Available online at www.sciencedirect.com
SCIENCE DIRECT®
Theoretical Computer Science 340 (2005) 334–348
Theoretical Computer Science
www.elsevier.com/locate/tcs

Palindromic factors of billiard words

J.-P. Borel[a,1], C. Reutenauer[b,*]

[a] *LACO, UMR CNRS 6090, 123 avenue Albert Thomas, F-87060 Limoges Cedex, France*
[b] *UQÀM, Département de Mathématiques, case postale 8888, succursale Centre-Ville, Montreal, Québec, Canada H3C 3P8*

Abstract

We study palindromic factors of billiard words, in any dimension. There are differences between the two-dimensional case, and higher dimension. Arbitrary long palindrome factors exist in any dimension, but arbitrary long palindromic prefixes exist in general only in dimension 2.

MSC: 68R15

Keywords: Words; Languages; Sturmian; Billiard; Palindromes

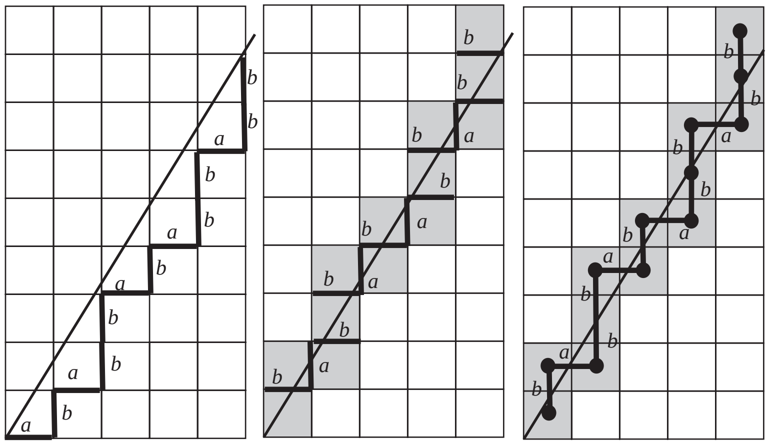
1. Introduction and notations

1.1. Billiard and Christoffel words in dimension 2

Let α be a positive irrational number. In several ways one may associate to it a Sturmian word on the alphabet $\mathcal{A} := \{a, b\}$. We use here a geometrical approach. Consider the grid $\mathcal{G}$ on the first quadrant of the plane: it is the set of vertical half-lines with integer x-coordinate and of horizontal half-lines with integer y-coordinate. The line $\mathcal{D}$ through the origin O and slope α divides $\mathcal{G}$ into two parts. We construct the word u_α and the billiard word c_α (cutting sequence) as follows:

* Corresponding author. Tel.: +1 514 9873000x3228; fax: +1 514 9878274.
E-mail addresses: borel@unilim.fr (J.-P. Borel), christo@math.uqam.ca (C. Reutenauer).
[1] Research partially supported by Région Limousin.

0304-3975/$ - see front matter
doi:10.1016/j.tcs.2005.03.036



(1) (2) (3)

Fig. 1.

Denoting by a the horizontal segment and by b the vertical one, u_α encodes the discrete path immediately under the half-line $\mathcal{D}$; hence in the example $u_\alpha = ababbababbabb\ldots$ (Fig. 1(1)). Looking at the squares crossed by $\mathcal{D}$ (grey in Fig. 1(2)) and their blackened sides, the billiard word c_α encodes the sequence of sides crossed by $\mathcal{D}$ (a for a vertical side, b for an horizontal one). Here $c_\alpha = babbababbabb\ldots$. Equivalently, c_α encodes the discrete path joining the center of the crossed squares (see Fig. 1(3)). It is easily seen that $u_\alpha = ac_\alpha$.

The words u_α are called *Christoffel words* [7] and are particular Sturmian words. Regarding factors, this does not restrict generality. See [1,5,15] for the theory of Sturmian words. Christoffel words are known since Bernoulli and have many applications in mathematics and physics; they are related to continued fractions, Farey sequences, and the Stern–Brocot tree (see e.g. [12] for the latter).

1.2. Billiard words in dimension $\geqslant 3$

Let $\mathcal{D}$ be the half-line of origin O, in k-dimensional space, and parallel to vector $(\alpha_1, \alpha_2, \ldots, \alpha_k)$, with α_i positive. We consider the sequence of k-cubes crossed by $\mathcal{D}$, and *facets* joining each cube to the next: a facet is a subset of the cube formed by all points having a fixed integer ith coordinate.

This ith coordinate will be encoded by a_i, and thus we obtain a sequence on the alphabet $\mathcal{A} = \{a_1, a_2, \ldots, a_k\}$, encoding the facets crossed by the half-line $\mathcal{D}$.

This works soon as one has

$$\frac{\alpha_i}{\alpha_j} \notin \mathbb{Q} \tag{1}$$

for any $i \neq j$: indeed in this case, each facet is crossed in its interior, so that the corresponding intersection point has a unique integer coordinate, its ith coordinate. [2]

In this way we obtain the *billiard word* $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$, or c_α, if we denote $\alpha := (\alpha_1, \alpha_2, \ldots, \alpha_k)$. Note that, as in Fig. 1(3), c_α encodes also the discrete path joining the centers of the k-cubes crossed by $\mathcal{D}$. We use this interpretation in the sequel.

