

THE SCENERY FLOW FOR GEOMETRIC STRUCTURES ON THE TORUS: THE LINEAR SETTING

P. ARNOUX* A. M. FISHER**

Abstract

The authors define the scenery flow of the torus. The flow space is the union of all flat 2-dimensional tori of area 1 with a marked direction (or equivalently, the union of all tori with a quadratic differential of norm 1). This is a 5-dimensional space, and the flow acts by following individual points under an extremal deformation of the quadratic differential. The authors define associated horocycle and translation flows; the latter preserve each torus and are the horizontal and vertical flows of the corresponding quadratic differential.

The scenery flow projects to the geodesic flow on the modular surface, and admits, for each orientation preserving hyperbolic toral automorphism, an invariant 3-dimensional subset on which it is the suspension flow of that map.

The authors first give a simple algebraic definition in terms of the group of affine maps of the plane, and prove that the flow is Anosov. They give an explicit formula for the first-return map of the flow on convenient cross-sections. Then, in the main part of the paper, the authors give several different models for the flow and its cross-sections, in terms of:

- stacking and rescaling periodic tilings of the plane;
- symbolic dynamics: the natural extension of the recoding of Sturmian sequences, or the S -adic system generated by two substitutions;
- zooming and subdividing quasi-periodic tilings of the real line, or aperiodic quasicrystals of minimal complexity;
- the natural extension of two-dimensional continued fractions;
- induction on exchanges of three intervals;
- rescaling on pairs of transverse measure foliations on the torus, or the Teichmüller flow on the twice-punctured torus.

Keywords Modular surface, Continued fractions, Sturmian sequences, Plane tilings, Teichmüller flow, Substitution dynamical system

2000 MR Subject Classification 32G15, 30F60, 53D25, 37D40

Chinese Library Classification O186.15, O174.51 **Document Code** A

Article ID 0252-9599(2001)04-0427-44

§0. Introduction

We begin by recalling some well-known relationships. First, there is the one-to-one correspondence between closed orbits of the geodesic flow on the modular surface and conjugacy classes of hyperbolic toral automorphisms. (This can be seen directly from the definitions (see Remark 1.3 in §1 below).) Secondly, one knows that it is possible to code this geodesic flow using continued fractions and via circle rotations (cf. [9, 42, 2, 7]). Thirdly, there is a

Manuscript received August 4, 2000.

*Institut de Mathématiques de Luminy (UPR 9016), 163 Avenue de Luminy, case 907 13288 Marseille

Cedex 9, France. **E-mail:** arnoux@iml.univ-mrs.fr

Dept Mat IME-USP, Caixa Postal 66281, CEP 05315-970 São Paulo, Brazil. **E-mail: afisher@ime.usp.br

strong relation between hyperbolic toral automorphisms and rotations by certain quadratic integers; these rotations appear as the first return map for the flow along the stable (or unstable) foliation of the toral automorphism.

In this paper we construct a fiber bundle, with torus fiber, over the unit tangent bundle $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$ of the modular surface, and an Anosov flow on this fiber bundle, that will make all these relations completely explicit. In particular, this flow will project to the usual geodesic flow; over a closed geodesic, it will be a suspension of the corresponding toral automorphism.

We first give in §1 a simple and explicit algebraic model for this flow. For this we make use of the group $SA(2, \mathbb{R})$ of special affine maps of the plane (the real affine maps of determinant 1), which is isomorphic to a subgroup of $SL(3, \mathbb{R})$. This algebraic presentation allows us to easily show that the flow is Anosov, hence ergodic, with a natural invariant measure.

In the later sections, we show that the scenery flow appears in a variety of guises: it can be represented as a flow on the space of points of the plane modulo a lattice of covolume 1, or on the space of Sturmian (=quasiperiodic of minimum complexity) tilings of the line; these can be considered as the simplest kind of aperiodic quasicrystals. The flow has a cross-section with a purely combinatoric description; we relate this to adic systems in the sense of Vershik (cf. [47, 48, 49]). We also give a purely arithmetic interpretation in terms of inhomogeneous continued fractions, using the Ostrowsky representation of integers or real numbers. Lastly, we show how this flow can be viewed as the restriction to a particular stratum of the Teichmüller flow on the twice punctured torus.

0.1. Organization of This Paper

In §1, we give the formal algebraic definitions, fix our notation and prove basic properties. In §2 we interpret this algebraic model geometrically, in terms of periodic plane tilings, and we give explicit formulas for the first return map of the scenery flow to two particular cross-sections, related to the additive and multiplicative continued fractions respectively.

In §3 we give a simple Markov partition for the additive cross-section. In §4, we give another viewpoint on this Markov partition, using Sturmian sequences. This part is the longest section of the paper, due to the need to recall some classical results on these sequences, which geometrically are just the symbolic itineraries generated by an irrational rotation with respect to its natural partition, and which have equally simple and useful purely combinatorial definitions explained below. Part of the interest of having these quite different descriptions is that one can see how the commutation relations between the scenery flow and the vertical and horizontal translation flows appear at the purely symbolic level, as relations between a substitution and a shift map.

In §5, we then use this to analyse dynamics for the Sturmian tilings of the line. In §6, we show how the geometric coordinates of a point in the flow space can be recovered from the combinatorics explicitly, via the inhomogeneous continued fraction (the Ostrowsky expansion of real numbers). This gives an arithmetic model of the scenery flow.

In §7 and §8 we model the scenery flow by means of Rauzy induction on the interval, and, equivalently, in terms of the Teichmüller flow on the twice punctured torus.

In §9, we show how the measures on the fibers that naturally appear in this work can be thought of as Gibbs measures for pressure 0, a generalization to nonstationary shift spaces of the Parry-Shannon measure for subshifts of finite type (the measure of maximal entropy). In the final section §10, we indicate some possible generalizations.

In each of the sections §2 to §8, our main result will be to build a flow, or a map, on some

space, and to exhibit a conjugacy with the flows and maps built in §1 and §2.

0.2. Underlying Ideas

Although the simplicity of the algebraic presentation might well hide the larger picture we wish to show here, we hope that the next sections will make clear that we are discussing (for a very explicit example) some quite general underlying ideas that certainly have a role to play in a variety of related situations.

The first idea is to consider a parameter space for a family of dynamical systems, and to define a map on the parameter space corresponding to inducing a given dynamical system on a subset. Here, we consider rotations R_θ , by an irrational angle θ , on the circle of length 1. The induction on a suitable subinterval produces, in dynamical space, another rotation; this process (also called *renormalization*) yields, on parameter space, the continued fraction map. This renormalization can be viewed also in a symbolic way: the initial system is related to the induced system by a substitution.

The second idea is to build the natural extension (in the sense of Rohlin, cf. [41]) for this procedure; here, it corresponds to giving a rule for “un-inducing” a given rotation. The resulting sequence of choices gives a second parameter. So taken together, one has a pair of parameters, which correspond to a pair of dynamical systems (two circle rotations), or, equivalently, to a pair of linear foliations on the 2-torus, with slopes equal to angles of rotation.

The torus provides a dynamical space on which both rotations can act simultaneously, as cross-sections to transverse linear flows. The operation of inducing on the one and un-inducing on the other has a nice geometrical interpretation on the torus, given by *box renormalization*: a cutting and stacking of a pair of boxes built from the rotations, and which form a fundamental domain for the torus.

The next step is to construct a fiber bundle over the set of parameters, with torus fibers. This places together in a single space all possible pairs of rotations of various angles. We then define a map on this bundle, given by the renormalization. Building a suspension of this map, we have the **scenery flow**. The origin of this name is that this flow gives the effect of traveling in a landscape: one zooms in toward the small-scale geometry of the first dynamical system, the forward landscape, while simultaneously zooming out of the second landscape.

In other words, on the fibers this has a hyperbolic dynamics, as it is expanding in one direction and contracting in the other. The renormalisation operator is a discrete-time version of the flow, with times indexed by returns to a cross-section for the flow on the base space. However this does not produce a map in the usual sense; one is moving each time to a different fiber. This leads to the notion of *mapping family*: a sequence of maps along a sequence of spaces (cf. [3]).

In the particular case where the combinatorics of the map, i.e. the continued fraction expansion of the rotation, is periodic, inducing returns us to the original rotation, and the hyperbolic mapping family can be thought of as a hyperbolic map in the usual sense; in fact one has, in this case, a very familiar object: a hyperbolic toral automorphism; this is the map referred to in the second paragraph. The pair of boxes mentioned above is in this case also familiar: it is a Markov partition.

One of the interests of the fiber bundle construction and the concept of mapping family is that one can then extend to a union of dynamical systems (organized as a bundle), to non-periodic combinatorics, and to sequences of maps, results that previously have been

stated only for a specific combinatorics, for single maps, or for periodic systems.

In terms of the symbolic dynamics, this takes us from the substitution dynamical systems, which form a countable set (corresponding to closed trajectories of the geodesic flow), to S -adic systems, that is, to systems generated by a finite number of substitutions in a set S .

In summary, essentially what we are doing here is to unify the parameter space and dynamical space of a family of dynamical systems (in this case, circle rotations). This single model brings together all of the dynamics: the circle rotations (of all angles), their extension to continuous time, i.e. to a flow along a foliation, and the dynamics of inducing (renormalization) together with its inverse. This includes also all Anosov toral maps, together with their “completion” to Anosov mapping families. This dynamics has, furthermore, been extended to continuous time; and this gives the scenery flow.

As we will see, this unified object is entirely natural and interesting to study; also the unified point of view simplifies the overall picture, and leads to interesting further questions.

0.3. Some Generalizations

These ideas generalize immediately and quite completely to a number of related situations; part of this is already implicit in the work of Veech (cf. [46]) regarding the moduli spaces of Riemann surfaces, where interval exchange transformations replace rotations, with Rauzy induction replacing the continued fraction transformation. It is also possible to carry out a similar construction for Hecke groups, by making use of the Rosen continued fraction (cf. [5]).

We mention that, although some parts also go through for automorphisms of the 3-torus and translations on the 2-torus, here the picture is still far from complete. It is a place where the different viewpoints studied in this paper could prove useful, because the symbolic dynamics, in specific instances, points to facts that are very difficult to grasp from a purely geometric or arithmetic viewpoint (cf. for example [39, 26]).

For the present case (circle rotations, Anosov maps) one can generalize in a different direction, extending these ideas to the nonlinear setting. We carry this out in a series of forthcoming papers, developing for the general (non-periodic) case theorems previously known for single maps: the stable manifold theorem, structural stability and openness, shadowing and the existence of Markov partitions, the existence of Gibbs states, and the smooth classification of de la Lave-Marco-Moriyon and Cawley. We study the higher genus situation as well, both for the linear and nonlinear case. In particular, we unify and extend theorems previously known in more restricted situations regarding unique ergodicity.

0.4. Related Work

In this paper we bring a different perspective to bear on several classical topics, which have been investigated by many authors: the continued fraction map, and the relation with inducing or renormalization for rotations; the modular space and its geodesic and horocycle flows; linear Anosov maps and their codings.

