A NOTE ON ABELIAN RETURNS IN ROTATION WORDS

N. RAMPERSAD, M. RIGO, AND P. SALIMOV

ABSTRACT. Pursuing the study started by Rigo, Salimov and Vandomme, we use elementary number-theoretic techniques to characterize rotation words having a finite set of abelian returns to all prefixes. We also make the connection between the three gap theorem and the number of semi-abelian returns for Sturmian words, simplifying some arguments developed by Puzynina and Zamboni.

1. INTRODUCTION

In this paper we study abelian return words in rotation words. The usual definition of *return word* is as follows: Given a factor v of an infinite word $\mathbf{w}$, a return word to v is, roughly speaking, a factor of $\mathbf{w}$ that separates two consecutive occurrences of v. In the abelian setting, we define an equivalence relation on words of the same length by saying that two such words are *abelian equivalent* if one can be obtained by permuting the letters of the other. An *abelian return word* to v is then a factor of $\mathbf{w}$ that separates two consecutive occurrences of members of the abelian equivalence class of v.

This notion of abelian return word was developed by Puzynina and Zamboni [8], who gave a very nice characterization of the class of Sturmian words based on abelian return words, namely: A recurrent infinite word is Sturmian if and only if each of its factors has either two or three abelian returns. Rigo, Salimov, and Vandomme [10] studied the set of abelian returns to prefixes of Sturmian words. Their work used the definition of Sturmian words as *rotation words*. That is, given an irrational number α and a partition of the unit circle into sub-intervals $I_0 = [0, 1 - \alpha)$ and $I_1 = [1 - \alpha, 1)$, a Sturmian word encodes the trajectory of a point ρ that is repeatedly rotated by a distance α around the unit circle. One writes a 0 when the trajectory passes through I_0 and a 1 when the trajectory passes through I_1. The resulting infinite sequence of 0's and 1's is called a *Sturmian word*. The main result of Rigo, Salimov, and Vandomme is that the set of abelian returns to prefixes of a Sturmian word is finite if and only if the initial point ρ is non-zero. The main tool in their proof is Kronecker's Theorem on the distribution of the points $\{n\alpha\}$ on the unit circle (where $\{\cdot\}$ denotes the operation that takes the fractional part of a real number).

In this work we consider the more general class of rotation words. In addition to the parameters α and ρ given above, we now add a third parameter β and partition the unit circle into intervals $I_0 = [0, 1 - \beta)$ and $I_1 = [1 - \beta, 1)$. The infinite word encoding the trajectory of the point ρ under rotation by α as described above is now called a *rotation word*. Adamczewski [1] has given a detailed study of these words. The main result of the present paper gives a characterization analogous to that of Rigo, Salimov, and Vandomme of the points ρ for which the resulting rotation word has the property that the set of abelian returns to prefixes is finite. Again, our main tool is Kronecker's Theorem.

We also apply the classical "three gap theorem" (see [3]) to provide an alternative proof of one direction of the characterization of Sturmian words given by Puzynina and Zamboni [8]. Note that the three gap theorem was also used very recently to prove that in reversible Christoffel factorizations of Sturmian words, only 2 or 3 distinct Christoffel words may occur [4].

2000 *Mathematics Subject Classification.* Primary 68R15; Secondary 68Q68, 11B85 .

The first author is supported by an NSERC Discovery Grant.

The third author is supported by the Russian President's grant no. MK-4075.2012.1, the Russian Foundation for Basic Research grant no. 12-01-00089 and by a University of Liège post-doctoral grant.

2. Definitions and preliminary results

Let $\mathbf{x} = x_0x_1\cdots$ be an infinite word over a finite alphabet. 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 $i \le j$. The factor $x_ix_{i+1}\cdots x_j$ of $\mathbf{x}$ is denoted by $\mathbf{x}[i,j]$. The notation $\mathbf{x}[i,i]$ is shortened to $\mathbf{x}_i$.

Let $A = \{a_1,\dots,a_k\}$ 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 $\Psi(w) = (|w|_{a_1},\dots,|w|_{a_k})$. 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 $\Psi(u) = \Psi(v)$. The *abelian complexity* of $\mathbf{x}$ is the function that maps $n \in \mathbb{N}$ to the number of factors of length n that are pairwise abelian inequivalent.

Let $\mathbf{x}$ be an infinite word. If, for each factor u of $\mathbf{x}$, there exist infinitely many i such that $\mathbf{x}[i, i+|u|-1] = u$ (resp. $\mathbf{x}[i, i+|u|-1] \sim_{ab} u$), then $\mathbf{x}$ is said to be *recurrent* (resp. *abelian recurrent*). If $\mathbf{x}$ is recurrent (resp. abelian recurrent) and if, for each factor u of $\mathbf{x}$, the distance between any two consecutive occurrences of factors equal to u (resp. abelian equivalent to u) is bounded by a constant depending only on u, then $\mathbf{x}$ is said to be *uniformly recurrent* (resp. *abelian uniformly recurrent*).

