Some New Results on Palindromic Factors of Billiard Words

Jean-Pierre Borel[1,⋆] and Christophe Reutenauer[2]

[1] LACO, UMR CNRS 6090, 123 avenue Albert Thomas,
F-87060 Limoges Cedex, France
borel@unilim.fr
[2] UQÀM, Département de Mathématiques, case postale 8888,
succursale Centre-Ville, Montreal (Québec), H3C 3P8, Canada
christo@math.uqam.ca

Abstract. For heuristic reasons billiard words may have more palindromic factors than any other words. Many results are already known, concerning the palindromic factors and the palindromic prefixes of Sturmian words and billiard words on two letters. We give general results concerning multidimensional billiard words, which describe very different situations. In some cases, these words have arbitrary long palindromic prefix factors. In other cases, at the opposite, they have finitely many distinct palindromic factors.

Keywords: Words, languages, Sturmian, Billiard, palindromes.

AMS classification: 68R15.

1 Introduction

1.1 Palindromic Factors

We consider finite or infinite words on the finite alphabet $\mathcal{A} := \{a_1, a_2, \ldots, a_k\}$, with $k \geq 2$. In the following we use a geometrical approach of words, and k can be viewed as the *dimension* of the words. A finite word v is called a *palindromic word* when it is equal to its reversal. We denote by $\widetilde{v}$ the reversal of word v.

We can make two preliminary remarks:

1. The palindromic property corresponds to a symmetry property of the finite word. An infinite word on $\mathcal{A}$ corresponds to a trajectory, or a curve, in the k-dimensional space $I\!R_+^k$, so that infinite words with many palindromic factors may correspond to curves which are locally invariant by many central symmetries. The best candidates are lines, which correspond to the so-called *billiard words*.
2. Letters are trivial palindromic words of length 1. In dimension 2, any word of length bigger than 2 has a non trivial palindromic factor. In the other sense, in any higher dimension, there exist infinite words without any palindromic factors, e.g. $(a_1a_2a_3)^\infty$.

⋆ Research partially supported by Région Limousin

C. De Felice and A. Restivo (Eds.): DLT 2005, LNCS 3572, pp. 180–188, 2005.

1.2 Billiard Words

There exist many different ways to define billiard words, which are special cases of general Sturmian words, especially in dimension 2. Here we choose the geometrical one. We consider a vector $\alpha := (\alpha_1, \alpha_2, \ldots, \alpha_k)$ with α_i positive. Let $\mathcal{D}$ be the half-line of origin O and parallel to $(\alpha_1, \alpha_2, \ldots, \alpha_k)$. We construct the *billiard word*, or *cutting sequence*, c_α as follow.

1.2.1 Billiard Words in Dimension $k = 2$

There are three different ways to define c_α, see Fig.1.a:

1. by looking at the horizontal and vertical segments on the grid $\mathcal{G}$, which is the set of vertical half-lines with integer x-coordinate and of horizontal half-lines with integer y-coordinate. We denote by a_1 the horizontal segement and by a_2 the vertical one. Then we encode the discrete path immediately under the half-line $\mathcal{D}$, and we obtain the *Christoffel word* $u_\alpha = a_1 c_\alpha$;
2. by moving from O to infinity on the half-line $\mathcal{D}$, we encode by a_1 a crossing point (black point) with a vertical line and by a_2 a crossing point with an horizontal line. This gives the infinite word c_α;
3. by looking at the centers (white points) of the unit squares crossed by $\mathcal{D}$. These centers are ordered, two consecutive centers correspond to joining squares, so that the vector joining these points is one of the vector of the canonical basis (e_1, e_2). We encode by a_i the vector e_i, and we obtain the infinite word c_α.