Although our point of view seems to be new, we are building on this classical mathematics; and our work is closely related to that of many people. We mention in particular the studies of coding of the geodesic flow in [9, 42, 2, 47]; of arithmetic expansions by [37, 43, 29] among others; we note that various of the present ideas are, in a different terminology, explored in the works of Vershik and collaborators (cf. [47, 49, 50]), and also in [2]. The generalized continued fractions we consider have been already studied in [27].

We mention some works related to symbolic dynamics of rotations: [23, 22, 15, 38, 13, 34]; this is but part of a long list of related papers.

Regarding the study of tilings of the line and of the plane, we wish to mention especially the work of deBruijn (cf. [17, 18]) on the Fibonacci tiling, which is the simplest periodic case. In §5, we generalize to the non-periodic case the results deBruijn obtains for the Fibonacci tiling. The interesting recent work of Robinson (cf. [40]) on Penrose tilings is also related to what we do here.

For background on hyperbolic toral automorphisms and the construction of Markov partitions, we refer the reader to the fundamental paper of Adler and Weiss (cf. [10]).

For other related references see also [3].

§1. The Algebraic Model

We denote by $G = SA(2, \mathbb{R})$ the **special affine group**, the group of measure-preserving and orientation preserving affine maps of the plane. This is a generalization to two dimensions of the affine group (the $ax + b$ group) on the line. An element g of G is the composition of a linear map M of determinant one with a translation by a vector \vec{v} . This composition can be taken in either order. For $g \in G$ the linear part M is well-defined, but the vector depends on whether one first translates or first applies the linear part. For $(x, y) \in \mathbb{R}^2$, we write the action of $g \in G$ on the right; for the first decomposition, we have $(x, y) \cdot g = (x, y) \cdot M + \vec{v}$, in which case the vector is the image of the origin; for the second $(x, y) \cdot g = ((x, y) + \vec{w}) \cdot M$.

Notation 1.1. We will denote by $(M, \vec{v}) \in SL(2, \mathbb{R}) \times \mathbb{R}^2$ the element g defined by $(x, y) \cdot g = (x, y) \cdot M + \vec{v}$, and by $(\vec{w}, N) \in \mathbb{R}^2 \times SL(2, \mathbb{R})$ the element g defined by $(x, y) \cdot g = ((x, y) + \vec{w}) \cdot N$.

The point here is that G is not a direct product, but rather a semidirect product, with \mathbb{R}^2 as the normal subgroup. Both ways of presenting g are natural, and each will turn out to be useful.

The group G is isomorphic to the subgroup of $SL(3, \mathbb{R})$ of matrices whose last column is $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, via the isomorphism:

$$(M, \vec{v}) \mapsto \begin{pmatrix} M & 0 \\ \vec{v} & 1 \end{pmatrix}.$$

Remark 1.1. We have the following relationship between these two notations for $g \in G$: writing \vec{v}, \vec{w} also as row vectors (as we did for the point (x, y)) and M as a matrix, then $(\vec{w}, M) = (M, \vec{w} \cdot M)$, where $\vec{w} \cdot M$ is matrix multiplication.

The group multiplication is given by $(M, \vec{v}) \cdot (M', \vec{v}') = (M \cdot M', \vec{v} \cdot M' + \vec{v}')$ in the first notation, and by $(\vec{v}, M) \cdot (\vec{v}', M') = (\vec{v} + \vec{v}' \cdot M^{-1}, M \cdot M')$ in the second.

We have written the action of G on the right and chosen to use row vectors for elements of \mathbb{R}^2 in order to later get standard formulas for actions and for quotients by discrete groups.

The group G projects naturally (by taking the linear part) onto $SL(2, \mathbb{R})$.

The kernel of this projection is the group of translations of the plane, which is isomorphic to \mathbb{R}^2 . As mentioned above, G is the semi-direct product $SL(2, \mathbb{R}) \ltimes \mathbb{R}^2$, with the usual action of $SL(2, \mathbb{R})$ on row vectors in \mathbb{R}^2 , i.e. multiplying the row vector on the right by the matrix.

We will write Γ for the subgroup $SA(2, \mathbb{Z})$ of elements of G with integral entries (for the vector as well as for the matrix). (This makes sense in both notations, since $SL(2, \mathbb{Z})$

preserves \mathbb{Z}^2 ; note that Γ is clearly a subgroup of G). This subgroup acts on G by left multiplication. The object of interest in this paper is the quotient $E = \Gamma \backslash G$. An element here is a coset $(SL(2, \mathbb{Z}) \cdot M, \vec{v} + \mathbb{Z}^2 \cdot M)$. The space E is five-dimensional, and projects naturally onto $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$. The points that project to $SL(2, \mathbb{Z}) \cdot M$ are just the translates of the lattice $\mathbb{Z}^2 \cdot M$; hence the fiber over $SL(2, \mathbb{Z}) \cdot M$ is $\mathbb{R}^2 / (\mathbb{Z}^2 \cdot M)$, which is the torus defined by the lattice in the plane spanned by the rows of M . Therefore E is a fiber bundle over $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$, with torus fibers.

We remark that this fiber bundle has base $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$, and not the modular surface $SL(2, \mathbb{Z}) \backslash \mathbb{H}$; in the latter case, there would be a problem because $SL(2, \mathbb{Z})$ acts on \mathbb{H} with fixed points, giving rise to singular points on the modular surface (as is well known) and there would be singular fibers in the bundle. This does not happen for the action of $SL(2, \mathbb{Z})$ on $SL(2, \mathbb{R})$.

Remark 1.2. An element of E can be viewed as a translation of a lattice $\mathbb{Z}^2 \cdot M$ of covolume 1 in \mathbb{R}^2 , with the translation vector defined up to an element of the lattice. There are two natural ways to give coordinates to this translation vector: we can either decompose it with respect to the canonical basis of \mathbb{R}^2 , which gives as coordinates a row vector \vec{v} , or use any basis for the lattice $\mathbb{Z}^2 \cdot M$, for example the image by M of the canonical basis, which gives a different row vector \vec{w} . This amounts to choosing one of the two possible notations (M, \vec{v}) or (\vec{w}, M) ; the relation noted in Remark 1.1 gives the obvious change of basis. Most of the time we will use the first notation, as it is better suited to a study of the translation flows. The second notation will be more natural when we restrict attention to a single torus fiber, $\mathbb{Z}^2 \cdot M \backslash \mathbb{R}^2$ (cf. Remark 1.3 below).

The group G acts on itself by translation on the right, and this induces a right action on E ; this allows us to define the main object of study in this paper:

Definition 1.1. *The scenery flow is the flow defined on E by the action on the right of the 1-dimensional subgroup $g_t = \left(\begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}, (0, 0) \right)$.*

We remark that this is a lift to E of the classical **modular flow**, the geodesic flow on $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R}) = PSL(2, SL(2, \mathbb{Z})) \backslash PSL(2, \mathbb{R})$, which is naturally identified with the unit tangent bundle of the modular surface (the hyperbolic plane modulo $SL(2, \mathbb{Z})$).

The reason for the name will become clear in §5; in §8, we will show that this flow can also be considered as the **Teichmüller mapping flow**.

Certain other one-dimensional subgroups of G are of special interest. We give the following names to the corresponding \mathbb{R} -actions:

- the **positive horocycle flow** $h_s^+ = \left(\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, (0, 0) \right)$;
- the **negative horocycle flow** $h_s^- = \left(\begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}, (0, 0) \right)$;
- the **vertical translation flow** $T_u^+ = \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, (0, u) \right)$;
- the **horizontal translation flow** $T_u^- = \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, (u, 0) \right)$.

It is immediate from the definition that the translation flows preserve each fiber of the projection to $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$. Thus one can consider each translation flow as a collection of linear flows on tori. In fact these flows are familiar: each fiber carries a natural quadratic differential, and these are what are known as the horizontal and vertical flows of the quadratic differential.

The horocycle flows project to the usual flows on $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$, whence the names. Since these latter flows are completely defined by their linear part, we will use the same names for the corresponding flows on $SL(2, \mathbb{R})$.

Remark 1.3. Let us consider for example a closed orbit for the modular flow, i.e. the geodesic flow on the space $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$. By definition of cosets, associated to this orbit are a real matrix M and an integer matrix A such that $M \cdot g_T = A \cdot M$, where T is the length of the closed geodesic, A is determined up to conjugacy in $SL(2, \mathbb{Z})$, and M is determined up to multiplication on the right by g_t , and on the left by $SL(2, \mathbb{Z})$. Looking at the fiber above $SL(2, \mathbb{Z}) \cdot M$, the image of a point (\vec{v}, M) by g_T is $(\vec{v}, M \cdot g_T) = (\vec{v}, A \cdot M)$. This is equivalent, by action on the left of $(\vec{0}, A^{-1})$, to $(\vec{v} \cdot A, M)$. This shows that g_T preserves the fiber over $SL(2, \mathbb{Z}) \cdot M$, with the action by A on the fiber being a hyperbolic map. This is what we referred to in the introduction: the return map to the fiber is an Anosov toral automorphism. (Here it is easier to use the (\vec{v}, M) notation, as mentioned in Remark 1.2.)

A basic property of these flows is that they satisfy the following commutation relations:

Proposition 1.1. *The flows h_s^+ and T_u^+ commute. Similarly, h_s^- and T_u^- commute. These flows satisfy the following commutation relations with the scenery flow:*

$$\begin{aligned} h_s^+ g_t &= g_t h_{se^{-t}}^+, & h_s^- g_t &= g_t h_{se^t}^-; \\ T_s^+ g_t &= g_t T_{se^{-t/2}}^+, & T_s^- g_t &= g_t T_{se^{t/2}}^-. \end{aligned}$$

These relations are immediately checked by computation. The first ones are the classical commutation relations between geodesic and horocycle flow. They imply that together h_s^+ and T_u^+ generate a 2-dimensional foliation, which is the stable foliation for the scenery flow, while h_s^- and T_u^- generate the unstable foliation. This implies, by the classical argument due to Hopf (cf. [24, 25]):

Proposition 1.2. *The flow g_t on E is ergodic for the measure on E given by the Haar measure on G .*

Proof. The set E has finite measure for the natural Haar measure on G . (The reason is that it is a fiber bundle, with fibers of area 1, over $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$; and this is isomorphic to the unit tangent bundle of the modular surface, which has finite volume.)

The tangent bundle to E splits in three parts, invariant by the scenery flow: the tangent to the flow itself, and the stable and unstable foliations defined above. It is then easy to check that the scenery flow is Anosov, which implies that it is ergodic, by Hopf's argument: the only point that requires checking is that the foliations are absolutely continuous, and this is true since these foliations are defined by the action of 2-parameter subgroups of G .

§2. Cross-Sections for the Scenery Flow: the Plane Tiling Model

We will want to make explicit computations for the scenery flow. For this purpose we will define a hypersurface in $SA(2, \mathbb{R})$, transverse to the geodesic flow, and study the first return map of the flow to this surface.