It is known that the number of finite factors of length n of c_α, in dimension 2, is equal to $n+1$. This well-known property characterises the Sturmian words in dimension 2; in dimension 3, the number of finite factors of length n is equal to n^2+n+1, see [2]; in dimension k, it is

$$\sum_{i=0}^{\min(k-1,n)} \binom{k-1}{i} \binom{n}{i} i!,$$

see [3].

1.3. Finite billiard words

Let $M := (m_1, m_2, \ldots, m_k) \in \mathbb{N}^k$, where the m_i are pairwise relatively prime. The segment OM crosses several k-cubes and one defines, as before, a finite word c_M on the same alphabet, called the (*finite*) *billiard word* associated to M. One has

$$|c_M|_{a_i} = m_i - 1, \quad 1 \leqslant i \leqslant k,$$
$$|c_M| = \sum_{i=1}^{k} m_i - k.$$

Note that, as usual, $|v|$ is the length of word v, and $|v|_a$ its a-degree. Observe that c_M is a palindrome, that is, equal to its reversal. We denote by $\widetilde{v}$ the reversal of word v. For a palindrome, one has $v = \widetilde{v}$ by definition.

2. Main results

2.1. Dimension 2

Everything is known in this case, and palindromic factors and prefixes of Sturmian words have been intensively studied; even, they characterize Sturmian words, see [10,11,14].

2.1.1. Palindromic prefixes

Theorem 2.1. *The palindromic prefixes of the infinite billiard word c_α are finite billiard words; for all $n > 0$ they are the prefixes of length $p_n + q_n - 2$, for all the main and intermediate convergents p_n/q_n of the continued fraction expansion of the real number α.*

This result is stated in [4,8,9], in a slightly different formulation.

[2] Note that the previous condition holds if the coordinates α_i are $\mathbb{Q}$-linearly independents, which is necessary for some results.

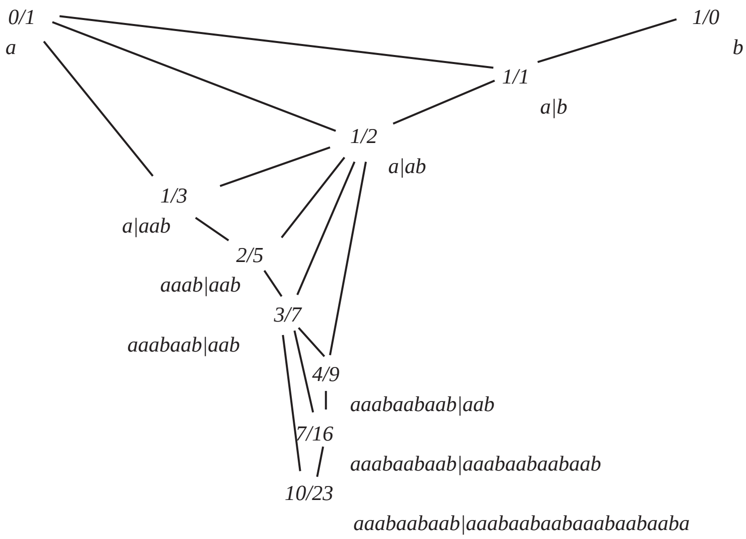


Fig. 2.

Consider an example. Let

$$u_\alpha = ac_\alpha = aaabaabaabaaabaabaabaaabaabaabaab\ldots$$

obtained by taking α in the interval $\frac{7}{16} < \alpha < \frac{10}{23}$. These two numbers have the continued fraction expansion:

$$\frac{7}{16} = \frac{1}{2 + \frac{1}{3 + \frac{1}{2}}} \quad \text{and} \quad \frac{10}{23} = \frac{1}{2 + \frac{1}{3 + \frac{1}{3}}}$$

denoted by $[0; 2, 3, 2]$ and $[0; 2, 3, 3] = [0; 2, 3, 2, 1]$. In other words, $\alpha = [0; 2, 3, 3, \ldots]$. The sequence of intermediate and main convergents is

$$\frac{0}{1}, \frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \frac{4}{9}, \frac{7}{16}, \frac{10}{23}, \ldots$$

which may be read on the Stern–Brocot tree of $\frac{10}{23}$, see Fig. 2

In Fig. 2, we have indicated the factorization of each Christoffel word, coming from the words above. To each such Christoffel word, which will be of the form aub, associate its factor u. These words u are exactly the palindromic prefixes of c_α (here, those of length $\leqslant 33$):

$$c_\alpha = aabaabaabaaabaabaabaaabaabaabaab\ldots$$

a	aa	aaa	$aaabaab$	$aaabaabaab$	$aaabaabaabaaabaabaabaaa$	$aaabaabaabaaabaabaabaaabaabaabaab$
ε	a		$aabaa$	$aabaabaa$	$aabaabaabaaabaabaabaa$	$aaababaabaaabaabaabaaabaabaabaa$

We have listed the Christoffel words on the first line, and the associated billiard words on the second. Note that there exists palindromic factors which are not prefixes, as b, aba, aaa or $baab$. But they are central factors of the previous palindromes.