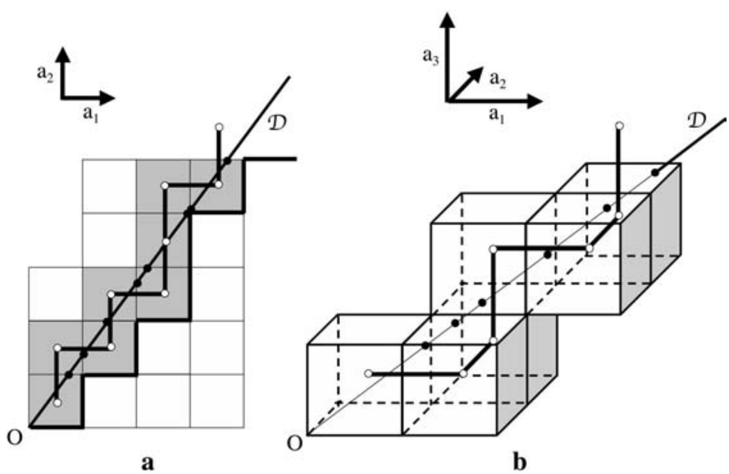

Fig. 1.

This works as soon as the half-line $\mathcal{D}$ has no integer point except for the origin, i.e., $\dfrac{\alpha_1}{\alpha_2}$ is an irrational number.

The same construction can be made using any half-lines $\mathcal{D}$, this gives general Sturmian words. These words have been intensively studied, see e.g. [1], [5],

[17], and are related to continued fraction expansions, Farey sequences, and the Stern-Brocot tree, [13] or [6]. The Christoffel words first appear in [7].

1.2.2 Billiard Words in Dimension $k \geq 3$

We consider some kinds of points in the k-dimensional space.

Definition 1.1. *Let $M = (x_1, x_2, \dots, x_k) \in \mathbb{R}_+^k$ be a k-dimensional point with positive coordinates. Such a point is called:*

- *a 2 -integer point when at least two coordinates x_j are positive integers;*
- *a visible point whenever there are no 2-integer points N on the segment OM, except for the endpoints O and M;*
- *an integer point when all its coordinates are positive integers;*
- *a visible integer point when it is both visible and integer.*

We consider the *facets* of the unit k-cubes: a facet is a subset of the k-cube formed by all points having a fixed i-th coordinate. The methods 2. and 3. above can be generalized in any dimension. This works as soon as $\mathcal{D}$ crosses each facet in its interior, i.e., the half-line $\mathcal{D}$ has no 2-integer point except for O. This property corresponds to

$$\frac{\alpha_i}{\alpha_j} \notin \mathbb{Q}, 1 \leq i < j \leq k. \tag{1}$$

This condition already holds if we have:

$$\text{the } \alpha_i\text{'s are } \mathbb{Q}\text{-linearly independent} \tag{2}$$

and we say that $(\alpha_1, \alpha_2, \dots, \alpha_k)$ is *totally irrational*, see [2]. These two conditions are the same only in dimension 2.

1.2.3 Finite Billiard Words

Let $M := (m_1, m_2, \dots, m_k) \in \mathbb{N}^k$, where the m_i are pairwise coprime. 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:

$$\begin{cases} |c_M|_{a_i} = m_i - 1, 1 \leq i \leq k \\ |c_M| = \displaystyle\sum_{i=1}^{k} m_i - k \end{cases}$$

Note that, as usual, $|v|$ is the length of word v, and $|v|_a$ its a-degree. Observe that c_M is a palindrome.

1.2.4 Billiard Words with Intercept

The same construction can be made for any half-line $\mathcal{D}$ parallel to $(\alpha_1, \alpha_2, \dots, \alpha_k)$ and starting from any point S as soon as $\mathcal{D}$ does not contain any 2-integer point. This condition is assumed in the following. By method 2. or 3., we construct the *billiard word with intercept* denoted by $c_{\alpha,S}$. So we have $c_\alpha = c_{\alpha,O}$, and by translation it suffices to consider the starting points S on the facets of the unit k-cube at the origin, i.e., with at least one zero coordinate s_j and the other ones between 0 and 1.