The first step will be to define coordinates on E . A good way to do this is to make a careful choice of fundamental domain for the action of $SA(2, \mathbb{Z})$ on $SA(2, \mathbb{R})$. We shall in fact make use of two different fundamental domains. The first one is simpler to use for our proofs; it will be called Ω_a (the a refers to its relationship to the additive continued fraction algorithm). The second choice is related to more classical approaches, and will be denoted by Ω_m (m because it is related to the multiplicative, or ordinary, continued

fraction algorithm). We will use for these two domains the construction given in detail in [7], following ideas developed in [45].

The idea for the construction of these domains is to consider $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$ as the space of lattices in the plane of covolume 1. Indeed, we can consider a matrix in $SL(2, \mathbb{R})$ as a pair of row vectors which generate an integer lattice of covolume 1; the left action by $SL(2, \mathbb{Z})$ amounts to choosing another basis of the same lattice. A point in E is then defined by giving first a lattice, and then a point modulo this lattice.

Be careful: we will be defining fundamental domains at two levels: for the lattice $\mathbb{Z}^2 \cdot M$ acting on the plane \mathbb{R}^2 , and also for the group $SA(2, \mathbb{Z})$ acting on $SA(2, \mathbb{R})$; this point may seem confusing at first. These domains are related: given a lattice in the plane, we first give a rule to define a fundamental domain for this lattice; this rule will then also define a fundamental domain for the action of Γ on G , since it will allow us to put coordinates on the space of lattices, and also will let us choose a unique representative for a class of points modulo this lattice, by taking the unique representative contained in the fundamental domain.

Notation 2.1. *In the sequel, for a real number x , we will denote by $\lfloor x \rfloor$ the greatest integer less or equal to x , and by $\lceil x \rceil$ the least integer greater than or equal to x . We will write $\{x\} = x - \lfloor x \rfloor$ for the fractional part of x .*

We will denote by $[a, b]$ the closed interval $a \leq x \leq b$, and by $]a, b[$ the open interval $a < x < b$.

2.1. The Domain Related to the Additive Continued Fraction

For each lattice, we want to find a canonical fundamental domain. This is usually done by taking the unit square for some particular basis of the lattice. However, to study the vertical and horizontal flows, it is better to have a fundamental domain whose boundary segments are parallel to the axes (and hence to these flows).

An L -shaped form, made of two joined rectangles with sides parallel to the axes, will always tile the plane. In fact, any lattice has a fundamental domain with this shape. (One of the two boxes in this L -shape may be degenerate, as follows: if the lattice contains a vertical vector, it is possible that a rectangle degenerates to a line, or even to a point, which is the case for the lattice \mathbb{Z}^2 .) In fact, the lattice has in general a (countably) infinite number of such fundamental domains. Therefore we will add an additional condition on the width of the rectangles that will ensure a unique such choice.

Fig. 2.1. The L -Shaped Domain

This picture may remind the reader of the construction by Adler and Weiss (cf. [10]) (and independently Ken Berg) of Markov partitions for hyperbolic toral automorphisms. This is no accident, as that situation turns out to correspond to the periodic orbits of the modular flow.

Definition 2.1. We denote by $\Omega_{a,1}$ the subset of $SA(2, \mathbb{R})$ given by pairs (M, \vec{v}) , with $M = \begin{pmatrix} l_0 & h_1 \\ -l_1 & h_0 \end{pmatrix}$ and $\vec{v} = (x, y)$, where M satisfies the following inequalities:

$$0 < l_0 < 1 \leq l_1 < 1 + l_0, \quad h_0, h_1 > 0$$

and \vec{v} satisfies the inequalities :

$$-l_0 \leq x < l_1 \quad \text{and} \quad 0 \leq y < h_0 \text{ if } x < 0, \text{ or } 0 \leq y < h_1 \text{ if } x \geq 0.$$

We denote by $\Omega_{a,0}$ the subset defined by imposing on M the condition

$$0 < l_1 < 1 < l_0 < 1 + l_1$$

while keeping the above condition on \vec{v} . We define $\Omega_a = \Omega_{a,0} \cup \Omega_{a,1}$.

This definition is best understood by way of a picture. We note that $SL(2, \mathbb{Z}) \cdot M$ defines a lattice in \mathbb{R}^2 , with basis (l_0, h_1) and $(-l_1, h_0)$; the vectors \vec{v} satisfying the condition of the definition form a fundamental domain for this lattice, as shown in Figure 2.1. This fundamental domain can have its wider rectangle on the right or on the left. We partition Ω_a into two corresponding subsets $\Omega_{a,1}$ (with wider rectangle on the right) and $\Omega_{a,0}$ (with wider rectangle on the left).

We have the following proposition.

Proposition 2.1. The set Ω_a is, up to a set of measure 0, a fundamental domain for the action of Γ on G .

Proof. It is proved in [7] that each lattice of unit covolume, with the exception of a set of measure 0, has exactly one fundamental domain consisting of two rectangles of the given width and height; this is what we need.

Fig. 2.2. The Box Renormalization

In summary, a point (M, \vec{v}) in $SA(2, \mathbb{Z}) \backslash SA(2, \mathbb{R})$ has a representative in the group $SA(2, \mathbb{R})$ with coordinates $(l_0, l_1, h_0, h_1, x, y)$, satisfying the above inequalities, where the first four give the matrix entries of M ; the rows of M are the basis for a lattice in the plane; these four numbers also determine an L -shaped fundamental domain for the lattice, and (x, y) gives the coordinates of a point in that subset of the plane \mathbb{R}^2 , with the origin $(0, 0)$ located as in Fig.2.1.

Remark 2.1. It is possible to give an exact fundamental domain (rather than just up to measure zero) (cf. [7] for the details). However we do not include that here as the formulas become tedious to write down. The problem is caused by lattices which contain vertical vectors. It is easy to give examples where the fundamental domain of such a lattice is made of just one rectangle, with arbitrary width; these are degenerate cases, which go to infinity under the scenery flow. They correspond, by projection to the linear part, to geodesics on

the modular surface that go to the cusp. Such degenerate cases will come up again in the next sections.

We now analyze the scenery and translation flows with the help of this fundamental domain. For the scenery flow, the idea is the following. Begin with a lattice whose fundamental domain consists of two rectangles, the first a rectangle of width 1 and the second with width < 1 , together with a point in one of these boxes; to fix ideas, let us suppose $l_0 = 1$. The action of the scenery flow is something like the “Baker’s Transformation”: it multiplies the abscissa by $e^{t/2}$, and the height by $e^{-t/2}$. At some point we leave the domain Ω_a , either because l_1 becomes 1, or because the l_0 no longer satisfies the condition $l_0 < 1 + l_1$. One sees that the limiting condition is $l_1 = 1/2$. After leaving Ω_a , the next return to Ω_a has the following description: we return by choosing a new fundamental domain for the lattice, cutting the widest rectangle in two parts and restacking one of the parts on the thinnest rectangle. We then choose a new representative of the point in this fundamental domain (cf. Fig. 2.2).

To be precise, we give a cross-section for the scenery flow by fixing $\sup(l_0, l_1) = 1$, calling Δ_a the set defined by this condition.

Definition 2.2. We denote by $\Delta_{a,1}$ the subset of $SA(2, R)$ given by pairs (M, \vec{v}) , with $M = \begin{pmatrix} l_0 & h_1 \\ -1 & h_0 \end{pmatrix}$ and $\vec{v} = (x, y)$, where M satisfies the following inequalities :

$$0 < l_0 < 1, \quad h_0, h_1 > 0$$

and \vec{v} satisfies the inequalities :

$$-l_0 \leq x < 1 \quad \text{and} \quad 0 \leq y < h_0 \text{ if } x < 0, \quad \text{or } 0 \leq y < h_1 \text{ if } x \geq 0.$$

We denote by $\Delta_{a,0}$ the subset given by pairs (M, \vec{v}) , with $M = \begin{pmatrix} 1 & h_1 \\ -l_1 & h_0 \end{pmatrix}$, and the symmetrical conditions.

We define $\Delta_a = \Delta_{a,0} \cup \Delta_{a,1}$, and we write Φ_a for the first return map of the scenery flow to Δ_a .

The set Δ_a is in the topological boundary of Ω_a , and it is of dimension 4. Recall that to points in Ω_a we have assigned coordinates $(l_0, l_1, h_0, h_1, x, y)$; on the subset Δ_a we simplify these to (l, h, x, y, ϵ) , where $\epsilon = 0$ or 1, indicating that we are in the subsets $\Delta_{a,\epsilon}$, and where l and h are the width and height of the narrowest rectangle. The width and height of the wider rectangle are then determined, since by definition of the cross-section Δ_a the width is 1, while its height is $1 - lh$, as the fundamental domain for the lattice has area $\det M = 1$.

One can compute explicitly the return map Φ_a ; we do this in the next proposition. In the next section, we will use this stacking construction in an essential way. We mention that the formula itself is of interest for its arithmetic.

Proposition 2.2. In these coordinates, the first return map Φ_a of the scenery flow on

Δ_a is given as follows on $\Delta_{a,1}$:

- if $l < 1/2$, ϵ is not changed and

$$(l, h, x, y) \mapsto \left(\frac{l}{1-l}, (h+1-lh)(1-l), \frac{x}{1-l}, y(1-l) \right) \quad \text{if } x < 1-l,$$

$$(l, h, x, y) \mapsto \left(\frac{l}{1-l}, (h+1-lh)(1-l), \frac{x-1}{1-l}, (y+h)(1-l) \right) \quad \text{if } x \geq 1-l;$$

- if $l > 1/2$, ϵ is changed to $1-\epsilon$ and

$$(l, h, x, y) \mapsto \left(\frac{1-l}{l}, (1-lh)l, \frac{x}{l}, yl \right) \quad \text{if } x < 1-l,$$

$$(l, h, x, y) \mapsto \left(\frac{1-l}{l}, (1-lh)l, \frac{x-1}{l}, (y+h)l \right) \quad \text{if } x \geq 1-l;$$

- the case $l = 1/2$ is degenerate.

The map Φ_a is defined by similar formulas on $\Delta_{a,0}$. It preserves the Lebesgue measure (for the coordinates (l, h, x, y) on Δ_a); the total mass of this measure is infinite.

Proof. The computation comes directly from the geometry. We have depicted in Fig.2.2 the case $l > 1/2$, with the other cases being similar. The formulas on $\Delta_{a,0}$ are left to the reader; one finds the same result for the coordinates l, h, y , but the condition $x \geq 1-l$ is replaced by $x \leq l-1$, and $\frac{x-1}{1-l}$ is replaced by $\frac{x+1}{1-l}$. The Jacobian matrix of the map, at each point where it is continuous, is a triangular matrix of determinant 1, hence it preserves Lebesgue measure, and the computation shows that the total mass of Δ_a for this measure is infinite.

