

Interactions between Dynamics, Arithmetics and Combinatorics: the Good, the Bad, and the Ugly

Valérie Berthé, Sébastien Ferenczi, and Luca Q. Zamboni

ABSTRACT. The aim of this paper is to survey possible generalizations of the well-known interaction between Sturmian sequences and rotations of \mathbb{T}^1 , using the Euclid algorithm of continued fraction approximation. We investigate mainly similar relations between Arnoux-Rauzy sequences and rotations of \mathbb{T}^2 , between codings of \mathbb{Z}^2 -actions on \mathbb{T}^1 and multidimensional Sturmian sequences, between exchanges of intervals and some sequences of linear complexity, and give an account of several tentative generalizations of continued fractions, using the notions of substitution and induction.

1. Introduction: the Rauzy program

A fundamental tool in arithmetics is the *Euclid algorithm* of continued fraction approximation. It gives a clear solution to the problem of approximating an irrational number α by a sequence of rational numbers $\frac{p_n}{q_n}$; not only does it give an approximation of good quality, with

$$\left| \alpha - \frac{p_n}{q_n} \right| \leq 0 \left(\frac{1}{q_n^2} \right),$$

but also the solution is unique, and in some sense canonical: every algorithm giving at least this quality of approximation is either the Euclid algorithm or some acceleration, using a subsequence of the rational numbers used by the former.

Given two numbers (α, β) , we could of course approximate them separately; but a more difficult problem is *simultaneous approximation*: find three sequences of integers (p_n, q_n, r_n) , such that the couples $(\frac{p_n}{q_n}, \frac{r_n}{q_n})$ are a good rational approximation of (α, β) . Given a norm $\| \cdot \|$ on \mathbb{R}^2 , the quality of the approximation is measured by

$$\frac{1}{q_n} \| \|q_n(\alpha, \beta)\| \|,$$

where the quantity between triple bars is the *distance to the nearest integer* of the vector $q_n(\alpha, \beta)$, namely

$$\min \{ \| \|q_n\alpha - p, q_n\beta - r\| \|, \quad (p, r) \in \mathbb{Z}^2 \}.$$

1991 *Mathematics Subject Classification*. Primary 37B10; Secondary 37E05, 11A63, 11J70, 11K60, 68R15.

This work has been supported by the ACI NIM 2004 Numération.

According to [CASSE] for instance, we cannot hope to get a quality better than $0(q_n^{-\frac{3}{2}})$ when α, β belong to a cubic real number field with $1, \alpha, \beta$ linearly independent over \mathbb{Q} ; α, β are then said to be *badly approximable numbers*.

But we do not know one algorithm giving the best quality in general. Our purpose here is not to give a review of the many different proposed algorithms, but to present a new approach to this question which was initiated by G. Rauzy in the 1980's.

This came from a classical result of Coven, Hedlund and Morse, giving an interaction between the field of *word combinatorics* and the field of *dynamical systems*, which is described extensively in section 2. The combinatorial objects involved are the *Sturmian sequences*, while the dynamical systems are the *rotations* of the torus \mathbb{T}^1 , involving irrational numbers α ; and Rauzy devised a new proof of the main result (Theorem 2.6 below) which made the interaction work through another mathematical object, the Euclid algorithm, which thus appeared naturally without being presupposed. Hence came the idea which can be called the *Rauzy program*: find generalizations of the Sturmian/rotations interaction which would naturally generate approximation algorithms; and, if we start from dynamical systems defined by d numbers, there is a chance that we shall get an algorithm of simultaneous approximation for these numbers.

Of course, the first system which comes to mind is a rotation of \mathbb{T}^2 ; and it is indeed through one of them that the Rauzy program was first realized, this is the *Tribonacci case* described in section 3. Here, not only we have an unexpected link between two very different objects, but also, following a natural geometric argument, we find an algorithm which gives the best possible approximation of one pair of cubic irrationals. More generally, such an arithmetic algorithm of simultaneous approximation can be defined for *Arnoux-Rauzy sequences*, but this algorithm is only defined on a set of measure 0. However, even after restricting ourselves to this set, most generalizations of the Tribonacci situation have failed so far, and for the moment the Rauzy program for rotations in \mathbb{T}^k seems to make little or no progress: that is why we consider this case as the *bad guy* in our story.

A clever idea to go around this difficulty is to consider two rotations of \mathbb{T}^1 instead of one of \mathbb{T}^2 ; this gave birth to a burgeoning part of the theory, exposed in section 4. The *ugly* name we gave to this section, although clearly exaggerated, refers to some esthetic difficulties in building two-dimensional sequences by iteration of patterns.

In our exposition, we did not follow the order indicated by the subtitle of this paper, because we love happy endings, and, after having described two cases where things are still rather involved, we have preferred to end with a *good* situation where the interaction works well. This is the case of *d-interval exchange transformations* studied in section 5. These were known since [OSE], after an idea attributed to Arnold [ARNOL]; but it was again Rauzy [RAU-1, RAU-2] who saw the operation of *induction* as a generalization of continued fractions: the basic idea is to take the first return map of a d -interval exchange transformation on a suitable interval, getting a new d -interval exchange transformation, and relate the parameters of the new one to those of the old one by a matrix equality. This idea had many far-reaching consequences during the last twenty years, and we shall not give more

than a very brief overview of them, with more emphasis on the original Rauzy program (involving links with word combinatorics) than on the, admittedly extremely deep, results which added techniques of differential geometry to the Rauzy idea.

2. The Sturmian case

This section is devoted to the well-known but fundamental case which started the Rauzy program. We detail the corresponding interactions between sequences, rotations, substitutions and continued fractions: we introduce two tools particularly suited to this framework, that is, the *induction* in section 2.4 and the *hat algorithm* in section 2.5. We survey in section 2.7 some recent and dramatic developments in Diophantine approximation and transcendence connected with Sturmian sequences.

2.1. Words, transformations, codings. Let \mathcal{A} be a finite set, called the *alphabet*, and its elements will be called *letters*.

DEFINITION 2.1. A *word* is a finite string $w_1 \cdots w_k$ of elements of \mathcal{A} ; the concatenation of two words w and w' is denoted multiplicatively, by ww' . A word $w_1 \cdots w_k$ is said to *occur* at place i in the infinite sequence or finite word u if $u_i = w_1, \dots, u_{i+k-1} = w_k$; we denote by $N(w, u)$ the number of these occurrences, and we say also that w is a *factor* of u .

For a sequence u , the *language* $L(u)$ is the set of all words occurring in u , while $L_n(u)$ is the set of all words of length n occurring in u ; the *complexity* of u is the function $p_u(n)$ which associates to each $n \in \mathbb{N}$ the cardinality of $L_n(u)$.

A sequence u is *uniformly recurrent* if every word occurring in u occurs in u infinitely many times with bounded gaps.

A word w in $L(u)$ is called *right special* (resp. *left special*) if there exist distinct letters $a, b \in \mathcal{A}$ such that both $wa, wb \in L(u)$, (resp. $aw, bw \in L(u)$). If $w \in L(u)$ is both right special and left special, then w is called *bispecial*.

For a word $w = w_1 \dots w_s$, $|w|$ denotes its length and $|w|_i$ the number of occurrences of the letter i in w ; we denote by \bar{w} its *retrograde* or *mirror image* $w_s \dots w_1$.

Prefixes, suffixes and *palindromes* are defined as in the usual language.

The sequence u is *M-balanced* if for all factors w and v of equal length we have $-M \leq N(a, w) - N(a, v) \leq M$ for all $a \in \mathcal{A}$. The sequence u is *uniformly balanced* if there exists M such that u is M -balanced.

A sequence u defines naturally a *symbolic dynamical system* defined as the set of sequences on the same alphabet as u whose set of factors is included in $L(u)$, endowed with the shift map S ; for more details, see section 5.1.4 of [PYT]. Let us recall a few definitions about topological dynamical systems:

DEFINITION 2.2. A system (X, T) is *minimal* if all the orbits are dense. It is *uniquely ergodic* if there is only one invariant probability measure.

For a subset E of X , the *first return time* $r_E(x)$ of a point $x \in E$ is $\min\{n > 0; T^n x \in E\}$. If the return time is always finite, we define the *induced map* on E by $T_E(x) = T^{r_E(x)}(x)$.

DEFINITION 2.3. Given a transformation T on a set X , a point $x \in X$, and a partition $\mathcal{P} = \{P_0, \dots, P_{k-1}\}$ of X , we define the (one-sided) sequence $\mathcal{P}(x, T)$ by $\mathcal{P}(x, T)_n = i$ whenever $T^n x \in P_i$, $n \in \mathbb{N}$; $\mathcal{P}(x, T)$ is called a *coding* of the transformation T .

2.2. The fundamental result. *Throughout this section, α is assumed to be irrational.* We consider the *rotation* or *translation* R_α defined on the torus \mathbb{T}^1 by

$$R_\alpha x = x + \alpha \pmod{1};$$

it is minimal because of the irrationality of α .

We want to find codings of R_α ; among possible partitions of the torus, we consider partitions by two intervals; even all these are not equivalent (we shall consider more general binary codings of rotations in section 5.1), as the partition by the two points 0 and $1 - \alpha$ is the only one such that, if we look at R_α on the fundamental domain $[0, 1[$, the translation $R_\alpha x - x$ is constant on each of its atoms. Indeed, R_α can be represented as a *two-interval exchange*

$$R_\alpha x = \begin{cases} x + \alpha & \text{if } x \in [0, 1 - \alpha[\\ x + \alpha - 1 & \text{if } x \in [1 - \alpha, 1[. \end{cases}$$

DEFINITION 2.4. A sequence $u = (u_n)_{n \in \mathbb{N}}$ is a *natural coding* of the rotation R_α if there exists x such that $u = \mathcal{P}(x, R_\alpha)$ for $\mathcal{P} = \{[0, 1 - \alpha[, [1 - \alpha, 1\}$ or $\mathcal{P} = \{[0, 1 - \alpha],]1 - \alpha, 1]\}$.

The fundamental result we state below is a combinatorial characterization of these natural codings.

DEFINITION 2.5. A sequence u is *Sturmian* if $p_u(n) = n + 1$ for all n .

THEOREM 2.6. *A sequence is a natural coding of an irrational rotation if and only if it is Sturmian.*

This result is of course the basis of the interest in Sturmian sequences, for which we refer the reader to the two excellent surveys in [LOT-1], Chapter 2 and [PYT], Chapter 6. Theorem 2.6 is in fact the juxtaposition of two results, one in [MOR-HED] characterizing the natural coding of rotations by a property similar to the 1-balance, and one in [COV-HED] establishing the equivalence between 1-balanced and Sturmian sequences.

The nontrivial part of Theorem 2.6 is of course the “if” part; the easy part is a direct consequence of the following lemma which is easily proved and will be crucial in all that follows.

LEMMA 2.7. *The finite word $w_0 \dots w_s$ is a factor of $\mathcal{P}(x, T)$ if and only if $\bigcap_{i=0}^s T^{-i} P_{w_i} \neq \emptyset$.*

We deduce from this lemma that, in the case of natural codings of rotations, the language of $\mathcal{P}(x, R_\alpha)$ does not depend on x , and that a natural coding of R_α is a Sturmian sequence.

The original proofs in [MOR-HED] and [COV-HED] use only word combinatorics; a later proof by Rauzy, described in [ARN-RAU], uses arithmetics, through the Euclid algorithm of continued fraction approximation, to link the combinatorial properties of the sequence and the dynamics of the system. Though it is not our purpose to give the details of that proof, we found useful, to help the reader to understand later generalizations, to give the outline of two variants of Rauzy’s proof (for the nontrivial direction) in sections 2.4 and 2.5.

Among Sturmian sequences one plays a particular part.

DEFINITION 2.8. A Sturmian sequence is *homogeneous* or *characteristic* if all its prefixes are left special factors.

In particular the natural coding $\mathcal{P}(\alpha, R_\alpha)$ is a homogeneous Sturmian sequence, and it is the only one among all $\mathcal{P}(x, R_\alpha)$.

In our sketch of the proof of Theorem 2.6, we detail two methods, based on continued fraction expansions, for generating homogeneous Sturmian words, and we extend them to the nonhomogeneous case in section 2.6. Let us observe that there exists an important literature devoted to the generation methods of Sturmian sequences, mainly in the homogeneous case [LOT-1]; see for instance [BRO-1] for a survey of some descriptions of homogeneous Sturmian words and [BRO-2] for a survey of some of the applications to number theory.

2.3. Substitutions and the Fibonacci sequence. We shall make extensive use of substitutions in this study, let us recall the basic definitions:

DEFINITION 2.9. A *substitution* is an application from an alphabet \mathcal{A} into the set \mathcal{A}^* of finite words on \mathcal{A} ; it extends to a morphism of \mathcal{A}^* by concatenation by $\sigma(w w') = \sigma(w)\sigma(w')$.

It is called *primitive* if there exists k such that a occurs in $\sigma^k(b)$ for any $a \in \mathcal{A}$, $b \in \mathcal{A}$.

A *fixed point* of σ is an infinite sequence u with $\sigma(u) = u$.

The *incidence matrix* of a substitution σ is defined by $M = (m_{ij})$ where m_{ij} is the number of occurrences of the letter i in the word $\sigma(j)$.

One of the most famous examples of substitutions is the *Fibonacci* substitution, $\sigma(0) = 01$, $\sigma(1) = 0$; its fixed point, the infinite sequence beginning by $\sigma^n(0)$ for all n , is called the *Fibonacci sequence*; the famous Fibonacci numbers F_n are just the lengths of the images of 0, $F_n = |\sigma^n(0)|$, satisfying the relations $F_n = F_{n-1} + F_{n-2}$ with $F_{-1} = 1$ and $F_0 = 1$. An old result, whose proof is a nontrivial exercise (5.4.11 of [PYT]), is that

PROPOSITION 2.1. *The Fibonacci sequence is Sturmian.*

The symbolic systems associated to fixed points of substitutions, or, more generally, the *S-adic* systems associated to sequences u whose language is generated by a family of substitutions (such as in section 2.4 below), have been extensively studied (see Chapter 12 in [PYT]); many results in this study can be read as providing *semi-topological conjugacies* between such symbolic systems and well-known geometric dynamical systems. For example, Theorem 2.6 and Proposition 2.1 imply that *the system associated to the Fibonacci sequence is semi-topologically conjugate to a rotation of \mathbb{T}^1* . Let us recall that two dynamical systems are said to be *semi-topologically conjugate* if there exist two sets X_1 and Y_1 which are at most countable such that $B_1 = X \setminus X_1$, $B_2 = Y \setminus Y_1$, and f is a bicontinuous bijection from B_1 onto B_2 which conjugates S and T .

2.4. Proof of Theorem 2.6: the Rauzy induction.

2.4.1. *Step R1.* We first prove that the language of a Sturmian sequence is generated by some catenation rules: starting from a Sturmian sequence u , we build a nested sequence of words W_n in its language, such that each factor of u is a factor of some W_n ; the W_n are built by recursion formulas, which may be expressed by using an infinite composition of a finite number of substitutions (we recall that such a generation process is called *S-adic*); the order of composition of the substitutions is ruled by a parameter α associated to u . This part has become quite standard, and can be proved using either the graphs of words [ARN-RAU], leading

to the so-called *Rauzy rules* or standard rules [**LOT-1**], or by a *recoding* argument [**ARN-FIS**, **PYT**], as detailed below.