Definition 1. Let u be a factor 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 $i \ge 0$ 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 $j \in \{1,\dots,|w|-1\}$.

Puzynina and Zamboni [8] 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.

If u is a prefix of $\mathbf{x}$, 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

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

The coding of rotations is a particular tool for constructing infinite words over a finite alphabet. Let $\mathcal{C}$ 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* $R_\alpha : \mathcal{C} \to \mathcal{C}$, defined for a real number α, maps x to $\{x+\alpha\}$. An *interval* $I = [a,b]$ (resp. *half-interval* $I = [a,b)$) of $\mathcal{C}$ is the set of points $\{R_\delta(a) \mid 0 \le \delta \le \gamma\}$ (resp. $\{R_\delta(a) \mid 0 \le \delta < \gamma\}$), where γ is the unique real number such that $R_\gamma(a) = b$ and $0 < \gamma < 1$. This quantity γ associated with I is denoted by $|I|$. For instance, if $0 \le b < a < 1$, then $[a,1] \cup [0,b)$ is denoted by $[a,b)$ and $|[a,b)| = 1 - a + b$.

Definition 2. Let $\alpha \in (0,1)$ and $\rho \in [0,1)$. Let I_1 be a half-interval of $\mathcal{C}$. The *rotation word* $\mathbf{r} = r(\alpha, I_1, \rho)$ is the word $r_0r_1\cdots$ satisfying, for all $i \ge 0$,

$$r_i = \begin{cases} \mathtt{1} & \text{if } R_\alpha^i(\rho) \in I_1; \\ \mathtt{0} & \text{otherwise.} \end{cases} \tag{1}$$

Let $\beta \in (0,1)$. If $I_1 = [1-\beta, 0)$, then the corresponding rotation word is usually denoted by $r(\alpha, \beta, \rho)$.

It is clear that if α is rational, then $\mathbf{r}$ is periodic. So from now on we will only consider $\alpha \in (0,1) \setminus \mathbb{Q}$ and we can make use of Kronecker's theorem.

Theorem 1 (Kronecker)**.** *Let α be irrational. Let $\rho \in \mathcal{C}$. The set of points $\{R_\alpha^i(\rho) \mid i \in \mathbb{N}\}$ is dense in $\mathcal{C}$.*

A straightforward consequence of Theorem 1 is the following.

Corollary 2. *Let α be irrational. Let $c > 0$. Let $\rho \in \mathcal{C}$. There exists a constant n depending only on c such that, for all (half-)intervals I in $\mathcal{C}$ satisfying $|I| \geqslant c$,*

$$I \cap \{R_\alpha^i(\rho) \mid 0 \leqslant i \leqslant n\} \neq \emptyset.$$

Let $\mathbf{r} = r(\alpha, I_1, \rho)$ be a rotation word. Define $I_0 = \mathcal{C}\backslash I_1$. For a binary word $v = v_0v_1\cdots v_m$, we define the set I_v of $\mathcal{C}$ as

$$I_v := I_{v_0} \cap R_\alpha^{-1}(I_{v_1}) \cap \cdots \cap R_\alpha^{-m}(I_{v_m}). \tag{2}$$

Hence $\mathbf{r}[i, i+m] = v$ if and only if $R_\alpha^i(\rho) \in I_v$. See [7, Section 2.1.2]. Note that I_v is in general a finite union of half-intervals.

Example 1. For the sake of simplicity we take rational values and show that the set I_v given in (2) is in general a finite union of half-intervals. Take $\alpha = 1/4$ and $I_1 = [9/10, 1)$. The set I_{00} is given by

$$I_0 \cap R_\alpha^{-1}(I_0) = [0, 9/10) \cap [3/4, 13/20) = [0, 13/20) \cup [3/4, 9/10)$$

and $I_{01} = [13/20, 3/4)$, $I_{10} = I_1 = [9/10, 1)$. For instance, one can easily see that I_{0000} is the union

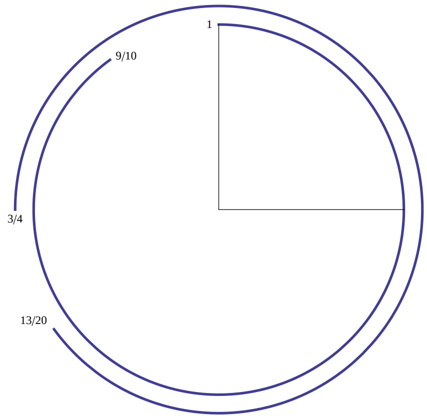

FIGURE 1. The intervals I_0 and $R_\alpha^{-1}(I_0)$.

of four pairwise disjoint half-intervals: $[0, 3/20)$, $[1/4, 3/5)$, $[1/2, 13/20)$ and $[3/4, 9/10)$. Note that any of these four half-intervals can be obtained as the intersection of two half-intervals of the kind $R_\alpha^{-i_1}(I_{j_1})$ and $R_\alpha^{-i_2}(I_{j_2})$. Such an observation will be used in the proof of our main result.