Fig. 2.3. The Additive Continued Fraction Map

As a consequence we have

Theorem 2.1. *The scenery flow is measurably isomorphic to the special flow built over the map $\Phi_a : \Delta_a \rightarrow \Delta_a$ (with invariant Lebesgue measure), with return time function r_Δ whose value at a point (l, h, x, y) in Δ_a is equal to $r_\Delta(l, h, x, y) = \inf(-\ln l, -\ln(1-l))$. The isomorphism from the special flow to the scenery flow is an almost surely 1-1 topological extension, and can be made a topological isomorphism by suitable identifications on the boundary of the domain of the special flow.*

Remark 2.2. Our reason for choosing this particular cross-section Δ_a is that, if we

project on the first coordinate l , we get the “tent map” f (cf. Fig. 2.3) defined by

$$l \mapsto \frac{l}{1-l} \quad \text{if } l < 1/2,$$

$$l \mapsto \frac{1-l}{l} \quad \text{if } l > 1/2,$$

which is isomorphic to the one-sided shift given by additive continued fraction expansions. Projecting the invariant measure on the first coordinate, we get the invariant Gauss measure dx/x on the interval $[0, 1]$ (which has infinite total mass). We will consider below another cross-section for which we get the more familiar transformation $l \mapsto \{1/l\}$, (the **Gauss map**), which is isomorphic to the shift for multiplicative continued fractions. This second cross-section is better for arithmetic purposes, while the first one gives simpler combinatorics; the second map is also nicer measure-theoretically as there the mass is finite, but has the disadvantage of having countably many branches. We observe that the usual Gauss map is an “acceleration” of the tent map f , in the sense that $f^{\lfloor 1/x \rfloor}(x) = \left\{ \frac{1}{x} \right\}$.

A key observation is that the map Φ_a can be considered as a skew product in several different ways, each of which leads to a different interpretation for the formulas, as will become clear in the later sections.

Thus, projecting to the first two coordinates, we get the first return map of the geodesic flow on the modular surface to a cross-section; projecting further, onto the first coordinate, we get the additive continued fraction map. One can also project on the coordinates (l, x) , obtaining then a certain two-dimensional continued fraction. The map Φ_a is in fact measurably isomorphic to the natural extension (in the sense of Rohlin, cf. [41]) of this particular continued fraction; we will return to this point later.

The first return map of the vertical translation flow is simpler to describe; it preserves each torus fiber, and it is a linear flow. A cross-section is given by $y = 0$, and we have:

Proposition 2.3. *The first return map of the vertical translation flow to the cross-section, for chosen (l_0, l_1, h_0, h_1) , is given by*

$$x \mapsto x + l_1 \quad \text{if } x < 0,$$

$$x \mapsto x - l_0 \quad \text{if } x > 0.$$

This map is just a rotation of angle l_1 on a circle of length $l_0 + l_1$. There are similar formulas for the horizontal translation flow; here we take for cross-section union of the left boundaries of the rectangles, that is the points where $x = -l_0$ and $x = 0$. We note that the first return map for the vertical flow only depends on l_0, l_1, x , while for the horizontal flow it depends on h_0, h_1 , and y .

2.2 The Domain Related to the Standard (Multiplicative) Continued Fraction

This domain Ω_m is defined in a similar way, except that, instead of imposing the condition $\sup(l_0, l_1) < 1 + \inf(l_0, l_1)$, we ask only that the width of the wider rectangle be greater than 1, and that the width of the other be less than 1, together with the further condition that the narrowest rectangle be the shortest, i.e. that $h_0 > h_1$ if and only if $l_0 > l_1$. In that case, there is a largest rectangle, which is both wider and taller than the other rectangle.

Definition 2.3. *We denote by $\Omega_{m,1}$ the domain of $SA(2, R)$ given by pairs (M, \vec{v}) , with $M = \begin{pmatrix} l_0 & h_1 \\ -l_1 & h_0 \end{pmatrix}$ and $\vec{v} = (x, y)$, where M satisfies the following inequalities :*

$$0 < l_0 < 1 \leq l_1, \quad 0 < h_0 < h_1,$$

and \vec{v} satisfies the inequalities

$$-l_0 \leq x < l_1 \quad \text{and } 0 \leq y < h_0 \text{ if } x < 0, \text{ or } 0 \leq y < h_1 \text{ if } x \geq 0.$$

We denote by $\Omega_{m,0}$ the domain defined by imposing on M the conditions:

$$0 < l_1 < 1 < l_0, \quad 0 < h_1 < h_0$$

and keeping the same condition on \vec{v} . We define $\Omega_m = \Omega_{m,0} \cup \Omega_{m,1}$.

It is proved in [7] that this is also, up to a set of measure 0, a fundamental domain for the action of $SA(2, \mathbb{Z})$ on $SA(2, \mathbb{R})$. We can once again give explicit formulas for the first return map of the scenery flow, starting from the cross-section $\sup(l_0, l_1) = 1$. The algorithm is the following. Beginning with an L -shape whose largest rectangle is of width 1, apply the scenery flow until the smallest rectangle has width 1. Then choose a new fundamental domain, by cutting the largest rectangle in subrectangles of width 1, and stacking these over the smallest one, as depicted in Figure 2.4.

We consider now a cross-section Δ_m to the scenery flow, defined to be the subset of Ω_m where $\sup(l_0, l_1) = 1$.

Fig. 2.4. The Multiplicative Box Renormalization

Definition 2.4. We denote by $\Delta_{m,1}$ the subset of $SA(2, \mathbb{R})$ given by pairs (M, \vec{v}) , with $M = \begin{pmatrix} l_0 & h_1 \\ -1 & h_0 \end{pmatrix}$ and $\vec{v} = (x, y)$, where M satisfies the following inequalities:

$$0 < l_0 < 1, \quad 0 < h_0 < h_1,$$

and \vec{v} satisfies the inequalities :

$$-l_0 \leq x < 1 \quad \text{and } 0 \leq y < h_0 \text{ if } x < 0, \text{ or } 0 \leq y < h_1 \text{ if } x \geq 0.$$

We denote by $\Delta_{m,0}$ the subset given by pairs (M, \vec{v}) , with $M = \begin{pmatrix} 1 & h_1 \\ -l_1 & h_0 \end{pmatrix}$, and the conditions :

$$-1 \leq x < l_1 \quad \text{and } 0 \leq y < h_0 \text{ if } x < 0, \text{ or } 0 \leq y < h_1 \text{ if } x \geq 0.$$

We define $\Delta_m = \Delta_{m,0} \cup \Delta_{m,1}$, and we write Φ_m for the first return map of the scenery flow on Δ_m .

One sees from the definition that $\Delta_m \subset \Delta_a$. In fact, we have that $\Delta_{m,\epsilon} = \Delta_{a,\epsilon} \cap \Phi_a(\Delta_{a,1-\epsilon})$. One easily checks that the first return map Φ_m exchanges $\Delta_{m,0}$ and $\Delta_{m,1}$; it is the induced map of Φ_a on Δ_m . More precisely, for any point p of Δ_m , $\Phi_m(p) = \Phi_a^n(p)$, where n is the smallest integer such that $\Phi_a^n(p)$ is not in the same component $\Delta_{a,\epsilon}$ as p .

Taking for coordinates (l, h, x, y) , where l, h are respectively the width and height of the smallest rectangle, and with x, y defining the translation vector, we check that $n = \lfloor \frac{1}{l} \rfloor$. One can then work out explicit formulas:

Proposition 2.4. *The first return map Φ_m of the scenery flow on Δ_m is defined on $\Delta_{m,1}$ by*

$$\begin{aligned} (l, h, x, y) &\mapsto \left(\left\{ \frac{1}{l} \right\}, l - hl^2, \frac{x}{l}, yl \right) && \text{if } x < l \left\{ \frac{1}{l} \right\}, \\ (l, h, x, y) &\mapsto \left(\left\{ \frac{1}{l} \right\}, l - hl^2, -\left\{ \frac{1-x}{l} \right\}, yl + hl + (l - hl^2) \left\lfloor \frac{1-x}{l} \right\rfloor \right) \\ &&& \text{if } x \geq l \left\{ \frac{1}{l} \right\}, \end{aligned}$$

and similarly on $\Delta_{m,0}$. This map preserves the Lebesgue measure $dl dh dx dy$, which has a finite total mass $2 \ln 2$.

Proof. The formula comes from the geometric interpretation. Since Φ_m is the induced map of Φ_a on Δ_m , it preserves the Lebesgue measure; this can also be checked directly by computing the Jacobian of the map at a continuity point. The total mass is obtained by a straightforward computation of the corresponding integral.

This tells us that we now have a model of the scenery flow which is based on the classical Gauss map:

Theorem 2.2. *The scenery flow is measurably isomorphic to the special flow built over the map $\Phi_m : \Delta_m \rightarrow \Delta_m$ (with invariant Lebesgue measure), with return time the function r'_Δ whose value on a point (l, h, x, y) in Δ_m is equal to $r'_\Delta(l, h, x, y) = -\ln l$. The isomorphism from the special flow to the scenery flow is an almost surely 1-1 topological extension, and can be made a topological isomorphism by suitable identifications on the boundary of the domain of the special flow.*

Remark 2.3. We note that the map defined in the Proposition has, once again, a nice skew product structure: the projection on the first coordinate is the continued fraction map $l \mapsto \{1/l\}$ that takes l to the fractional part of its inverse, with the first two coordinates giving a natural extension of this map, $(l, h) \mapsto (\{1/l\}, l - hl^2)$. The projection of the invariant measure to the first coordinate gives the usual Gauss measure $\frac{dx}{1+x^2}$, and this provides a way of understanding that measure.

The projection on the first and the third coordinates gives rise to a two-dimensional continued fraction map, of which the complete map is the natural extension. We mention that a closely related two-dimensional skew continued fraction has been studied by Shunji Ito in [27]. The difference between Ito's algorithm and the present one lies essentially in the choice of coordinates; one can get Ito's continued fraction by taking a different transverse surface in place of Δ . This is part of a general phenomenon; the modular flow provides a unified framework in which to understand many classical variants of continued fractions considered in the literature; thus, among others, the Usual continued fraction, Optimal continued fraction, Nearest Integer continued fraction, α -continued fraction, and Backwards continued fraction and their natural extensions can be obtained by choosing an appropriate cross-section of the modular flow. See [30] for a general presentation.

Which cross-section one should take depends on one's viewpoint, and on what use one wants to make of the resulting formulas. Thus, as we will see below, a choice which may be more natural for its symbolic dynamics or geometry can lead to quite obscure arithmetic formulas, and vice versa.

We mention that some even more exotic continued fractions (for example the Rosen continued fraction, cf. [5]) are also related to flows, but on different spaces: in that case, it is the geodesic flow on a Hecke surface.

§3. Symbolic Dynamics for the Scenery Flow: Markov Coding

3.1. Overview

Our aim in this and the next section is to give symbolic descriptions of the scenery flow and the related translation flows. The basic idea is to consider a cross-section and its associated return map, and to obtain symbolic dynamics by choosing an appropriate partition of the cross-section. The remarkable fact is that in doing this we shall make use of the different flows, obtaining in that way very different symbolic systems. We described earlier the commutation relations which are satisfied by the scenery and translation flows; here, we will recover these commutation relations at the symbolic level.

In short, we will associate to the scenery flow a subshift of finite type, for which the natural measure will be infinite, and to the vertical flow the shift on the collection of all Sturmian sequences; this last set, to be defined precisely in §4.2, can be considered as a generalization of substitution minimal sets, and gives rise to finite measure, uniquely ergodic systems of entropy zero.

