcal{E}_u(\mathbf{t}) \mid u \in \mathrm{Pref}(\mathbf{t}) : |u| \equiv 1 \pmod 2\}$ is finite. Then there exist u and v distinct prefixes of odd length of the Thue–Morse word $\mathbf{t}$ such that $\mathcal{E}_u(\mathbf{t}) = \mathcal{E}_v(\mathbf{t})$. Since $\mathcal{APR}_{\mathbf{t}}$ is finite, we can moreover assume that $\theta_{\mathbf{t},u} = \theta_{\mathbf{t},v}$. Indeed, infinitely many sequences of the kind $\mathcal{E}_u(\mathbf{t})$ are equal and thus defined on the same alphabet $A_{\mathbf{t},u}$. For all such sequences, there are finitely many morphisms of the kind $\theta_{\mathbf{t},u}$ associating with each element of $A_{\mathbf{t},u}$ an element of the finite set $\mathcal{APR}_{\mathbf{t}}$. So we can impose the extra condition on $\theta_{\mathbf{t},u}$. Let

$$I(w) := \{i \in \mathbb{N} \mid \mathbf{t}[i, i+|w|-1] \sim_{ab} w\}$$

denote the set of occurrences of factors of $\mathbf{t}$ abelian equivalent to a word w. We have $I(u) = I(v)$ as $\theta_{\mathbf{t},u} = \theta_{\mathbf{t},v}$. Without loss of generality, we may suppose that $|u| = 2k+1$, $|v| = 2\ell+1$ with $k < \ell$. We have $\Psi(u) = (k, k+1)$ or $\Psi(u) = (k+1, k)$ and $\Psi(v) = (\ell, \ell+1)$ or $\Psi(v) = (\ell+1, \ell)$. Let a_u (resp. a_v) denote the letter having $k+1$ (resp. $\ell+1$) occurrences in the prefix u (resp. v). Note that a_u and a_v are respectively the last letters of u and v.

For any odd position j, recalling that $\mathbf{t}[2m, 2m+1] \in \{\texttt{10}, \texttt{01}\}$, we have

$$\mathbf{t}_j = a_u \Leftrightarrow \mathbf{t}[j, j+|u|-1] \sim_{ab} u \Leftrightarrow \mathbf{t}[j, j+|v|-1] \sim_{ab} v \Leftrightarrow \mathbf{t}_j = a_v$$

where the central equivalence comes from the fact that $I(u) = I(v)$. As there exists at least one such j, we have $a_u = a_v =: a$.

For any even position j, we have

$$\mathbf{t}_{j+|u|-1} = a \Leftrightarrow j \in I(u) \Leftrightarrow j \in I(v) \Leftrightarrow \mathbf{t}_{j+|v|-1} = a$$

since $I(u) = I(v)$. Using this observation, we can show by induction that

$$\mathbf{t}_{|u|-1+n(|v|-|u|)} = a$$

for all $n \in \mathbb{N}$. In other words, there exists a constant infinite arithmetical subsequence in $\mathbf{t}$, which is a contradiction, since it is well-known that the Thue-Morse word does not contain any such subsequence. Indeed, for $n = 0$, it is clear that the last letter of u is a. Suppose now that the result holds true for $n \geq 0$. We have $\mathbf{t}_{|u|-1+n(|v|-|u|)} = a$. Since $|u|, |v|$ are odd, $n(|v| - |u|)$ is an even number and belongs to $I(u) = I(v)$. Therefore $\mathbf{t}_{|v|-1+n(|v|-|u|)} = a$ and $|v| - 1 + n(|v| - |u|) = |u| - 1 + (n + 1)(|v| - |u|)$. □

Remark 39. Using the same notation as in the previous proof, we show that the set $\{\mathcal{E}_u(\mathbf{t}) \mid u \in \mathrm{Pref}(\mathbf{t}) : |u| \equiv 0 \pmod 2\}$ is infinite. Proceed by contradiction. Then there exist u and v distinct prefixes of even length of $\mathbf{t}$ such that $\mathcal{E}_u(\mathbf{t}) = \mathcal{E}_v(\mathbf{t})$ and $\theta_{\mathbf{t},u} = \theta_{\mathbf{t},v}$. Hence $I(u) = I(v)$. Note that $2\mathbb{N} \subseteq I(u) = I(v)$. Suppose that $|u| = 2k$, $|v| = 2\ell$, with $k < \ell$. Since a prefix of even length has a Parikh vector of the kind (r, r), $2i+1$ is in $I(u)$ if and only if $\mathbf{t}_i = \mathbf{t}_{i+k}$. Similarly, $2i + 1$ is in $I(v)$ if and only if $\mathbf{t}_i = \mathbf{t}_{i+\ell}$. From $I(u) \setminus 2\mathbb{N} = I(v) \setminus 2\mathbb{N}$, we deduce that, for all $i \in \mathbb{N}$, $\mathbf{t}_i = \mathbf{t}_{i+k}$ implies $\mathbf{t}_i = \mathbf{t}_{i+\ell}$ and conversely. This leads to the contradiction that $\mathbf{t}$ is ultimately periodic of period $\ell - k$. Indeed, suppose to the contrary that for some i, $\mathbf{t}_{i+k} \neq \mathbf{t}_{i+\ell}$. In this case, either $\mathbf{t}_{i+k}$ or $\mathbf{t}_{i+\ell}$ is equal to $\mathbf{t}_i$. From our last deduction, we get that all three letters $\mathbf{t}_i$, $\mathbf{t}_{i+k}$, $\mathbf{t}_{i+\ell}$ are equal.