We follow here the argument of [**BER-HOL-ZAM**]. Let τ_0 and τ_1 be the substitutions defined by:

$$\tau_0(0) = 0, \tau_0(1) = 01, \quad \tau_1(0) = 10, \tau_1(1) = 1.$$

Let u be a Sturmian sequence (w.l.o.g. we take $\mathcal{A} = \{0, 1\}$). The language $L_2(u)$ is either equal to $\{00, 01, 10\}$ or to $\{11, 01, 10\}$. Hence exactly one of the words ii ($i \in \{0, 1\}$) is a factor of u and there is a unique sequence u' such that $u = S^b(\tau_i(u'))$, where $b = 0$ if u begins in i and $b = 1$ otherwise, and S denotes the shift on $\{0, 1\}^{\mathbb{N}}$. One proves furthermore that u' is again a Sturmian sequence using classical combinatorial arguments based on the 1-balance and palindrome properties of Sturmian sequences (see [**COV-HED**] or [**PYT**]).

By iterating this process, one obtains two sequences $(i_n), (b_n) \in \{0, 1\}^{\mathbb{N}}$ and a sequence of Sturmian sequences $(u^{(n)})_{n \geq 1}$, with (i_n) not eventually constant, such that for each n

$$u = S^{b_1} \circ \tau_{i_1} \circ \dots \circ S^{b_n} \circ \tau_{i_n}(u^{(n)}).$$

We then write

$$i_1 i_2 \dots = 0^{a_1} 1^{a_2} 0^{a_3} 1^{a_4} \dots$$

with $a_i \geq 1$ for $i \geq 2$.

Let α be the irrational real number in $]0, 1[$ whose continued fraction expansion is given by the sequence $(a_n)_{n \geq 1}$

$$\alpha = [0; a_1 + 1, a_2, a_3, \dots] = \frac{1}{1 + a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{a_4 + \dots}}}}.$$

It is easily seen by induction on k that for $i \in \{0, 1\}$

$$(|\tau_0^{a_1} \circ \dots \circ \tau_{k-1}^{a_k}(i)|_0, |\tau_0^{a_1} \circ \dots \circ \tau_{k-1}^{a_k}(i)|_1) = \begin{cases} (q_k - p_k, p_k) & i = k \pmod{2} \\ (q_{k-1} - p_{k-1}, p_{k-1}) & i \neq k \pmod{2}, \end{cases}$$

where the sequence $(\frac{p_n}{q_n})_{n \in \mathbb{N}}$ denotes as usual the sequence of convergents of α . We then can deduce that the frequency of occurrence of the letter 1 in u (which is known to exist thanks to the 1-balance property) is equal to α .

2.4.2. Step R2. Now, for a given α and w.l.o.g. the partition $\mathcal{P} = \{[0, 1 - \alpha[, [1 - \alpha, 1[\}$, we shall generate the homogeneous Sturmian sequence $\mathcal{P}(\alpha, R_\alpha)$ in a similar S -adic way. This computation, though it is in the folklore, always gives rise to a few technical difficulties, so we give it completely; this is a particular case of a proof which can be found in [**BER-HOL-ZAM**] (see also for similar proofs [**ARN-FER-HUB**] or [**ARN-FIS**]), and of the more general concept of Rauzy induction (section 5.2).

For computational reasons, we apply to every point the translation $y \mapsto y - \alpha$. The rotation R_α becomes the following exchange of two intervals denoted by $E^{(0)}$:

$$\begin{cases} E^{(0)}(y) &= y + \alpha & \text{if } y \in P_0^{(0)} := [-\alpha, 1 - 2\alpha[, \\ E^{(0)}(y) &= y + \alpha - 1 & \text{if } y \in P_1^{(0)} := [1 - 2\alpha, 1 - \alpha[. \end{cases}$$

Let us compute $\mathcal{P}(\alpha, R_\alpha)$, which is also $\mathcal{P}^{(0)}(0, E^{(0)})$ with partition $\mathcal{P}^{(0)} = \{P_0^{(0)}, P_0^{(1)}\}$.

We set $\delta_{-1} = 1 - \alpha$ and $\delta_0 = \alpha$. Let a_1 be the largest nonnegative integer k such that $\delta_{-1} = k\delta_0 + \delta_1$, with $0 \leq \delta_1 < \delta_0$, which means $a_1 + 1 = \lfloor \frac{1}{\alpha} \rfloor$ and $\delta_1 = 1 - (a_1 + 1)\alpha$. Let I be the interval $[-\delta_0, \delta_1[= [-\alpha, 1 - (a_1 + 1)\alpha[$. Let $\mathcal{P}^{(1)} = \{P_0^{(1)}, P_1^{(1)}\}$ with $P_0^{(1)} := [-\delta_0, -\delta_0 + \delta_1[$ and $P_1^{(1)} := [-\delta_0 + \delta_1, \delta_1[$.

We make now what is called the *symmetric Rauzy induction*: for a point y in I , let $t(y)$ be the first return time of y in I , that is, $t(y) = \min\{k \geq 1; (E^{(0)})^k(y) \in I\}$, and $E^{(1)}(y) = (E^{(0)})^{t(y)}(y)$ the induced map of $E^{(0)}$ on I . It is possible to compute t and $E^{(0)}$:

- if y is in $P_0^{(1)} = [-\alpha, 1 - (a_1 + 2)\alpha[$, then $E^{(0)}(y) = y + \alpha$ is in $[0, 1 - (a_1 + 1)\alpha[\subset I$, and so $t(y) = 1$ and $E^{(1)}(y) = y + \alpha$;
- if y is in $P_1^{(1)} = [-\delta_0 + \delta_1, \delta_1[= [1 - (a_1 + 2)\alpha, 1 - (a_1 + 1)\alpha[$, then $(E^{(0)})^l(y) = y + l\alpha$ is in $[-\alpha + l\alpha + \delta_1, l\alpha + \delta_1[\subset [\delta_1, \delta_{-1}[\subset I^c$ for $1 \leq l \leq a_1$, then $(E^{(0)})^{a_1-1}(y) \in [1 - 2\alpha, 1 - \alpha[$; we thus get $(E^{(0)})^{a_1+1}(y) = y + \alpha - 1 \in [-\alpha, 0[\subset I$; hence $t(y) = a_1 + 1$ and $E^{(1)}(y) = y + (a_1 + 1)\alpha - 1$.

Let $v^{(0)}$ be the sequence $\mathcal{P}^{(0)}(0, E^{(0)})$ and $v^{(1)}$ be $\mathcal{P}^{(1)}(0, E^{(1)})$. Suppose we know $v^{(1)}$ and we want to find $v^{(0)}$; $(E^{(0)})^n(0)$ will be in I for $n = 0$, $n = t(0)$, $n = t_2(0) := t(0) + t(E^{(0)}(0))$, the second return time of 0 in I , ..., $n = t_k(0) := t_{k-1}(0) + t((E^{(0)})^{k-1}(0))$, which are the return times of 0 in I for all positive k . Moreover,

- if $v_k^{(1)} = 0$, then $(E^{(1)})^k(0) = (E^{(0)})^{t_k(0)}(0)$ is in $P_0^{(1)} \subset P_0^{(0)}$, and, by last paragraph, we know that $t((E^{(1)})^k(0)) = 1$ and hence $t_{k+1}(0) = t_k(0) + 1$;
- if $v_k^{(1)} = 1$, then $(E^{(1)})^k(0)$ is in $P_1^{(1)} \subset P_0^{(0)}$, if $a_1 \geq 1$, and so we know, again by last paragraph, that $(E^{(0)})^{t_k(0)+l}(0)$ is in $P_0^{(0)}$ for $1 \leq l \leq a_1 - 1$, $(E^{(0)})^{t_k(0)+a_1}(0)$ is in $P_1^{(0)}$ and that $t((E^{(1)})^k(0)) = a_1$, and so $t_{k+1}(0) = t_k(0) + a_1 + 1$.

From this analysis, it results that $v_{t_k(0)}^{(0)} = v_k^{(1)}$ for all k ; when two consecutive $t_k(0)$ differ by more than 1, then they differ by $a_1 + 1$, $v_k^{(1)} = 0$, and the digits $v_{t_k(0)+1}^{(0)}, \dots, v_{t_k(0)+a_1-1}^{(0)}$ are all equal to 0, and $v_{t_k(0)+a_1}^{(0)} = 1$. In short, we say that $v^{(0)}$ is deduced from $v^{(1)}$ by the substitution $\tau_0^{a_1}$ (if $a_1 = 0$, then $\tau_0^{a_1}$ reduces to the identity), that is, $v^{(0)} = \tau_{a_1}(v^{(1)})$.

Furthermore, $v^{(1)} = \mathcal{P}^{(1)}(0, E^{(1)})$ with

$$\begin{cases} E^{(1)}(y) &= y + \alpha & \text{if } y \in P_0^{(1)}, \\ E^{(0)}(y) &= y + (a_1 + 1)\alpha - 1 & \text{if } y \in P_1^{(0)} \end{cases}$$

is the natural coding $\mathcal{P}(\alpha_1, R_{\alpha_1})$ with $\alpha_1 = \frac{\alpha}{1 - a_1\alpha} = \frac{1}{T(\alpha) + 1}$ where T denotes the *Gauss map*

$$T :]0, 1[\rightarrow]0, 1[, \quad x \mapsto \{1/x\}.$$

We can now iterate this process: we repeat the previous construction with $E^{(1)}$, δ_1 , α_1 , $\mathcal{P}^{(1)}$ replaced by $E^{(n)}$, δ_n , α_n , $\mathcal{P}^{(n)}$ defined with the following recurrence

formulas, for $n \geq 1$, with $E^{(n)}$ associated to α_n in the same way as $E^{(0)}$ to α :

$$\begin{aligned}\alpha_n &= \frac{1}{T^{n-1}(\alpha) + 1}, \\ a_n &= \left\lfloor \frac{1}{T^{n-1}(\alpha)} \right\rfloor \\ \delta_{n-1} &= a_{n+1}\delta_n + \delta_{n+1} \\ \mathcal{P}^{(n)}(0, E^{(n)}) &= \tau_{a_n}(\mathcal{P}^{(n+1)}(0, E^{(n+1)})),\end{aligned}$$

with, if n is even, $E^{(n)} : [-\delta_n, \delta_{n-1}[\rightarrow [-\delta_n, \delta_{n-1}[$ given by

$$E^{(n)}(z) = \begin{cases} z + \delta_n & \text{if } z \in [-\delta_n, -\delta_n + \delta_{n-1}[\\ z - \delta_{n-1} & \text{if } z \in [-\delta_n + \delta_{n-1}, \delta_{n-1}[\end{cases};$$

and if n is odd, $E^{(n)} : [-\delta_{n-1}, \delta_n[\rightarrow [-\delta_{n-1}, \delta_n[$ given by

$$E^{(n)}(z) = \begin{cases} z + \delta_{n-1} & \text{if } z \in [-\delta_{n-1}, -\delta_{n-1} + \delta_n[\\ z - \delta_n & \text{if } z \in [-\delta_{n-1} + \delta_n, \delta_n[\end{cases}.$$

As the a_n for $n \geq 2$, are all strictly positive, it is easily seen that the length of $\tau_0^{a_1} \circ \dots \circ \tau_1^{a_{2n}}(0)$ tends to infinity. Hence

$$\mathcal{P}(\alpha, R_\alpha) = \lim_{n \rightarrow +\infty} (\tau_0^{a_1} \circ \dots \circ \tau_1^{a_{2n}})(0).$$

Moreover, the definitions of the a_n allow us to recognize the partial quotients of the continued fraction approximation of α , and the a_n are defined by $\alpha = [0; a_1 + 1, a_2, \dots]$.

2.4.3. Step R3. Hence, given u and its a_n and α from step R1, u has the same language as $\mathcal{P}(\alpha, R_\alpha)$, and thus by Lemma 2.7 as all the $\mathcal{P}(x, R_\alpha)$; hence, by a standard approximation argument, u itself is some $\mathcal{P}(x, R_\alpha)$.

A by-product of this proof is to provide a semi-topological conjugacy between R_α and an S-adic system, or, in short, an *S-adic presentation* of R_α . We can also check what the reader suspects already: the angle α of the rotation corresponding to the Fibonacci sequence is the *golden ratio number* $\phi = \frac{\sqrt{5}-1}{2}$; but note that here the interval (of the partition for the natural coding) coded by 1 has to be the shorter one, so it is $[0, 1 - \phi[$, and we need to exchange 0 and 1 if we want to be coherent with the usual notations.

2.5. Proof of Theorem 2.6: the hat algorithm. We follow [RIS-ZAM] and the same scheme as in the previous section; we first associate a sequence of digits $(a_n)_{n \geq 1}$ to a given Sturmian sequence u (step H1). We then give a generation method for a characteristic Sturmian word associated with (a_n) (step H2). In step H3 we connect the a_n with the continued fraction expansion of α . We end as in step R3.

2.5.1. Step H1. Starting from a Sturmian sequence u (on $\mathcal{A} = \{0, 1\}$), we prove that for each n , there is one left special factor of u of length n , l_n ; moreover, l_m is a prefix of l_n for $m < n$, hence there is an infinite (Sturmian) sequence c such that $c_0 \dots c_{n-1} = l_n$ for all n ; c is a characteristic Sturmian sequence.

For an increasing sequence of integers $(k_n)_{n \in \mathbb{N}}$, $c_0 \dots c_{k_n}$ is also right special, hence bispecial; then $c_0 \dots c_{k_n+1}$ is left special, while $c_0 \dots c_{k_n}(1 - c_{k_n+1})$ is a factor of u but cannot be left special. We describe the sequence $(c_{k_n+1})_{n \in \mathbb{N}}$ by $a_1 \geq 0$, $a_2 > 0$, $\dots a_n > 0$, \dots the length of the consecutive strings of 0's and 1's in this sequence, starting from a_1 zeros (at the beginning we consider the empty word as bispecial, hence $c_{k_0+1} = c_0$).

2.5.2. *Step H2.* The knowledge of the a_n allows us to build explicitly the homogeneous sequence c by the *hat algorithm*: we produce an infinite sequence

$$d = V_1^{a_1} V_2^{a_2} \dots V_n^{a_n} \dots$$

over the alphabet $\{0, 1, \hat{0}, \hat{1}\}$ as follows: start from $V_1 = \hat{0}$; then let V_{2n} be the shortest, or last (in the order of occurrence), suffix of $V_1^{a_1} V_2^{a_2} \dots V_{2n-1}^{a_{2n-1}}$ beginning by $\hat{1}$ (if there is no such suffix, which may happen at the initial step, take the shortest suffix of $\hat{1} V_1^{a_1} V_2^{a_2} \dots V_{2n-1}^{a_{2n-1}}$ beginning by $\hat{1}$) deprived of every hat except its initial one; let V_{2n+1} be the shortest suffix of $V_1^{a_1} V_2^{a_2} \dots V_{2n}^{a_{2n}}$ beginning by $\hat{0}$ (if there is no such suffix, take the shortest suffix of $\hat{0} V_1^{a_1} V_2^{a_2} \dots V_{2n-1}^{a_{2n-1}}$ beginning by $\hat{0}$) deprived of every hat except its initial one. Then c is just d deprived of all its hats, while the bispecial words are all the prefixes of c ending before a hat, and deprived of all their hats.