Didier gives a characterization of the coding of a rotation with a partition into m intervals of length greater than α by using Sturmian words and cellular automata [5]. Sturmian words are particular rotation words of the kind $r(\alpha, [1-\alpha, 0), \rho)$ or $r(\alpha, (1-\alpha, 0], \rho)$ where α is irrational. Let us recall some well-known facts about them. An infinite word $\mathbf{x} \in A^\omega$ is called C*-balanced*, for some $C > 0$, if all factors u, v of $\mathbf{x}$ with $|u| = |v|$ satisfy $||u|_a - |v|_a| \leq C$ for all letters $a \in A$. If $C = 1$, then we simply say that $\mathbf{x}$ is balanced.

Theorem 3. [7, Theorem 2.1.5] *A binary infinite word* $\mathbf{x} \in \{\mathtt{0}, \mathtt{1}\}^\omega$ *is Sturmian if and only if it is aperiodic and balanced.*

Let $\mathbf{x} = r(\alpha, [1-\alpha, 0), \rho)$ be a Sturmian word. As a consequence of the above result, for each $n \geq 1$, the set $\{|u|_\mathtt{1} \mid u \in \mathrm{Fac}(\mathbf{x}) \cap \{\mathtt{0}, \mathtt{1}\}^n\}$ contains exactly two elements. Otherwise stated, the abelian complexity of a Sturmian word is constant and takes only the value 2. So, we can speak of "heavy" and "light" factors of length n of the Sturmian word $\mathbf{x}$. The heavy ones contain one more $\mathtt{1}$ than the light ones. Denote the set of heavy factors of length n by $H(n)$. Let us define the following two sets $I_H(n) := \bigcup_{v \in H(n)} I_v$ and $I_L(n) = \mathcal{C}\backslash I_H(n)$. These sets are indeed intervals [10].

$$I_H(n) = [1 - \{n\alpha\}, 1) \quad \text{and} \quad I_L(n) = [0, 1 - \{n\alpha\}). \tag{3}$$

A factor $x_i \cdots x_{i+n-1}$ is heavy if and only if $R_\alpha^i(\rho) \in I_H(n)$.

Rigo, Salimov and Vandomme prove the following two results.

Theorem 4. [10, Theorem 19] *Let* $\mathbf{x} = r(\alpha, \alpha, \rho)$ *be a Sturmian word. The set* $\mathcal{APR}_\mathbf{x}$ *is finite if and only if* $\rho \neq 0$.

Proposition 5. [10, Prop. 26] *If* $\mathbf{x}$ *is an abelian recurrent word such that* $\mathcal{APR}_{\mathbf{x}}$ *is finite, then* $\mathbf{x}$ *has bounded abelian complexity.*

Remark 1. A rotation word $\mathbf{r} = r(\alpha, I_1, \rho)$ with α irrational and I_1 non-empty is uniformly recurrent and in particular, abelian recurrent. Indeed, let v be a factor of $\mathbf{x}$ occurring at position i, i.e., $R_\alpha^i(\rho) \in I_v$. If I_v is a finite union of pairwise disjoint half-intervals, we denote by ℓ the length of the smallest such interval. Due to Kronecker's theorem, there exist s and t such that $\{s\alpha\} \in (0, \ell/2)$ and $\{t\alpha\} \in (1-\ell/2, 1)$. Therefore, for each $x \in I_v$, there exists some $j \le \max\{s,t\}$ such that $R_\alpha^j(x) \in I_v$. Otherwise stated, the factor v occurs with gaps bounded by $\max\{s,t\}$.

2.1. **Rotation words with bounded abelian complexity.** Adamczewski [1] observed that codings of rotations have different combinatorial and arithmetic behaviors depending on whether the parameters (α, β) satisfy $\beta \in \mathbb{Z} + \alpha\mathbb{Z}$. It mainly follows from the following result obtained by Kesten [6].

Theorem 6. *Let* $\alpha \in (0,1)$ *and* $0 \le a < b \le 1$. *There exists a constant* K *such that*

$$\text{for all } M,\ \#\{i \le M \mid \{i\alpha\} \in [a,b)\} - M(b-a) < K \tag{4}$$

if and only if $b - a = \{j\alpha\}$ *for some integer* j.