Each of these two symbolic descriptions is particularly suited to one of the flows, in the sense that the first return map of the flow is conjugate to the shift on the related set; however, in each case the other flow also enters the picture. In the framework of the subshift of finite type, the return map of the scenery flow appears as the (bilateral) shift, while the return map of the vertical translation flow appears as an adic transformation (in the sense of Vershik, cf. [49]) on the unilateral shift. In the framework of Sturmian sequences, the return map of the vertical translation flow appears as the shift, while the return map of the scenery flow now appears as the natural extension of the coding map on Sturmian sequences defined in §4.4. We will obtain in this way a completely combinatorial model, using only classical notions of theoretical computer science.

Some points need special attention: first, of course, the geometric meaning of the shift map depends entirely on the symbolic space on which it acts, that is, on which subset of the full shift is defined by the allowed words. Second, we use different cross-sections for the scenery flow and the translation flows, and this difference will appear at the symbolic level as well. This explains why, for example, the scenery flow gives a *bilateral* shift, while the vertical flow gives an adic transformation on the corresponding *unilateral* shift.

The plan of this section and the next one is as follows: in §3.2, we give symbolic dynamics for the first return map of the scenery flow on the additive section Δ_a , as a subshift of finite type on four symbols. In §4.1, we define, in an informal way, another symbolic description of the additive cross-section, using the translation flows; this will help motivate the three parts which follow. In §4.2, we give symbolic dynamics for the vertical translation flow on each torus fiber. As mentioned before, this will amount to giving symbolic dynamics for circle rotations. In §4.3, we use the combinatorial properties of the sequences obtained there (Sturmian sequences) to define a coding map, to which we associate a symbolic dynamics. In §4.4, we show that the natural extension of this coding map is conjugate to the map Φ_a . This gives a purely combinatorial definition of the coding defined in §4.1, and shows the relation with the initial coding of §3. In §4.5, we explain how we can express the first return map of the vertical flow as an adic transformation (generalized odometer) on a nonstationary

subshift of finite type (cf. [3]). In §4.6, we sketch the analogous results for the map Φ_m , first return map of the scenery flow on the multiplicative section Δ_m : this map is conjugate to a subshift of finite type on a countable number of symbols. We write down the conjugacy explicitly later, in §6.

Remark 3.1. As often happens when symbolic dynamics is introduced for a dynamical system, one gets problems for coding particular points, since the original space is continuous while the topology of the symbolic space is totally disconnected, i.e. is a Cantor set. We encounter two different problems of this type, one related to the dynamics on the base space, and the other related to the dynamics on individual fibers.

First, there is a problem caused by the pairs of rectangles with commensurable heights or widths. (These correspond to fibers in which one of the translation flows has only periodic orbits, and in the base space to orbits of the geodesic flow that go to the cusp at infinity.) It is possible to include these points, but one needs then in the combinatorics to treat particular cases, which would substantially complicate the picture.

Hence we will always make the simplifying assumption in what follows that **the pairs of rectangles have incommensurable heights and widths**; formally, instead of the set $E = SA(2, \mathbb{Z}) \backslash SA(2, \mathbb{R})$, we consider the subset E^* of points $(l_0, l_1, h_0, h_1, x, y)$ such that l_0/l_1 and h_0/h_1 are irrational. This removes an invariant set of measure 0.

Second, some ambiguities arise when coding points in the orbit of the boundary of the rectangles; this is also a set of measure 0, on which the coding is not unique. We identify explicitly the symbolic sequences corresponding to these boundary points.

3.2. Symbolic Dynamics for the Scenery Flow

The cross-section Δ_a we defined in the preceding section admits a natural partition in two sets $\Delta_{a,0}$ and $\Delta_{a,1}$. This is however not a generating partition for the map Φ_a , since it only depends on the first coordinates. Therefore, if $p = (l, h, x, y, \epsilon)$ and $p' = (l, h, x', y', \epsilon)$ are points of Δ_a that differ only in the x, y coordinates, the images $\Phi_a^n(p)$ and $\Phi_a^n(p')$ will belong to the same partition element for all $n \in \mathbb{Z}$. In other words, this partition does not separate the points in a torus fiber. It does however separate the different fibers; by projecting this partition to $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$, one can prove the following:

Proposition 3.1. *The first return map, to the projection of Δ_a , of the geodesic flow on the modular surface, restricted to the orbits that are not asymptotic to the cusp (in the past and in the future), is topologically conjugate to the shift on the subset of $\{0, 1\}^{\mathbb{Z}}$ consisting in sequences that are not eventually constant (in the past and in the future).*

Via this conjugacy the measure on the flow space corresponds to the natural Lebesgue measure on the cross-section, which in turn maps isomorphically to an infinite invariant measure on the shift space.

We now define a finer partition which will generate.

Definition 3.1. *We write $\Delta_{a,0}^0$ (resp. $\Delta_{a,0}^1$) for the set $\{(l, h, x, y, 0) \in \Delta_a | x < 0\}$ (resp. $x \geq 0$). We denote by $\Delta_{a,1}^0$ (resp. $\Delta_{a,1}^1$) the set $\{(l, h, x, y, 1) \in \Delta_a | x < 0\}$ (resp. $x \geq 0$).*

In the geometric model, $\Delta_{a,0}^0$ corresponds to pairs of rectangles where the left rectangle is the widest, while the point corresponding to (x, y) is in the left rectangle; for $\Delta_{a,0}^1$, the point is in the right rectangle. Note that, by our convention, the central boundary of the partition (corresponding to $x = 0$) is considered to be a part of $\Delta_{a,0}^1$ or $\Delta_{a,1}^1$, not $\Delta_{a,0}^0$ or $\Delta_{a,1}^0$.

Write \mathcal{A} for the alphabet on four letters $\mathcal{A} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, and denote by ν the map $\nu : \Delta_a \rightarrow \mathcal{A}$ that takes a point in $\Delta_{a,0}^0$ (resp. $\Delta_{a,0}^1, \Delta_{a,1}^0, \Delta_{a,1}^1$) to $(0, 0)$ (resp.

$(0, 1), (1, \mathbf{0}), (1, 1)$). The sequence $(\nu(\Phi_a^n(p)))_{n \in \mathbb{Z}}$ gives our symbolic dynamics for the map Φ_a . We have

Theorem 3.1. *Let (Λ, S) be the subshift of finite type in $\mathcal{A}^{\mathbb{Z}}$ defined by the conditions : $(0, 1)$ is not followed by $(0, \mathbf{0})$ or $(1, \mathbf{0})$, and $(1, \mathbf{0})$ is not followed by $(1, 1)$ or $(0, 1)$. The map $p \mapsto (\nu(\Phi_a^n(p)))_{n \in \mathbb{Z}}$ is one-to-one from E^* to an explicit shift-invariant subset of Λ , and conjugates Φ_a to the shift S on Λ .*

Proof. We first note that the sequences obtained from Φ_a are indeed in Λ . The reason is that, on $\Delta_{a,0}$, the map Φ_a stacks a part of the left rectangle on the right rectangle; hence, if the point was initially in the right rectangle, its image will also be in the right rectangle. Symmetrically, in $\Delta_{a,1}$, a point in the left rectangle will stay in the left rectangle. These are exactly the conditions that define Λ .

We next prove that the partition is generating. The idea is to first notice that the itinerary of the orbit with respect to the partition $\{\Delta_{a,0}, \Delta_{a,1}\}$ defines the size of the two rectangles: the positive orbit defines the width of the rectangles, while the negative orbit defines the height; hence, if two points in Δ_a have the same symbolic dynamics, they have the same first two coordinates. Next, note that the map Φ_a is locally a dilation on the x coordinate; thus, if two points have different x coordinates, then after sufficient iteration, the difference will be increased to more than 1, at which time they will not belong to the same element of the partition. The same holds for the y coordinate, using the inverse of Φ_a . Hence the partition generates. In §6, we give complete arithmetic formulas for the inverse map (recovering the geometric coordinates from the symbolic dynamics), and this can be used to give a second, and more constructive, proof of this fact.

This map is not onto: we cannot get the ultimately constant sequences, because in that case we would always be stacking in the same direction, and this is only possible if one of the rectangles has width (or height, for a sequence that is constant for all $n < N$) zero. In fact, sequences that are ultimately constant in the future (resp. in the past) correspond to the pairs of rectangles with commensurable widths (resp. heights) that we have excluded. We also cannot get sequences which contain ultimately only $(0, \mathbf{0})$ and $(1, \mathbf{0})$; this would mean that, after some time, the point (x, y) is always on the left. This is possible only if $x = 0$, and by our convention, as noted above, the point then belongs to the right rectangle, and the corresponding sequence only contains $(0, 1)$ and $(1, 1)$. (We could consider such sequences as a second (not admissible) coding for points on the central boundary.) These restrictions only remove a small set in Λ .

Except for these bad points, we get all of Λ . One way to prove this is to show that the above partition is a Markov partition for Φ_a ; this involves easy, but somewhat tedious, computations. A second way involves proving that the projection of the partition on the coordinates (l, x) gives a Markov partition in the standard sense for the noninvertible projected map, hence all allowed strings occur, and then using the fact that Φ_a is the natural extension. We give a similar (and easier) explicit proof in §4.4.

We remark that this coding is well behaved with respect to the projection to the base space $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$: if we project the alphabet \mathcal{A} to $\{0, 1\}$ by forgetting the second letter of each pair, we recover the symbolic dynamics for the geodesic flow that was given in the above proposition.

§4. Symbolic Dynamics for the Scenery Flow: Sturmian Sequences and Adic Transformations

4.1. Another Symbolic Description of the Scenery Flow

We can describe in another way the points of E^* by a pair of symbolic sequences. Let p be a point in $\Delta_a \cap E^*$, with coordinates $= (l_0, l_1, h_0, h_1, x, y)$. We consider the positive orbit $\{T_u^+(p) | u \geq 0\}$ of p under the vertical flow. In the model given in §2 (tiling of the plane), this orbit consists in vertical segments contained in the two boxes; we write $(u_0, u_1, \dots) \in \{0, 1\}^{\mathbb{N}}$ for the sequence that describes the order in which the orbit crosses the two rectangles. In a similar way, let $(v_n)_{n \in \mathbb{N}}$ be the sequence that describes the order in which the orbit $T_u^-(p)$ of p under the horizontal flow crosses the two rectangles (cf. Fig. 4.1). We claim that these two sequences completely define the point p .

Fig.4.1. The Horizontal and Vertical Orbits

An informal proof goes like this: suppose that the frequency of 1 in u is well defined; this frequency is equal to the ratio of the width of rectangle 1 to the total width; but since the width of the widest rectangle is 1, because the point is in Δ_a , both widths are completely determined. Moreover, because the ratio is irrational, by definition of E^* , the sequence u defines the coordinate x . (This would be false if the ratio of width were rational: in that case, the orbits of the vertical flow would be closed, so there would only be a finite number of possible sequences u , all of them periodic, and the sequence u would not specify the coordinate x .) In a similar way, the sequence v defines the ratio of heights, hence, using the fact that the total area is 1, it completely defines the heights, and also the coordinate y .