EXAMPLE 2.1. For the Fibonacci sequence u , the a_n are $0, 1, 1 \dots$ and we have to exchange 0 and 1 (see the remark after section 2.4.3); we check that it is equivalent to keep the original 0 and 1 with the a_n equal to $1, 1, 1 \dots$, and we get

$$\hat{0}\hat{1}0\hat{0}10\hat{1}0010\hat{0}1010010\hat{1}001001010010 \dots;$$

not only do we retrieve this famous sequence much faster than by using the Fibonacci substitution, but, by stopping before each hat, we get the bispecial words.

2.5.3. *Step H3.* We know that for a given α and w.l.o.g. the partition $\mathcal{P} = \{[0, 1 - \alpha], [1 - \alpha, 0]\}$, the coding of any x is a Sturmian sequence using only Lemma 2.7; it remains just to identify its a_n ; this is done by straightforward computations, using only Lemma 2.7; the possible values of c_{k_n+1} are then dictated by inequalities in which we recognize those defining the partial quotients of the continued fraction approximation of α , and the a_n are finally defined by the same formula as in step R2.

2.5.4. *Step H4.* As step R3.

2.6. Nonhomogeneous Sturmian sequences. With both the induction and the hat algorithm, we get an explicit determination of the characteristic Sturmian sequence associated to a number α , using the partial quotients of the continued fraction approximation of α . The same methods can be used also to get any nonhomogeneous sequence $\mathcal{P}(x, R_\alpha)$, and this defines naturally another algorithm, which was first defined by Ostrowski [OST-1, OST-2], see also the survey [BER]. This numeration system is based on a numeration scale provided by the denominators of the convergents in the continued fraction expansion of α .

Let us set $\delta_n := q_n \alpha - p_n$, for all n in \mathbb{N} . Let us first recall (see for instance [SOS]) that every real number x , with $-\alpha \leq x < 1 - \alpha$, can be uniquely expanded

as

$$x = \sum_{k=1}^{+\infty} c_k \delta_{k-1},$$

where $0 \leq c_1 \leq a_1 - 1$, $0 \leq c_k \leq a_k$, for $k \geq 2$,

$$c_k = 0 \text{ if } c_{k+1} = a_{k+1},$$

$c_k \neq a_k$, for infinitely many even and infinitely many odd indices.

In [BER-HOL-ZAM] we find the following result (see also [ARN-FER-HUB, ARN-FIS], [VER-SID], [MASUI-SUG-YOS]):

THEOREM 2.10. *Let $\alpha = [0; a_1 + 1, a_2, \dots, a_n, \dots]$. Let u be the Sturmian sequence $\mathcal{P}(x, R_\alpha)$. There exists a sequence of integers $(c_n)_{n \in \mathbb{N}}$ where*

$$\forall n, \begin{cases} 0 \leq c_n \leq a_n, \\ c_{n+1} = a_{n+1} \Rightarrow c_n = 0, \end{cases}$$

and a sequence of Sturmian sequences $(u^{(k)})$ such that

$$\forall k, \quad \omega = S^{c_1} \tau_0^{a_1} \circ S^{c_2} \tau_1^{a_2} \circ S^{c_3} \tau_0^{a_3} \circ \dots \circ S^{c_k} \tau_{k-1}^{a_k} (u^{(k)}),$$

and

$$x = \sum_{k=1}^{+\infty} c_k (-1)^{k-1} \delta_{k-1} = \sum_{k=1}^{+\infty} c_k (q_{k-1} \alpha - p_{k-1}).$$

Note that the above formula can also be interpreted by using the hat algorithm for the homogeneous sequence; namely, with the notations and construction of section 2.5.2, we have

$$\mathcal{P}(x, R_\alpha) = V_1^{a_1 - c_1} V_2^{a_2 - c_2} \dots V_n^{a_n - c_n} \dots$$

The Ostrowski numeration system is particularly well suited to the study and description of the combinatorial properties of Sturmian sequences; see for instance the references and examples quoted in the survey [BER].

Let us just briefly mention an application considered in [BER-HOL-ZAM] concerning powers of prefixes in Sturmian sequences. The initial critical exponent of a Sturmian sequence u is defined as the supremum of all real numbers $p > 0$ for which there exist arbitrary long prefixes of u of the form w^p , for w being a factor. It is possible to compute for a given Sturmian sequence its initial critical exponent in terms of Ostrowski's representation and to characterize those irrational rotations R_α for which there exists a Sturmian sequence with initial critical exponent equal to 2: such α 's form a set of zero measure among irrational numbers with unbounded partial quotients. Let us note that the initial critical exponent does not behave in a uniform way on a given Sturmian shift space, that is, on the whole symbolic dynamical system generated by a given Sturmian sequence: for instance, the Fibonacci sequence u begins in no $\frac{1+\sqrt{5}}{2} < 3$ power at all, while every sequence in the symbolic dynamical system generated by u which does not belong to its shift orbit begins in arbitrarily long cubes.

2.7. Application: Sturmian sequences and Diophantine approximation. A domain which has been very active recently is the construction of transcendental numbers through their expansions; an excellent though already outdated survey can be found in [ALL].

Beginning with arithmetic expansions, an old conjecture due to Borel [BOR] states that *every algebraic number is normal*. The opposite case is the case of numbers whose expansion in some base k has low complexity, beginning with those with Sturmian expansion; and indeed, these numbers were proved to be *transcendental*, first in [MAH] for homogeneous Sturmian expansions associated with a quadratic parameter, in [LOX-vdP] for general homogeneous Sturmian expansions, see also [ADAMS-DAV], and then in [FER-MAU] for general Sturmian expansions in base 2 and their natural generalizations to base k . The proofs are based on a combinatorial reformulation of Ridout's theorem [RID].

But all this now pales compared with the last result of [ADA-BUG-LUC]: every irrational number whose expansion has at most linear complexity is transcendental, which implies in particular that the b -adic expansion of an algebraic irrational number cannot be generated by a finite automaton, result until then known as the Loxton and van der Poorten conjecture; the combinatorial criterion of [FER-MAU] is refined thanks to a clever use of the Schmidt Subspace Theorem; such results are also extended to algebraic bases associated to Salem and Pisot numbers, as well as to the Hensel expansions of p -adic numbers [ADA-BUG].

When we look at continued fraction expansions, the problems seem more difficult; but it was proved in [ALL-DAV-QUE-ZAM] that a number is transcendental if the partial quotients in its continued fraction expansion $[0; a_1, \dots, a_n, \dots]$ form a Sturmian sequence. Real numbers ξ with a Sturmian continued fraction expansion have interesting Diophantine approximation properties [ROY-1, ROY-2, BUG-LAU]: this is due to the fact that characteristic Sturmian sequences begin with infinitely and sufficiently many prefixes that are palindromes (this provides simultaneous rational approximations to ξ and ξ^2) and that they admit many prefixes of the form www' with w' prefix of w (this provides very good quadratic approximations to ξ).

More precisely, for a real number χ , let the exponent $\hat{\lambda}_2(\chi)$, which measures the quality of simultaneous rational approximations to ξ and ξ^2 , be defined as the supremum of all real numbers λ such that for all sufficiently large X , the inequalities

$$1 \leq |x_0| \leq X, \quad |x_0\xi - x_1| \leq X^{-\lambda}, \quad |x_0\xi^2 - x_2| \leq X^{-\lambda}$$

admit a nonzero integer solution x_0, x_1, x_2 . A real number ξ is *extremal* according to [ROY-1, ROY-2] if it is neither rational nor quadratic, and $\hat{\lambda}_2(\xi)$ equals $\frac{\sqrt{5}-1}{2}$; this value is then optimal since being equal to the best upper bound for the exponent $\hat{\lambda}_2(\chi)$ according to [DAV-SCH]. Extremal points are shown to be transcendental (by applying the Schmidt Subspace theorem) and can be considered as very close concerning this approximation behavior to quadratic real numbers (for which the exponent $\hat{\lambda}_2(\chi)$ equals 1). It is proved in [ROY-1, ROY-2] that extremal numbers can be produced by using real numbers whose continued fraction is given by the Fibonacci sequence on $\{1, 2\}$. Furthermore, the spectrum of values taken by the exponent $\hat{\lambda}_2(\chi)$ is considered in [BUG-LAU] for the real numbers whose continued fraction is a characteristic Sturmian sequence; see also [ROY-3].

Finally, [BLO-THE-VLA] gives an example of application of related Sturmian products of matrices in a totally different context.

3. The bad: rotations on \mathbb{T}^2

3.1. Natural codings. For a translation $(x, y) \mapsto (x + \alpha, y + \beta)$ in \mathbb{T}^2 , there are many possible codings, and surprisingly, the simplest partitions, using rectangles, for example, do not give the best results, as discussed below. Hence we have been led to give a more general definition of the notion of natural coding, which does not exclude rectangles but allows more interesting shapes of atoms.

DEFINITION 3.1. Let \mathbb{T}^d denote the d -dimensional torus \mathbb{R}^d/L where L is a lattice (discrete maximal subgroup) in \mathbb{R}^d . A *translation* on \mathbb{T}^d is a map R_α where $\alpha \in \mathbb{R}^d$, and $R_\alpha : \mathbb{T}^d \rightarrow \mathbb{T}^d$ is defined by $R_\alpha(x) = x + \alpha \pmod{L}$. We say a sequence u is a *natural coding* of a translation R_α of \mathbb{T}^d if there exists a fundamental domain Ω for L together with a partition $\Omega = \Omega_1 \cup \Omega_2 \cup \dots \cup \Omega_k$ such that on each Ω_i the map R_α is a translation by a vector α_i and u is the coding of a point $x \in \Omega$ with respect to the partition $\mathcal{P} = \{\Omega_1, \dots, \Omega_k\}$.

If the coordinates $\alpha_1, \dots, \alpha_d$ of the translation vector α are \mathbb{Q} -independent with 1, then R_α is minimal and uniquely ergodic, which is the case to which we restrict ourselves here.

Let us see why we are not simply considering boxes in \mathbb{T}^d for the atoms of the partition. First, recall that a subset A of \mathbb{T}^d with (Lebesgue) measure $\mu(A)$ is said to be a *bounded remainder set* for the translation $R_\alpha : x \mapsto x + \alpha$, if there exists $C > 0$ such that

$$\forall N, \quad |\text{Card}\{0 \leq n < N; n\alpha \in A\} - N\mu(A)| \leq C.$$

This property can be easily reformulated in combinatorial terms as a property of uniform balance for the sequence $u(A)$ defined by $u(A)_n = 1$ if and only if $n\alpha \in A$ for $n \in \mathbb{N}$ ([ADA-3]). The property of 1-balance for Sturmian sequences is thus related to the fact that the intervals of the partition for a natural coding (Definition 2.4) have length in $\alpha\mathbb{Z} + \mathbb{Z}$, and are thus bounded remainder sets according to [KES]. Moreover, in the Sturmian case, the induced map of R_α on the atoms of the coding partition is again a rotation, and this fact is very important: only because of this can the induction in section 2.4.2 be iterated (of course, inducing on any interval would give a three-interval exchange, and continue to give three-interval exchanges if we induce again on so-called admissible intervals; for more details, see section 5). According to [FER-1, RAU-4], if the induced map of a translation on a set A is still a translation, then this set A is a bounded remainder set. This seems to indicate that, in the more elementary generalizations of the Sturmian/rotation interaction, the atoms of a natural coding of a translation should be preferably chosen as bounded remainder sets; hence they cannot be boxes in \mathbb{T}^d , as there are no nontrivial rectangles which are bounded remainder sets for ergodic translations on the torus [LIA].

Second, Sturmian sequences are the sequences with lowest complexity function among nonperiodic sequences. If we want to obtain codings with the smallest possible complexity function, then codings of translations with respect to boxes will not be suited. Indeed one could consider more generally according to [SCH-SZA-WIN,

STE-WIN, WIN] an ergodic translation on a compact abelian group C with normalized Haar measure μ , and code trajectories with respect to continuity sets of C , that is, to sets with topological boundary of Haar measure 0. A Hartman sequence is then defined as a binary sequence $u \in \{0, 1\}^{\mathbb{N}}$ with $u_n = 1$ if and only if $T^n(0_C) \in M$, for a fixed Hartman set M . When $C = \mathbb{T}^d$ and the translation is some R_α (assumed to be ergodic), then we recover binary codings of Kronecker sequences (that we discuss in section 5.1 for the one-dimensional torus). Hartman sequences have been introduced as a generalization of Sturmian sequences: it is thus natural to study the growth rate of the complexity function; it is proved in **[STE-WIN]** to be subexponential, and when the Hartman set M is a box in the torus \mathbb{T}^d , then the complexity grows polynomially of maximal order d . But by considering Rauzy fractals as described below in section 3.2 and 3.3 as Hartman sets, then one gets linear complexity.

Third, it is a reasonable assumption that our natural codings could be used to construct Markov partitions for automorphisms of the torus, and this will indeed be true in the founding particular case studied in section 3.2 where a particular substitution, the *Tribonacci substitution*, is proved (Theorem 3.4 below) to be a natural coding of a translation. It is then possible to construct a Markov partition for the toral automorphism of \mathbb{T}^3 whose matrix is the incidence matrix of the Tribonacci substitution, this construction being based on the *Rauzy fractal*, see e.g. **[PYT]** Chapter 7 or **[ITO-OHT]**. But it is known **[BOW]** that Markov partitions cannot be boxes for hyperbolic automorphisms of the torus in dimension 3 or more.

3.2. The Tribonacci substitution. This is the case where everything goes well, and the Rauzy program is completely fulfilled; though it may seem a very isolated case, we cannot overstate its importance.

Let τ be the substitution $\tau(0) = 01$, $\tau(1) = 02$, $\tau(2) = 0$; τ is usually called the *Rauzy substitution* but we prefer the term *Tribonacci substitution*, as, if $T_n = |\tau^{n-2}(0)|$, we have $T_n = T_{n-1} + T_{n-2} + T_{n-3}$, starting from $T_0 = T_1 = 0$, $T_2 = 1$.

Unsurprisingly, we shall be led to consider the polynomial $X^3 - X^2 - X - 1$ which is the characteristic polynomial of its incidence matrix $M_\tau = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$; it has three roots, $\beta = 1, 83929\dots > 1$, and two complex roots $\alpha = (-0, 419\dots) - (0, 606\dots)i$ and $\bar{\alpha}$, with $|\alpha| < 1$, hence β is a Pisot number.