Let L, N, n be integers. Note first that

$$\#\{L+1 \le i \le L+n \mid \{i\alpha\} \in [a,b)\} - n(b-a)$$

$$= \#\{i \le L+n \mid \{i\alpha\} \in [a,b)\} - (L+n)(b-a) - \#\{i \le L \mid \{i\alpha\} \in [a,b)\} + L(b-a).$$

Therefore, if there exists K such that (4) holds, we get

$$|\#\{L+1 \le i \le L+n \mid \{i\alpha\} \in [a,b)\} - n(b-a)| < 2K.$$

Now observe that $|\#\{L+1 \le i \le L+n \mid \{i\alpha\} \in [a,b)\} - \#\{N+1 \le i \le N+n \mid \{i\alpha\} \in [a,b)\}|$ is exactly $||\mathbf{r}[L+1, L+n]|_1 - |\mathbf{r}[N+1, N+n]|_1|$ where $\mathbf{r} = r(\alpha, [a,b), 0)$. By adding and subtracting $n(b-a)$, this quantity is easily seen to be bounded by $4K$ for all L, N, n. In other words, if (4) holds then $\mathbf{r}$ is $4K$-balanced. Conversely, if $\mathbf{r} = r(\alpha, [a,b), 0)$ is C-balanced for some C, then one can show along the lines of the proof of Proposition 7 in [2] that (4) holds.

Consequently, the rotation word $r(\alpha, [a,b), 0)$ is C-balanced, for some C, if and only if $b - a = \{j\alpha\}$ for some integer j.

Lemma 1. [9] *An infinite word has bounded abelian complexity if and only if it is* C*-balanced for some* $C > 0$.

Remark 2. Considering a rotation word where the interval I_1 is of the kind $[1-\beta, 0)$ and where the starting point ρ may vary has the same degree of freedom as considering a rotation word where the starting point is 0 and the interval may vary. Otherwise stated, a rotation word $r(\alpha, \beta, \rho) = r(\alpha, [1-\beta, 0), \rho)$ is in fact of the kind $r(\alpha, [a,b), 0)$. Indeed, if $0 \le \rho < 1-\beta$, then $r(\alpha, \beta, \rho) = r(\alpha, [1-\beta-\rho, 1-\rho), 0)$. If $1 - \beta \le \rho < 1$, then $r(\alpha, \beta, \rho)$ is the conjugate of $r(\alpha, [1-\rho, 2-\beta-\rho), 0)$.

The next result is merely a consequence of Kesten's theorem.

Proposition 7. *Let* $\alpha, \beta \in (0,1)$. *Let* $\rho \in [0,1)$. *The rotation word* $\mathbf{r} = r(\alpha, \beta, \rho)$ *has bounded abelian complexity if and only if* $\beta = \{m\alpha\}$ *for some integer* m.

As a consequence of this result and Proposition 5, if $\mathbf{r} = r(\alpha, \beta, \rho)$ is such that β is not of the kind $\{m\alpha\}$ for some integer m, then $\mathcal{APR}_{\mathbf{r}}$ is infinite. The question is therefore to determine which rotation words of the kind $\mathbf{r} = r(\alpha, \{m\alpha\}, \rho)$ are such that $\mathcal{APR}_{\mathbf{r}}$ is finite.

3. Finiteness of $\mathcal{APR}_{\mathbf{r}}$ for rotation words

Theorem 8. *Let α be irrational. Let $m \geq 1$ be an integer. Let $\mathbf{r} = r(\alpha, \{m\alpha\}, \rho)$ be a rotation word. The set $\mathcal{APR}_{\mathbf{r}}$ is finite if and only if $\rho \notin \{\{-i\alpha\} \mid 0 \leqslant i < m\}$.*

Proof. Let $\mathbf{r} = r(\alpha, \{m\alpha\}, \rho) = r_0 r_1 r_2 \cdots$ be a rotation word. We define m infinite words by periodic decimation of period m. For $j \in \{0, \ldots, m-1\}$, we set

$$\mathbf{r}^{(j)} = r_j r_{j+m} r_{j+2m} \cdots .$$

Note that each $\mathbf{r}^{(j)}$ is a Sturmian word of the kind $r(\{m\alpha\}, [1 - \{m\alpha\}, 1), \rho + \{j\alpha\})$ that can also be written as $r(\{m\alpha\}, I_1^{(j)}, \rho)$ where $I_1^{(j)} = R_\alpha^{-j}([1 - \{m\alpha\}, 1))$.