Remark 40. Using the same notation as in the previous remark, there exist no prefixes u, v of $\mathbf{t}$ such that $|u|$ is even, $|v|$ is odd and $I(u) = I(v)$. (The symmetric case can be treated in the same way.) Assume that $|u| = 2k$, $|v| = 2\ell + 1$ for some positive integers $k \neq \ell$. We get $\Psi(u) = (k, k)$ and $\Psi(v) = (\ell, \ell + 1)$ or $\Psi(v) = (\ell + 1, \ell)$. Let a denote the letter that has $\ell + 1$ occurrences in v. As $v \in \mathrm{Pref}(\mathbf{t})$, $\mathbf{t}_{|v|-1} = a$. Note that, for all even positions j, if j is in $I(v)$, then $\mathbf{t}_{j+|v|-1} = a$. Moreover, for all even j, we have $j \in I(u)$. Since $I(u) = I(v)$, we get $2\mathbb{N} \subseteq I(v)$ and thus $\mathbf{t}_{j+|v|-1} = a$ for all even j. Therefore, for all even $j \geq |v| - 1$, we have $\mathbf{t}_j = a$ and also $\mathbf{t}_{j+1} = \mathtt{1} - a$ since $\mathbf{t}$ is made up of blocks $\mathtt{01}$ or $\mathtt{10}$. This means that the Thue–Morse word is ultimately periodic of period 2 which is a contradiction.

6 Acknowledgment

The authors are very pleased to contribute to this special issue in honor of Jean-Paul Allouche. This is a nice way to wish him a happy birthday with some combinatorial flavor. In particular, the first author is very grateful to Jean-Paul Allouche who is always enthusiastic, positive and ready to usefully comment on his research work.

References

[1] J.-P. Allouche and J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence. Sequences and their Applications (Singapore, 1998), 1–16, *Springer Ser. Discrete Math. Theor. Comput. Sci.*, Springer, London, (1999).

[2] J.-P. Allouche and J. Shallit, *Automatic Sequences, Theory, Applications, Generalizations*, Cambridge University Press (2003).

[3] J. Cassaigne and F. Nicolas, Factor complexity, in V. Berthé and M. Rigo, eds., *Combinatorics, Automata and Number Theory*, Encyclopedia Math. Appl., **135**, Cambridge Univ. Press, 2010, pp. 163–247.

[4] M. Boshernitzan, A condition for minimal interval exchange maps to be uniquely ergodic, *Duke Math. J.* **52** (1985), 723–752.

[5] E. M. Coven and G. A. Hedlund, Sequences with minimal block growth, *Math. Systems Theory* **7** (1973), 138–153.

[6] F. M. Dekking, Strongly non-repetitive sequences and progression-free sets, *J. Comb. Theory Ser. A* **27** (1979), 181–185.

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

[8] M. Denker, C. Grillenberger, and K. Sigmund, Ergodic Theory on Compact Spaces, *Lecture Notes in Mathematics*, Vol. **527**, Springer-Verlag, (1976).

[9] F. Durand, A characterization of substitutive sequences using return words, *Discrete Math.* **179** (1998), 89–101.

[10] S. Fischler, Palindromic prefixes and episturmian words, *J. Combin. Theory Ser. A* **113** (2006), 1281–1304.

[11] C. Holton and L. Q. Zamboni, Descendants of primitive substitutions, *Theory Comput. Syst.* **32** (1999), 133–157.

[12] J. Justin and L. Vuillon, Return words in Sturmian and episturmian words, *Theor. Inform. Appl.* **34** (2000), 343–356.

[13] M. Lothaire, *Combinatorics on Words*, Cambridge Mathematical Library, Cambridge University Press, 1997.

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

[15] J. F. Morgenbesser, J. Shallit, and T. Stoll, Thue-Morse at multiples of an integer, *J. Number Theory* **131** (2011), 1498–1512.

[16] M. Morse and G. A. Hedlund, Symbolic dynamics II: Sturmian sequences, *Amer. J. Math.* **62** (1940), 1–42.

[17] S. Puzynina and L. Q. Zamboni, Abelian returns in Sturmian words, *J. Combin. Theory Ser. A* **120** (2013), 390–408.

[18] G. Richomme, K. Saari, and L. Q. Zamboni, Abelian complexity of minimal subshifts, *J. London Math. Soc.* **83** (2011), 79–95.

[19] P. Salimov, Uniform recurrence of a direct product, *Discrete Math. Theor. Comput. Sci.* **12** (2010), 1–8.

[20] L. Vuillon, A characterization of Sturmian words by return words, *European J. Combin.* (2001) **22**, 263–275.

2010 *Mathematics Subject Classification*: Primary 68R15; Secondary 68Q68.
Keywords: return word, recurrence, abelian equivalence, Sturmian word.

(Concerned with sequences A001911, A003622, A003849, A010060, and A022342.)

Received June 2 2012; revised version received September 24 2012. Published in *Journal of Integer Sequences*, March 2 2013.

Return to Journal of Integer Sequences home page.