Discrete Mathematics and Theoretical Computer Science DMTCS vol. **9:2**, 2007, 195–212

A geometrical approach to Palindromic Factors of Standard Billiard Words

Jean-Pierre Borel[1†]

[1] *XLim, UMR 6090 Université de Limoges - CNRS - 123 avenue Albert Thomas F-87060 LIMOGES CEDEX - FRANCE*

received 13 Jan 2005, revised 20 Jun 2005, accepted 30 Jun 2005.

Many results are already known, concerning the palindromic factors and the palindomic prefixes of Standard billiard words, i.e., Sturmian words and billiard words in any dimension, starting at the origin. We give new geometrical proofs of these results, especially for the existence in any dimension of Standard billiard words with arbitrary long palindromic prefixes.

Keywords: Words, Languages, Sturmian, Billiard, Palindromes.

1 Introduction

1.1 Standard Billiard Words in dimension $k = 2$

Let ρ be any positive number. There are many ways to associate to ρ a family of Sturmian words, which are special words on the two letters alphabet $\mathcal{A} := \{a_1, a_2\}$. Here we choose the geometrical one.

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 half-line $\mathcal{D}$ starting at the origin O with slope ρ divides $\mathcal{G}$ into two parts. It will be more convenient for a multidimensional generalization, to consider a positive vector $\alpha = (\alpha_1, \alpha_2)$ parallel to $\mathcal{D}$, thus we have $\rho = \frac{\alpha_2}{\alpha_1}$. We construct the *Christoffel word* $u_\rho = u_{\alpha_1,\alpha_2}$ and the *Standard billiard word* $c_\rho = c_{\alpha_1,\alpha_2}$ as shown in Fig.1.a:
Coding by a_1 the horizontal black segment on the grid and by a_2 the vertical one as in Fig.1.a ,u_{α_1,α_2} encodes the discrete path immediately under the half line $\mathcal{D}$; hence we have $u_{\alpha_1,\alpha_2} = a_1a_2a_1a_2a_1a_2a_2a_1a_2\ldots$ Looking at the squares crossed by $\mathcal{D}$, i.e., the grey squares in Fig.1.a, the Standard billiard word c_{α_1,α_2} encodes the sequence of horizontal and vertical segments joining the two centers (i.e., white points in Fig. 1.a) of the consecutive grey unit squares. Consider the

[†] research partially supported by Région Limousin.

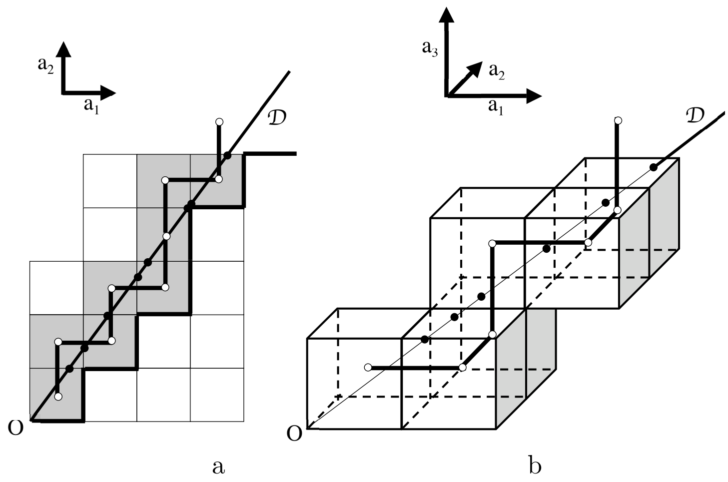

Fig. 1: Two and three-letters Billiard Words

first path. Cancelling the first horizontal segment, and then translating by vector $(-\frac{1}{2}, \frac{1}{2})$, we obtain the new path, hence we have $u_{\alpha_1,\alpha_2} = a_1 c_{\alpha_1,\alpha_2}$.

Consider now a mobile point P, starting from the origin O and moving on $\mathcal{D}$ up to infinity. This point intersects successively horizontal and vertical parts of the grid, the corresponding position are the black points in Fig.1.a. We encode now by a_1 an intersection point (black point) on a vertical line of the grid, and by a_2 an intersection point on an horizontal line. As P goes from O up to infinity, we get an infinite word on $\mathcal{A}$ which is the Standard billiard word c_{α_1,α_2}. So these words are also called *cutting sequences* or *billiard words* starting at the origin, and denoted by $c_{\ldots}$. These two geometrical constructions of Standard billiard words are useful in any dimension. In this paper, we shall use the second one.

General Sturmian words can be obtained in the same way, using arbitrary lines $\mathcal{D}$ of positive slope ρ. The set of factors of these words depends only on the slope ρ, so that factors of Sturmian words are factors of the Christoffel word. In this paper, we consider only *Standard* words, i.e., those corrresponding to the half-lines starting from he origin.

Christoffel words have been known since Bernoulli. They have been introduced in [8], and have many applications in mathematics and physics : they are related to continuous fraction expansions, Farey sequences, and the Stern-Brocot tree, see [6] and [14] for the later. General presentations of Christoffel and Sturmian words can be found in [1], [5] or [17].