Having m Sturmian words at our disposal, we can define as in (3), intervals $I_H(n)$ and $I_L(n)$ corresponding respectively to the heavy and light factors of length n in $\mathbf{r}^{(j)}$. Since Sturmian words with the same slope have the same language of factors, these intervals are the same for all j and (3) becomes

$$I_H(n) = [1 - \{nm\alpha\}, 1) \quad \text{and} \quad I_L(n) = [0, 1 - \{nm\alpha\}). \tag{5}$$

• We will first assume that ρ does not belong to $\{\{-i\alpha\} \mid 0 \leqslant i < m\}$ and we will show that $\mathcal{APR}_{\mathbf{r}}$ is finite.

Consider a factor $r_{im+j} r_{(i+1)m+j} \cdots r_{(i+n-1)m+j}$ of length n occurring in $\mathbf{r}^{(j)}$. This factor is heavy if and only if $R_\alpha^{im+j}(\rho) \in I_H(n)$ or, equivalently, if

$$R_\alpha^{im}(\rho) \in R_\alpha^{-j}(I_H(n)) = \left[R_\alpha^{-nm-j}(0), R_\alpha^{-j}(0)\right) =: I_H^{(j)}(n). \tag{6}$$

As usual, $I_L^{(j)}(n)$ denotes $\mathcal{C} \setminus I_H^{(j)}(n)$.

Now consider an arbitrary factor v occurring in $\mathbf{r}$ of length $|v| = dm + \ell$ with $d \geq 0$ and $0 \leq \ell < m$ and starting in position im for some $i \geq 0$,

$$v = \mathbf{r}[im, (i+d)m + \ell - 1].$$

Considering positions in v congruent to the same value modulo m, and doing so for each congruence class, this factor v can be seen as the shuffle of ℓ factors of length $d+1$ occurring in $\mathbf{r}^{(0)}, \ldots, \mathbf{r}^{(\ell-1)}$:

$$v_0 = r_{im} r_{(i+1)m} \cdots r_{(i+d)m}, \ \ldots, \ v_{\ell-1} = r_{im+\ell-1} r_{(i+1)m+\ell-1} \cdots r_{(i+d)m+\ell-1}$$

and $m - \ell$ factors of length d occurring in $\mathbf{r}^{(\ell)}, \ldots, \mathbf{r}^{(m-1)}$:

$$v_\ell = r_{im+\ell} r_{(i+1)m+\ell} \cdots r_{(i+d-1)m+\ell}, \ \ldots, \ v_{m-1} = r_{im+m-1} r_{(i+1)m+m-1} \cdots r_{(i+d-1)m+m-1}.$$

From the above discussion, for $t \in \{0, \ldots, \ell-1\}$, v_t is a heavy factor of length $d+1$ in $\mathbf{r}^{(t)}$ if and only if

$$R_\alpha^{im}(\rho) \in I_H^{(t)}(d+1).$$

In the same way, for $t \in \{\ell, \ldots, m-1\}$, v_t is a heavy factor of length d in $\mathbf{r}^{(t)}$ if and only if

$$R_\alpha^{im}(\rho) \in I_H^{(t)}(d).$$

Consequently, the number of $\mathtt{1}$'s occurring in v is completely determined by the position of $R_\alpha^{im}(\rho)$ with respect to the m intervals $I_H^{(0)}(d+1), \ldots, I_H^{(\ell-1)}(d+1), I_H^{(\ell)}(d), \ldots, I_H^{(m-1)}(d)$. Consider, for all possible choices $A_0, \ldots, A_{m-1} \in \{H, L\}$, the following 2^m subsets of $\mathcal{C}$

$$\bigcap_{j=0}^{\ell-1} I_{A_j}^{(j)}(d+1) \cap \bigcap_{j=\ell}^{m-1} I_{A_j}^{(j)}(d).$$

Each such nonempty subset is a finite union of some nonempty pairwise disjoint half-intervals. The family of all the half-intervals occurring in any of these subsets is denoted by $\mathcal{I}(d, \ell)$. Clearly, this family defines a partition of $\mathcal{C}$ (this follows from the fact that $I_H^{(j)}(n)$ and $I_L^{(j)}(n)$ is already a partition of $\mathcal{C}$). Let $d \geq 0$ and ℓ be such that $0 \leq \ell < m$. From the above discussion, if there exists some $I \in \mathcal{I}(d, \ell)$ such that $R_\alpha^{im}(\rho) \in I$ and $R_\alpha^{jm}(\rho) \in I$, then the factors $\mathbf{r}[im, (i+d)m + \ell - 1]$ and $\mathbf{r}[jm, (j+d)m + \ell - 1]$ are abelian equivalent. In particular, if $R_{m\alpha}^{i}(\rho) = R_\alpha^{im}(\rho) \in I$ and $\rho \in I$, then $\mathbf{r}[im, (i+d)m + \ell - 1]$ is abelian equivalent to the prefix $\mathbf{r}[0, dm + \ell - 1]$ of length $dm + \ell$ of $\mathbf{r}$.