2 Factors of Billiard Words

2.1 Some Notations

For a given vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, we consider the subspace $\mathcal{D}^\perp$ of the k-dimensional space, and the orthogonal projection $\mathcal{P}$ of the open unit k-cube centered at the origin onto $\mathcal{D}^\perp$. We denote by b_j, $1 \leq j \leq k$, the orthogonal projection onto $\mathcal{D}^\perp$ of the vectors e_j of the canonical basis of $I\!R^k$. C denotes the orthogonal projection of the center of the first k-unit cube in the grid, i.e., $\vec{OC} = \frac{1}{2}\sum_{j=1}^{k} b_j$, and by $\tilde{S}$ the crossing point of $\mathcal{D}$ and $\mathcal{D}^\perp$.

Definition 2.1. *A finite sequence $H_0, H_1, H_2, \ldots, H_n$ of points in $\mathcal{P}$ is called a b-trajectory if for each $1 \leq i \leq n$, there exists some $j = j(i)$ such that $\overrightarrow{H_{i-1}H_i} = b_j$.*

2.2 *b*-Trajectories and Factors of Billiard Words

Such a b-trajectory can be characterized by its origin H_0 and its coding word $a_{j(1)}a_{j(2)}\ldots a_{j(n)}$ in $\mathcal{A}$. Let $\mathcal{F}_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ be the language, i.e., the set of factors, of all the billiard words $c_{\alpha_1,\alpha_2,\ldots,\alpha_k,S}$.

Theorem 2.1. *1. For a given length n and almost all points H_0 in $\mathcal{P}$, there exists a unique b-trajectory of length n and starting from H_0.*
2. A finite word on $\mathcal{A}$ is in $\mathcal{F}_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ if and only if it encodes some b-trajectory in $\mathcal{P}$.
3. In the totally irrational case (2) and for any S, a finite word on $\mathcal{A}$ is a factor of $c_{\alpha_1,\alpha_2,\ldots,\alpha_k,S}$ if and only if it encodes some b-trajectory in $\mathcal{H}$.

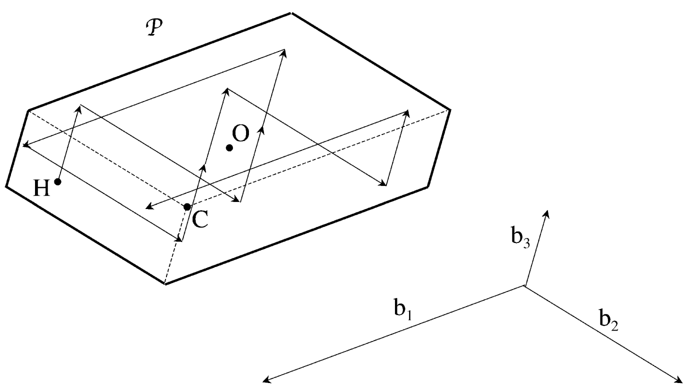

Fig. 2.

In the Figure 2 above, in dimension 3, $\mathcal{P}$ is an hexagon, and the b-trajectory of length 11 starting from H is represented.

The exceptional points in the first part of the theorem above can be characterized. They correspond to the ambiguous cases, i.e., to half-lines $\mathcal{D}$ parallel to

$(\alpha_1, \alpha_2, \ldots, \alpha_k)$ which contain some 2-integer point. By iterating the Part 1. of the theorem, there exists a unique infinite b-trajectory starting from a given non-exceptional point in $\mathcal{P}$. It can be easily proved that the orthogonal projections of the centers of the unit k-cubes crossed by $\mathcal{D}$ remain in the translated polyedron $\mathcal{P}_S := \mathcal{P} + \sum_{j=1}^{k} s_j b_j$ whose center is $\tilde{S}$. Thus the billiard word $c_{\alpha,S}$ encodes the infinite b-trajectory starting from the point $C - \tilde{S}$, i.e., $\sum_{j=1}^{k} (\frac{1}{2} - s_j) b_j$.