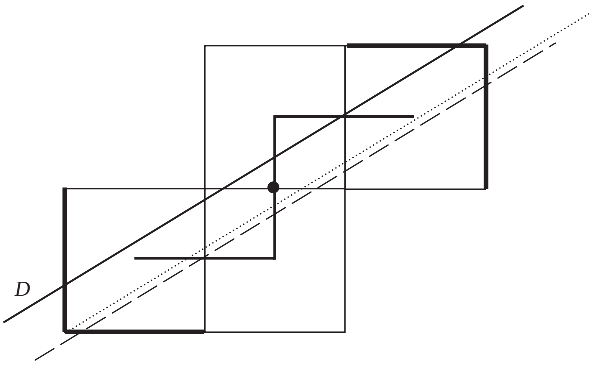


Fig. 3.

2.1.2. Palindromic factors

It is well-known that the language $\mathcal{L}$ of factors of c_α is stable under reversal: $v \in \mathcal{L} \Rightarrow \widetilde{v} \in \mathcal{L}$. Let v be any palindromic factor of c_α. It is sometimes possible to extend v into another factor ava or bvb, and to iterate. In the following, we call *central factor* of a finite word u any factor v such that $u = v_1 v v_2$, with v_1 and v_2 of the same length.

Theorem 2.2. *Each palindromic factor of a billiard word c_α is a central factor of some palindromic prefix of c_α.*

This result is obtained in [8]. Note that palindromic factors characterize Sturmian words, see [11].

Proof. Let v be a maximal palindromic factor of c_α in a nonprefix position; maximal means that nor ava, nor bvb is a factor of c_α. Consider the figure representing the sequence of squares encoded by v.

Since v is a factor of c_α, either avb or bva is a factor of c_α. We consider the first case, as in Fig. 3, corresponding to $v = aba$. Then line $\mathcal{D}$ enters the figure through a vertical line, stays inside it, and leaves the figure through an horizontal line.

By symmetry under the center of the figure (since v is a palindrome), there exist a parallel line entering by an horizontal and leaving by a vertical segment (in the figure, the dotted line with long segments). Hence the parallel line beginning at the lower left point of the figure stays in the figure (in the figure, the dotted line). This means that v is a prefix of c_α. $\square$

2.2.

Thus there are infinitely many palindromic prefixes in an infinite billiard word, and palindromic factors are factors of the palindromic prefixes. They appear infinitely often, since billiard words are recurrent (even uniformly recurrent). The work of Laurent Vuillon [17] gives even precise information on their appearance, through the notion of *first return* of a factor.

2.3. Dimension $\geqslant 3$

2.3.1. Prefix integer point

Let $M := (m_1, m_2, \ldots, m_k) \in \mathbb{N}^k$ and H the orthogonal projection of M on the line through O parallel to $(\alpha_1, \alpha_2, \ldots, \alpha_k)$.

Definition 2.1. A point in $\mathbb{R}^k$ is a 2-integer point if at least two of its coordinates are integers.

Definition 2.2. $M = (m_1, m_2, \ldots, m_k)$ is called an integer prefix of $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ if triangle OMH does not contain any 2-integer point, except O and M.

2.3.2. Palindromic prefixes

Denote by π_{ij} the mapping that associates to a word u on $\{a_1, a_2, \ldots, a_k\}$ the word on $\{a_i, a_j\}$ obtained by erasing all other letters. Then $\pi_{ij}(c_{\alpha_1,\alpha_2,\ldots,\alpha_k})$ is the billiard word c_{α_i,α_j}.

Theorem 2.3. *A prefix v of $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ is palindromic if and only if each $\pi_{ij}(v)$ is a palindromic prefix of c_{α_i,α_j}.*

Theorem 2.4. • *For almost all $(\alpha_1, \alpha_2, \ldots, \alpha_k) \in \mathbb{R}_+^k$, in the sense of Lebesgue, the word $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ has only finitely many palindromic prefixes.*

• *There exist $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ such that $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ has infinitely many palindromic prefixes.*

We shall prove these results in the sequel. We give, for the last property, a proof that will imply that the corresponding lines are dense. In the first case, the number of palindromic factors may be very small: we give an example in dimension 3 where the only nonempty palindromic prefix is the first letter.

According to Theorem 2.3, in order to have palindromic prefixes, there must be some synchronization between the palindromic prefixes of the words $\pi_{ij}(c_{\alpha_1,\alpha_2,\ldots,\alpha_k})$, hence between the corresponding convergents.

2.3.3. Palindromic factors

As said in the abstract, the situation is the same in any dimension.

