Some properties of abelian return words (long abstract)

M. Rigo*, P. Salimov†, E. Vandomme‡

June 5, 2012

Abstract

We investigate some properties of abelian return words as recently introduced by S. Puzynina and L. Q. Zamboni in [17]. In particular, we obtain a characterization of Sturmian words with non-null 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. We end this paper by considering the notion of abelian derived sequence. It turns out that, for the Thue–Morse word, the set of abelian derived sequences is infinite.

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 carried on in the abelian context? Up to which extend, classical results can 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 $n \geq 0$ onto the number of distinct factors of length n occurring in an infinite word $\mathbf{x}$. It provides a classification criterion for families of infinite words. 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 $n \geq 0$ onto the number of distinct abelian classes partitioning the set of factors of length n occurring in $\mathbf{x}$. This latter notion was already introduced in the seventies [5]. Some other important questions in combinatorics on words, like avoiding abelian repetitions were initiated at the same period, see for instance [6]. See also the reference [18] on abelian complexity and containing many relevant bibliographic pointers.

Return word is a classical notion in combinatorics on words or symbolic dynamical systems. For instance, F. Durand obtained a characterization of primitive substitutive sequences in

*University of Liège

†The second author is supported by the Russian Foundation for Basic Research grant no. 11-01-00997 and by a University of Liège post-doctoral grant.

‡University of Liège and Institut Fourier, Grenoble

terms of return words and derived sequences [8]. 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 S. Puzynina and L. 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 in [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, in [17], is considered the set of abelian returns with respect to each factor taken separately. In particular, an important contribution of [17] 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.*

In this talk, 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. Then, we limit 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-)null 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 5). 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 $\mathbf{f}$, [7, 9]. By contrast, the set of abelian returns to all prefixes for the characteristic word of slope $1/\tau^2$ is infinite even if this latter word is simply $\mathtt{0}\mathbf{f}$. 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. Finally, we introduce the notion of abelian derived sequence and study their first properties. This abstract does not contain any proof, full text is available here `http://www.discmath.ulg.ac.be/papers/abelian-submission.pdf`

1.1 Abelian returns

In this talk, we will distinguish two cases: abelian return to a prefix or 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. Let $\Psi : A \to \mathbb{N}^k$ be the Parikh mapping. We denote by $|w|_{a_i}$ the number of occurrences of the letter a_i in w and $\Psi(w) = (|w|_{a_1}, \ldots, |w|_{a_k})$. 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 factorized as $\mathbf{x} = u_1u_2u_3\cdots$ 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 $\mathbf{x}$.

Let $\mathbf{x}$ be an infinite word. If, for all factors 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 all factors 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*. Note that uniform recurrence implies obviously abelian uniform recurrence. We will show in Proposition 2 that the converse does not hold.

Definition 1. Let u be a prefix of an abelian uniformly recurrent word $\mathbf{x}$. We say that a non-empty 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 slightly more general situation where u is not restricted to be a prefix of $\mathbf{x}$. In [17], this notion is called 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 2. 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 and 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 1. Let $\mathbf{x}$ be an abelian recurrent word. The set $\mathcal{AR}_{\mathbf{x},u}$ is finite, for all factors u of $\mathbf{x}$, if and only if $\mathbf{x}$ is abelian uniformly recurrent.

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

Theorem 1. *Let $\mathbf{x}$ be a recurrent word. The set $\mathcal{AR}_{\mathbf{x}}$ is finite if and only if $\mathbf{x}$ is periodic.*

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

2 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 lead 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 non-null intercept.

Lemma 1. *If $\mathbf{x}$ is a uniformly recurrent and abelian periodic word, then the set $\mathcal{APR}_{\mathbf{x}}$ is finite.*

In Lemma 1, 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 $\mathbf{x}$. 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.

Proposition 3. *A uniformly recurrent word $\mathbf{x}$ is periodic if and only there exists some prefix u such that all the abelian returns in $\mathcal{APR}_{\mathbf{x},u}$ have length 1.*

We already know from Lemma 1 that the Thue–Morse word has a finite set of abelian returns to all its prefixes. Here we describe precisely this set.

Lemma 2. *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 $\ell \in \{0, \ldots, n-2\}$, $\mathbf{x}[i, i+\ell] \not\sim_{ab} \mathbf{x}[j, j+\ell]$.*

Remark 2. From this lemma, we can derive a necessary condition for a word to be an abelian return to a prefix. A word $w = w_1 \cdots w_n$ of length n is an abelian return to a prefix only if 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 4. *The set $\mathcal{APR}_{\mathbf{t}}$ of abelian returns to prefixes for the Thue–Morse word $\mathbf{t}$ contains exactly the words*

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

2.1 Sturmian words

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 $\{x+\alpha\}$.

By *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 [2, 14] 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 $I_1 = (1-\alpha, 1]$).

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 1. 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 5. *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.*

Lemma 3. *A Sturmian word is not abelian periodic.*

Theorem 6. *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`.

2.2 Link with abelian complexity

The *abelian complexity* of an infinite word $\mathbf{x}$ is the function $p_\mathbf{x}^{ab} : \mathbb{N} \to \mathbb{N}$ that maps $n \ge 0$ onto the number of distinct classes of abelian equivalence 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 $||u|_a - |v|_a| \le C$ for all $a \in A$.

Proposition 7. *If* $\mathbf{x}$ *is an abelian recurrent word such that* $\mathcal{APR}_\mathbf{x}$ *is finite, then* $\mathbf{x}$ *has bounded abelian complexity.*

3 Abelian derived sequence

As it was studied in [8] 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 appearance. The next result permits us to define such a sequence.

Lemma 4. *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 2, the factorization of $\mathbf{x}$ with elements in $\mathcal{APR}_{\mathbf{x},u}$ is not necessarily unique.

Definition 3. We define a map $\mu_{\mathbf{x},u} : \mathcal{APR}_{\mathbf{x},u} \to \{\mathtt{1}, \ldots, \#(\mathcal{APR}_{\mathbf{x},u})\} =: A_{\mathbf{x},u}$ 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 2. Consider the Thue–Morse word where the first few symbols are

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

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

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

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

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

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

$$\mathcal{E}_u(\mathbf{t}) = \mathtt{12314212521612314216125212314212521612521231421612}\cdots$$

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

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

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

We investigate the first properties of these derived sequences. In particular, we get the following.

Proposition 8. *For the Thue–Morse word $\mathbf{t}$, the set $\{\mathcal{E}_u(\mathbf{t}) \mid u \in \mathrm{Pref}(\mathbf{t})\}$ is infinite.*

References

[1] J.-P. Allouche, 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, J. Shallit, *Automatic sequences, Theory, applications, generalizations*, Cambridge University Press (2003).

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

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

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

[10] N. J. Fine, The distribution of the sum of digits (mod p), *Bull. Amer. Math. Soc.* **71** (1965), 651-–652.

[11] M. Denker, C. Grillenberger, K. Sigmund, Ergodic theory on compact spaces, *Lecture Notes in Mathematics*, Vol. **527**, Springer-Verlag, (1976).

[12] G. H. Hardy, E. M. Wright, *An introduction to the theory of numbers*, fifth edition, The Clarendon Press, Oxford University Press, (1979).

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

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

[17] S. Puzynina, L. Q. Zamboni, Abelian returns in Sturmian words, `http://arxiv.org/abs/1204.5755`

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

[19] M. Rigo, P. Salimov, E. Vandomme, Some properties of abelian return words, *submitted for publication.*

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