Moreover, in the totally irrational case (2), the set of the orthogonal projections of the centers of the unit k-cubes crossed by $\mathcal{D}$ is dense in $\mathcal{P}_S$. This proves the Part 3. of the theorem.

It is possible to give a characterization of those points H_0 which are starting points of a b-trajectory, for a given finite word v on $\mathcal{A}$.

Proposition 1 *For any finite word v on the alphabet $\mathcal{A}$, the set $\mathcal{P}_v$ of the starting points of the b-trajectories with coding word v is an open polyedral convex set of dimension $k-1$. The diameter of this set tends to zero as the length of v tends to infinity.*

For many v the set $\mathcal{P}_v$ is an empty set, the non-empty sets correspond to those words v of the language $\mathcal{F}_\alpha$. In [2] and [3], the study of the structure of these sets allows to obtain the complexity of the language of the billiard words in the totally irrational case, for 3-dimensional words and in any dimension respectively.

2.3 Application to Palindromic Factors

We say that a factor w of a finite word v is a *central factor* when we have $v = v_1 w v_2$ with $|v_1| = |v_2|$, and we call *center of* v the central factor of v of length 1 or 2, as $|v|$ is odd or even.

Proposition 2
- *For each letter a_j in $\mathcal{A}$, the set $\mathcal{P}_{a_j}$ contains the point $-\frac{1}{2} b_j$, thus it is non-empty.*
- *The only non-empty set $\mathcal{P}_{a_j a_j}$ corresponds to the letter a_{j_0} coding the unique b-trajectory of length 1 starting at the origin.*

Let $\mathcal{H}$ be the closure of the set of points in the infinite b-trajectory corresponding to the billiard word $c_{\alpha,S}$. With the total irrationality hypothesis (2), $\mathcal{H}$ is the closure of $\mathcal{P}$, and we obtain that any billiard word has infinitely many palindromic factors, and more precisely:

- for any even integer n, a unique palindromic factor of length n, whose center is the only pair of letters $a_{j_0} a_{j_0}$ which belongs to the language $\mathcal{F}_{\alpha_1, \alpha_2, \ldots, \alpha_k}$;
- for any odd integer n, and for each letter a_j in the alphabet $\mathcal{A}$, a unique palindromic factor of length n in which the letter a_j is in central position.

The general situation (1) is more complicate.

Theorem 2.2. – *The billiard word $c_{\alpha,S}$ contains arbitrary long palindromic factors of even length if and only if O is in $\mathcal{H}$.*
– *The billiard word $c_{\alpha,S}$ contains arbitrary long palindromic factors of odd length and of center a_j if and only if $\frac{1}{2}b_j$ is in $\mathcal{H}$.*

As an example in dimension $k = 3$, we take for α the vector $(2, \sqrt{5}, 1 + \sqrt{5})$ and consider the billiard word c_α starting at the origin. Then $\frac{1}{2}b_1$ is in $\mathcal{H}$, but O, $\frac{1}{2}b_2$ and $\frac{1}{2}b_3$ are not in $\mathcal{H}$. Moreover the billiard word does not contain any palindromic factor of even length: in this case, $j_0 = 3$, but $\mathcal{H}$ does not intersect $\mathcal{P}_{a_3a_3}$. So the word a_3a_3 is not a factor of this billiard word, the number of factors of length 2 is equal to 6 instead of 7 in the general case.

We prove that there always exist arbitrary long palindromic factors in any billiard word in dimension 3. This result is false in higher dimension, and it is possible to find a billiard word with 4 letters, with a finite number of palindromic factors.

3 Palindromic Prefix Factors of Billiard Words

In this section we consider only billiard words c_α starting from the origin.