THEOREM 3.2. [RAU-3] *The fixed point u of τ is a natural coding of the translation by $(\frac{1}{\beta}, \frac{1}{\beta^2})$ on \mathbb{T}^2 .*

Hence the translation $R_{(\frac{1}{\beta}, \frac{1}{\beta^2})}$ is semi-topologically conjugate to the symbolic dynamical system associated to the fixed point u of the Tribonacci substitution.

The proof uses a description of the rotation as an *exchange of pieces* in \mathbb{R}^2 . The fundamental domain Ω is one of the famous Rauzy fractals, see for example section 7.4 of **[PYT]**: we can build Rauzy fractals for a whole class of rotations in the same cubic field, but the simplest one Ω_R is a fundamental domain for the rotation $z \rightarrow z + \alpha^2$ modulo $\mathbb{Z} + \mathbb{Z}\alpha$; our present fundamental domain Ω is the image by an affine map of Ω_R . So, up to this affine map, we can work on Ω_R and the latter translation.

The domain Ω_R is defined by drawing first the polygonal line $\vec{e}_{u_1} + \vec{e}_{u_2} + \dots + \vec{e}_{u_n} + \dots$ where $\vec{e}_i, 1 \leq i \leq 3$ is the canonical basis of \mathbb{R}^3 , and then projecting its

vertices on the contracting plane of M_τ parallel to the expanding direction $(1, \frac{1}{\beta}, \frac{1}{\beta^2})$ of M_τ . The atoms of the corresponding partition Ω_{Ri} are defined as the projection of those points $\vec{e}_{u_1} + \dots + \vec{e}_{u_n}$ for which $u_n = i$; the image of Ω_{Ri} by the translation is the projection of those points $\vec{e}_{u_1} + \dots + \vec{e}_{u_n}$ for which $u_{n+1} = i$. The translation vector on Ω_{Ri} is thus given by the projection of the canonical vector \vec{e}_i ; the lattice L for which Ω_R is a fundamental domain is generated by the projections of the vectors $\vec{e}_i - \vec{e}_j$, $i, j = 1, 2, 3$. In other words, we get a tiling of the plane by periodic translates according to the lattice L of Ω_R .

We thus approximate a line in \mathbb{R}^3 (the expanding direction of M_τ) by points with integer coordinates whose choice is determined by a combinatorial information, that is, the Tribonacci substitution. Observe that in dimension 2, the choice of such a discrete approximation of a line by points is canonical: we choose points with integer coordinates in such a way that we always follow the direction of the line. An important fact is that the polygonal line remains at a bounded distance from the expanding direction (this is a direct consequence of the fact that β is a Pisot number), and this means that the couple $(\frac{1}{\beta}, \frac{1}{\beta^2})$ is well approximated by the rational numbers $(\frac{T_n}{T_{n+1}}, \frac{T_{n-1}}{T_{n+1}})$; indeed, it is proved in [CHE-HUB-MES] (though this was probably already known to Rauzy) that this provides the best possible simultaneous approximation of $(\frac{1}{\beta}, \frac{1}{\beta^2})$ if we use the distance to the nearest integer defined by a particular norm, the so-called *Rauzy norm*; recall that if \mathbb{R}^d is endowed with the norm $\| \cdot \|$, and if $\theta \in \mathbb{T}^d$, then an integer $q \geq 1$ is a best approximation of θ if $\|q\theta\| < \|k\theta\|$ for all $1 \leq k \leq q-1$; the quality of the approximation is in $O(T_n^{-\frac{3}{2}})$, which is not better than could be expected in view of [CASSE]; the best possible constant

$$\inf\{c; q^{\frac{1}{2}}\|q\left(\frac{1}{\beta}, \frac{1}{\beta^2}\right)\| < c \text{ for infinitely many } q\}$$

is proved in [CHE-HUB-MES] to be equal to $(\beta^2 + 2\beta + 3)^{-\frac{1}{2}}$.

This result is connected with the following one quoted in [LAG-1, CHEV] and due to [ADAMS-1, ADAMS-2]: if $1, \alpha, \beta$ is a \mathbb{Q} -basis of a non-totally real cubic field, the best simultaneous approximations of (α, β) with respect to a given norm are a subset of a third-order linear recurrence with constant coefficients whose polynomial is given by the fundamental unit of $\mathbb{Q}(\alpha, \beta)$. See also [ITO-FUJ-HIG-YAS, ITO, CHEV] and [LAG-1] for closely related results for elements of cubic fields.

It was proved independently [ARN] that the fixed point of Tribonacci is also the *natural coding of a 6-interval exchange transformation on the circle*, see section 5.2 below for the definitions. Together with Rauzy's result, this gives a semi-topological conjugacy between two geometric systems, one living on a space of dimension 1 and the other living on a space of dimension 2; this provides also the only known nontrivial interval-exchange transformation (either on the circle, or on $[0, 1[$, by going to 7 intervals) which has continuous eigenfunctions.

3.3. Arnoux-Rauzy sequences. Following the success of the Tribonacci case, the Rauzy program began to be carried on to a category of infinite sequences generalizing the Sturmian ones and including the Tribonacci sequence.

DEFINITION 3.3. A sequence u is an *Arnoux-Rauzy sequence* if it is recurrent (that is, all its factors occur infinitely often), and, for all n , has complexity $2n + 1$ and only one left special and one right special factor of length n .

This definition implies that the sequence is on three letters, which we shall denote by $0, 1, 2$, and that the right (resp. left) special factors w have three extensions $w0, w1, w2$ (resp. $0w, 1w, 2w$).

Let us note that Arnoux-Rauzy sequences can be similarly defined over any alphabet of larger size, say d ; one thus obtains infinite words of complexity $(d - 1)n + 1$. Contrary to the Sturmian case, these words are not characterized by their complexity function any more. For instance, codings of non-degenerated three-interval exchanges have also complexity $2n + 1$ (see section 5.3).

One particular example in dimension d is the *d-Bonacci* substitution defined by $\tau(0) = 01, \tau(1) = 02, \dots, \tau(d - 1) = 0d, \tau(d) = 0$, for which the situation in section 3.2 can be generalized [SOL]. Note that the *infini-Bonacci* substitution is defined in [FER-2], but it is related to a dyadic translation.

The founding paper [ARN-RAU] proved that *each Arnoux-Rauzy sequence is a natural coding of a 6-interval exchange transformation on the circle*, and that their language can be generated similarly to section 2.4 above, with three elementary substitutions. Each Arnoux-Rauzy sequence defines also a natural algorithm of simultaneous approximation which we may call the *Arnoux-Rauzy algorithm*: it can be deduced from [ARN-RAU], see also [CAS-CHE, RIS-ZAM, ZAM]. This associates to each Arnoux-Rauzy sequence a couple of real numbers (α, β) , which are the frequencies of two letters in the sequence, and are used in defining the lengths of the exchanged intervals. The steps of this algorithm consist in defining three numbers $(\alpha_n, \beta_n, \gamma_n)$ such that one is larger than the sum of the other two, and then subtracting this sum from the largest one; this implies a first weakness of the Arnoux-Rauzy class: the set of all couples thus generated is a subset S of \mathbb{T}^2 of measure zero.

Still, to each couple in S we can associate an Arnoux-Rauzy sequence (where two of the symbols have frequencies given by this couple); so it was conjectured by Rauzy that *each Arnoux-Rauzy sequence is a natural coding of a translation*, whose angle is the associated $(\alpha, \beta) \in S$. In this direction, a version of the hat algorithm adapted for the Arnoux-Rauzy case is described in [WOZ-ZAM].

Rauzy's conjecture is true of course for the Tribonacci sequence, and this result was extended in [ARN-ITO] to all Arnoux-Rauzy sequences which are fixed points of substitutions. Indeed it is possible to associate a generalized Rauzy fractal to any *Pisot unimodular* substitution; see for instance [CAN-SIE-1, CAN-SIE-2, SIR-WAN] and the references in [PYT] and [LOT-2]. (A substitution is unimodular if the determinant of its incidence matrix equals ± 1 , and of Pisot type if the eigenvalues of its incidence matrix satisfy the following: there exists a dominant eigenvalue α such that for every other eigenvalue λ , one gets $\alpha > 1 > |\lambda| > 0$). More generally, for any unimodular morphism of Pisot type the measure-theoretical isomorphism with a translation on the torus (or equivalently the existence of a periodic tiling of the plane by the Rauzy fractal) is conjectured. A large literature is devoted to this question, which is surveyed in [PYT], Chapter 7. In particular, the study of the topological properties of the Rauzy fractal is mainly due to [RAU-3], where the Rauzy fractal is shown to be connected with simply connected interior;

see also [MES-1, MES-2, ITO-KIM] for a study of its fractal boundary. Similar compact sets with fractal boundary are also considered as geometrical representations of the β -shift when β is a Pisot unit [THU, AKI-1, AKI-2, PRA].

Moreover, it has been recently proved in [HUB-MES] that, in the case of Arnoux-Rauzy sequences which are fixed points of substitutions and under some additional conditions on the eigenvalues of the matrix of the substitution, the Arnoux-Rauzy algorithm still provides the best simultaneous approximation of (α, β) (this is not known yet for any other example of (α, β)).

However, Rauzy's conjecture was disproved in [CAS-FER-ZAM]:

THEOREM 3.4. *There exists an Arnoux-Rauzy sequence which is not a natural coding of a rotation on the two-dimensional torus \mathbb{T}^2 .*

This was done by building combinatorially an Arnoux-Rauzy sequence which is not uniformly balanced, and showing that, if the conjecture was true, the pieces of the fundamental domain would have to be bounded remainder sets, as discussed in section 3.1. Such a sequence is not a fixed point of a substitution and the partial quotients in its algorithm of simultaneous approximation are unbounded.

Hence the Arnoux-Rauzy sequences are not the good framework to carry on the Rauzy program, and, unfortunately, there has been no other attempts in finding natural codings of rotations on \mathbb{T}^2 . Nevertheless, Arnoux-Rauzy sequences generalize many combinatorial properties of Sturmian sequences: see for instance [CAST-MIG-RES], where they appear naturally in a generalization of the Fine and Wilf's theorem for three periods. The family of Arnoux-Rauzy words has been itself extended to the family of *episturmian* words [JUS-PIR-1]; see in particular [JUS-PIR-2] which considers a numeration system for the episturmian sequences playing the part that the Ostrowski one plays for Sturmian sequences.

4. The ugly: action of two rotations on \mathbb{T}^1

It is classical in Diophantine approximation to consider at the same time, via the transference principle, problems of simultaneous approximation and of minimization of linear forms, as discussed for instance in [CASSE].

The purpose of this section is to deal now with double rotation sequences, that is, with some particular codings of a free \mathbb{Z}^2 -action by two rotations. We thus consider the *double rotation* (R_α, R_β) defined on the torus \mathbb{T}^1 by the \mathbb{Z}^2 -action

$$(m, n) \cdot x = R_\alpha^m R_\beta^n(x) = x + m\alpha + n\beta \pmod{1};$$

throughout this section, 1, α , β are supposed to be rationally independent; the \mathbb{Z}^2 -action is thus free.

Given any partition of the torus and a point x we can here again define a (two-dimensional) sequence coding the orbit of x under this \mathbb{Z}^2 -action by $\mathcal{P}(x, (R_\alpha, R_\beta))$ by $\mathcal{P}(x, (R_\alpha, R_\beta))(x)_{(m,n)} = i$ whenever $R_\alpha^m R_\beta^n x \in P_i$, for $(m, n) \in \mathbb{Z}^2$; the sequence $\mathcal{P}(x, (R_\alpha, R_\beta))(x)$ is called a *coding* of the \mathbb{Z}^2 -action by the double rotation (R_α, R_β) .

All the notions and results presented in this section can naturally be extended to a higher dimension framework. We have chosen to restrict ourselves to the two-dimensional case for clarity issues.

4.1. Codings of \mathbb{Z}^2 -actions by rotations. We propose here two different possible natural codings, on a three-letter and on a two-letter alphabet, respectively. We shall see in section 4.2 that the two-letter one is of particular interest concerning the (rectangular) complexity function, whereas the second one has a natural geometric interpretation as a coding of a standard discrete plane (see section 4.4).

DEFINITION 4.1. A two-dimensional sequence $u = (u_{(m,n)})_{(m,n) \in \mathbb{Z}^2}$ is a *natural coding* on a three-letter (resp. two-letter) alphabet of the \mathbb{Z}^2 -action by the double rotation (R_α, R_β) if there exists x such that $u = \mathcal{P}(x, (R_\alpha, R_\beta))$ for $P = \{[0, 1 - \alpha[, [1 - \alpha, 0[\}$ or $P = \{[0, 1 - \alpha],]1 - \alpha, 0]\}$ (resp. $P = \{[0, \alpha[, [\alpha + \beta[, [\alpha + \beta, 1[\}$ or $P = \{[0, \alpha],]\alpha, \alpha + \beta],]\alpha + \beta, 1]\}$, by assuming furthermore $\alpha + \beta < 1$). We also call two-letter (resp. three-letter) natural codings *two-letter* (resp. *three-letter*) *two-dimensional Sturmian sequences*.

Let us observe that we lose here the symmetry between the parts played by α and β for two-letter natural codings. The lines of such a two-dimensional sequence are Sturmian sequences whereas the columns are binary codings of rotations, as described in section 5.1.

4.2. Combinatorial properties. Let us see in what respect the natural codings introduced in Definition 4.1 can be considered as reasonable two-dimensional generalizations of the Sturmian sequences.

Let us first define the two-dimensional analogues of the combinatorial notions introduced in section 2.1. We consider here finite rectangular arrays of consecutive

letters, that is, rectangular words $w = \begin{matrix} w_{0,n-1} & \cdots & w_{m-1,n-1} \\ \vdots & & \vdots \\ w_{0,0} & \cdots & w_{m-1,0} \end{matrix}$; we say here that

w has size (m, n) . The word w is said to occur at place (i, j) in the two-dimensional sequence u if $u_{i+k, j+l} = w_{k, l}$, for $0 \leq k \leq m - 1$ and $0 \leq l \leq n - 1$. The *rectangular complexity* of the two-dimensional sequence u is the function $p_u(m, n)$ which associates to each $(m, n) \in \mathbb{N}^2$, m and n being nonzero, the cardinality of the set $L_{(m,n)}(u)$ of rectangular factors of size (m, n) occurring in u .

The analogue of Lemma 2.7 also holds here: the word $w = \begin{matrix} w_{0,n} & \cdots & w_{m,n} \\ \vdots & & \vdots \\ w_{0,0} & \cdots & w_{m,0} \end{matrix}$

is a factor of the coding $\mathcal{P}(x, (R_\alpha, R_\beta))$ if and only if

$$(4.1) \quad \bigcap_{0 \leq i \leq m, 0 \leq j \leq n} R_\alpha^{-i} R_\beta^{-j} P_{w_{i,j}} \neq \emptyset.$$

We first deduce that, because of the minimality of the \mathbb{Z}^2 -action, for a given (α, β) , the language of rectangular factors of $\mathcal{P}(x, (R_\alpha, R_\beta))$ is here again the same for every x , provided that the atoms of the partition have nonzero measure. Then we can deduce results concerning the counting of rectangular factors of a given size (Proposition 4.1 below). We can not only deduce topological results from (4.1) but also metrical results such as the following: the frequencies of rectangular factors of size (m, n) of a two-dimensional Sturmian sequence take at most $\min(m, n) + 5$ values [BER-VUI-2].