Theorem 2.5. *Each factor of $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ is a factor of some palindromic factor of $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$. In particular, arbitrary long palindromic factors exist.*

3. Integer prefix point, up-down method and synchronization

3.1. Integer prefix point

We consider $\mathcal{D}$ as before: it is the half-line of origin O and parallel to vector $\alpha := (\alpha_1, \alpha_2, \ldots, \alpha_k)$, with the condition $(\alpha_i/\alpha_j) \notin \mathbb{Q}$ for any $i \neq j$.

Proposition 3.1. *Let $M = (m_1, m_2, \ldots, m_k) \in \mathbb{N}^k$. Let H be the orthogonal projection of M onto $\mathcal{D}$, and T the intersection of $\mathcal{D}$ and the facets of the k-dimensional parallelepiped $\mathcal{P}$ of long diagonal OM. Then the following conditions are equivalent*:

(i) *M is an integer prefix point of $(\alpha_1, \alpha_2, \ldots, \alpha_k)$.*

(i′) *There is no 2-integer point in the triangle OMT, except O and M.*

(ii) *Triangle OMH intersects the integer k-cubes in their facets, except for the points O and M.*

(iii) *The finite billiard word c_M is a prefix of the infinite billiard word c_α.*

Proof. Let M' be any point on HM and consider the finite word v' encoding the intersection of the segment OM' with the facets of the k-dimensional grid in $\mathbb{R}^k$. When $M = H$, $v' = v$ is a prefix of c_α. When $M' = M$, $v' = c_M$.

Property (iii) means that v' is constant when M' varies on segment MH: each segment OM' meets the k-cubes by the same facets. Hence, that segment never contains a 2-integer point. Hence (ii) is true, and (i) is equivalent to (ii).

Now, triangle OMT is contained in OMH, thus (i) implies (i′). Conversely, if (i′) is true, when $\mathcal{D}$ leaves $\mathcal{P}$ it enters some k-cube of which M is a vertex. This means that H is interior to this k-cube, and each point of triangle MTH, not on MT, is in the interior of this k-cube, hence has no integer coordinate, and this proves (i). □

In dimension 2, these integer prefix points are well-known, and 2-integer points are exactly integer points. Consider the billiard word c_α, with $\alpha = (\alpha_1, \alpha_2)$ and α_1/α_2 irrational. Let $\mathcal{D}$ be the half-line of origin O and slope α_2/α_1.

Proposition 3.2. *Let v be a prefix of c_α and M the point (m_1, m_2), with $m_i := |v|_i + 1$. The following conditions are equivalent*:

(1) *v is a palindrome.*

(1′) *$v = c_M$.*

(2) *Triangle OMH contains no integer point except O and M.*

(3) *Distance MH is minimal among all distances $M'H'$, where M' is an integer point, on the same side of $\mathcal{D}$ as M, and such that its orthogonal projection H' onto $\mathcal{D}$ is between O and H.*

(4) *m_2/m_1 is an intermediate or main convergent of α_2/α_1.*

Equivalence of (2)–(4) is in [6], and the equivalence between (1) and (4), i.e. Theorem 2.1, is in [8].

3.2. *Up and down*

3.2.1. *On prefix integer points*

Let $M = (m_1, m_2, \ldots, m_k)$ be a prefix integer point of $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, and $i \neq j$ in $\{1, \ldots, k\}$. Let $M_{ij} \in \mathbb{Z}^2$ be the projection (m_i, m_j), and $\alpha_{ij} = (\alpha_i, \alpha_j)$.

Proposition 3.3. *M is a prefix integer point of α if and only if each M_{ij} is a prefix integer point of α_{ij}.*

Proof. Denote also by π_{ij} the projection $\mathbb{R}^k \to \mathbb{R}^2$ which sends α onto (α_i, α_j). Then $\pi_{ij}(\mathcal{D})$ is the half-line $\mathcal{D}_{ij}$, $\pi_{ij}(T) = T_{ij}$, T_{ij} is the point where $\mathcal{D}_{ij}$ leaves the rectangle of diagonal OM_{ij}. Hence:

$$OM_{ij}T_{ij} = \pi_{ij}(OMT).$$

If in triangle OMT there is no other point than O and M, having at least two integer coordinates, the same holds for all triangles $OM_{ij}T_{ij}$, and hence M_{ij} is a prefix integer point of (α_i, α_j).

If however there is some 2-integer point in triangle OMT, different from O, M, let i, j corresponding to its integer coordinates. Then, by projection, there is some integer point in triangle $OM_{ij}T_{ij}$, different from O and M_{ij}. □

3.2.2. On palindromes

Proposition 3.4. *Let v be a finite word on $\mathcal{A} = \{a_1, a_2, \ldots, a_k\}$. Then v is a palindrome if and only if all its projections π_{ij} are palindromes.*

Proof. This is clearly necessary. Suppose now that v is not palindrome: $v \neq \widetilde{v}$. Then $v = wa_iw'$, $\widetilde{v} = wa_jw''$, where w is the longest common prefix of v and $\widetilde{v}$, and thus $i \neq j$. Then:

$$\pi_{ij}(v) = \pi_{ij}(w)a_i\pi_{ij}(w') \neq \pi_{ij}(w)a_j\pi_{ij}(w'') = \pi_{ij}(\widetilde{v}) = \widetilde{\pi_{ij}(v)}$$

which means that $\pi_{ij}(v)$ is not a palindrome. □

Theorem 2.3 is an immediate consequence, since for v prefix of $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$, $\pi_{ij}(v)$ is a prefix of $\pi_{ij}(c_{\alpha_1,\alpha_2,\ldots,\alpha_k}) = c_{\alpha_i\alpha_j}$.