1.2 Standard Billiard Words in dimension $k \geq 3$

Let $\mathcal{D}$ be the half-line of origin O in the k-dimensional space, and parallel to vector $(\alpha_1, \alpha_2, \ldots, \alpha_k)$, with positive α_i. We consider the sequence of k-dimensional unit cubes, or "k-cubes", crossed by $\mathcal{D}$, and the "k-facets" joining each k-cube to the next : a k-facet is the subset of the k-cube formed by all the points having a fixed integer i-th coordinate, there are 2 k-facets for each i, and so each k-cube has $2k$ k-facets.

In Fig.1.b, in dimension $k = 3$, the half-line $\mathcal{D}$ crosses the six 3-cubes which are represented in the figure.

Looking again to the sequence of intersection points (the black points in Fig.1.b), and coding these points by a_i when the i-th coordinate is an integer for the corresponding k-facet, we obtain an infinite word on the alphabet $\mathcal{A}_k := \{a_1, a_2, \ldots, a_k\}$. This word is the Standard billiard word associated to $\mathcal{D}$. This word also encodes, as in dimension 2, the sequence of segments between the centers of consecutive crossed k-cubes (the white points in Fig.1.b).

This Standard billiard word is denoted by $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$. The example given in Fig.1.b corresponds to $c_{\alpha_1,\alpha_2,\alpha_3} = a_1a_2a_3a_1a_2a_3\ldots$

This works as soon as one has:

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

Indeed in this case, each k-facet is crossed in its interior, so that the corresponding intersection point belongs to a unique k-facet, so that the index i is unique and well-defined. This condition already holds if we have:

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

and these two conditions are the same only in dimension 2. In many cases in the following, we use the stronger condition (2), and we say that $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ is *totally irrational*, see [2].

1.3 Classical Results

Standard billiard words and Sturmian words have been intensively studied in any dimension. These studies specially consider the language of these words, i.e., the set of factors, and the complexity of these words, i.e., the numbers of factors of length n, for any n.

In dimension $k = 2$ there are exactly $n + 1$ factors of length n, and this can be used to characterize Sturmian words; moreover, they appear as the words of minimal complexity, except for ultimate periodic words. In dimension $k = 3$, there are $n^2 + n + 1$ factors of length n, see [2], and

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

in any dimension k, see [3]. Hypothesis (2) is a necessary condition for these results: it is possible to get some Standard billiard words having exactly 6 factors of length 2, in dimension 3, using only hypothesis (1) (unpublished).

2 Palindromic Factors and Palindromic Prefixes: main Results

Looking for palindromic factors of Standard billiard words, the situation is the same in any dimension. But the situation is quite different if we only consider palindromic prefixes.

Theorem 1 *With the strong irrationality hypothesis (2), the Standard billiard word $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$ contains:*

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

This Theorem can be found in [7] in the general case. Proof is an easy generalisation of the proof in dimension 2. We say that a factor w of a finite word u is a *central factor* when we have $v = v_1wv_2$ 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. This result implies that whenever we have two palindromic factors of a given Standard billiard word, and with the same center, then the shorter one is a central factor of the longuer one.

In dimension 2, any palindromic factor of a Standard billiard word is a central factor of a palindromic prefix of this Standard billiard word. This result is proved in [10], and a geometrical proof is given in [7]. This result cannot be extended to dimension $k \geq 3$, according to the finite number of palindromic prefixes in general case, see the Theorem below.

It is also possible to obtain a characterization of Sturmian words in dimension 2, using properties on the first occurence of each palindromic factor in the word, see [13]. Note that condition (2) is necessary to get these results: if we suppose only (1), it is possible to obtain some Standard billiard words in dimension $k = 3$ without any palindromic factors of even length, and in dimension $k = 4$ some Standard billiard words having only finitely many palindromic factors (unpublished).

Theorem 2 *With the strong irrationality hypothesis (2),*

- *in dimension $k = 2$, each Standard billiard word has infinitely many palindromic prefixes; the length of these palindromic prefixes can be deduced from the sequence of main and intermediate convergents, using the continued fraction expansion of the slope ρ of the line $\mathcal{D}$;*
- *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 is a negligible set, looking to the Lebesgue measure on the k-dimensional unit sphere. However, this set is dense on the positive part of this unit sphere.*

In dimension 2, many proofs of this Theorem have been given, for example in [4], [10], [11], [12], [16], with various approaches. In the multidimensional case $k \geq 3$, it has been proved in [7]. Essentially, the problem is to "synchronize" (see [7]) the convergents of the continued fraction expansion of the ratios $\frac{\alpha_i}{\alpha_j}$. The existence of such a synchronization is related to probabilistic considerations on these convergents, and is an exceptional case. The density follows from an explicit construction of "good" vectors $(\alpha_1, \alpha_2, \ldots, \alpha_k)$.

In this paper, we show that we can obtain these results by using only the geometrical approach of Standard billiard words : the two results above are proved in dimension 2, and the existence and the density property in dimension $k \geq 3$, for Standard billiard words with infinitely many palindromic prefixes.