PROPOSITION 4.1. [BER-VUI-2] *If $\max(\alpha, 1 - \alpha) \leq \max(\beta, 1 - \beta)$, then the rectangular complexity of a two-letter two-dimensional Sturmian sequence u*

satisfies:

$$\forall(m, n), p_u(m, n) = mn + n.$$

We assume $\alpha + \beta < 1$. The rectangular complexity of a three-letter two-dimensional Sturmian sequence u satisfies:

$$\forall(m, n), p_u(m, n) = mn + m + n.$$

In the two-letter case, if we do not require the hypothesis $\max(\alpha, 1 - \alpha) \leq \max(\beta, 1 - \beta)$, then we can only say that the complexity is equal to $mn + n$ for m or n large enough. In fact, according to (4.1), rectangular factors are in one-to-one correspondence with subsets of the torus of the form $R_\alpha^{-i} R_\beta^{-j} P_{w_{i,j}}$ which are finite intersections of intervals; when these subsets are themselves intervals, it is sufficient to count their endpoints to get their number; but they are not necessarily intervals; nevertheless, $R_\alpha^{-i} R_\beta^{-j} P_{w_{i,j}}$ is an interval as soon as one of the atoms of the partition has length smaller than or equal to $\max(\alpha, 1 - \alpha)$ and to $\max(\beta, 1 - \beta)$ (this is the case for a three-letter coding); indeed, if I and J are two intervals of \mathbb{T}^1 whose intersection is not connected, then the sum of their lengths is strictly larger than 1. For more details on these connectedness issues, see [ALE-BER, BER-VUI-2].

These sequences are easily seen to be *uniformly recurrent* (that is, for every n , there exists an integer N such that every square factor of size (N, N) contains every factor of size (n, n)) and *nonperiodic* (a two-dimensional sequence is *periodic* if it is invariant under translation, i.e., if it admits a nonzero vector of periodicity).

When one works with multidimensional sequences, fundamental problems in the definition of the objects appear; for instance, how can one define a multidimensional complexity function? A possible notion of complexity consists in counting the rectangular factors of given size; this is what we have chosen here so far. For a more general definition of complexity, see [SAN-TIJ-2]. A two-dimensional sequence is *periodic* if it is invariant under translation, i.e., if it admits a nonzero vector of periodicity. Note that the fact that the lattice of periodicity vectors has rank 2 is characterized by a bounded rectangle complexity function. Nevertheless, there is no characterization of periodic sequences by means of the complexity function: one can construct two-dimensional sequences with a nonzero periodicity vector and a very large complexity function; namely, consider the sequence $(u_{m+n})_{(m,n) \in \mathbb{Z}^2}$, where the unidimensional sequence $(u_n)_{n \in \mathbb{Z}}$ has maximal complexity. Conversely Nivat has conjectured the following:

CONJECTURE 4.1. If there exists (m_0, n_0) such that $P(m_0, n_0) \leq m_0 n_0$, then the two-dimensional sequence u is periodic.

The conjecture was proved for factors of size $(2, n)$ or $(n, 2)$ in [SAN-TIJ-3]. A more general conjecture is given in [SAN-TIJ-1, SAN-TIJ-3, SAN-TIJ-2]. But in higher dimension cases, counter-examples to the conjecture can be produced [SAN-TIJ-3]. A weakened version of the conjecture has been proved in [EPI-KOS-MIG, QUA-ZAM]: if there exists (m, n) such that $p_u(m, n) \leq \frac{mn}{12}$, then the sequence is periodic. See also the survey [CAS-2].

It is interesting, with respect to this conjecture, to consider the “limit” case of sequences of rectangle complexity function $mn + 1$. Such sequences are fully described in [CAS-1] and are all proved to be non-uniformly recurrent. The two-letter two-dimensional Sturmian sequences of complexity $mn + n$ are conjectured to be the uniformly recurrent sequences with smallest complexity function (it remains

to give a more precise meaning to the term “smallest” for parameters belonging to \mathbb{N}^2). Indeed no other example of uniformly recurrent nonperiodic sequence seems to be known.

4.3. Induction and generalized continued fractions. The question now is to try to extend the approach of section 2 using the interplay between a multi-dimensional version of the continued fraction algorithm and a notion of induction in order to understand the small scale structure of the double rotation. Several problems occur. First, there is no canonical notion of a generalization of Euclid’s algorithm. To remedy to the lack of a satisfactory tool replacing continued fractions, several approaches are possible: one can consider best simultaneous approximations as in [LAG-1, LAG-2] but we then lose unimodularity, and the sequence of best approximations heavily depends on the chosen norm; see also [WIN] for the use of filters. We chose here to focus on unimodular continued fractions such as Jacobi-Perron continued fractions, as illustrated in Theorem 4.3 below.

A second issue is to extend to \mathbb{Z}^2 the notion of first-return map (based on the order structure of \mathbb{Z}). We hence need to introduce a suitable notion of induction on a subinterval of \mathbb{T} . We thus propose to work with equivalence relations whose classes are the orbits of the \mathbb{Z}^2 -action; we consider the induced equivalence relation obtained by restriction to a subset.

Third, we strongly use the fact in the Sturmian case that the induced map of a rotation on an arbitrary subinterval of the circle is usually not a rotation but something similar, that is, an exchange of three intervals (see section 5). Let us recall that a necessary (but not sufficient condition) for a subinterval to lead to an induced map which is still an exchange of two intervals is that it is a bounded remainder set. Surprisingly enough, we will find here a suitable interval I on which the induction of the \mathbb{Z}^2 -action is again generated by a \mathbb{Z}^2 -action by a double rotation $(R_{\alpha', \beta'})$, that is, the induced equivalence relation on I coincides with the equivalence relation of the \mathbb{Z}^2 -action $(R_{\alpha', \beta'})$.

DEFINITION 4.2. Let I be a subset of \mathbb{T} . We define the *induced equivalence relation* on I as the restriction to the set I of the equivalence relation defined by the following: two points are equivalent if and only if they belong to the same orbit under the \mathbb{Z}^2 -action.

If I is the induction interval, we thus consider the set of (m, n) such that $R_\alpha^m R_\beta^n x \in I$; let us note that this subset is not a sublattice of \mathbb{Z}^2 .

We are particularly interested in the case where there exists a free \mathbb{Z}^2 -action by a double rotation on the set I such that both equivalence relations do coincide. In this case, we say that this new \mathbb{Z}^2 -action generates the induction of the initial action.

THEOREM 4.3. [ARN-BER-ITO] *Assume that $1, \alpha, \beta$ are rationally independent. Let $\gamma = k_0 + k_1\alpha + k_2\beta$ be a real number in $(0, 1)$ with k_0, k_1, k_2 relatively prime integers. Then, there exist α' and β' such that $1, \alpha', \beta'$ are rationally independent and such that the induction of the \mathbb{Z}^2 -action by (R_α, R_β) on any interval of length γ has the same equivalence classes as the \mathbb{Z}^2 -action by $(R_{\alpha'/\gamma}, R_{\beta'/\gamma})$.*

Unlike the Sturmian case, the generators of the \mathbb{Z}^2 -action are not canonically defined; there is a large choice for the induction procedure. Any unimodular continued fraction expansion can be used for that purpose.

We now can follow the scheme of section 2.4 and 2.5 in order to generate two-dimensional Sturmian sequences in a generalized S -adic way. It is indeed possible to generate three-letter alphabet two-dimensional Sturmian sequences by using two-dimensional substitutions governed by the Jacobi-Perron algorithm according to the induction process of Theorem 4.3 ([ARN-BER-ITO], see also [ARN-BER-SIE]).

4.4. Geometric interpretation. Let us consider now an application in discrete geometry of the dynamical systems introduced here, and of their symbolic codings. Sturmian sequences are known to be codings of digitizations of an irrational straight line, see for instance the survey [KLE-ROS]. One thus could expect from a generalization of Sturmian words that they correspond to a digitization of a hyperplane with irrational normal vector. This is indeed the case.

More precisely, let us thus consider the following digitization scheme corresponding to arithmetic planes introduced in [REV], see also [AND-ACHA-SIB]: given $\vec{v} \in \mathbb{R}^d$, $\mu, \omega \in \mathbb{R}$, the *lower* (resp. *upper*) discrete hyperplane $\mathfrak{P}(\vec{v}, \mu, \omega)$ is the set of points $\vec{x} \in \mathbb{Z}^d$ satisfying $0 \leq \langle \vec{x}, \vec{v} \rangle + \mu < \omega$ (resp. $0 < \langle \vec{x}, \vec{v} \rangle + \mu \leq \omega$). The parameter μ is called the *translation parameter* whereas ω is called the *thickness*. Moreover, if $\omega = \max\{v_i\} = |\vec{v}|_\infty$ (resp. $\omega = \sum v_i = |\vec{v}|_1$) then $\mathfrak{P}(\vec{v}, \mu, \omega)$ is said to be *naive* (resp. *standard*). Let us observe that the vector \vec{v} can belong either to \mathbb{R}^n or to \mathbb{Z}^n . In the first case it is reasonable to assume that $1, v_1, \dots, v_n$ are \mathbb{Q} -independent, whereas in the other case that $\gcd(v_1, \dots, v_n) = 1$.

In the standard case, one thus approximates a plane with irrational normal vector $\vec{v} \in \mathbb{R}^3$ by square faces oriented along the three coordinates planes; for each of the three kinds of faces, one defines a distinguished vertex; the standard discrete plane $\mathfrak{P}(\vec{v}, \mu, |\vec{v}|_1)$ is then equal to the set of distinguished vertices; after projection on the plane $x + y + z = 0$, along $(1, 1, 1)$, one obtains a tiling of the plane with three kinds of diamonds, namely the projections of the three possible faces. One can code this projection over \mathbb{Z}^2 by associating with each diamond the name of the projected face. These sequences are in fact three-letter two-dimensional Sturmian sequences (see e.g. [BER-VUI-1, BER-VUI-2, ARN-BER-ITO, ARN-BER-SIE]).

Similarly, a naive plane is well-known to be functional on a coordinate plane [REV], that is, for any point P of this coordinate plane, there exists a unique point in the naive plane obtained by adding to P a third coordinate. One can extend this notion of functionality to a larger family of arithmetic discrete planes, recovering a dynamical description of these planes in terms of codings of a \mathbb{Z}^2 -action, and apply it to the enumeration of some local configurations [BER-FIO-JAM].

More generally, many natural properties of discrete planes can be studied by using this dynamical approach such as their properties of self-similarity, periodicity, algorithmic generation, as illustrated in the survey [BRI-COE-KLE]. One of the main issues is the connectivity question, that is, the determination of the *connectivity number* defined as the infimum of those $\omega (= \omega(\vec{v}))$ for which the discrete plane $\mathfrak{P}(\vec{v}, 0, \omega)$ remains connected (with a notion of discrete connectivity to be defined beforehand). An explicit solution is provided in [BRI-BAR] for some vectors \vec{v} .

5. The good: interval exchanges

We consider here two natural generalizations of the Sturmian natural codings introduced in Definition 2.4. We first can still code a rotation but according to

any two-interval partition: this is what we call a *binary coding of a rotation* (section 5.1), and this is very close to considering the *natural coding of a 3-interval exchange transformation* (section 5.4). A second approach consists in modifying the map which acts on a natural fundamental domain of a rotation, that is, the two-interval exchange transformation; we thus consider d -interval exchange transformations (Definition 5.2) in the remaining sections.

5.1. Binary codings of rotations and discrepancy.

DEFINITION 5.1. A sequence $u = (u_n)_{n \in \mathbb{N}}$ is a *binary coding* of the rotation R_α if there exist β and x in \mathbb{T}^1 such that $u = \mathcal{P}(x, T)$ for $\mathcal{P} = \{[0, \beta[, [\beta, 1[$ or $\mathcal{P} = \{]0, \beta],]\beta, 1]\}$.

Lemma 2.7 still holds here but the main difference is that the sets $\cap_{i=0}^s T^{-i} P_{w_i}$ are not necessarily intervals (as for double rotations as seen in section 4.2): in the case where $\max(\beta, 1 - \beta) \leq \max(\alpha, 1 - \alpha)$, then they are connected; similarly, they are also connected for s large enough [ALE-BER]. Hence when $\beta \notin \alpha\mathbb{Z} + \mathbb{Z}$, the complexity for a binary coding is equal to $2n$ for n large enough [ROT]. Several continued fraction algorithms have been introduced to study these sequences [ADA-1, ADA-2, DID] that allows one to generate them in an S -adic way; for a list a references, see [BER] and Chapter 12 in [PYT]. One can also generate them via the hat algorithm; since this generation process is very close to the one detailed below in section 5.2 for three-interval exchanges, we do not give it explicitly here. Indeed, one key feature of the binary codings of rotations is that they are connected to interval exchange transformations via the induction (see section 5.4 below and [BER-CHE-FER, ADA-1] for more details): the induced map on the interval $[0, \beta]$ of a rotation R_α is a three-interval exchange, and conversely any three-interval exchange can be seen as an exduction of a rotation; see also [GUI-MAS-PEL] for a realization of three-interval exchanges as cut-and-project schemes.

One motivation to their study is that binary codings for rotations provide a good combinatorial insight on the local star discrepancy for Kronecker sequences. The *local star discrepancy* for the sequence $(n\alpha)_{n \in \mathbb{N}}$ is defined as

$$\Delta_N^*(\alpha, \beta) = \left| \sum_{n=0}^{N-1} \chi_{[0, \beta[}(\{n\alpha\}) - N\beta \right|,$$

whereas the *star discrepancy* is defined as

$$D_N^*(\alpha) = \sup_{0 < \beta < 1} \Delta_N^*(\alpha, \beta).$$

Most of the discrepancy results concerning Kronecker sequences were obtained by using the Ostrowski numeration system (see [KUI-NIE] and [DRM-TIC]). A similar approach has been developed in [ADA-2], inspired by [RAU-5, RAU-6], based on the S -adic expansion of binary codings obtained via the Rauzy induction for three interval exchanges [ADA-1]. Indeed, an algorithm is proposed in [ADA-2], which computes $\limsup \frac{\Delta_n^*(\alpha, \beta)}{\log n}$ when α is a quadratic number and $\beta \in \mathbb{Q}(\alpha)$; furthermore, examples are produced such that

$$\limsup \frac{\Delta_n^*(\alpha, \beta)}{\log n} = \limsup \frac{D_n^*(\alpha)}{\log n}$$

which leads to the following question asked in [ADA-2]: if α is a quadratic number, does there exist $\beta \in \mathbb{Q}(\alpha)$ and $\beta \notin \alpha\mathbb{Z} + \mathbb{Z}$, such that $\limsup \frac{\Delta_n^*(\alpha, \beta)}{\log n} =$

$\limsup \frac{D_n^*(\alpha)}{\log n}$; and if the answer is positive, how to determine explicitly such a β with respect to α ?