We will give a formal proof of this in the next parts, and explicit arithmetic formulas in §5. We just remark here that the sequence u can be described as symbolic dynamics of the first return map of the vertical flow to the cross-section defined in §2, with respect to the natural partition into right and left rectangles; we only need to project the point p to the point $(l_0, l_1, h_0, h_1, x, 0)$ in the cross-section for the vertical flow.

The action of the scenery flow (or more exactly of its return map) on these symbolic coordinates, the pair of sequences (u, v) , is not the shift map. Indeed, the sequences u and v

are not changed locally by that flow, but only by the stacking operation; if we suppose that p is in $\Delta_{a,1}$, the rightmost part of rectangle **1** is stacked onto rectangle **0**, and we see that the effect on the sequence u is a deflation, erasing each **1** following a **0**, while the effect on v is an inflation, replacing each **1** by **10**. (A more subtle question is what will be the first letter of the new word; we study that in detail in §4.3 and §4.4.)

The purpose of the next three parts is to give precise definitions and proofs; we will see that this can be done in a purely combinatorial way, although it can help to keep in mind the geometry when considering the special cases that occur in some proofs. (These correspond to the boundaries of the rectangles and their iterates.)

4.2. Sturmian Sequences: Basic Definitions

In this section, we build symbolic dynamics for the vertical flow. Let us consider the restriction of this flow to a torus fiber. It is clear from the formula that the dynamics depends only on x , l_0 , and l_1 . The first return map is a rotation, as noted above, with its domain naturally divided into two intervals, which we label **0** and **1**. We associate to the point x the sequence $(\epsilon_n)_{n \in \mathbb{Z}}$, where ϵ_n is the symbol **0** or **1** of the interval that contains the n th iterate of x . This gives natural symbolic dynamics for the vertical flow: if we denote by S the shift map on infinite sequences, which sends the sequence u to the sequence v defined by $v_n = u_{n+1}$, it is clear from the definition that the shift map S is topologically semiconjugate to the first return map. (That is, there is a continuous onto map from the symbolic space to the circle such that the diagram commutes). The sequences we get are well-known (cf. [23, 22, 15]); we recall some definitions and facts. We shall refer to the literature for some of the proofs.

In this section, we normalize the rotation to the circle \mathbb{R}/\mathbb{Z} of length 1, which we shall parametrize by $[0, 1[$ or $]0, 1]$. (The original notation will prove more convenient in §6). We write R_α for the rotation $x \mapsto x + \alpha \pmod{1}$. We first give a precise definition for the sequences described above.

Notations 4.1. Let $\alpha \in [0, 1[$. To code the rotation R_α symbolically, it will be convenient to partition the circle \mathbb{R}/\mathbb{Z} in two intervals determined by the points 0 and $1 - \alpha$ (preimage of 0 by the rotation of angle α). There are two consistent ways to carry this out, and it will be useful to keep track of both possibilities. Denote by I_0 the interval $[0, 1 - \alpha[$, and by I_1 the interval $]1 - \alpha, 1[$. In the same way, we denote by J_0 the interval $]0, 1 - \alpha]$, and J_1 the interval $]1 - \alpha, 1]$, the only difference between I_a and J_a being whether it is closed on the left or the right. We call I (resp. J) the map with values in $\{\mathbf{0}, \mathbf{1}\}$ and defined by $I(x) = \mathbf{a}$ if $x \in I_a$ (resp. $J(x) = \mathbf{a}$ if $x \in J_a$). We write $\mathcal{I} = \{I_0, I_1\}$ resp. $\mathcal{J} = \{J_0, J_1\}$ for these partitions of the circle.

Definition 4.1. We say that a sequence $(u_n)_{n \in \mathbb{N}}$ taking values in $\{\mathbf{0}, \mathbf{1}\}$ is **generated** by the rotation R_α if there exists a point β such that either for all n , $u_n = I(R_\alpha^n(\beta))$ or for all n , $u_n = J(R_\alpha^n(\beta))$.

In other words, the sequence is given as the itinerary of a point for R_α with respect to the partition. Since there is no canonical way to code the endpoints, we allow all consistent codings, choosing them so as to be all closed either on the right or on the left. Note that points in the past orbit of 0, and only those (remember that the coding u_n is defined for times in \mathbb{N} and not in \mathbb{Z}), will have two possible codings.

Notation 4.2. We will say that the sequence u is a **rotation sequence** if (and only if) it is generated by some rotation R_α . An equivalent definition is that there exist numbers $\alpha, \beta \in [0, 1[$ such that for all n either we have $u_n = \lfloor (n+1)\alpha + \beta \rfloor - \lfloor n\alpha + \beta \rfloor$, or we have

$u_n = [(n+1)\alpha + \beta] - [n\alpha + \beta]$; these are the sequences given by partitions \mathcal{I}, \mathcal{J} respectively.

A remarkable fact is that these rotation sequences can be given a completely combinatorial description. To explain this we first need a few definitions.

Definition 4.2. Let $u = (u_n)_{n \in \mathbb{N}}$ be a sequence taking values in a finite alphabet \mathcal{A} . The language \mathcal{L}_u of u is the set of all the finite words $u_i u_{i+1} \cdots u_{i+k}$ that occur in u . The complexity $p(n)$ of u is the function that counts the number of words of length n occurring in u : $p(n) = \#(\mathcal{L}_u \cap \mathcal{A}^n)$.

Lemma 4.1. If the complexity of the sequence u satisfies $p(n) \leq n$ for some n , then u is eventually periodic.

Proof. The given condition implies that for some n we have $p(n) = p(n+1)$, since $p(1)$ is at least 2 for a non-constant sequence; p is non-decreasing because each word occurring in u can be extended. Thus, each word of length n can be extended in exactly one way. This means that, knowing n letters, we then will know the next one. But, since there is only a finite number of admissible words of length n , one of these will occur at a second location in u ; hence from the first occurrence of that word on, the sequence is periodic.

We will be interested in the non-periodic sequences of minimal complexity. By the lemma, these are the sequences of complexity $p(n) = n + 1$.

The basic fact, known to Hedlund and Morse (cf. [23, 22, 15]), is that there are three equivalent characterizations of these sequences, two purely combinatorial and the third geometrical. We need first this:

Definition 4.3. A sequence u taking values in a finite alphabet is said to be **balanced** if for any two words U and V of the same length occurring in u and for any element \mathbf{a} in the finite alphabet, the number of \mathbf{a} 's in U and V differ by at most 1.

Proposition 4.1. The following are equivalent for a sequence $(u_n)_{n \in \mathbb{N}}$ on two symbols :

- (i) the sequence has complexity $p(n) = n + 1$;
- (ii) it is a non-eventually-periodic balanced sequence ;
- (iii) it is generated by a rotation R_α for some α irrational.

For logical definiteness we take the first of these conditions as our definition:

Definition 4.4. A sequence satisfying (i) above is called a **Sturmian sequence**.

Proof of Proposition 4.1. The proof of (i) \Leftrightarrow (ii) is combinatorial; it can be found in [23].

We show (iii) \Rightarrow (i). Let u be a rotation sequence associated to α, β with α irrational and write U for a word $U_0 U_1 \cdots U_{n-1}$ of length n . This occurs at the beginning of the sequence u if and only if β belongs to all the intervals $R_\alpha^{-k}(I_{U_k})$ (or similarly with I replaced by J); more generally the word U can occur somewhere in the sequence u if and only if this intersection is not empty. It follows that there are as many possible words of length n in a rotation sequence as elements in the join of the partitions $\mathcal{I}, R_\alpha^{-1}\mathcal{I}, R_\alpha^{1-n}\mathcal{I}$. But this finer partition is determined by the orbits of the endpoints of the intervals, the points $0, 1 - \alpha = R_\alpha^{-1}(0), \dots, R_\alpha^{-n}(0)$. Since α is irrational these are distinct, and so there are $n + 1$ endpoints and hence $n + 1$ components. Thus we have $p(n) = n + 1$, proving (i).

We next give a proof of (iii) \Leftarrow (ii), which is quite short, and which helps show the importance of the second combinatorial condition. We count the number of 1's in the word of length p of a rotation sequence, beginning with u_n ; this is $u_n + \cdots + u_{n+p-1} = [(n+p)\alpha + \beta] - [n\alpha + \beta]$, and this number, as a function of n , can take on only two values. One concludes that a rotation sequence is balanced.

Proving (iii) from the other statements is more difficult; a detailed proof is given in [23].

A proof can also be obtained from the ideas developed below: a Sturmian sequence can be infinitely recoded in a similar way to the continued fraction; one can then prove that this Sturmian sequence is one of the rotation sequences associated to the number α defined by that continued fraction. See §6 for explicit arithmetic formulas, which yield a constructive proof of the proposition.

By (iii) and Weyl's equidistribution theorem it follows that for a Sturmian sequence u the limiting frequency of any symbol in the sequence exists. An (elementary but not altogether easy) exercise is to prove this directly from part (ii), the property of being balanced.

We will be interested in the set of all Sturmian sequences, which we will denote by Σ . This forms a subset of the full one-sided symbol space $\{0, 1\}^{\mathbb{N}}$. As is usual, we give this space the product topology, with respect to which it is homeomorphic to the Cantor set.

Remark 4.1. The set Σ is not a closed subset of the space $\{0, 1\}^{\mathbb{N}}$. It is contained in the set of all balanced sequences, which is closed (if a sequence is not balanced, this can be checked on a finite subsequence, so a non-balanced sequence cannot be a limit of balanced sequences). One can prove that every balanced sequence is a limit of Sturmian sequences, hence the closure of Σ is the set of all balanced sequences; what needs to be added is the countable set of all the eventually periodic balanced sequences.

We recall the shift dynamics on the space $\{0, 1\}^{\mathbb{N}}$: the image Su of a sequence u is the sequence v such that, for all integers $n \geq 0$, $v_n = u_{n+1}$. The subset Σ inherits this dynamics.

We are also interested in a much smaller subset of $\{0, 1\}^{\mathbb{N}}$. Given a Sturmian sequence u , we write $\Omega = \overline{\{S^n u\}}$ for the closure of the orbit of u under the shift S . The next proposition shows these are the smallest shift-invariant subsets of Σ .

Proposition 4.2. *Let Ω be the orbit closure of a Sturmian sequence. The dynamical system (Ω, S) is minimal, and S is one-to-one on Ω except for a single point which has two preimages.*

Proof. If the sequence is Sturmian, by (ii) of Proposition 4.1 it is not eventually periodic; the shifted sequence keeps this property, and is clearly balanced. Certainly all its shifted sequences share this property. By (ii) \Rightarrow (i) of Proposition 4.1 therefore, since $p(n) = n + 1$, every word that occurs in the sequence occurs an infinite number of times. This implies that every word that occurs in the sequence can be extended in at least one way on the left. Because there are $(n + 1)$ words of length n (for each n), there is exactly one word L_n that can be extended on the left in two ways. The word L_n is a prefix of L_{n+1} , and this sequence of finite length words increases to a unique infinite word l . This word has two preimages in Ω , with all the others having exactly one preimage.