3 A geometrical Characterization of Palindromic Prefixes

3.1 Some kinds of Integer Points

Definition 1 *Let $M = (x_1, x_2, \ldots, x_k) \in R_+^k$ 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 in some cases M;*
- *an integer point whenever all its coordinates are positive integers;*
- *a visible integer point when it is both visible and integer.*

The condition (1) of irrationnality, which allows to define the Standard billiard word $c_{\alpha_1, \alpha_2, \ldots, \alpha_k}$, corresponds to the non-existence of 2-integer points on the half-line $\mathcal{D}$, except for the origin O.

A visible integer point M is a point whose coordinates m_j are pairwise coprimes positive integers, and this property corresponds to the following geometrical property: the segment OM crosses the k-cubes only in the interior of k-facets. As before, we can encode the finite sequence of these intersection on the alphabet $\mathcal{A}$. The corresponding word, of length $\sum_{j=1}^{k}(m_j - 1)$, is called a *finite Standard billiard word* and denoted by c_M. It has been pointed by Jean Berstel, in the beginning of the nineties, that:

Proposition 1 *For any visible integer point M, the finite Standard billiard word c_M is a palindromic word.*

Proof The geometrical proof is immediate. Consider the mid-point I of the segment OM, and all the intersection points between OM and the k-facets. I is a center of symmetry of the figure (see Fig.2). The two segments OM and MO have the same encoded sequence of intersection points with the grid, so the reverse word $\widetilde{c_M}$ and c_M are the same, and this word is a palindrome.
□

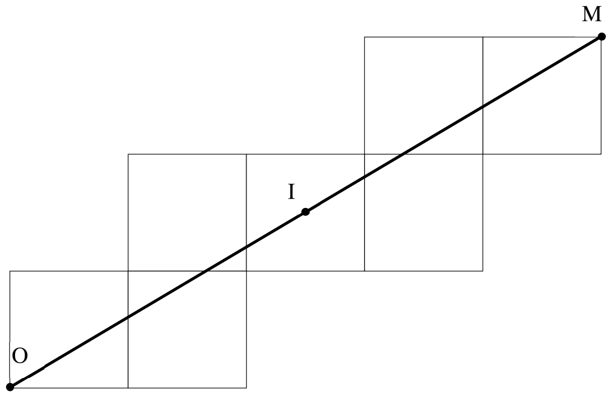

Fig. 2: Symmetry and Palindromicity

The following definition has been introduced in [7], and the corresponding points play a crucial role in the sequel.

Definition 2 *Let us consider a vector $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ satisfying condition (1), an integer point M, and the point H which is the second intersection point between $\mathcal{D}$ and the rectangle parallelepided whose edges are parallel to the axes, and whose OM is the main diagonal. 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.*

It can be noticed that, in dimension 2, we can replace in the definition above the point H either by the orthogonal projection H' of M on $\mathcal{D}$, or by the intersection point H'' between $\mathcal{D}$ and the sphere of center O and radius OM: the segment HH'' contains H', and it remains in the unit cube whose M is a vertex (see Fig.3).

3.2 Palindromic Prefixes and Integer Prefix Points

The two following results give the relation between the palindromic prefix factors of Standard billiard words and the integer prefix points of the corresponding half-line.

Proposition 2 *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 $c_{\alpha_1, \alpha_2, \ldots, \alpha_k}$.*

This result, which is proved in [7], has an immediate geometric proof: let M be any integer prefix point of $\mathcal{D}$. Consider H as above, and a point N moving continuously on the segment MH from M to H. Then the segment ON remains in the triangle OMH, so that OH and OM have the same encoding sequence of intersection points with the interiors of k-facets, and the palindromic word c_M encodes OH, hence it is a prefix factor of $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$.

The converse is a special case of the following.

Proposition 3 *Let $\mathcal{D}$ be any half-line parallel to vector $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ satisfying the irrationality condition (1), and u a palindromic prefix factor of the Standard billiard word $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$. Then the integer point M whose coordinates are the positive integers $m_j := |u|_{a_j} + 1$ is an integer prefix point of $\mathcal{D}$.*

We denote as usual by $|u|_a$ the number of occurences of the letter a in the finite word u, for any $a \in \mathcal{A}$.

Observe that, using the proposition above, this implies that $u = c_M$: indeed, u and c_M are two prefixes of the same Standard billiard word, with the same length.

Proof We prove this result at first in the two dimensional case, and then use the *projections* π_{ij} (see [7]), defined as follow for any $1 \leq i < j \leq k$:

- on finite and infinite words, by $u_{ij} := \pi_{ij}(u)$ is the word on $\mathcal{A}_{ij} := \{a_i, a_j\}$ obtained by cancelling in u any letter except a_i and a_j;
- on $R^{\,k}_{\,+}$, 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 $R^{\,k}_{\,+}$, by $\pi_{ij}(\alpha_1, \alpha_2, \ldots, \alpha_k) := (\alpha_i, \alpha_j)$.