5.2. Inductions for interval exchanges. As we have seen in section 2.2, a rotation of \mathbb{T}^1 may be seen, on the fundamental domain $[0, 1[$, as an exchange of two intervals; hence a natural generalization is the notion of d -interval exchange introduced in [OSE].

DEFINITION 5.2. A d -interval exchange transformation is given by a probability vector $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_d)$ together with a permutation π of $\{1, 2, \dots, d\}$. The map T is the piecewise translation defined by partitioning the interval $[0, \lambda_1 + \dots + \lambda_d[$ into d sub-intervals of lengths $\lambda_1, \lambda_2, \dots, \lambda_d$ and rearranging them according to the permutation π ; or, formally,

$$X_i = \left[\sum_{j < i} \lambda_j, \sum_{j \leq i} \lambda_j \right],$$

$$Tx = x + \sum_{\pi^{-1}j < \pi^{-1}i} \lambda_j - \sum_{j < i} \lambda_j$$

on X_i .

In this section, all interval exchange transformations are assumed to have the *i.d.o.c. property* [KEA]: the orbit of the discontinuities are infinite and distinct. We suppose also that the permutation π is *irreducible*: $\pi\{1, \dots, k\} \neq \{1, \dots, k\}$ for every $k < d$; then the i.d.o.c. property is implied by the rational independence of the λ_i , and is true for (Lebesgue)-almost every vector λ .

The *natural coding* of a d -interval exchange transformation is just the coding by the partition into intervals of continuity X_i , $1 \leq i \leq d$.

The basic induction operation is the *Rauzy induction* [RAU-1, RAU-2]: it is described in a new and elegant form in [YOC] or [MAR-MOU-YOC], and we shall just sum up this paper and encourage the reader to consult it.

In this paper, following an idea of [KER], they define interval exchange maps by a vector λ in \mathbb{R}^d and two permutations, π_0 and π_1 , and T is the piecewise translation defined by partitioning the interval $[0, \lambda_1 + \dots + \lambda_k[$ into k sub-intervals of lengths $\lambda_1, \lambda_2, \dots, \lambda_k$, ordered according to the permutation π_0^{-1} and rearranging them according to the permutation π_1^{-1} . Then the *Rauzy induction* is made on the interval $[0, \lambda_1 + \dots + \lambda_d - \lambda_{a_1-e}[$ where $a_0 = \pi_0^{-1}d$, $a_1 = \pi_1^{-1}d$ and $e = 0, 1$ is defined by $\lambda_{a_1-e} = \max(\lambda_{a_0}, \lambda_{a_1})$. The induced transformation is again a d -interval exchange map, associated to the vector $\lambda' = \mathcal{I}(\lambda)$, where $\lambda = V\lambda'$ for a nonnegative matrix V , and two permutations denoted by $(\pi'_0, \pi'_1) = R_e(\pi_0, \pi_1)$. Thanks to the i.d.o.c. condition, we can always iterate this operation, building vectors $\lambda^{(n)}$, permutations $\pi_0^{(n)}, \pi_1^{(n)}$, matrices V_n and numbers $e_n = 0, 1$. This generalizes the *additive* algorithm of continued fraction approximation, which is the one in section 2.4, but decomposed into elementary steps (instead of inducing on $[-\alpha, 1 - (a_1 + 1)\alpha[$ we could have induced successively on each $[-\alpha, 1 - (i + 1)\alpha[$, $1 \leq i \leq a_1$). As in section 2.4, we can use it to build natural codings of points, we can also use the fact that $\lambda = V_1 \dots V_n \lambda^{(n)}$ to give a simultaneous approximation of the $d - 1$ -uple $(\frac{\lambda_i}{\lambda_1 + \dots + \lambda_d}, 1 \leq i \leq d - 1)$ by rational numbers. However, up to our knowledge, the Rauzy induction has never been used concretely for these purposes.

But the main interest of the Rauzy induction appears in [VEE-1]: the transformation \mathcal{I} on the simplex preserves a positive infinite measure; the dynamics of this transformation, linked to the *Teichmüller flow* on a suitable surface, allowed Veech [VEE-1] and Masur [MAS] to prove the unique ergodicity of almost all (for the Lebesgue measure on the simplex, d and the permutation(s) being fixed) d -interval exchange transformations, and further measure-theoretic properties [VEE-3, VEE-4, VEE-5].

A *multiplicative* form of the Rauzy induction was given by Zorich [ZOR]: instead of looking at all steps $\lambda^{(n)}, \pi_0^{(n)}, \pi_1^{(n)}$, he looks only at the subsequence obtained by grouping together all the consecutive steps corresponding to a constant value of e_n . This induction transformation has a finite invariant measure, and this was used to prove deeper results, culminating in the *measure-theoretic weak mixing* of almost all d -interval exchange transformations [AVI-FOR]. A further acceleration of Zorich induction is given in [MAR-MOU-YOC] by grouping all the steps corresponding to sets of $d - 1$ values of a_{e_n} ; it is then used to extend to interval exchange maps some important properties of rotations concerning the smoothing of cocycles.

Another, not so well known, induction, is the *da Rocha induction* [daR, LOP-daR].

We build a family of intervals $[\beta_i - L_i^{(k)}, \beta_i + R_i^{(k)}]$, $i = 1, \dots, d - 1$, where the β_i are the discontinuities of T^{-1} , and the endpoints of these intervals are the iterates $T^{-n}0$ which are closer to the β_i than any previous iterate $T^{-k}0$, $0 \leq k < n$. The intervals at stage $k + 1$ are built by inducing on the union of the $d - 1$ intervals at stage k : we get an exchange on d intervals, this adds to the picture one new point near one β_i , and we cut the i -th interval so that this point becomes an extremity. The induction operation, acting on the vectors $L_1^{(k)}, \dots, L_{d-1}^{(k)}, R_1^{(k)}, \dots, R_{d-1}^{(k)}$, $k \geq 0$, (linked by some relations), preserves an explicit infinite measure, and this can be used also to prove the unique ergodicity of almost all d -interval exchange transformations, as well as to build the natural codings.

5.3. The hat algorithm. Given a d -interval exchange transformation satisfying the i.d.o.c. condition, we want to study the combinatorial properties of its natural coding. It is well known that the complexity is $p(n) = (d - 1)n + 1$, and that for all n large enough there are $d - 1$ left special factors and $d - 1$ right special factors of length n . Let us note that, since there might be several right special factors of a given length, such sequences are not Arnoux-Rauzy sequences. As in section 2.5, we want to compute the bispecial factors, by building them explicitly, together with the intervals corresponding to them through Lemma 2.7.

This has been done in [FER-HOL-ZAM-1] [FER-HOL-ZAM-2] for $d = 3$ and the permutation $(3, 2, 1)$; we explain it here in a slightly different way in view of its generalization.

5.3.1. Description for $d = 3$. We normalize our parameters ($\lambda_1 + \lambda_2 + \lambda_3 = 1$) and restrict them to the region of the simplex defined by $\lambda_1 < \frac{1}{2}$, $2\lambda_1 + \lambda_2 > 1$; this amounts to choosing the words of length two to be 12, 13, 21, 22, 31; when this is not the case, we relate our set of parameters, by elementary maps, to another set in the privileged region, and this allows us to deduce the bispecial factors in the

general case by using substitutions on the words. The i.d.o.c. condition is satisfied by forbidding some rational relations between λ_1 and λ_2 .

We build inductively pairs of words $(W_1^{(k)}, W_2^{(k)})$, where $W_i^{(k)}$ begins by the letter i and corresponds to the interval $[\beta_i - L_i^{(k)}, \beta_i + R_i^{(k)}]$, where $\beta_1 = \lambda_3$ and $\beta_2 = \lambda_3 + \lambda_2$ are the discontinuity points of T^{-1} . For each k , $W_i^{(k)}$ is a prefix of $W_i^{(k+1)}$, and a strict prefix for at least one $i = 1, 2$, and no strict prefix of $W_i^{(k+1)}$ can be bispecial if it contains $W_i^{(k)}$ as a strict prefix. Note that the sequence of words $W_i^{(k)}$, when $k \rightarrow +\infty$, converges to an infinite sequence which is the natural coding of T for the point β_i . And as in section 2.5, we build first a hatted pair $(W_i')^{(k)}$ and get the $W_i^{(k)}$ by removing the hats; but this time there are two kinds of hats, upper and lower.

Hence we describe the evolution of six objects, $W_1', W_2', L_1, L_2, R_1, R_2$; this evolution goes through three states called A, B, C . In state A we have $\bar{W}_i = W_i$ for $i = 1, 2$ (both unhatted words are palindromes) and $R_1 = R_2$; in state B for $i = 1, 2$ $\bar{W}_i = W_{3-i}$ and $L_1 + R_1 = L_2 + R_2$; in state C the unhatted words are palindromes and $L_1 = L_2$. In each state, the i.d.o.c. condition will ensure the inequalities we need to continue:

- At the beginning $W_1' = \underset{\wedge}{1}$, $W_2' = \hat{\underset{\wedge}{2}}$, $L_1 = \lambda_3$, $R_1 = \lambda_1 - \lambda_3 = R_2$, $L_2 = \lambda_2 + \lambda_3 - \lambda_1$, and we are in state A .
- If we are in state A
 - if $L_1 > R_1$, replace L_1 by $L_1 - R_1$ and extend (on the right) W_1' by the shortest suffix of W_1' beginning by $\hat{\underset{\wedge}{3}}$, removing all hats after this $\hat{\underset{\wedge}{3}}$, (if there is no such suffix in W_1' , which might happen at the beginning, use the shortest such suffix of $\hat{\underset{\wedge}{3}}W_1'$), leave all other objects unchanged and go to state A again.
 - If $L_2 > R_2$, replace L_2 by $L_2 - R_2$ and extend (on the right) W_2' by the shortest suffix of W_2' beginning by $\hat{\underset{\wedge}{2}}$, removing all hats after this $\hat{\underset{\wedge}{2}}$ (the initial $\hat{\underset{\wedge}{2}}$ is used either as a $\underset{\wedge}{2}$ or a $\hat{\underset{\wedge}{2}}$, removing the unused hat), leave all other objects unchanged and go to state A again.
 - The two operations above are independent and can be made, as many times as necessary, in any order; when these cannot be done any more,
 - if $L_1 < R_1$ and $L_2 < R_2$, replace R_1 by $R_1 - L_1$ and R_2 by $R_2 - L_2$; extend W_1' by the shortest suffix of W_2' beginning by $\underset{\wedge}{2}$, removing all hats after this one; extend W_2' by the shortest suffix of W_1' beginning by $\underset{\wedge}{1}$, removing all hats after this one; leave all other objects unchanged and go to state B .
- If we are in state B
 - if $R_1 > L_2$, then also $R_2 > L_1$; replace R_1 by $R_1 - L_2$ and R_2 by $R_2 - L_1$; extend W_1' by the shortest suffix of W_2' beginning by $\underset{\wedge}{1}$, removing all hats after this one; extend W_2' by the shortest suffix of W_1' beginning by $\underset{\wedge}{2}$, removing all hats after this one; leave all other objects unchanged and go to state A .
 - If $L_1 > R_2$, then also $L_2 > R_1$; replace L_1 by $L_1 - R_2$ and L_2 by $L_2 - R_1$; extend W_1' by the shortest suffix of W_2' beginning by $\underset{\wedge}{2}$,

removing all hats after this one; extend W'_2 by the shortest suffix of W'_1 beginning by $\hat{3}$, removing all hats after this one; leave all other objects unchanged and go to state C .

- State C is the dual state of state A ; if some $R > L$, we replace it by $R - L$, extend the corresponding W' by a lower hat suffix from itself, and stay in state C ; when we have both $L > R$, we replace them by $R - L$, extend each W'_i by an upper hat suffix from W'_{3-i} , and go to state B .

The evolution between states A, B, C can be described by a graph G which is very similar to the *graph of the Rauzy induction*, see figure 1 in [MAR-MOU-YOC] (only the loops are duplicated and the initial state is not the same). A multiplicative form of this process is described in [FER-HOL-ZAM-1] [FER-HOL-ZAM-2], by grouping together all the steps between two consecutive states B : the arithmetic objects considered in [FER-HOL-ZAM-1] are $\delta_k, \alpha'_k, \beta'_k$, where, at stage k , $|\delta_k|$ is the common $L + R$, and α'_k, β'_k are, according to k , either L_1, L_2 or R_1, R_2 .

5.3.2. Generalization. The additive algorithm can be generalized to d -interval exchange transformations (this is easier when the permutation is $(d, d - 1, \dots, 1)$ [FER-ZAM] but possible also in the other cases), with the parameters restricted to a suitable region (the general case being deduced by the same trick as for $d = 3$). We build $(d - 1)$ -uples of bispecial words, and control $2d - 2$ lengths L_i and R_i linked by $d - 2$ relations; the rules of construction (and the relations between the lengths) follow a path in a graph which has 9 vertices for $d = 4$, while the graph of the Rauzy induction has 7 vertices (though some similarities may not be accidental); then for $d = 5$ we have 28 vertices, and so on.

This process is a variant of the da Rocha induction: if the α_i are the discontinuities of T , the construction of bispecial factors amounts, through Lemma 2.7, to the construction of a family of intervals $[\beta_i - L_i^{(k)}, \beta_i + R_i^{(k)}]$, $i = 1, \dots, d - 1$, where the β_i are the discontinuities of T^{-1} , and the endpoints of these intervals are the iterates $T^{-n}\alpha_i$ which are closer to the β_i than any iterate $T^{-k}\alpha_i$, considered previously. In this process, like in da Rocha's, the intervals are built by induction on a non-connected union of intervals; and, while the da Rocha induction can be considered, from a geometric point of view, as dual to the Rauzy induction, this new induction might be its own dual.

5.4. The case of three intervals. For $d = 3$, there are five possible permutations π different from identity, but with (213), (132), (231) or (312), T becomes just a two-interval exchange, which is a rotation. Hence the study reduces to *symmetric* 3-interval exchange transformations, those using the permutation (321):

$$Tx = \begin{cases} x + 1 - \alpha & \text{if } x \in [0, \alpha[\\ x + 1 - 2\alpha - \beta & \text{if } x \in [\alpha, \alpha + \beta[\\ x - \alpha - \beta & \text{if } x \in [\alpha + \beta, 1[. \end{cases}$$

And T satisfies the i.d.o.c. condition if and only if we have the three inequalities, for every integer p and q

- $p\alpha + q\beta \neq p - q$,
- $p\alpha + q\beta \neq p - q - 1$,
- $p\alpha + q\beta \neq p - q + 1$.

5.4.1. *Induction.* But even in this case, rotations are not very far, as T defined above is just, after normalization, the induced map on the interval $[0, \beta' = \frac{1}{1+\beta}[$ of the rotation of angle $\alpha' = \frac{1-\alpha}{1+\beta}$. Note that, through the theory of *Kakutani equivalence*, there are many transformations (for example, the time-one of the *horocycle flow*) which are induced maps of rotations; but here, the induction set is an interval, and this is enough to transmit to T many properties of the rotations, and first of all the unique ergodicity, which is always satisfied under the i.d.o.c. condition (contrarily to what happens if $d > 3$).