4. Auxiliary results on continued fractions

4.1. A probabilistic result

Let $q = q_n$ the denominator of a main convergent of the real number α (see [13, Chapter X] or [16] for general results on continued fractions). Then by [13, Theorem 171, p. 140]:

$$\left|\alpha - \frac{p_n}{q_n}\right| \leqslant \frac{1}{q_nq_{n+1}}.$$

Thus, since $q_{n+1} = a_{n+1}q_n + q_{n-1}$, where all these numbers are integers and positive, one has

$$\|\alpha q_n\| \leqslant \frac{1}{q_{n+1}} \leqslant \frac{1}{q_n}.$$

We denote as usual by $\|x\|$ the distance of x to the closest integer. When x varies between 0 and 1, the inequality $\|qx\| \leqslant 1/q$ is satisfied in a union of $q + 1$ disjoint intervals, whose sum of lengths is $2/q$. Taking uniform probability, we deduce that the probability that x

in [0, 1] has q as denominator of some main convergent is bounded by $2/q$. Note that this probability exists since the set of all corresponding x is a finite union of intervals, except rational numbers. Note also that this argument does not work for intermediate convergents, since $\|qx\|$ may be bigger than $1/q$.

Proposition 4.1. *Let q be a positive integer $\geqslant 2$, and $0 < x < 1$. Then the probability P_q that q be a denominator of a main or intermediate convergent of x satisfies*:

$$P_q \leqslant \frac{2}{\sqrt{q}} + \frac{2}{\sqrt{q}-1}.$$

This result is certainly not optimal, but sufficient for our purposes. Note that P_q exists for the same reason as above.

Proof. Let q be the denominator of some intermediate convergent of α. Then one has, for denominators q_n and q_{n+1} of two successive main convergents: $q_n \leqslant q < q_{n+1}$, $q = aq_n + q_{n-1}$, $1 \leqslant a \leqslant a_n - 1$ where all these numbers are positive integers. Moreover, the intermediate convergent is

$$\frac{p}{q} = \frac{ap_n + p_{n-1}}{aq_n + q_{n-1}}.$$

We have two cases.

- If $a > \sqrt{q} - 1$, then $q = aq_n + q_{n-1} > (\sqrt{q} - 1)q_n$ and $q_n < q/(\sqrt{q} - 1)$. Moreover:

$$\|q_n\alpha\| < \frac{1}{q_{n+1}} \leqslant \frac{1}{q}$$

  and so α belongs to a set whose probability is $2/q$. Since q_n may take any integer value between 1 and $q/(\sqrt{q} - 1)$, the probability that x has q as denominator of main or intermediate convergent with $a > \sqrt{q} - 1$ is bounded by

$$\frac{2}{q} \sum_{1 \leqslant i < \frac{q}{\sqrt{q}-1}} 1 \leqslant \frac{2}{\sqrt{q}-1}.$$

- If $a \leqslant \sqrt{q} - 1$, then

$$q = aq_n + q_{n-1} < (a+1)q_n \leqslant \sqrt{q}\,q_n$$

  which implies $q_n > \sqrt{q}$, and

$$\|q\alpha\| < \|q_{n-1}\alpha\| < \frac{1}{q_n} < \frac{1}{\sqrt{q}}.$$

  Indeed, the first inequality is a consequence of the theory of continued fractions, since the points (p, q) associated to main and intermediate convergents on the same side of $\mathcal{D}$ are closer and closer to $\mathcal{D}$.

  Probability that x satisfies this inequality is bounded by $2/\sqrt{q}$.

We conclude by summing the two bounds above. □

4.2. Synchronization of convergents

Proposition 4.2. *For almost all positive real number α, the set of positive real numbers β, having an infinity of denominators of intermediate or main convergents in common with α, has Lebesgue measure* 0.

Proof. Let $f(n)$ be an increasing function, such that the series $\sum_{n=1}^{\infty} 1/f(n)$ converges. Then for almost all positive real numbers the sequence of partial quotients a_n of α satisfies $a_n \leqslant f(n)$ for large n, see [16]. We only consider these numbers α, for $f(n) := n^2$, hence there exists a positive constant C_α such that $a_n \leqslant C_\alpha n^2$.