3.2.1 The case of Dimension $k = 2$

Let u be a palindromic prefix factor of the Standard billiard word c_{α_1,α_2}, and M the integer point whose coordinates are $(|u|_{a_1} + 1, |u|_{a_2} + 1)$. u is a prefix factor of the Standard billiard word, hence $\mathcal{D}$ crosses the grey square on Fig.3, whose M is the right upper vertex. Fig.3 corresponds to the case $c_{\alpha_1,\alpha_2} = a_1a_2a_1a_1a_2a_1a_2\ldots$ and $u = a_1a_2a_1a_1a_2a_1$. The letter a_2 following u in c_{α_1,α_2} shows that $\mathcal{D}$ crosses the horizontal upper side of the grey square. We use the mid-point I of the segment OM as a center of symmetry for the figure. Consider the symetric point G of H with respect to I : this point is on the x-axis, on the first unit segment. The open segment OH crosses the grid $m_1 + m_2 - 2$ times, i.e., $|u|_{a_1} + |u|_{a_2}$ times. Hence the finite word v on $\mathcal{A}$ coding these intersections is a prefix factor of c_{α_1,α_2}, of length $|u|$: so $v = u$.

Using the symmetry, the finite word coding the intersections of the open segment GM with the grid is the reverse word $\widetilde{u}$ of u.

We made the hypothesis that u is a palindromic word, and so we have $v = u = \widetilde{u}$, and the two segments OH and GM have the same intersection coding sequence with the grid. We move continuously from OH to GM by a horizontal translation. Then, the moving segment cannot cross any integer point: if it were, the coding word of the segment is modified, and some factor a_2a_1 is replaced by a_1a_2, so that the coding word decreases strictly, corresponding to the lexicographic order. But the word is the same at the beginning and at the end of this translation

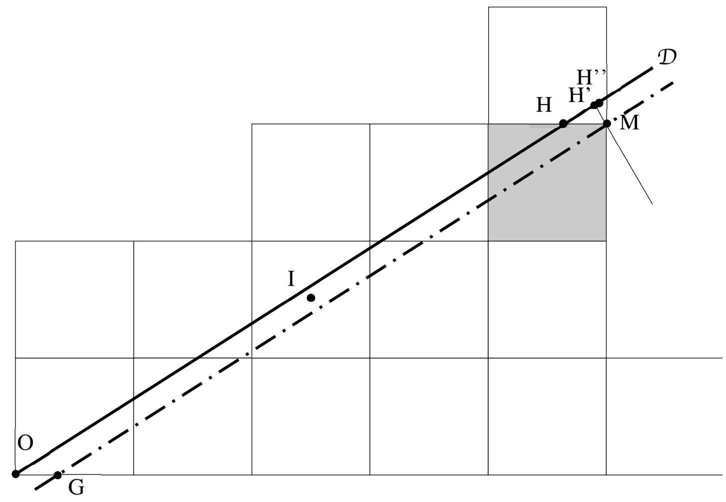

Fig. 3: M is an integer prefix point of $\mathcal{D}$

movement. Hence we obtain that there are no integer point in the parallelogram $OHMG$, except for O and M. Thus there are no integer point in the triangle OHM, except for O and M, and this proves that M is an integer prefix point of $\mathcal{D}$.

3.2.2 The case of Higher Dimensions

Consider now $k \geq 3$. Let u be a palindromic prefix factor of the Standard billiard word $c_{\alpha_1,\alpha_2,\ldots,\alpha_k}$, and M be the integer point whose coordinates are $m_j := |u|_{a_j} + 1$, $1 \leq j \leq k$. We must prove that there are no 2-integer points in the triangle OHM, except for O and M.

Suppose that such a 2-integer point exists, namely P, and let $1 \leq i < j \leq k$ such that the two coordinates p_i and p_j of P are integers. We use the corresponding projection π_{ij}. Then :

- u_{ij} is a palindromic prefix factor of c_{α_i,α_j};
- $\pi_{ij}(P)$ is the integer point P_{ij} of the plane Π_{ij} whose coordinates are (p_i, p_j);
- $\pi_{ij}(M)$ is the integer point M_{ij} of the plane Π_{ij} whose coordinates are $(|u|_{a_i} + 1, |u|_{a_j} + 1)$.

Observe that $|u|_{a_i} = |u_{ij}|_{a_i}$ and $|u|_{a_j} = |u_{ij}|_{a_j}$. So we can use the result above in dimension 2, M_{ij} is an integer prefix point of $\mathcal{D}_{ij} := \pi_{ij}(\mathcal{D})$, and the triangle $OH_{ij}M_{ij}$ does not contain any integer point, except for O and M_{ij}. This contradicts the fact that P_{ij} is both an integer point and in the interior of the triangle $OH_{ij}M_{ij}$, by projection : P is an interior point of the triangle OHM, and the projection $\pi_{ij}(OHM) = OKM_{ij}$ of this triangle is included in the triangle $OH_{ij}M_{ij}$, the point K being on the segment OH_{ij}. $\square$

3.2.3 Synthesis

Combining the two propositions above we get the following geometric characterization of the palindromic prefix factors. This characterization will be used in the sequel.

Proposition 4 *The palindromic prefix factors of the Standard billiard words are related to the integer prefix points M of the half-line $\mathcal{D}$, and are exactly the finite Standard billiard words c_M.*

4 Palindromic Factors in dimension $k = 2$

4.1 Palindromic Prefixes

As we have said, the palindromic prefix factors correspond to those points M such that there are no integer points in the triangle OMH, except for O and M. These points M of coordinates (m_1, m_2) exactly correspond to the main and intermediate convergents $\frac{m_2}{m_1}$ of the continued fraction expansion of $\rho := \frac{\alpha_2}{\alpha_1}$. This is an easy consequence of a geometrical presentation of the continued fractions developement theory, made by H.J.S.Smith and Felix Klein, [15], at the end of the XIXth century, see for example [9], and can be resumed as follow.