Similarly, starting from T , we can get a rotation by induction: the first step of the Rauzy induction leads to a 3-interval exchange with a permutation different from (321), that is a rotation.

And the study of the natural coding reduces, through substitutions as in section 2.4.2, to the study of the coding of a rotation $R_{\alpha'}$ by the partition of the circle by the points $0, 1 - \alpha'$, and β' ; the study of this coding is completely similar to the study of the binary coding in section 5.1.

5.4.2. *Structure theorem.* But, thanks to the hat algorithm, we have a full equivalent of Theorem 2.6:

THEOREM 5.3. [FER-HOL-ZAM-2] *A uniformly recurrent infinite word is a natural coding of a three-interval exchange transformation satisfying the i.d.o.c. condition with $2\alpha < 1$ and $2\alpha + \beta > 1$ if and only if it satisfies the following four conditions:*

- *the factors of length 2 of u are $\{12, 13, 21, 22, 31\}$,*
- *if w is a factor of u , its retrograde \bar{w} is also a factor of u ,*
- *for every n there are exactly two left special factors of length n , one beginning in 1 and one beginning in 2,*
- *if w is a bispecial factor ending in 1 and $w \neq \bar{w}$ then $w2$ is left special if and only if $\bar{w}1$ is left special.*

For the α and β outside the privileged triangle, the same trick as in section 5.2 applies. The proof follows closely the steps of section 2.5; the “only if” direction follows from geometric considerations, while in the “if” direction,

- we start from a sequence u satisfying the four conditions in Theorem 5.3, it determines an admissible path in the graph G of the hat algorithm (section 5.3), which can be conveniently described by a sequence $(n_k, m_k, \varepsilon_{k+1})$, $k \in \mathbb{N}$ for integers $n_k > 0$, $m_k > 0$, $\varepsilon_{k+1} = \mp 1$, satisfying a few conditions to eliminate improper expansions;
- with the hat algorithm, we build the bispecial words, and hence the two infinite sequences \mathcal{O}_1 (beginning with 1) and \mathcal{O}_2 (beginning with 2) for which all prefixes are left special;
- any admissible path in G defines a unique couple (α, β) in the privileged triangle, through an explicit algorithm of simultaneous approximation using the sequence $(n_k, m_k, \varepsilon_{k+1})$, the convergence and the i.d.o.c. condition being ensured by the elimination of improper expansions;
- the natural coding of the 3-interval exchange transformation T defined by the couple (α, β) has the same two sequences \mathcal{O}_i with left special prefixes as our original u ; hence u has the same language as (all) the codings

$\mathcal{P}(x, T)$, and, by a standard approximation argument, u itself is some $\mathcal{P}(x, T)$.

Thus the interaction between Sturmian sequences and rotations is completely reproduced here: we have a complete combinatorial characterization of the natural codings, and this comes through a new algorithm of simultaneous approximation. However, this algorithm is far from giving the best approximation of (α, β) : it is a variant of the Euclid algorithm in the direction α' (of the inducing rotation), but the quality is not better than $O(q_n^{-1})$ globally; its main interest is that it satisfies a form of *Lagrange's theorem*: ultimately periodic expansions correspond to quadratic lengths and to *substitutive* natural codings [**FER-HOL-ZAM-1**, **FER-HOL-ZAM-2**]. But the main consequences of this theorem come from the description of trajectories, which allows, by computing exactly the *return words* of one family of bispecial intervals, to give a complete description of the system as an S-adic system, or, equivalently, by *Rokhlin stacks*. This description in turn gives new results on the measure-theoretic properties of T : first, the computation of the *eigenvalues* of the associated spectral operator is done in [**FER-HOL-ZAM-3**], thereby answering two questions posed by Veech in [**VEE-1**]: a nontrivial 3-interval exchange transformation can be measure-theoretically isomorphic to a rotation (not related to the inducing rotation); then the *joinings* of 3-interval exchange transformations are studied in [**FER-HOL-ZAM-4**].

6. Concluding remarks

This field is in constant evolution, and we shall just make a few remarks on the open questions at the time of writing; for more information, the reader can consult for example Chapter 12 of [**PYT**] and the references in it.

Though even the Sturmian sequences still give rise to interesting questions, see sections 2.6 and 2.7, it is not surprising that the “bad” part is very rich in unsolved problems, but these seem to be quite difficult; the first question is to find good natural codings of rotations on \mathbb{T}^2 , and we do not even know what kind of atoms are bounded remainder sets, or which one give the lowest complexity. The Rauzy fractals are linked to the general problem of geometric representation of substitution dynamical systems. As for the Arnoux-Rauzy category of sequences, it is still worth studying, to find to which dynamical systems it can correspond, either through some notion of natural coding or through the weaker notion of measure-theoretic isomorphism.

The study of the “ugly” part started only recently and, even through a cursory glance, the reader will find it full of open problems; the Nivat conjecture and its variants form maybe the most tantalizing.

The domain of interval exchange transformation has given birth to a wealth of difficult conjectures, which have in most cases been brilliantly proved; the last survivor in this illustrious family is the (measure-theoretic) question of Veech on the *simplicity* of almost all interval exchange transformations [**VEE-2**] (see also [**FER-HOL-ZAM-4**]).

Some other related attempts to give a symbolic representation of geometric structures (that is, semi-conjugacies between geometric and symbolic dynamical systems) such as hyperbolic automorphisms of the torus [**KEN-VER**] in terms of symbolic dynamics and radix representations have been made: see for instance the survey [**SID**] devoted to arithmetics dynamics which aims at giving explicit

arithmetic expansions of real numbers with a dynamical meaning; the systems that are usually studied in this framework are translations of the torus, as considered in the present survey, automorphisms of the torus, or β -shifts.

References

- [ADA-1] B. Adamczewski, *Codages de rotations et phénomènes d'autosimilarité*, J. Théor. Nombres Bordeaux **14** (2002), 351–386.
- [ADA-2] B. Adamczewski, *Répartition des suites $(n\alpha)_{n \in \mathbb{N}}$ et substitutions*, Acta Arith. **112** (2004), 1–22.
- [ADA-3] B. Adamczewski, *Symbolic discrepancy and self-similar dynamics*, Ann. Inst. Fourier (Grenoble) **54** (2004), 1001–1034.
- [ADA-BUG] B. Adamczewski, Y. Bugeaud, *On the complexity of algebraic numbers*, preprint (2004).
- [ADA-BUG-LUC] B. Adamczewski, Y. Bugeaud, F. Luca, *Sur la complexité des nombres algébriques*, C. R. Math. Acad. Sci. Paris **339** (2004), 11–14.
- [ADAMS-1] W.W. Adams, *Simultaneous Diophantine approximations and cubic irrationals*, Pacific J. Math. **30** (1969), 1–14.
- [ADAMS-2] W.W. Adams, *Simultaneous Diophantine approximations to basis of real cubic field*, J. Number Theory **1** (1971), 179–194.
- [ADAMS-DAV] W. W. Adams, J. L. Davison, *A remarkable class of continued fractions*, Proc. Amer. Math. Soc. **65** (1977), 194–198.
- [AKI-1] S. Akiyama, *Self affine tiling and Pisot numeration system*, in *Number theory and its applications* (Kyoto, 1997), Dev. Math. **2** (1999), 7–17, Kluwer Acad. Publ., Dordrecht.
- [AKI-2] S. Akiyama, *On the boundary of self affine tilings generated by Pisot numbers*, J. Math. Soc. Japan **54** (2002), 283–308.
- [ALE-BER] P. Alessandri, V. Berthé, *Three distance theorems and combinatorics on words*, Enseign. Math. **44** (1998), 103–132.
- [ALL] J.-P. Allouche, *Nouveaux résultats de transcendance de réels à développement non aléatoire*, Gaz. Math. **84** (2000), 19–34.
- [ALL-DAV-QUE-ZAM] J.-P. Allouche, J. L. Davison, M. Queffélec, L. Q. Zamboni, *Transcendence of Sturmian or morphic continued fractions*, J. Number Theory **91** (2001), 39–66.
- [AND-ACHA-SIB] E. Andres, R. Acharya, C. Sibata, *Discrete Analytical Hyperplanes*, Graphical Models and Image Processing **59** (1997), 302–309.
- [ARNOL] V.I. Arnold, *Small denominators and problems of stability of motion in classical and celestial mechanics*, Usp. Math. Nauk. **18** (1963), 91–192 (in Russian), translated in Russian Math. Surveys **18** (1963), 86–194.
- [ARN] P. Arnoux, *Un exemple de semi-conjugaison entre un échange d'intervalles et une translation sur le tore*, Bull. Soc. Math. France **116** (1988), 489–500.
- [ARN-BER-ITO] P. Arnoux, V. Berthé, S. Ito, *Discrete planes, \mathbf{Z}^2 -actions, Jacobi-Perron algorithm and substitutions*, Ann. Inst. Fourier (Grenoble) **52** (2002), 1001–1045.
- [ARN-BER-SIE] P. Arnoux, V. Berthé, A. Siegel, *Two-dimensional iterated morphisms and discrete planes*, Theoret. Comput. Sci. **319** (2004), 145–176.
- [ARN-FER-HUB] P. Arnoux, S. Ferenczi, P. Hubert, *Trajectories of rotations*, Acta Arith. **87** (1999), 209–217.
- [ARN-FIS] P. Arnoux, A.M. Fisher, *The scenery flow for geometric structures on the torus: the linear setting*, Chinese Ann. Math. Ser. B **22** (2001), 427–470.
- [ARN-ITO] P. Arnoux, S. Ito, *Pisot substitutions and Rauzy fractals*, Bull. Belg. Math. Soc. Simon Stevin **8** (2001), 181–207.
- [ARN-RAU] P. Arnoux, G. Rauzy, *Représentation géométrique de suites de complexité $2n + 1$* , Bull. Soc. Math. France **119** (1991), 199–215.
- [AVI-FOR] A. Avila, G. Forni, *Weak mixing for interval exchange maps and translation flows*, preprint (2004), arXiv:math.DS/0406326.
- [BAR-LIA] G. Barat, P. Liardet, *Dynamical systems originated in the Ostrowski alpha-expansion*, Publ. Math. Debrecen, to appear.
- [BER] V. Berthé, *Autour du système de numération d'Ostrowski*, Bull. Belg. Math. Soc. Simon Stevin **8** (2001), 209–239.

- [BER-CHE-FER] V. Berthé, N. Chekhova, S. Ferenczi, *Covering numbers: arithmetics and dynamics for rotations and interval exchanges*, J. Anal. Math. **79** (1999), 1–31.
- [BER-FIO-JAM] V. Berthé, C. Fiorio, D. Jamet, *Generalized functionality for arithmetic discrete planes*, DGCI'05, Lecture Notes in Computer Sci., Springer, to appear.
- [BER-HOL-ZAM] V. Berthé, C. Holton, L. Q. Zamboni, *Initial powers of Sturmian sequences*, Acta Arith., to appear.
- [BER-VUI-1] V. Berthé, L. Vuillon, *Suites doubles de basse complexité*, J. Théor. Nombres Bordeaux **12** (2000), 179–208.
- [BER-VUI-2] V. Berthé, L. Vuillon, *Tilings and rotations on the torus: a two-dimensional generalization of Sturmian sequences*, Discrete Math. **223** (2000), 27–53.
- [BLO-THE-VLA] V. D. Blondel, J. Theys, A. A. Vladimirov, *An elementary counterexample to the finiteness conjecture*, Siam J. Matrix Anal. Appl. **24** (2003), 963–970.
- [BOR] É. Borel, *Sur les chiffres décimaux de $\sqrt{2}$ et divers problèmes de probabilités en chaîne*, C. R. Acad. Sci. Paris **230** (1950), 591–593.
- [BOW] R. Bowen, *Markov partitions are not smooth*, Proc. Amer. Math. Soc. **71** (1978), 130–132.
- [BRI-BAR] V. E. Brimkov, R. P. Barneva, *Connectivity of discrete planes*, Theoret. Comput. Sci. **319** (2004), 203–227.
- [BRI-COE-KLE] V. E. Brimkov, D. Coeurjolly, R. Klette, *Digital planarity-A review*, LIRIS laboratory, research report (2004).
- [BRO-1] T. C. Brown, *Descriptions of the characteristic sequence of an irrational*, Canad. Math. Bull. **34** (1991), 36–41.
- [BRO-2] T. C. Brown, *Applications of standard Sturmian words to elementary number theory*, Theoret. Comput. Sci. **273** (2002), 5–9.
- [BUG-LAU] Y. Bugeaud, M. Laurent, *Exponents of Diophantine approximation and Sturmian continued fractions*, Ann. Inst. Fourier (Grenoble), to appear.
- [CAN-SIE-1] V. Canterini, A. Siegel, *Automate des préfixes-suffixes associé à une substitution primitive*, J. Théor. Nombres Bordeaux **13** (2001), 353–369.
- [CAN-SIE-2] V. Canterini, A. Siegel, *Geometric representation of substitutions of Pisot type*, Trans. Amer. Math. Soc. **353** (2001), 5121–5144.
- [CAS-1] J. Cassaigne, *Two dimensional sequences with complexity $mn + 1$* , J. Auto. Lang. Comb. **4** (1999), 153–170.
- [CAS-2] J. Cassaigne, *Subword complexity and periodicity in two or more dimensions*, in *Developments in Language Theory. Foundations, Applications and Perspectives (DLT'99)*, Aachen (Allemagne), World Scientific, (2000), 14–21.
- [CAS-CHE] J. Cassaigne, N. Chekhova, *Fonctions de récurrence des suites d'Arnoux-Rauzy et réponse à une question de Morse et Hedlund*, preprint (2004).
- [CAS-FER-ZAM] J. Cassaigne, S. Ferenczi, L. Q. Zamboni, *Imbalances in Arnoux-Rauzy sequences*, Ann. Inst. Fourier (Grenoble) **50** (2000), 1265–1276.
- [CASSE] J. W. S. Cassels, *An introduction to diophantine approximation*, Cambridge Tracts in Mathematics and Mathematical Physics **45** (1972), Cambridge University Press.
- [CAST-MIG-RES] M. G. Castelli, F. Mignosi, and A. Restivo, *Fine and Wilf's theorem for three periods and a generalization of Sturmian words*, Theoret. Comput. Sci. **218** (1999), 83–94.
- [CHE-HUB-MES] N. Chekhova, P. Hubert, A. Messaoudi, *Propriétés combinatoires, ergodiques et arithmétiques de la substitution de Tribonacci*, J. Théor. Nombres Bordeaux **13** (2001), 371–394.
- [CHEV] N. Chevallier, *Best simultaneous diophantine approximations of some cubic algebraic numbers*, J. Théor. Nombres Bordeaux **14** (2002), 403–414.
- [COV-HED] E.M. Coven, G.A. Hedlund, *Sequences with minimal block growth*, Math. Systems Theory **7** (1972), 138–153.
- [DAV-SCH] H. Davenport, W. Schmidt, *Approximation to real numbers by algebraic integers*, Acta Arith. **15** (1969), 393–416.
- [DID] G. Didier, *Codages de rotations et fractions continues*, J. Number Theory **71** (1998), 275–306.
- [DRM-TIC] M. Drmota, R. F. Tichy, *Sequences, discrepancies and applications*, Springer, Berlin, 1997.
- [EPI-KOS-MIG] C. Epifanio, M. Koskas, F. Mignosi, *On a conjecture on bidimensional words*, Theoret. Comput. Sci. **299** (2003), 123–150.
- [FER-1] S. Ferenczi, *Bounded remainder sets*, Acta Arith. **61** (1992), 319–326.