We claim that there exists ε such that, for all $d \geq 0$ and all $\ell \in \{0, \ldots, m-1\}$, if ρ belongs to some $I \in \mathcal{I}(d, \ell)$, then $|I| \geqslant \varepsilon$. This is enough to ensure the finiteness of $\mathcal{APR}_{\mathbf{r}}$. Indeed, thanks to Corollary 2, there exists an integer n_ε such that, for all $\theta \in \mathcal{C}$, at least one of points in $\{\theta, R_{m\alpha}(\theta), \ldots, R_{m\alpha}^{n_\varepsilon}(\theta)\}$ belongs to I as well as ρ. From the conclusion obtained in the previous paragraph and considering θ of the kind $R_{m\alpha}^i(\rho)$, we conclude that, for all i, at least one of the factors of length $dm + \ell$ starting in position $im, (i+1)m, \ldots, (i+n_\varepsilon)m$ in $\mathbf{r}$ is abelian equivalent to $\mathbf{r}[0, dm + \ell - 1]$. Otherwise stated, the gap between any two occurrences of consecutive factors abelian equivalent to any prefix is bounded by mn_ε.

To conclude this part of the proof, we still need to prove the claim. We take $\varepsilon = \min S$ where

$$\begin{aligned} S = &\{|[\rho, R_\alpha^{-j}0]| \,:\, 0 \leqslant j < m\} \cup \{|[R_\alpha^{-j}0, \rho]| \,:\, 0 \leqslant j < m\} \\ &\cup \{|[R_\alpha^{-i}0, R_\alpha^{-j}0]| \,:\, 0 \leqslant i,j < m;\ i \neq j\} \cup \{|[R_\alpha^{-i}0, R_\alpha^{-j-m}0)| \,:\, 0 \leqslant i,j < m;\ i \neq j\} \\ &\cup \{|[R_\alpha^{-j-m}0, R_\alpha^{-i}0)| \,:\, 0 \leqslant i \neq j < m\}. \end{aligned}$$

We proceed by contradiction. Assume that there exists $I \in \mathcal{I}(n, \ell)$ for some n, ℓ such that $\rho \in I$ and $|I| < \varepsilon$. If I is one of the half-intervals belonging to

$$\mathcal{J} = \bigcup_{A \in \{H,L\}} \left\{ I_A^{(0)}(n+1), \ldots, I_A^{(\ell-1)}(n+1), I_A^{(\ell)}(n), \ldots, I_A^{(m-1)}(n) \right\},$$

then the fact that ρ belongs to I would imply that $|I| \geqslant \varepsilon$. Indeed, from (6), one endpoint of I would be of the kind $R_\alpha^{-j}0$, and hence I would contain an interval either of the kind $[\rho, R_\alpha^{-j}0)$ or $(R_\alpha^{-j}0, \rho]$. Since $|I| < \varepsilon$, we deduce that I is the intersection $[a, d)$ of two half-intervals $[a, b)$ and $[c, d)$ in $\mathcal{J}$ such that $|[a, d)| < \varepsilon$ and $\rho \in [a, d)$.

We consider all the possible cases:

(1) If $[c, d) = I_H^{(j)}(k)$ for some j, k, then we get the contradiction
$$|[a, d)| \geqslant |[\rho, d)| = |[\rho, R_\alpha^{-j}0)| \geqslant \varepsilon.$$
(2) If $[a, b) = I_L^{(j)}(k)$ for some j, k, then we get the contradiction
$$|[a, d)| \geqslant |[a, \rho)| = |[R_\alpha^{-j}0, \rho)| \geqslant \varepsilon.$$
(3) If $[c, d) = I_L^{(j_1)}(k)$ and $[a, b) = I_H^{(j_2)}(k)$ for some j_1, j_2, k, then $[d, c) = I_H^{(j_1)}(k)$ and $a = R_\alpha^{-j_2-km}0$, $b = R_\alpha^{-j_2}0$, $c = R_\alpha^{-j_1}0$, $d = R_\alpha^{-j_1-km}0$. We get the contradiction
$$|[a, d)| = |[b, c)| = |[R_\alpha^{-j_2}0, R_\alpha^{-j_1}0)| \geqslant \varepsilon.$$
(4) If $[c, d) = I_L^{(j_1)}(k_1)$ and $[a, b) = I_H^{(j_2)}(k_2)$ for some j_1, j_2 and $k_1 \neq k_2$, then due to (5), one of a, d is $R_\alpha^{-j_1-km}0$ and the another one is $R_\alpha^{-j_2-(k+1)m}0$. The length of $[a, d)$ is in S and this is again a contradiction.