Theorem 3 *Let $\mathcal{N}_-$ and $\mathcal{N}_+$ be the sets of all integer points strictly below (resp. above) the half-line $\mathcal{D}$ of irrational slope $\rho = \frac{\alpha_2}{\alpha_1}$. Let $\mathcal{C}_-$ and $\mathcal{C}_+$ be the convex hulls of these sets. Then:*

- *The boundaries of $\mathcal{C}_-$ and $\mathcal{C}_+$ are two infinite polygonal lines. For $\mathcal{C}_-$ (resp. $\mathcal{C}_+$), this line contains the positive part of the x-axis (resp. y-axis) starting from I (resp. from J), and an infinite sequence of finite segments, whose slopes are decreasing (resp. increasing) and are equal to the main convergents of ρ, bigger (resp. smaller) than ρ.*
- *The vertices $M = (m_1, m_2)$ of these polygonal lines are the integer points such that $\frac{m_2}{m_1}$ is a main convergent of the continued fraction expansion of ρ;*
- *The other integer points $M = (m_1, m_2)$ on the polygonal line are the integer points such that $\frac{m_2}{m_1}$ is an intermediate convergent of the continued fraction expansion of ρ;*

These boundaries are called the Newton polygonal line associated to the half-line $\mathcal{D}$. They are very close to $\mathcal{D}$, as seen in Fig.4, and closer when the partial quotients a_n of the slope are large integers. In Fig.4 above, the vertices of the Newton polygonal line are the black points, and the other integer points on the line are the white ones, except for those on the axis. It corresponds to the continued fraction expansion:

$$\rho = [0; 1, 2, 1, 2, \ldots] = \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{\cdots}}}}}$$

which has $\frac{0}{1}$, $\frac{1}{1}$, $\frac{2}{3}$, $\frac{3}{4}$, $\frac{5}{7}$, $\frac{8}{11}$, ..., as main convergents, and $\frac{1}{2}$ and $\frac{5}{7}$ as intermediate convergents in the corresponding area.

It is an easy consequence of this construction that any integer point M on the Newton polygonal line is an integer prefix point of $\mathcal{D}$: the corresponding triangle OMH is in the intermediate

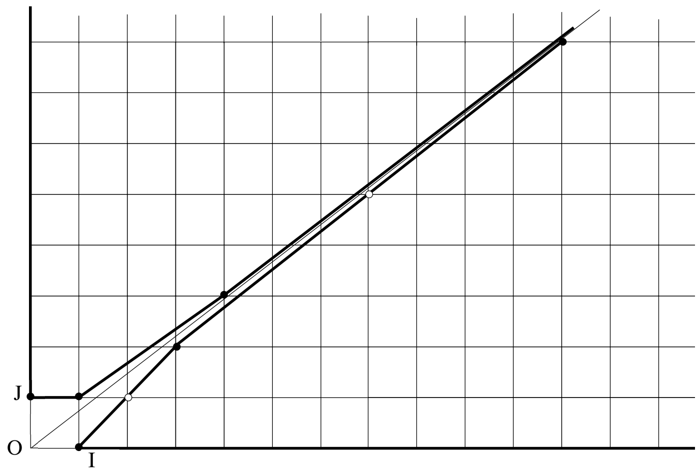

Fig. 4: The two Newton polygonal lines.

area between the two polygonal lines, so contains no integer points, except for O and M.

The converse property is true:

Proposition 5 *Let M be any integer point, not on the Newton polygonal line. Then the interior of the triangle OMH contains some integer point, so that M cannot be an integer prefix point.*

Proof Let M is any integer prefix point of $\mathcal{D}$ below the half-line (note that the other case can be treated in a similar way), and suppose that M is not on the Newton polygonal line. Then there exists an edge AB of the Newton line such that the two segments OM and AB intersects in their interior. There are no integer points, except for O, A or B, in the two following triangles:

1. the triangle $OH'M$: M is an integer prefix point, so that the rectangle $OH'MG'$ does not contain any integer point, except for O and M, as noticed before;
2. the triangle $OA'B'$, except for O and integer points on the segment AB, where A' and B' are the intercepts of the line (AB) with the x-axis and the half-line $\mathcal{D}$ respectively (the triangle is as in Fig.5, since the slope of AB is bigger than ρ). Indeed, AB is an edge of the Newton line, so all the integer points below $\mathcal{D}$ are under the line (AB), or on the segment AB.

From 1. we deduce that the point M is in the interior of the triangle BJA, where AJ is parallel to $\mathcal{D}$ and BJ is perpendicular to $\mathcal{D}$.

From 2. we deduce that there are no integer point in the triangle OBA, except for O and the integer points on AB, and the parallelogram $OCBA$ has no integer point in its interior, by symmetry with respect to the mid-point I of OB : the coordinates of I take their values in $\frac{1}{2}\mathbb{Z}$, so that the symmetric point of an integer point is another integer point.