The proof of minimality is more difficult; we do not give it now, as it will follow from the recoding given below.

We remark on the difference between Ω and the set Σ of all Sturmian sequences: as one can show, Ω consists of Sturmian sequences where 1 occurs with a fixed given frequency, while for Σ all possible irrational frequencies occur. This explains why, unlike the closure of Σ , Ω does not contain periodic sequences. It follows from the proposition that the set Σ is an uncountable disjoint union of the closed sets Ω ; we note that Σ is not itself closed.

Remark 4.2. One of the implications of Proposition 4.2 is that one could instead work with biinfinite sequences, indexed by \mathbb{Z} , since a countable number of sequences (in the positive orbit of 0) can be extended on the left in two ways, while all the others can be extended in only one way. Another way to understand this is: since this shift is a zero entropy system, the future almost surely determines the past. For clarity of exposition, it

is more convenient to work for the moment with one-sided sequences, however we will in §5 use biinfinite sequences.

Definition 4.5. We will call the unique sequence that has two preimages under the shift on Ω the **special word** or the **special sequence** of the system Ω .

This notion has a geometric interpretation, contained in the following:

Proposition 4.3. The dynamical system (Ω, S) is an almost surely 1 – 1 topological extension of a rotation (S^1, R_α) , via the coding given by the two natural partitions \mathcal{I} and \mathcal{J} ; the special word corresponds to the image α of the point 0, which has a non-ambiguous coding, while each of its preimages by the rotation has two possible codings.

The quotient of the set Ω by the equivalence relation that identifies for each $n \geq 1$ the two preimages of order n of the special word is, with its natural quotient topology, homeomorphic to the circle.

We remark that the minimality of an irrational circle rotation can be proved in two ways, therefore, as a consequence of Proposition 4.2 via this continuous projection to the circle, or directly from the geometry by an easy compactness argument; however the converse argument would not work, that is, the minimality of the map does not lift from the circle to the space Ω .

4.3. Recoding Sturmian Words: the Coding Map

Since we will be using sequences of several different types, we will from now on speak of Sturmian *infinite words*, or Sturmian *words* for short, instead of Sturmian sequences, and reserve the term *sequence* for the coding sequences to be defined below.

Definition 4.6. A **substitution** on the alphabet $\{0, 1\}$ is a map from $\{0, 1\}$ to the set $\{0, 1\}^*$ of finite words on the alphabet $\{0, 1\}$ that sends each letter to a non-empty word. It extends naturally to finite words and infinite sequences on this alphabet, replacing each letter by the corresponding word.

Recall that the set $\{0, 1\}^*$ admits a natural structure of free monoid for the concatenation operation; in this setting, a substitution appears as a non-erasing morphism of the free monoid.

In this section we show how to recode a Sturmian word on the alphabet $\{0, 1\}$ to another Sturmian word, using a substitution, the choice of which depends on the initial word. We then iterate this process, keeping track of the different substitutions used. This produces a sequence (the coding sequence) on another alphabet. We next show that the set of coding sequences is a subshift of finite type, and that the correspondence between Sturmian words and the associated coding sequences is almost surely one-to-one. Finally we recover, in a completely combinatorial way, the symbolic dynamics for the scenery flow described in §3.

Let u be a Sturmian word; from the definition, there are only three words of length 2 occurring in u . Since **01** and **10** must occur (for otherwise u would be eventually constant), this means that exactly one of the words **00** or **11** does not occur. (Geometrically this corresponds, of course, to having rotation angle $\alpha \in]\frac{1}{2}, 1[$ and $]0, \frac{1}{2}[$ respectively.)

Definition 4.7. A Sturmian word will be said of **type 0** if **11** does not occur, and of **type 1** if **00** does not occur.

Notation 4.3. We will denote by Σ_0 (resp. Σ_1) the set of Sturmian words of type 0 (resp. 1), and Σ^0 (resp. Σ^1) the set of Sturmian words that begin with **0** (resp. **1**); we will denote by Σ_0^0 (resp. Σ_0^1) the set of Sturmian words of type 0 that begin with **0** (resp. with **1**), with the corresponding notation for words of type 1.

Suppose that the word u is of type 0. Then every **1** is followed by a **0**, so we can certainly

recode u by using as new symbols the words $\mathbf{0}$ and $\mathbf{10}$. The interesting fact is that, possibly after removing the first letter, the word thus obtained will again be Sturmian. The next proposition states this precisely. We need first :

Defnition 4.8. We denote by σ_0, σ_1 the two substitutions, defined from $\{\mathbf{0}, \mathbf{1}\}$ to $\{\mathbf{0}, \mathbf{1}\}^*$ by

$$\begin{aligned}\sigma_0 : \quad \mathbf{0} &\mapsto \mathbf{0}, \\ &\mathbf{1} \mapsto \mathbf{10}; \\ \sigma_1 : \quad \mathbf{0} &\mapsto \mathbf{01}, \\ &\mathbf{1} \mapsto \mathbf{1}.\end{aligned}$$

Lemma 4.2. *If v is not balanced, then for any letter \mathbf{a} , $\sigma_0(\mathbf{a}v)$ is not balanced.*

Proof. We suppose that v is not balanced. It is not hard to prove that we can then find two words U and V of the same length and containing the same number of $\mathbf{0}$'s, such that the words $\mathbf{0}U\mathbf{0}$ and $\mathbf{1}V\mathbf{1}$ occur in v . Because $\mathbf{0}U\mathbf{0}$ occurs in v , there exists \mathbf{b} such that $\mathbf{b0}U\mathbf{0}$ occurs in $\mathbf{a}v$ (the only problem is the case when it is the initial word; this is why we needed \mathbf{a}). Because $\sigma_0(\mathbf{b})$ ends with $\mathbf{0}$ in all cases, the words $\mathbf{00}\sigma_0(U)\mathbf{0}$ and $\mathbf{10}\sigma_0(V)\mathbf{10}$ occur in $\sigma_0(\mathbf{a}v)$. But now it is clear that $\sigma_0(U)$ and $\sigma_0(V)$ have the same length and the same number of $\mathbf{0}$'s; hence the two words $\mathbf{00}\sigma_0(U)\mathbf{0}$ and $\mathbf{10}\sigma_0(V)\mathbf{1}$ occur in $\sigma_0(\mathbf{a}v)$; they have the same length; and their number of $\mathbf{0}$'s differs by two.

We can now explain how one recodes Sturmian words.

Proposition 4.4. *Let u be a Sturmian word of type 0.*

(i) *If u is not a special word, then either $u = \sigma_0(v)$, where v is a Sturmian word, or $u = S\sigma_0(v)$, where v is a Sturmian word that starts with $\mathbf{1}$ (but not both).*

(ii) *If u is a special word, we can write both $u = \sigma_0(v_0)$ and $u = S\sigma_0(v_1)$, where $Sv_0 = Sv_1$ is a special Sturmian word.*

If u is of type 1, the same property holds, by exchanging the symbols $\mathbf{0}$ and $\mathbf{1}$.

Proof. It is immediate that if u is of type 0 then we can express it in a unique way as $u = \sigma_0(v)$, as noted above.

(i) Suppose that u is not a special word. Then exactly one of $\mathbf{0}u$ and $\mathbf{1}u$ is Sturmian.

Supposing now that $\mathbf{0}u$ is Sturmian, we can write $\mathbf{0}u = \sigma_0(v')$, where v' starts with $\mathbf{0}$. Taking $v = Sv'$, then by the preceding lemma v is Sturmian, and $u = \sigma_0(v)$. If u begins with $\mathbf{1}$, then we certainly cannot write $u = S\sigma_0(v')$, with v' beginning with $\mathbf{1}$. If u begins with a $\mathbf{0}$, then v also begins with a $\mathbf{0}$. We can then replace the first letter of v by a $\mathbf{1}$, getting a new word w , and can then write $u = S\sigma_0(w)$. But if w is also Sturmian, since $Sw = Sv$, then Sv is a special word; it is easy to check that this then implies that u is also a special word.

Suppose next that $\mathbf{1}u$ is Sturmian. Then its preimage is $\mathbf{01}u$, since $\mathbf{1}$ is isolated, and we can write $\mathbf{01}u = \sigma_0(\mathbf{0}v)$, where v is Sturmian and begins with $\mathbf{1}$. It is then clear that $u = S\sigma_0(v)$, where v is Sturmian and begins with $\mathbf{1}$. As before, we can write $u = \sigma_0(w)$ where w is the word obtained from v by replacing the first letter by a $\mathbf{0}$. But again, if this word w is Sturmian, then u is a special word.

(ii) Suppose next that u is a special word of type 0. Then its preimages are $\mathbf{01}u = \sigma_0(\mathbf{01}v)$ and $\mathbf{10}u = \sigma_0(\mathbf{10}v)$, since u must begin with a $\mathbf{0}$. The two words $\mathbf{0}v$ and $\mathbf{1}v$ are Sturmian, by the lemma, so v is a special word; and we clearly have that $u = \sigma_0(\mathbf{0}v) = S\sigma_0(\mathbf{1}v)$.

This same proof works for words of type 1, by using σ_1 in place of σ_0 .

A direct consequence of this proposition is that second preimages of special words (under the shift map) recode to each other.

Corollary 4.1. *Let v be a special word of type 0 (respectively type 1). Then there is a special word v' such that $\mathbf{10}v = \sigma_0(\mathbf{10}v')$ and $\mathbf{01}v = \sigma_0(\mathbf{01}v')$ (resp. $\mathbf{10}v = \sigma_1(\mathbf{10}v')$ and $\mathbf{01}v = \sigma_1(\mathbf{01}v')$).*

We see that these second preimages of special words play an important role, and they deserve a name.

Definition 4.9. *The two shift preimages of order 2 of the special word of a Sturmian system are called the **fixed words** of the system.*

We will see later that there is an easy algorithm (Rauzy rules) to compute an arbitrary long prefix of a fixed word, and that algorithm will explain the reason for this terminology.

The content of the preceding corollary is that the two fixed words of a system recode to the two fixed words of (another) system, or equivalently, the image of a fixed word by either one of the substitutions σ_0, σ_1 is a fixed word, of type 0 or 1 respectively, which begins with the same letter.

We can use Proposition 4.4 to define a transformation on the space Σ .

Definition 4.10. *We define the **coding map** $\Phi: \Sigma \rightarrow \Sigma$ to take a Sturmian word u of type 0 (resp. 1) to the unique word v such that $u = \sigma_0(v)$ (resp. $u = S\sigma_1(v)$), or, if such a word does not exist, to the unique word v such that $u = S\sigma_0(v)$ (resp. $u = \sigma_1(v)$).*

Note that, in this definition, in case of ambiguity, we have made a choice, taking as image the word whose initial letter is $\mathbf{0}$.

Proposition 4.4 ensures that Φ is well-defined; its definition entails an arbitrary choice for the special word. It is easy to check that Φ is a 3 – 1 map; the partition $\{\Sigma_0^0, \Sigma_0^1, \Sigma_1^0, \Sigma_1^1\}$ defined above is a Markov partition. The map Φ is one-to-one from Σ_0^0 to Σ and from Σ_1^1 to Σ ; it is one-to-one from Σ_0^1 to Σ^1 , and from Σ_1^0 to Σ^0 .