- [FER-2] S. Ferenczi, *Substitutions on an infinite alphabet*, in preparation.
- [FER-HOL-ZAM-1] S. Ferenczi, C. Holton, L. Q. Zamboni, *The structure of three-interval exchange transformations I: an arithmetic study*, Ann. Inst. Fourier (Grenoble) **51** (2001), 861–901.
- [FER-HOL-ZAM-2] S. Ferenczi, C. Holton, L. Q. Zamboni, *The structure of three-interval exchange transformations II: a combinatorial description of the trajectories*, J. Analyse Math. **89** (2003), 239–276.
- [FER-HOL-ZAM-3] S. Ferenczi, C. Holton, L. Q. Zamboni, *The structure of three-interval exchange transformations III: ergodic and spectral properties*, J. Analyse Math. **93** (2004), 103–138.
- [FER-HOL-ZAM-4] S. Ferenczi, C. Holton, L. Q. Zamboni, *Joinings of three-interval exchange transformations*, Ergodic Theory Dynam. Systems, to appear.
- [FER-MAU] S. Ferenczi, C. Mauduit, *Transcendence of numbers with a low complexity expansion*, J. Number Theory **67** (1997), 146–161.
- [FER-ZAM] S. Ferenczi, L. Q. Zamboni, *Combinatorics of symmetric interval exchange transformations*, in preparation.
- [FRI-PAS-TIC-WIN] S. Frisch, M. Pašteka, R. F. Tichy, R. Winkler *Finitely additive measures on groups and rings*, Rend. Circolo Mat. di Palermo Series II **48** (1999), 323–340.
- [GUI-MAS-PEL] L.S. Guimond, Z. Masáková, E. Pelantová, *Combinatorial properties of infinite words associated with cut-and-project sequences*, J. Théor. Nombres Bordeaux **15** (2003), 697–725.
- [HUB-MES] P. Hubert, A. Messaoudi, *Best simultaneous diophantine approximations of Pisot numbers and Rauzy fractals*, preprint (2004).
- [ITO] S. Ito, *On periodic expansions of cubic numbers and Rauzy fractals*, in *Dynamical systems: From crystal to chaos*, Proceedings of the conference held in honor of Gérard Rauzy on his 60th birthday in Luminy-Marseille, July 6–10, 1998, World Scientific Publishing Co. Inc., River Edge, NJ, 2000.
- [ITO-FUJ-HIG-YAS] S. Ito, J. Fujii, H. Higashino, and S.-i. Yasutomi, *On simultaneous approximation to (α, α^2) with $\alpha^3 + k\alpha - 1 = 0$* , J. Number Theory **99** (2003), 255–283.
- [ITO-KIM] S. Ito, M. Kimura, *On Rauzy fractal*, Japan J. Indust. Appl. Math. **8** (1991), 461–486.
- [ITO-OHT] S. Ito, M. Ohtsuki, *Modified Jacobi-Perron algorithm and generating Markov partitions for special hyperbolic toral automorphisms*, Tokyo J. Math. **16** (1993), 441–472.
- [JUS-PIR-1] J. Justin, and G. Pirillo, *Episturmian words and episturmian morphisms*, Theoret. Comput. Sci. **276** (2002), 281–313.
- [JUS-PIR-2] J. Justin, and G. Pirillo, *Episturmian words: shifts, morphisms and numeration systems*, International Journal of Foundations of Computer Science **15** (2004), 329–348.
- [KEA] M.S. Keane, *Interval exchange transformations*, Math. Zeitsch. **141** (1975), 25–31.
- [KEN-VER] R. Kenyon, A. M. Vershik, *Arithmetic construction of sofic partitions of hyperbolic toral automorphisms*, Ergodic Theory Dynam. Systems **18** (1998), 357–372.
- [KER] S. Kerckhoff, *Simplicial systems for interval exchange maps and measured foliations*, Ergodic Theory Dynam. Systems **5** (1985), 257–271.
- [KES] H. Kesten, *On a conjecture of Erdős and Szűsz related to uniform distribution mod 1*, Acta Arith. **12** (1966/1967), 193–212.
- [KLE-ROS] R. Klette, A. Rosenfeld, *Digital straightness-A review*, Discrete Applied Mathematics **139** (2004), 197–230.
- [KUI-NIE] L. Kuipers, H. Niederreiter, *Uniform distribution of sequences*, Pure Appl. Math., Wiley, New York, 1974.
- [LAG-1] J.C. Lagarias, *Best simultaneous Diophantine approximations I, Growth rates of best approximations denominators*, Trans. Amer. Math. Soc. **72** (1982), 545–554.
- [LAG-2] J.C. Lagarias, *Best simultaneous Diophantine approximations II, Behavior of consecutive best approximations*, Pacific Journal of Mathematics **102** (1982), 61–88.
- [LIA] P. Liardet, *Regularities of distribution*, Compositio Mathematica **61** (1987), 267–293.
- [LOP-daR] A.O. Lopes, L.F.C. da Rocha, *Invariant measures for Gauss maps associated with interval exchange maps*, Indiana Univ. Math. J. **43** (1994), 1399–1438.
- [LOT-1] M. Lothaire, *Algebraic combinatorics on words*, Cambridge University Press (2002).
- [LOT-2] M. Lothaire, *Applied combinatorics on words*, Cambridge University Press, to appear.
- [LOX-vdP] J. H. Loxton, A. K. van der Poorten, *Arithmetic properties of certain functions in several variables III*, Bull. Austral. Math. Soc. **116** (1977), 15–47.

- [MAH] K. Mahler, *Arithmetische Eigenschaften der Lösungen einer Klasse von Funktionalgleichungen*, Math. Ann. **103** (1929), 573–587.
- [MAR-MOU-YOC] S. Marmi, P. Moussa, J.-C. Yoccoz, *The cohomological equation for Roth type interval exchange maps*, preprint (2004), arXiv:math.DS/0403518.
- [MASUI-SUG-YOS] K. Masui, F. Sugisaki, M. Yoshida, *HPS for $BO(\phi) = 2$* , preprint (2004).
- [MAS] H. Masur, *Interval exchange transformations and measured foliations*, Ann. of Math. **115** (1982), 169–200.
- [MES-1] A. Messaoudi, *Propriétés arithmétiques et dynamiques du fractal de Rauzy*, J. Théor. Nombres Bordeaux **10** (1998), 135–162.
- [MES-2] A. Messaoudi, *Frontière du fractal de Rauzy et système de numération complexe*, Acta Arith. **95** (2000), 195–224.
- [MOR-HED] M. Morse, G.A. Hedlund, *Symbolic dynamics II. Sturmian trajectories*, Amer. J. Math. **62** (1940), 1–42.
- [OSE] V.I. Oseledec, *The spectrum of ergodic automorphisms*, Dokl. Akad. Nauk. SSSR **168** (1966), 1009–1011.
- [OST-1] A. Ostrowski, *Bemerkungen zur Theorie der Diophantischen Approximationen I*, Abh. Math. Sem. Hamburg I (1922), 77–98.
- [OST-2] A. Ostrowski, *Bemerkungen zur Theorie der Diophantischen Approximationen II*, Abh. Math. Sem. Hamburg I (1922), 250–251.
- [PRA] B. Praggastis, *Numeration systems and Markov partitions from self-similar tilings*, Trans. Amer. Math. Soc. **351** (1999), 3315–3349.
- [PYT] N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, Lecture Notes in Math. **1794** (2002), Springer-Verlag.
- [QUA-ZAM] A. Quas, L. Q. Zamboni, *Periodicity and local complexity*, Theoret. Comput. Sci. **319** (2004), 229–240.
- [RAU-1] G. Rauzy, *Une généralisation du développement en fraction continue*, Séminaire Delange-Pisot-Poitou, 1976/1977, Théorie des Nombres, Fasc. 1, Exp. No.15,16, Paris, 1977.
- [RAU-2] G. Rauzy, *Échanges d'intervalles et transformations induites*, Acta Arith. **34** (1979), 315–328.
- [RAU-3] G. Rauzy, *Nombres algébriques et substitutions*, Bull. Soc. Math. France **110** (1982), 147–178.
- [RAU-4] G. Rauzy, *Ensembles à restes bornés*, Séminaire de Théorie des Nombres de Bordeaux, 1982-1983, exposé 25.
- [RAU-5] G. Rauzy, *Des mots en arithmétique*, in *Avignon Conference on Language Theory and Algorithmic Complexity (Avignon, 1983)*, Univ. Claude-Bernard, Lyon (1984), 103–113.
- [RAU-6] G. Rauzy, *Sequences defined by iterated morphisms*, in *Sequences (Naples/Positano, 1988)*, Springer, New York (1990), 275–286.
- [REV] J.-P. Réveillès, *Géométrie discrète, calcul en nombres entiers et algorithmiques*, Thèse d'État, Université Louis Pasteur, Strasbourg, France, 1991.
- [RID] D. Ridout, *Rational approximations to algebraic numbers*, Mathematika **4** (1957), 125–131.
- [RIS-ZAM] R. Risley, L. Q. Zamboni, *A generalization of Sturmian sequences; combinatorial structure and transcendence*, Acta Arith. **95** (2000), 167–184.
- [daR] L.F.C. da Rocha, *A new induction for interval exchange maps*, preprint (1993).
- [ROT] G. Rote, *Sequences with subword complexity $2n$* , J. Number Theory **46** (1994), 196–213.
- [ROY-1] D. Roy, *Approximation to real numbers by cubic algebraic integers I*, Proc. London Math. Soc. **88** (2004), 42–62.
- [ROY-2] D. Roy, *Approximation to real numbers by cubic algebraic integers II*, Annals of Math. **158** (2003), 1081–1087.
- [ROY-3] D. Roy, *On two exponents of approximation related to a real number and its square*, preprint (2004), arXiv:math.NT/0409232.
- [SAN-TIJ-1] J. W. Sander, R. Tijdeman, *Low complexity function and convex sets in \mathbb{Z}^k* , Math. Z. **233** (2000), 205–218.
- [SAN-TIJ-2] J.W. Sander, R. Tijdeman, *The complexity of functions on lattices*, Theoret. Comput. Sci. **246** (2000), 195–225.
- [SAN-TIJ-3] J. W. Sander, J. W., R. Tijdeman, *The rectangle complexity of functions on two-dimensional lattices*, Theoret. Comput. Sci. **270** (2002), 857–863.
- [SCH-SZA-WIN] J. Schmeling, E. Szabó, R. Winkler, *Hartman and Beatty sequences*, in *Algebraic Number Theory and Diophantine Analysis*, 405–421, Walter de Gruyter, Berlin (2000).

- [SID] N. Sidorov, *Arithmetic dynamics*, in *Topics in dynamics and ergodic theory*, S. Bezuglyi et al. (ed.), Cambridge University Press, Lond. Math. Soc. Lect. Note Ser. **310** (2003), 145–189.
- [SIR-WAN] V. F. Sirvent, Y. Wang, *Self-affine tiling via substitution dynamical systems and Rauzy fractals*, *Pacific J. Math.* **206** (2002), 465–485.
- [SOL] B. Solomyak, *Substitutions, adic transformations, and beta-expansions*, in *Symbolic dynamics and its applications* (New Haven, CT, 1991), Contemp. Math. **135** Amer. Math. Soc., Providence, RI (1992), 361–372.
- [SOS] V. T. Sós, *On the distribution mod 1 of the sequence $n\alpha$* , Ann. Univ. Sci. Budapest, Eötvös Sect. Math. **1** (1958), 127–134.
- [STE-WIN] C. Steiner, R. Winkler, *Complexity of Hartman sequences*, J. Théor. Nombres Bordeaux, to appear.
- [THU] W. P. Thurston, *Groups, tilings and finite state automata*, Lectures notes distributed in conjunction with the Colloquium Series, in *AMS Colloquium lectures*, 1989.
- [VEE-1] W.A. Veech, *Gauss measures for transformations on the space of interval exchange maps*, Ann. of Math. **115** (1982), 201–242.
- [VEE-2] W.A. Veech, *A criterion for a process to be prime*, Monatsh. Math. **94**, 4 (1982), 335–341.
- [VEE-3] W.A. Veech, *The metric theory of interval exchange transformations I: Generic spectral properties*, Amer. J. Math. **106** (1984), 1331–1359.
- [VEE-4] W.A. Veech, *The metric theory of interval exchange transformations II: Approximation by primitive interval exchanges*, Amer. J. Math. **106** (1984), 1361–1387.
- [VEE-5] W.A. Veech, *The metric theory of interval exchange transformations III: The Sah-Arnoux-Fathi invariant*, Amer. J. Math. **106** (1984), 1389–1422.
- [VER-SID] A. M. Vershik, N. A. Sidorov, *Arithmetic expansions associated with rotations of the circle and continued fractions*, St. Petersburg Math. Journ. **5** (1994), 1121–1136.
- [WIN] R. Winkler, *Ergodic group rotations, Hartman sets and Kronecker sequences*, Monatsh. Math. **135** (2002), 333–343.
- [WOZ-ZAM] N. Wozny, L. Q. Zamboni, *Frequencies of factors in Arnoux-Rauzy sequences*, Acta Arith. **96** (2001), 261–278.
- [YOC] J.-C. Yoccoz, *Continued fraction algorithms for interval exchange maps: an introduction*, Proceedings of the conference “Frontiers in Number Theory, Physics and Geometry”, Les Houches, 2003, to appear.
- [ZAM] L. Q. Zamboni, *Une généralisation du théorème de Lagrange sur le développement en fraction continue*, C. R. Acad. Sci. Paris, Série I, **327** (1998), 527–530.
- [ZOR] A. Zorich, *Finite Gauss measure on the space of interval exchange transformations. Lyapunov exponents*, Ann. Inst. Fourier (Grenoble) **46** (1996), 325–370.

LIRMM (CNRS, UMR 5506-UNIV. MONTPELLIER 2)-161 RUE ADA -34392 MONTPELLIER CEDEX 5 - FRANCE

E-mail address: berthe@lirmm.fr

INSTITUT DE MATHÉMATIQUES DE LUMINY (CNRS, UMR 6206) - CASE 907, 163 AV. DE LUMINY - F13288 MARSEILLE CEDEX 9 - FRANCE, AND FÉDÉRATION DE RECHERCHE DES UNITÉS DE MATHÉMATIQUES DE MARSEILLE (CNRS, FR 2291)

E-mail address: ferenczi@iml.univ-mrs.fr

DEPARTMENT OF MATHEMATICS - UNIVERSITY OF NORTH TEXAS - DENTON, TX 76203-1430 - USA

E-mail address: luca@unt.edu