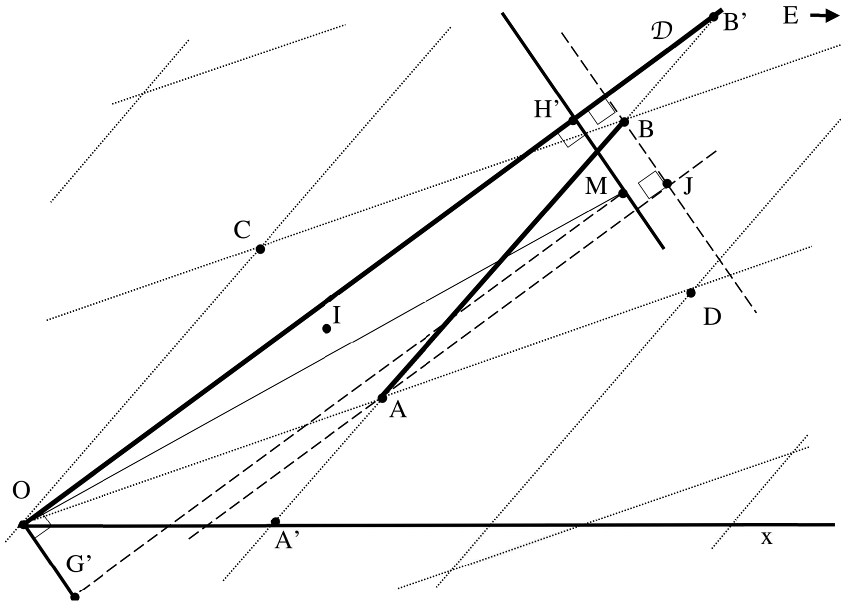

Fig. 5: Integer points and lattices.

This property implies that all the integer points in the plane are on the new grid built on this parallelogram (the pointed lines in Fig.5). The intersection points of this grid are integer points, and it is possible to have some other integer points on the pointed lines: this case corresponds to partial quotients $a_n > 1$, i.e., to the existence of other integer points on the segment AB than the endpoints A and B.

Let $ABED$ the fondamental domain of the grid obtained from $OCBA$ by translation of vector $\vec{OA}$ (for simplicity reasons, E is outside the figure). M is in the interior of the triangle AJB, and so is in the interior of the parallelogram $ABED$.

But there are no integer points in the interior of the grid, so we get a contradiction, and this proves the announced result: the palindromic prefix factors of the Standard billiard word correspond to the integer points on the Newton polygonal line associated with the half-line, i.e., to the continued fraction expansion of the slope ρ. □

4.2 Palindromic Factors

In dimension 2, any palindromic factor of a Standard billiard word is a central factor of some palindromic prefix factor. This result is proved in [10], and an easy geometrical proof is given in [7]. This property gives an other characterization of sturmian words (see [13]).

5 Palindromic Prefixes in dimension $k \geq 3$

In this section, we prove in a geometrical way that there exist, for any $k \geq 3$, some half-lines $\mathcal{D}$ satisfying the irrationnality condition (1), and having infinitely many integer prefix points. Moreover the proof implies the two following properties:

- the set of the half-lines having infinitely many integer prefix points is a dense subset of the positive part of the unit sphere in R^k;
- this set has the cardinality of the continuum.

In the following we consider some open cones $\mathcal{C}$ in the positive quadrant R^k_+ of the k-dimensional space, with vertex O. The proof is based on the two following principles.

Proposition 6 *Let M be any visible integer point. There exists an open cone $\mathcal{C}$ with vertex O, such that:*

- *the point M is in $\mathcal{C}$;*
- *the point M is an integer prefix point for any half-line $\mathcal{D} \in \mathcal{C}$ satisfying (1).*

Proof There are no 2-integer points on the open segment OM. Note that the set of 2-integer points is a closed subset of the positive k-dimensional space. So there exists an open cylinder (see Fig.6) whose central axis is the segment OM, and containing no 2-integer points, except for those close to the the origin and to M.

The condition (1) implies that there is no 2-integer point on the segment OH, thus the 2-integer points close to M are not inside the triangle OHM.

The other exceptional 2-integer points are close to the origin, they have two zero coordinates, so they cannot be in the cone $\mathcal{C}$ having the same base as the cylinder, see Fig.6, where the doted lines are inside both the cone and the cylinder.

Let $\mathcal{D}$ be any half-line in $\mathcal{C}$, the orthogonal projection H of M on $\mathcal{D}$ is contained in the open cylinder, and so is the triangle OHM, except for the two points O and M. Thus this triangle does not contain any 2-integer point, except for O and M, and this means that M is an integer prefix point of $\mathcal{D}$. $\square$

Proposition 7 *Let $\mathcal{C}$ be any open cone with vertex O. Then $\mathcal{C}$ contains at least one visible integer point M.*

By iteration we get immediately that $\mathcal{C}$ contains infinitely many visible integer points. The proof of this proposition needs some technical preliminary lemmas.