Let $S := \sum 1/\sqrt{q}$ where the sum is over all intermediate or main convergents of α. Then

$$S \leqslant \sum_{n=1}^{\infty} \frac{a_n}{\sqrt{q_n}} < 2C_\alpha \sum_{n=1}^{\infty} \frac{n^2}{\left(\frac{\sqrt{5}+1}{2}\right)^{n/2}} < \infty,$$

where (q_n) is the sequence of denominators of main convergents of α, and where in S we have grouped those q with $q_n \leqslant q < q_{n+1}$, hence $1/\sqrt{q} \leqslant 1/\sqrt{q_n}$; there are a_n such q. The last inequality follows from the fact that the denominators of main convergents are minimal for the golden number $\psi := [1; 1, 1, \ldots]$, and then equal to the Fibonacci sequence:

$$F_n = \frac{5+\sqrt{5}}{10}\left(\frac{\sqrt{5}+1}{2}\right)^n + \frac{5-\sqrt{5}}{10}\left(\frac{1-\sqrt{5}}{2}\right)^n > \frac{1}{2}\left(\frac{\sqrt{5}+1}{2}\right)^n.$$

Thanks to Proposition 4.1, the series $\sum P_q$ converges. Then, the lemma of Borel–Cantelli implies that the set of $\beta \in [0, 1]$ having an infinity of these q as denominator of convergent (main or intermediate) is of measure 0.

Since denominators of convergents depend only on the fractional part of β, so is the set of positive real numbers β. □

5. Proof of main results

5.1. Existence of arbitrarily long palindromic factors

Proposition 5.1. *Let the α_j's be $\mathbb{Q}$-linearly independents. The word $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ contains arbitrarily long palindromic factors. More precisely*:

- *it contains arbitrarily long palindromic factors of even length,*
- *it contains for any i arbitrary long palindromic factors of odd length and central letter a_i.*

In the proof below, we use two-sided infinite words, and billiard words. Their definition is straightforward.

Proof. Let $C := (\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})$. Consider the line $\mathcal{D}'$ passing through C and parallel to $(\alpha_1, \alpha_2, \ldots, \alpha_k)$. The associated billiard word c is well-defined, since the quotients α_i/α_j

are irrational for $i \neq j$. Moreover, due to the linear independence over $\mathbb{Q}$, $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ has the same factors as c. Observe that C is a center of symmetry for the integer lattice, and for the above line $\mathcal{D}'$. Hence the right infinite word defined by the half-line $\mathcal{D}'_+$ after C is the reversal of the left infinite word defined by the symmetric half-line $\mathcal{D}'_-$ before C. Thus $c = \widetilde{v}v$. Hence for each prefix w of v, $\widetilde{w}w$ is a palindrome of even length which is factor of c, hence of $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$. We conclude since w is arbitrary long.

For the second assertion, we argue similarly, by replacing C by the point $(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}, 0, \frac{1}{2}, \ldots, \frac{1}{2})$ with a 0 in ith position. □

5.2. Palindromic prefixes: general case

Let $k \geqslant 3$ and v a palindromic prefix of $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ and $M = (m_1, m_2, \ldots, m_k)$ its integer prefix point corresponding to v. Then, by Proposition 3.3, (m_2, m_1) and (m_3, m_1) are integer prefix points of (α_3, α_1) and (α_3, α_1). Hence m_1 is a denominator of some intermediate or main convergent of α_2/α_1 and α_3/α_1.

Hence for almost all (α_2, α_3), and a fortiori for almost all $(\alpha_1, \alpha_2, \ldots, \alpha_k)$, m_1 is bounded. This means that the number of occurrences of letter a_1 in v is bounded, and thus v is of bounded length, since $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ has infinitely many occurrences of letter a_1. Even, for these $(\alpha_1, \alpha_2, \ldots, \alpha_k)$, there are only finitely many palindromic prefixes.

This proves the first part of Theorem 2.4. □

5.3. An example

Let

$$(\alpha, \beta, \gamma) = \left(1, \frac{15+\sqrt{5}}{10}, \frac{1+\sqrt{5}}{2}\right).$$

The expansion into continued fractions are

$$\frac{\beta}{\alpha} = \frac{15+\sqrt{5}}{10} = [1; 1, 2, 1, 1, 1, \ldots] \quad \text{and}$$
$$\frac{\gamma}{\alpha} = \frac{1+\sqrt{5}}{2} = [1; 1, 1, 1, 1, 1, \ldots].$$

The sequence of denominators of intermediate and main convergents are respectively

$$(1, 1, 2, 3, 4, 7, 11, \ldots) \quad \text{and} \quad (1, 2, 3, 5, 8, 13, \ldots)$$

and all convergents are main convergents, except that corresponding to 2 in the left-hand case. The remaining denominators satisfy the same recursion $q_{n+2} = q_{n+1} + q_n$ and hence the values that appear cannot be equal: in the first case, denominators are $q_n = F_{n-1} - F_{n-5}$ for $n \geqslant 6$, and $q_n = F_n$ in the second one. Only $(1, 2, 3)$ are common to both sequences: indeed $F_{n-1} = F_n - F_{n-2} < F_n - F_{n-4} < F_n$ for $n \geqslant 4$.