• We now consider the converse. Suppose that $\rho = R_\alpha^{-k}0$ for some $0 \leqslant k < m$. Note that ρ always lies in the interval $I_L^{(k)}(n) = [R_\alpha^{-k}0, R_\alpha^{-k-nm}0)$. Hence, for all n, the prefix of length n of $\mathbf{r}^{(k)}$ is light. We will show how this fact allows one to find arbitrarily large gaps between consecutive occurrences of factors abelian equivalent to some prefix.

Fix $\ell > 0$ and define

$$\delta = \min\{|[R_\alpha^{-j}0, R_\alpha^{im}\rho]| \,:\, 0 \leqslant j < m,\ 0 < i \leq \ell\}.$$

Thanks to Kronecker's Theorem, there exists n such that $1 - \{nm\alpha\} < \delta$. In view of (5), this means that the m intervals $I_L^{(j)}(n) = [R_\alpha^{-j}0, R_\alpha^{-nm-j}0)$ have length $1 - \{nm\alpha\} < \delta$ for all $j \in \{0, \ldots, m-1\}$. Therefore, by definition of δ, for all $i \in \{1, \ldots, \ell\}$ and all $j \in \{0, \ldots, m-1\}$, we get that $R_\alpha^{im}\rho$ does not belong to $I_L^{(j)}(n)$, i.e., the words

$$r_{im+j} r_{(i+1)m+j} \cdots r_{(i+n-1)m+j}$$

are heavy. The factors $\mathbf{r}[m, m + mn - 1]$,..., $\mathbf{r}[\ell m, \ell m + mn - 1]$ are the shuffle of m of these heavy words. But the prefix of length nm of $\mathbf{r}$ is the shuffle of the prefixes of length n of the words $\mathbf{r}^{(0)}, \ldots, \mathbf{r}^{(m-1)}$ and we already know that the prefix of length n of $\mathbf{r}^{(k)}$ is light. Consequently none of the factors $\mathbf{r}[m, m+mn-1]$,..., $\mathbf{r}[\ell m, \ell m+mn-1]$ are abelian equivalent to $\mathbf{r}[0, mn-1]$, showing that there exists an element in $\mathcal{APR}_{\mathbf{r}}$ of length at least $(\ell-1)m$. Since ℓ can be chosen arbitrarily large, $\mathcal{APR}_{\mathbf{r}}$ is infinite. □

Example 2. Let $\tau = (1+\sqrt{5})/2$ be the golden mean and set $\alpha = 1/\tau^2 = 0.381966\cdots$. Consider the rotation word $\mathbf{r} = r(\alpha, \{2\alpha\}, \alpha)$. We have

$$\mathbf{r} = 110111101101111011110110111101101111\cdots$$

Observe that if $\mathbf{f} = r(\alpha, \{\alpha\}, \alpha)$ is the Fibonacci word, we have

$$\mathbf{f} = 010010100100101001010010010100100\cdots,$$

and furthermore, the word $\mathbf{r}$ can be obtained from $\mathbf{f}$ by the rule $r_i = f_i + f_{i+1}$ for all $i \geq 0$. Alternatively, we have $\mathbf{r} = h(\mathbf{f})$ where h is the substitution that maps $0 \to 11$ and $1 \to 0$.

By Theorem 8, the set $\mathcal{APR}_{\mathbf{r}}$ is finite. To show this, let $\mathcal{I}(d,\ell)$ be as in the as in the proof of Theorem 8. Then there should exist ϵ such that for all $d \geq 0$ and $\ell \in \{0,1\}$, if $\rho \in I$ for some $I \in \mathcal{I}(d,\ell)$, then $|I| \geq \epsilon$. To determine this ϵ, note that since $\rho = \alpha$ in this case, the set S defined in the proof of Theorem 8 is

$$S = \big\{\{\alpha\}, 1-\{\alpha\}, \{2\alpha\}, 1-\{2\alpha\}, \{3\alpha\}, 1-\{3\alpha\}\big\},$$

and therefore, we have $\epsilon = \min S = \{3\alpha\} = 0.1458980\cdots$.