3.1 Dimension 2

The question of palindromic prefix factors of billiard words has been studied in dimension 2, see for example [9], [10], [11], [12] and [16].

It can be easily shown that the palindromic prefix factors of billiard words are finite billiard words, and correspond to the continued fraction expansion (see [8] or [14]) of the slope $\rho := \frac{\alpha_2}{\alpha_1}$ of the half-line $\mathcal{D}$.

Theorem 3.1. *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 $\frac{p_n}{q_n}$ of the continued fraction expansion of the real number ρ.*

This result is stated in [4], [9], [10], in a slightly different formulation. A purely geometrical proof of this result can also be given, by using a geometrical approach of the theory of continued fraction expansions, made by H.J.S. Smith and Felix Klein, [15], at the end of the XIXth century, see for example [8].

3.2 Higher Dimensions

3.2.1 Main Result

Now the dimension k is greater than 2. In the general case, the billiards words have only finitely many palindrome prefix factors. Note that these palindromic prefixes are finite billiard words as in dimension 2.

Theorem 3.2. *With Hypothesis (1), in dimension $k \geq 3$, the set of vectors $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ such that $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ has infinitely many palindromic prefixes if a negligible set, with respect to the Lebesgue measure on the k-dimensional unit sphere. However, this set is dense on the positive part of this unit sphere.*

As an example in dimension 3, if we choose $(\alpha_1, \alpha_2, \alpha_3) = (1, \frac{15+\sqrt{5}}{10}, \frac{1+\sqrt{5}}{2})$, the corresponding billiard word $c_{\alpha_1,\alpha_2,\alpha_3}$ is: $a_2a_3a_1a_2a_3a_2a_3a_1a_2a_3a_2a_1a_3a_2a_3a_1a_2a_3a_2a_3a_1a_2a_3a_2a_1a_3a_2a_3a_2a_1a_3a_2a_1a_3a_2a_3a_2a_1a_3a_2a_2a_3\ldots$ and a_2 is the only one palindromic prefix factor of $c_{\alpha_1,\alpha_2,\alpha_3}$.

3.2.2 Integer Prefix Points

We use the notion of integer prefix point: the integer point M is called an *integer prefix point* of $\mathcal{D}$ when the triangle OHM does not contain any 2-integer points, except for O and M, where H is the orthogonal projection of M onto the line $\mathcal{D}$. These points correspond to the palindromic prefix factors of the billiard words.

Proposition 3 *Let $M = (m_1, m_2, \ldots, m_k)$ be any integer point and $\mathcal{D}$ any half-line parallel to vector $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ satisfying the irrationality condition (1). The following properties are the same:*

- *M is an integer prefix point of $\mathcal{D}$;*
- *c_M is a palindromic prefix factor of $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$.*

3.2.3 Up and Down Method

We can use the *projections* to study these points. For $1 \leq i < j \leq k$, the projections π_{ij} are defined by:

- on finite and infinite words, by $u_{ij} := \pi_{ij}(u)$ is the word on $\mathcal{A}_{ij} := \{a_i, a_j\}$ obtained by erasing in u any letters excepted a_i and a_j;
- on $\mathbb{R}_+^k$, by $\pi_{ij}(x_1, x_2, \ldots, x_k) := (x_i, x_j)$, which belongs to the plane Π_{ij};
- on the set of vectors $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ in $\mathbb{R}_+^k$, by $\pi_{ij}(\alpha_1, \alpha_2, \ldots, \alpha_k) := (\alpha_i, \alpha_j)$.

It is very simple to prove that M is an integer prefix point of $\mathcal{D}$ if and only if all its projections $M_{ij} := \pi_{ij}(M)$ are integer prefix points of $\mathcal{D}_{ij} := \pi_{ij}(\mathcal{D})$, thus correspond to some convergent of the ratio $\frac{\alpha_j}{\alpha_i}$. Hence the first coordinate m_1 of M is the denominator of some convergent for any irrational numbers $\frac{\alpha_j}{\alpha_1}$, $2 \leq j \leq k$. This is a *synchronization* property of convergents.