The billiard word $c_{\alpha,\beta,\gamma}$ is

$$c_{\alpha,\beta,\gamma} = bcabcbcabcbacbcabcbcabcbacbcbacbacbcbacbbc\ldots$$

whose projections onto $\{a, b\}$ and $\{a, c\}$ are

$$c_{\alpha,\beta} = babbabbababbabbabbababbabb\ldots$$

and

$$c_{\alpha,\gamma} = caccacaccaccacaccacaccacc\ldots .$$

The synchronization principle shows that palindromic prefixes of $c_{\alpha,\beta,\gamma}$ corresponds to palindromic prefixes of $c_{\alpha,\beta}$ and $c_{\alpha,\gamma}$ having the same occurrence of letters a. Since in dimension 2, palindromic prefixes correspond to denominators of convergents (main or intermediate), with $q = 1+$ numbers of a's, only common values of q are possible, and hence the number of a's can be only 0, 1 or 2.

Hence, looking for these prefixes, we find for $c_{\alpha,\beta}$ the words b, bab et $babbab$ and for $c_{\alpha,\gamma}$ the words c, cac et $caccac$. Moving up to $c_{\alpha,\beta,\gamma}$, we find, according to the number of a's (which is 0, 1 or 2):

- for 0, b or bc,
- for 1, $bcabc$,
- for 2, $bcabcbcabcb$.

Hence, only b is a palindromic prefix of $c_{\alpha,\beta,\gamma}$.

5.4. *Palindromic prefixes of arbitrarily long length*

We need the following result on continued fractions due to Lagrange [13, Theorem 184, p. 153].

Proposition 5.2. *Let x be a real number and p/q a rational number such that*

$$\left| x - \frac{p}{q} \right| < \frac{1}{2q^2}.$$

Then p/q is a main convergent of x.

Let n be any fixed integer. We consider any numbers $\alpha_1, \alpha_2, \ldots, \alpha_k$ whose continued fractions expansion are given until the order n. We deduce from it the existence of a positive constant K bigger than any convergents p/q of all these α_i, and bigger than any ratios $p/q\alpha_i$, where p/q is a convergent of α_i.

The corresponding denominators are denoted by $q_j^{(i)}$, $1 \leqslant i \leqslant k$ and $0 \leqslant j \leqslant n$. The expansion is extended two ranks more:

- $n + 1$: for any i,

$$q_{n+1}^{(i)} = a_{n+1}^{(i)} q_n^{(i)} + q_{n-1}^{(i)},$$

where $a_{n+1}^{(i)}$ is chosen in such a way that $q_{n+1}^{(i)} =: \pi_{n+1}^{(i)}$ are distinct prime numbers. This is possible since $q_n^{(i)}$ and $q_{n-1}^{(i)}$ are relatively prime, by using Dirichlet's theorem. We may even assume that

$$a_{n+1}^{(i)} > A := 4K^2.$$

- $n+2$: since the $\pi_{n+1}^{(i)}$ are distinct prime numbers, we may find by the Chinese remainder theorem an arbitrary big integer Q such that:

$$Q \equiv q_n^{(i)} \pmod{\pi_{n+1}^{(i)}}.$$

Thus we may find $a_{n+2}^{(i)}$ such that:

$$Q = q_{n+2}^{(i)} = a_{n+2}^{(i)} q_{n+1}^{(i)} + q_n^{(i)}.$$

This construction is iterated, and leads to numbers $\alpha_1, \alpha_2, \ldots, \alpha_k$, such that from some rank, they have the synchronization property for all steps of type "$n+2$" above. For such a rank, the convergent of each α_i can be written

$$\frac{p_{n+2}^{(i)}}{q_{n+2}^{(i)}} = \frac{P^{(i)}}{Q}$$

as the denominator $q_{n+2}^{(i)}$ does not depend on i, and the index $n+2$ is cancelled for simplicity reason, for a given n.

Lemma 1. *$P^{(i)}/P^{(j)}$ is a main convergent of $\frac{\alpha_i}{\alpha_j}$ if $i \neq j$.*

Proof. Since the coefficients $a_{n+3}^{(i)}$ are of type "$n+1$" above, we have the corresponding inequality:

$$\left|\alpha_i - \frac{P^{(i)}}{Q}\right| = \left|\alpha_i - \frac{p_{n+2}^{(i)}}{q_{n+2}^{(i)}}\right| < \frac{1}{q_{n+3}^{(i)} q_{n+2}^{(i)}} < \frac{1}{a_{n+3}^{(i)} q_{n+2}^{(i)2}} < \frac{1}{AQ^2}.$$

Hence

$$\frac{\alpha_i}{\alpha_j} - \frac{P^{(i)}}{P^{(j)}} = \frac{(Q\alpha_i - P^{(i)})P^{(j)} - (Q\alpha_j - P^{(j)})P^{(i)}}{Q\alpha_j P^{(j)}}$$

$$\left|\frac{\alpha_i}{\alpha_j} - \frac{P^{(i)}}{P^{(j)}}\right| < \frac{P^{(j)} + P^{(i)}}{AQ^2\alpha_j P^{(j)}} < \left(\frac{P^{(j)}}{Q} + \frac{P^{(i)}}{Q}\right) \frac{P^{(j)}}{Q\alpha_j} \frac{1}{A} \frac{1}{P^{(j)2}}$$

$$< \frac{2K^2}{A} \frac{1}{P^{(j)2}} < \frac{1}{2P^{(j)2}}$$

which implies the lemma by Proposition 5.2. □

Hence $(P^{(1)}, P^{(2)}, \ldots, P^{(k)})$ is an integer prefix point of $(\alpha_1, \alpha_2, \ldots, \alpha_k)$, and the prefix of length $\sum P^{(i)} - k$ of $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ is a palindrome. These points are in infinite number, since this happens for $n+2$, $n+4$, $n+6$ and so on.