5.1 Non-lacunary Sets of Integers

We consider now infinite subsets, or sequences, N_j of the set of positive integers N^*, whose elements are enumerated in an increasing order by $n_{j,r}$, $r \in N^*$. We call such a sequence an

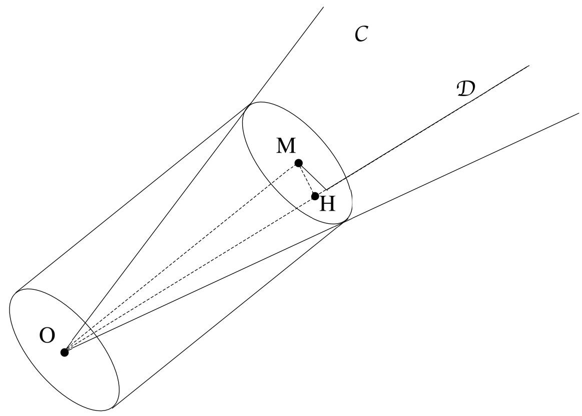

Fig. 6: Cylinder without 2-integer points and the associated cone

unlacunary sequence whenever the following property is true:

$$\lim_{r\to+\infty} \frac{n_{j,r+1}}{n_{j,r}} = 1 \tag{3}$$

Thus an unlacunary sequence is a sequence such that for any $\varepsilon > 0$, the interval $[x, (1+\varepsilon)x]$ contains at least one point in the sequence, for sufficient large x. This property is denoted by *unlac.*

Lemma 1 *The sequence* P *of prime numbers is unlacunary.*

This result is an easy consequence of the Prime Number Theorem, which is equivalent to $p_n \sim n \ln n$ if we denote by p_n the nth prime number in the increasing order. However the Prime Number Theorem is not easy to prove, so we shall not use this lemma.

Lemma 2 *Let* p *and* q *be two distinct prime numbers. Then the sequence* $\boldsymbol{N}_{p,q}$ *of those integers having only* p *and* q *as prime factors, is unlacunary.*

Proof We consider the increasing sequence of the logarithms of the numbers in $\boldsymbol{N}_{p,q}$, i.e., all real numbers $n \ln p + m \ln q$ with non negative integer coefficients n and m, except for $n = m = 0$. We have to prove that the difference between two consecutive numbers in this new sequence tends to 0, thus the same property for the increasing sequence of real numbers:

$$n + m\frac{\ln q}{\ln p}.$$

It is well known that $\frac{\ln q}{\ln p} =: \beta$ is an irrational number, so that the sequence of the fractional parts $\{m\beta\}$,with positive integers m, is dense in the unit interval $[0, 1]$.

Lemma 3 *Let* β *be any positive irrational number. Then the increasing sequence of all numbers* $n + m\beta$*,* $(n, m) \in \boldsymbol{N}^2$*, is such that the difference of two consecutive terms tends to* 0.

Proof It is sufficient to show that for any positive integer N, any interval of the form$[\frac{K}{N}, \frac{K+1}{N}]$ contains some number $n + m\beta$, for sufficiently large K : this property implies that the upper limit of the difference of two consecutive terms is bounded by $\frac{1}{N}$ for any positive integer N, so is 0.

Let N be any positive integer, and consider the N intervals $I_k := [\frac{k}{N}, \frac{k+1}{N}]$, $0 \leq k \leq N-1$. Each of them contains the fractional part $\{m_k\beta\}$, for some positive integer $m_k > 0$, using the density of this sequence. We can write $\{m_k\beta\} = m_k\beta - n_k$ with some positive integer $n_k > 0$.

Choose $K > N(1 + \max_{0 \leq k \leq N-1} n_k)$. Using the Euclidian division, $K = AN + r$, the remainder r satisfies $0 \leq r \leq N-1$ and the quotient A is such that $A > \max_{0 \leq k \leq N-1} n_k$. Then we have :

$$\frac{K}{N} = A + \frac{r}{N} < (A - n_r) + m_r\beta < A + \frac{r+1}{N} = \frac{K+1}{N}$$

which implies the expected result, since $A - n_r$ and m_r are positive integers. □

As said before, the ratio $\frac{\ln q}{\ln p}$ is an irrational real number, which proves lemma 2. □

Lemma 4 *Let $\boldsymbol{N}_j$, $1 \leq j \leq k$, be k unlacunary sequences, and $\mathcal{C}$ any open cone with vertex O. Then $\mathcal{C}$ contains infinitely many points of the direct product $\bigotimes_{j=1}^{k} \boldsymbol{N}_j$.*

Proof The open cone $\mathcal{C}$ contains some point $X = (x_1, x_2, \ldots, x_k)$ and some neighbourhood of X. Thus it contains some k-cube whose main diagonal is of the form XY with $Y = (1+\varepsilon)X := ((1+\varepsilon)x_1, (1+\varepsilon)x_2, \ldots, (1+\varepsilon)x_k)$ for some $\varepsilon > 0$. Then we use the *unlac* property for all the sequences $\boldsymbol{N}_j$: there exists some λ such that the interval $[\lambda x_j, (1+\varepsilon)\lambda x_j]$ contains some element n_j of $\boldsymbol{N}_j$ for all j (we consider only a finite number of unlacunary sequences). The new k-cube whose main diagonal is $X'Y'$, with $X' := \lambda X$ and $Y' := \lambda Y$, is the image of the former one by the homothety of center O and ratio λ. So this new cube is in the interior of $\mathcal{C}$. But the integer point $(n_1, n_2, \ldots, n_k)$ is inside the new k-cube, so is inside $\mathcal{C}$. □

Note that the proof can be made for arbitrary large values of λ, so that we get infinitely many points in the direct product $\bigotimes_{j=1}^{k} \boldsymbol{N}_j$, which are contained in $\mathcal{C}$.