Next, we find n_ϵ such that for all $\theta \in \mathcal{C}$, at least one of the points in $\{\theta, R_{2\alpha}(\theta), \ldots, R_{2\alpha}^{n_\epsilon}(\theta)\}$ belongs to I. A computer calculation shows that if $n_\epsilon = 12$, then when the points in $\{\theta, R_{2\alpha}(\theta), \ldots, R_{2\alpha}^{n_\epsilon}(\theta)\}$ are arranged in order around the unit circle, the largest gap between consecutive points is $0.124611797\cdots$, which is less than ϵ. Therefore, the interval I must contain one of the points. From the proof of Theorem 8 we conclude that for any prefix of $\mathbf{r}$, the longest abelian return has length at most $2 \cdot 12 = 24$. This is not the optimal bound however, since computer calculations suggest that the longest abelian return actually has length 5.

4. Finiteness of semi-abelian returns and the three gap theorem

Puzynina and Zamboni [8] proved the following:

Theorem 9. *A binary recurrent infinite word $\mathbf{x}$ is Sturmian if and only if each factor u of $\mathbf{x}$ has two or three semi-abelian returns in $\mathbf{x}$.*

They give a combinatorial proof of this result. Here we show that one direction of this result is a consequence of the *three gap theorem* (see [3]):

Theorem 10 (Three gap theorem)**.** *Let ρ be a real number, let $\alpha \in (0,1)$ be irrational, and let I be a proper subinterval of $(0,1)$. The gaps between the successive integers j such that $\{j\alpha+\rho\} \in I$ take at most three values, one being the sum of the other two.*

We now use the number-theoretic approach of Rigo, Salimov, and Vandomme [10], together with the three gap theorem to prove:

Proposition 11. *Let $\mathbf{x}$ be a Sturmian word. Each factor u of $\mathbf{x}$ has two or three semi-abelian returns in $\mathbf{x}$.*

Proof. Let $I_0 = [0, 1-\alpha)$ and $I_1 = [1-\alpha, 1)$. Let $\mathbf{x} = x_0x_1\cdots$ be the Sturmian word defined by

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

For a binary word $v = v_0\cdots v_k$, recall (2) that

$$I_v = I_{v_0} \cap R_\alpha^{-1}(I_{v_1}) \cap \cdots \cap R_\alpha^{-k}(I_{v_k}),$$

so that $\mathbf{x}[i, i+k] = v$ if and only if $R_\alpha^i(\rho) \in I_v$. For each length k there are exactly two abelian equivalence classes of words of length k in $\mathbf{x}$: As usual, let us denote these $H(k)$ and $L(k)$, where the words in $H(k)$ have one more 1 than those in $L(k)$.

Rigo, Salimov, and Vandomme [10] proved that

- $\mathbf{x}[i, i+k] \in H(k)$ if $R_\alpha^i(\rho) \in [1-\{k\alpha\}, 1)$; and,

- $\mathbf{x}[i, i+k] \in L(k)$ if $R_\alpha^i(\rho) \in [0, 1 - \{k\alpha\})$.

It now follows from Theorem 10 that the gaps between successive positions i such that $\mathbf{x}[i, i+k] \in H(k)$ take either two or three values. In other words, for any factor $u \in H(k)$, the set of abelian returns to u in $\mathbf{x}$ consist of words of either two or three different lengths. The same reasoning applies to factors in $L(k)$. However, from the argument given in [8, Section 6], the factor u has at most one semi-abelian return of each length. Thus u has either two or three semi-abelian returns in $\mathbf{x}$. □

References

[1] B. Adamczewski, Codages de rotations et phénomènes d'autosimilarité, *J. Théor. Nombres Bordeaux* **14** (2002), 351–386.

[2] B. Adamczewski, Balances for fixed points of primitive substitutions, *Theoret. Comput. Sci.* **307** (2003), 47–75.

[3] P. Alessandri, V. Berthé, Three distance theorems and combinatorics on words, *Enseign. Math. (2)* **44** (1998), 103–132.

[4] M. Bucci, A. De Luca, L. Q. Zamboni, Reversible Christoffel factorizations, `arXiv:1211.3049` [cs.DM]

[5] G. Didier, Combinatoire des codages de rotations, *Acta Arith.* **85** (1998), 157–177.

[6] H. Kesten, On a conjecture of Erdős and Szüsz related to uniform distribution mod 1, *Acta Arith.* **12** 1966/1967, 193–212.

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

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

[9] G. Richomme, K. Saari, L. Q. Zamboni, Abelian complexity in minimal subshifts, *J. London Math. Soc.* **83** (2010), 79–95.

[10] M. Rigo, P. Salimov, E. Vandomme, Some properties of abelian return words, *J. Integer Seq.* (to appear).

(N. Rampersad) Department of Math/Stats, University of Winnipeg, 515 Portage Ave., Winnipeg, MB, R3B 2E9, Canada.

(M. Rigo, P. Salimov) Department of Mathematics, University of Liège, Grande traverse 12 (B37), B-4000 Liège, Belgium. `M.Rigo@ulg.ac.be`