5.4.1. An example

Let

$$\alpha = \left(\frac{\sqrt{5}-1}{2}, \frac{\sqrt{5}+1}{2}, \frac{\sqrt{5}+3}{2}, \frac{\sqrt{5}+5}{2}, 1\right) \in \mathbb{R}^5.$$

The denominators of convergents of α_i/α_5 are the same, and correspond to the Fibonacci sequence. The α_i are not independent over $\mathbb{Q}$, but α_i/α_j is irrational if $i \neq j$. The word $w = c_{\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5}$ is

$$w = dcdbcdedcbdcadbcdedcbdcdedcbadcdbcdedcbdcdab$$
$$cdedcdbcdedcbdacdbcdedcddcde\ldots .$$

The number 2 in the Fibonacci sequence corresponds to the approximations $(\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \frac{7}{2}, \frac{2}{2})$, hence to the prefix of length $13 = 1 + 3 + 5 + 7 + 2 - 5$. This prefix is $dcdbcdedcbdca$, which is not a palindrome (but almost: replace the last letter a by d, which is the following letter in w).

Number 8 in the Fibonacci sequence corresponds to the approximations $(\frac{5}{8}, \frac{13}{8}, \frac{21}{8}, \frac{29}{8}, \frac{8}{8})$, hence to the prefix of length $71 = 5 + 13 + 21 + 29 + 8 - 5$. This prefix is

$$dcdbcdedcbdcadbcdedcbdcdedcbadcdbcdedcb$$
$$dcdabcdedcdbcdedcbdacdbcdedcbdcd.$$

It is a palindrome.

This occurs again with 34, which results in a palindromic prefix of length 317, and from there, with periodicity 3 on the Fibonacci sequence. In this way are obtained all palindromic prefixes of w, which are curiously all of odd length, with central letter e, except for the two first d and dcd.

Acknowledgements

The authors thank Srecko Brlek for a preliminary computing in dimension 3, and Aldo de Luca for an important mail exchange on this subject.

References

[1] J.-P. Allouche, J. Shallit, Automatic Sequences, Cambridge University Press, Cambridge, 2003.

[2] P. Arnoux, C. Mauduit, I. Shiokawa, J.I. Tamura, Complexity of sequences defined by billiard in the cube, Bull. Soc. Math. France 122 (1994) 1–12.

[3] Y. Baryshnikov, Complexity of trajectories in rectangular billiards, Comm. Math. Phys. 174 (1995) 43–56.

[4] J. Berstel, A. de Luca, Sturmian words, Lyndon words and trees, Theoret. Comput. Sci. 178 (1997) 171–203.

[5] J. Berstel, P. Seebold, Sturmian words, in: M. Lothaire (Ed.), Algebraic Combinatorics on Words, Cambridge University Press, Cambridge, 2002.

[6] J.-P. Borel, F. Laubie, Quelques mots sur la droite projective réelle, J. Théorie Nombres Bordeaux 5 (1993) 23–51.

[7] E.B. Christoffel, Observatio arithmetica, Ann. Mat. 6 (1875) 148–152.

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

[9] A. de Luca, Combinatorics of standard Sturmian words, Structures in Logic and Computer Science, Lecture Notes in Computer Science, Vol. 1261, 1997, pp. 249–267.

[10] X. Droubay, Palindromes in the Fibonacci word, Inform. Process. Lett. 55 (1995) 217–221.

[11] X. Droubay, G. Pirillo, Palindromes and Sturmian words, Theoret. Comput. Sci. 223 (1999) 73–85.

[12] R.L. Graham, D.E. Knuth, O. Patashnik, Concrete Mathematics, second ed., Addison-Wesley, Reading, MA, 1994.

[13] G.H. Hardy, E.M. Wright, An Introduction to the Theory of Numbers, fourth ed., Clarendon Press, Oxford, 1975.
[14] G. Pirillo, A new characteristic property of the palindrome prefixes of a standard Sturmian word, Sem. Lotharingien Combin. 43, 1999 (electronic, see http://www.mat.univie.ac.at/∼slc/).
[15] N. Pytheas Fogg, in: V. Berthé, et al. (Ed.), Substitutions in Dynamics, Arithmetics and Combinatorics, Lecture Notes in Mathematics, Vol. 1794, 2002.
[16] A.M. Rockett, P. Szüsz, Continued Fractions, World Scientific Pub. Co, Singapore, 1992.
[17] L. Vuillon, A characterization of Sturmian words by return words, European J. Combin. 22 (2001) 263–275.