5.2 Proof of Proposition 7

We apply lemma 4, using $\boldsymbol{N}_j := \boldsymbol{N}_{p_{2j-1}, p_{2j}}$ for $1 \leq j \leq k$, where p_j is the jth prime number in increasing order. Then the cone $\mathcal{C}$ contains some points of the direct product $\bigotimes_{j=1}^{k} \boldsymbol{N}_j$, and such a point is a visible point: any two of its coordinates are coprime. □

An other proof can be made, using $\boldsymbol{N}_j = \boldsymbol{P}$ for any j. We must consider some open cone $\mathcal{C}'$ in $\mathcal{C}$ such that any half-line $(\alpha_1, \alpha_2, \ldots, \alpha_k)$ in $\mathcal{C}'$ have distincts α_j's. Such a cone exists, and contains some point in $\boldsymbol{P}^{\otimes n}$, whose coordinates are distinct prime numbers. It is clear that this point is a visible point.

5.3 Existence of Lines with an Infinity of Integer Prefix Points

We can now give the geometrical proof of the main announced result, by combining propositions 6 and 7. We start from any visible point M_0, then there exists some open cone $\mathcal{C}_0$ such that M_0 is an integer prefix point of any half-line $\mathcal{D}$ in $\mathcal{C}_0$ and satisfying (1). Then we take a new visible

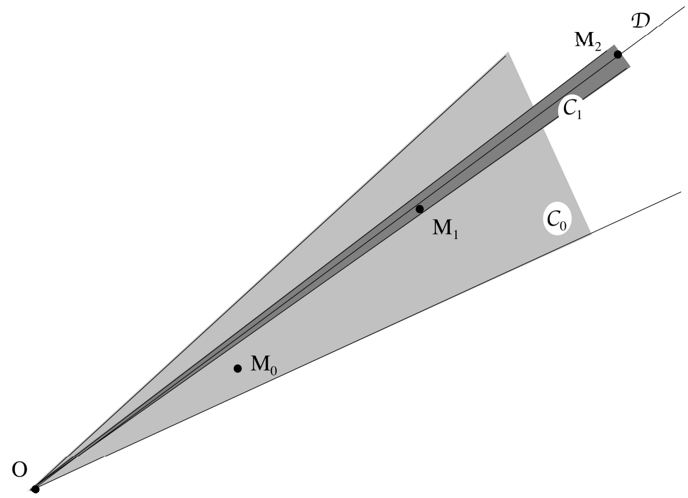

Fig. 7: Successive cones

point M_1 inside this cone, this new point is an integer prefix point of any half-line $\mathcal{D}$ satisfying (1) in a new cone $\mathcal{C}_1$, included in $\mathcal{C}_0$ (see Fig.7): using proposition 6 we get an open cone $\mathcal{C}'_1$, and take $\mathcal{C}_1 := \mathcal{C}_0 \cap \mathcal{C}'_1$. Then we iterate, taking inside $\mathcal{C}_1$ a new visible point M_2, and then the open cone $\mathcal{C}_2$, and so on.

The infinite sequence of the open cones $\mathcal{C}_n$ is a decreasing one, so there exists some half-line $\mathcal{D}$ in the limit set $\bigcap_{n=0}^{\infty} \mathcal{C}_n$ (more precisely, it can be proved that the limit set is exactly equal to some half-line starting from the origin), and all the integer points M_n are integer prefix points of $\mathcal{D}$. $\square$

More precisely, this method gives the existence of a large set of lines with infinitely many integer prefix points. We start from any open cone $\mathcal{C}_0$ as above. Then there exist two visible points M_1 and M'_1 in this cone. Then by iteration, we get two new open cones, and two new visible points in each cone, and so on. This proves that the set of those lines in $\mathcal{C}_0$ with infinitely many integer prefix points, has the cardinality of the continuum.

Acknowledgements

The author thanks the referees for their interesting suggestions.

References

[1] J.-P. Allouche, J. Shallit, *Automatic sequences: Theory and Applications*, Cambridge University Press, 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] J-P. Borel, C. Reutenauer, Palindromic factors of billiard words, *Theoret. Comput. Sci.*, **340-2** (2005), pp. 334-348.

[8] E. B. Christoffel, Observatio arithmetica, *Annali di Matematica*, **6** (1875), 148-152.

[9] H. Davenport, *The Higher Arithmetic : An Introduction to the Theory of Numbers, 7th ed.*, Cambridge Univ. Press, Cambridge, 1999.

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

[11] A. de Luca, Combinatorics of standard sturmian words. In: *Structures in Logic and Computer Science*, Lecture Notes Comput. Sci., **1261**, Springer-Verlag, 1997, 249-267.

[12] X. Droubay, Palindromes in the Fibonacci word, *Inf. Proc. Letters*, **55** (1995), 217-221.

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

[14] R.L. Graham, D.E. Knuth, O. Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, MA, 2nd ed., 1994.

[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. In: *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 *et al.* ed., 2002.