A classical theorem due to Lagrange, see [14], implies that M is an integer prefix point of $\mathcal{D}$ when M is very close to $\mathcal{D}$. Using this property, we can prove in some special cases that M is an integer prefix point of $\mathcal{D}$ when all the M_{1j}, $2 \leq j \leq k$, are integer prefix points of the half-lines $\mathcal{D}_{1j}$.

3.2.4 A Probabilistic Result on Continued Fraction Expansions

The following result shows that the synchronization property of convergents cannot appear infinitely many, for almost all vectors $(\alpha_1, \alpha_2, \ldots, \alpha_k)$.

Proposition 3.1. *For almost all positive real numbers 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*.*

This proposition can be proved by using some classical probabilistic results on continued fraction expansion, see [18]. Then we get an upper bound of the probability that a given integer q is the denominator of some convergent, see the proposition below, and use the well-known Borel-Cantelli lemma.

Proposition 3.2. *Let q be a positive integer ≥ 2, and $0 < x < 1$. Then the probability P_q that q be a denominator of a main or intermediate convergent of x satisfies:*

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

3.2.5 The Exceptional Case

There are two different proofs for the existence of billiard words with infinitely many palindromic prefix factors. The first one consists of an iterative construction of the numerators and denominators of the convergents of the ratios $\frac{\alpha_j}{\alpha_1}$, $2 \leq j \leq k$, such that we can use the Lagrange theorem. The second one is a purely geometrical proof.

Both proofs give the density of these words, and the first one allows to choose billiard words respecting the total irrationality property (2).

References

1. J.-P. Allouche, J. Shallit, *Automatic sequences: Theory and Applications*, 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. Séébold, Sturmian words, in M. Lothaire, *Algebraic combinatorics on words*, Cambridge University Press, 2002.
6. J.-P. Borel, F. Laubie, Quelques mots sur la droite projective réelle, *J. Théorie des Nombres de Bordeaux*, **5**, 1993, 23-51.
7. E. B. Christoffel, Observatio arithmetica, *Annali di Matematica*, **6**, 1875, 148-152.
8. H. Davenport, *The Higher Arithmetic: An Introduction to the Theory of Numbers, 7th ed.*, Cambridge Univ. Press, 1999.
9. A. de Luca, Sturmian words: structure, combinatorics, and their arithmetics, *Theoret. Comput. Sci.*, **183**, 1997, 45-82.
10. A. de Luca, Combinatorics of standard sturmian words, *Structures in Logic and Computer Science, Lecture Notes Comput. Sci.*, **1261**, 1997, 249-267.
11. X. Droubay, Palindromes in the Fibonacci word, *Inf. Proc. Letters*, **55**, 1995, 217-221.

12. X. Droubay, G. Pirillo, Palindromes and Sturmian words, *Theoret. Comput. Sci.*, **223**, 1999, 73-85.
13. R.L. Graham, D.E. Knuth, O. Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, MA, 2nd ed., 1994.
14. G.H. Hardy, E.M. Wright, *An Introduction to the Theory of Numbers*, Oxford at the Clarendon Press, 5th ed., 1979.
15. F. Klein, *Ausgewählte Kapitel der Zahlentheorie*, Teubner, Liepzig, 1907.
16. G. Pirillo, A new characteristic property of the palindrome prefixes of a standard Sturmian word, *Sém. Lotharingien Combinatoire*, **43**, 1999, (electronic, see http://www.mat.univie.ac.at/~slc/).
17. N. Pytheas Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*, Lecture Notes in Math. **1794**, V. Berthé and al. ed., 2002.
18. A. M. Rockett, P. Szüsz, *Continued Fractions*, World Scientific Pub. Co, 1992.