We will use this partition to associate, to any Sturmian word, the itinerary of its orbit by Φ .

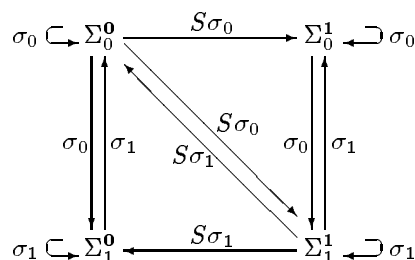


Fig.4.2. The Transition Graph for Additive Coding

Definition 4.11. *The **coding sequence** for a Sturmian word u is the sequence which takes values in the set $\mathcal{A} = \{(0, \mathbf{0}), (0, \mathbf{1}), (1, \mathbf{0}), (1, \mathbf{1})\}$, obtained by recoding u an infinite number of times by the map Φ .*

The symbolic dynamics thus defined can be explained as follows: beginning with a Sturmian word u , we find a new Sturmian word $u^{(1)}$. We can then iterate the process, recoding $u^{(1)}$. In this way we obtain an infinite sequence of Sturmian words $u^{(i)} = \Phi^i(u)$ such that $u^{(i)} = \sigma_a u^{(i+1)}$ or $u^{(i)} = S\sigma_a u^{(i+1)}$.

We remark on the choice we made in defining Φ . If one of the words $u^{(i)}$ is special, we have a choice for the recoding; these are the two preimages of a special word, and Φ in that case

chooses the word that starts with $\mathbf{0}$. At the next step, one of the recoded words is a fixed word, and then from this point on, the choice of recoding is completely determined (by the above corollary). As is easily seen, the alternative sequence of words consists of preimages of special words for a finite time, and then (when the type of the recoding substitution changes), of fixed words.

It is clear that a special word can appear in the process if and only if we start with a word in the positive orbit of a special word by the shift; in this case, there are two possible sequences of words, and both, after a finite time, consist of fixed words. The sequence of words obtained by applying Φ consists eventually of fixed words starting with $\mathbf{1}$. (At the first step, we obtain a word that begins with $\mathbf{0}$, but after a finite number of steps, we get a word beginning with $\mathbf{1}$.)

It is natural to ask whether the partition generates, and what are the admissible coding sequences (that is, the coding sequences obtained by Φ).

Theorem 4.1. (i) *The admissible coding sequences are contained in a one-sided subshift of finite type, on the alphabet \mathcal{A} , defined by the following condition :*

$(0, \mathbf{1})$ is not followed by $(0, \mathbf{0})$ or $(1, \mathbf{0})$ and $(1, \mathbf{0})$ is not followed by $(0, \mathbf{1})$ or $(1, \mathbf{1})$.

They can be obtained as sequences of vertices for paths in the graph shown in Fig. 4.2, where we have represented all the possible transitions, labelling each edge by the corresponding transformation.

(ii) *all sequences in this finite type subshift are obtained, except for*

—the eventually constant sequences,

—the sequences of the form $U(0, \mathbf{0})(0, \mathbf{1})^a(1, \mathbf{1})v$ or $U(1, \mathbf{1})(0, \mathbf{1})^a(1, \mathbf{1})v$, where U is any admissible word on the alphabet \mathcal{A} , v is a sequence that contains only the symbols $(0, \mathbf{0})$ and $(1, \mathbf{0})$, and $a \in \mathbb{N}$ is an integer.

(iii) *The partition almost generates, that is, it separates all Sturmian words, except for the pairs of preimages of order k , $k > 2$, of special words, whose coding sequence consists eventually of blocks of type alternatively $(0, \mathbf{0})^n$ and $(1, \mathbf{1})^n$.*

Proof. (i) The fact that all admissible sequences are of the given type is clear: if a word is of type 0 and begins by $\mathbf{1}$, it must be recoded using σ_0 , so the recoded word must begin also with $\mathbf{1}$; hence $(0, \mathbf{1})$ cannot be followed by $(0, \mathbf{0})$ or $(1, \mathbf{0})$. In the same way, $(1, \mathbf{0})$ cannot be followed by $(0, \mathbf{1})$ or $(1, \mathbf{1})$.

(ii) An eventually constant path on the graph ends on words which are all of the same type. If all the recoded words are of type 0, the initial word u must contain arbitrarily long strings of 0's, and this is impossible for a balanced word. Applying the same reasoning to the recoded words $u^{(n)}$, we see that the sequence cannot be eventually constant.

The fact that all the other sequences defined by the graph are admissible is a direct consequence of the fact that the partition by the $\Sigma_t^{\mathbf{a}}$ is a Markov partition for Φ , as we remarked above; the only trouble comes from the convention in the definition of Φ . Namely, it is easily checked, from the corollary give above, that fixed words are recoded to fixed words beginning with the same letter. Hence the coresponding coding sequence consists only of the letters $(0, \mathbf{1})$ and $(1, \mathbf{1})$ if the initial letter is $\mathbf{1}$, or $(0, \mathbf{0})$ and $(1, \mathbf{0})$ if the initial letter is $\mathbf{0}$. Suppose now that a coding sequence consists eventually only of letters $(0, \mathbf{0})$ and $(1, \mathbf{0})$; by inspection of the graph, we see that the last letter not of this type must be $(1, \mathbf{1})$, which may be preceded a finite number of times by $(0, \mathbf{1})$, and then $(0, \mathbf{0})$ or $(1, \mathbf{1})$. But a detailed study of cases shows that the corresponding word is a special word, which can be recoded, using Φ , in a sequence that consists eventually only of the letters $(0, \mathbf{1})$ and

$(1, 1)$; hence, by our convention, these sequences are not admissible. Note that sequences of type $(0, 1)^a(1, 1)v$, where v contains only $(0, \mathbf{0})$ and $(1, \mathbf{0})$, are admissible: they code for the preimages of order 1 of special words that begin with 1. (These are the ones that we removed when defining the images of special words by Φ .)

(iii) Suppose that the sequence contains only letters $(0, \mathbf{0})$ and $(1, 1)$. One checks by induction that, from the knowledge of the first letter of the recoded word $u^{(n)}$, we can deduce only the first letter of the initial word; the second letter is not possible to determine. One can then check that the two preimages of order 3 of the special word, **a10l** and **a01l**, admit this sequence as symbolic dynamics, verifying the claim. A similar proof applies when the coding sequence is eventually of this type.

Remark 4.3. In the cases where the coding sequence corresponds to two Sturmian words, these words differ by exactly two letters, since this is the case for the preimages of order 2 or more of a special word.

Writing Λ^+ for the one-sided subshift of finite type defined by the graph above, we can rephrase Theorem 4.1 as follows:

Corollary 4.2. *The map that takes a Sturmian word to its coding sequence semi-conjugates the dynamical system (Σ, Φ) to the subshift of finite type (Λ^+, S) ; the semi-conjugacy is one-to-one, except on the negative orbit of fixed points, and its image contains all of Λ^+ , except the eventually constant sequences and the sequences specified in part (ii) of the Theorem 4.1.*

If two Sturmian words belong to the same Sturmian system Ω , then the recoded words (their images by Φ) also belong to the same system. Thus they have the same type, and will be recoded by the same substitution, the only difference being that perhaps a shift appears in the coding process. Therefore recoding makes sense at the level of systems, as follows.

Definition 4.12. *The coding sequence for a Sturmian system Ω is the sequence $(i_n)_{n \in \mathbb{N}}$ of indices of substitutions, with $i_n = 0$ or 1, obtained by recoding the fixed words of Ω .*

This sequence is uniquely defined; it is what we call the *additive coding* of Ω . We can write the sequence as $\sigma_0^{a_0} \sigma_1^{a_1} \cdots \sigma_0^{a_{2n}} \sigma_1^{a_{2n+1}} \cdots$, by grouping together the strings of σ_0 and σ_1 . The sequence (a_n) of integers, which are all strictly positive except perhaps for a_0 , is called the *multiplicative coding* for the system. This sequence has a natural interpretation in terms of rotations. In particular, as we will see in §6, if $a_0 \neq 0$, then (a_n) is the continued fraction expansion of $\alpha/(1 + \alpha)$, where α is the angle of the rotation associated to the Sturmian sequence.

If we know the additive coding $(i_n)_{n \in \mathbb{N}}$ of a system Ω , it is possible to obtain the fixed words. Namely, the fixed word beginning with the symbol $\mathbf{0}$ is infinitely recoded in fixed words also beginning with $\mathbf{0}$. Hence it admits as a prefix all the finite words $\sigma_{i_0} \sigma_{i_1} \cdots \sigma_{i_n}(\mathbf{0})$.

These words are in fact easy to compute, using the fact that we have, for any substitution ρ , that $\rho(\sigma_0(\mathbf{0})) = \rho(\mathbf{0})$, and $\rho(\sigma_1(\mathbf{0})) = \rho(\mathbf{01}) = \rho(\mathbf{0})\rho(\mathbf{1})$.

Remark 4.4. Now we can explain why we have termed these the *fixed words*; it is a generalization of the standard situation, where one has a single substitution ρ . Indeed, as the reader who is familiar with the classical theory of substitution dynamical systems will note, the above procedure is very similar to the process by which one constructs what is known in that subject as the fixed word of ρ , i.e. the infinite word which is a fixed point for the action of ρ on sequence space. Indeed, if the sequence $(i_n)_{n \in \mathbb{N}}$ happens to be periodic, of period p , then we can reduce to that case as the p words we obtain are fixed for the composed

substitutions $\rho = \sigma_{i_0}\sigma_{i_1}\cdots\sigma_{i_{p-1}}$ (and its cyclic permutations). What we are doing here can be considered as a generalization of that idea to an infinite, and not necessarily periodic, sequence of substitutions.

We summarize this as follows.

Proposition 4.5 (Rauzy Rules). *Let Ω be a Sturmian system with coding sequence $(i_n)_{n \in \mathbb{N}}$. Define two sequences of finite words $(U_n)_{n \in \mathbb{N}}$ and $(V_n)_{n \in \mathbb{N}}$ by recurrence : $U_0 = \mathbf{0}$, $V_0 = \mathbf{1}$; if $i_n = 0$, then $U_{n+1} = U_n$, $V_{n+1} = V_n U_n$; if $i_n = 1$, then $U_{n+1} = U_n V_n$, $V_{n+1} = V_n$. Then U_n (resp. V_n) is a sequence of prefixes of the fixed word with initial letter $\mathbf{0}$ (resp. $\mathbf{1}$).*

4.4. Recoding Sturmian Words: the Natural Extension of the Coding map, and the Symbolic Dynamics for the Scenery Flow

An important consequence of the previous section is that the coding sequences for Sturmian words appear as the non-eventually-constant sequences in a subshift of finite type.

We could account for the eventually constant sequences on this graph by coding periodic rotations, but we will not do so as it is tedious and not so useful to write down the details.

This dynamics is very different from the one discussed in §3.2, given by the shift on Sturmian words. The Sturmian words gave symbolic dynamics for the rotations; the coding sequence, on the other hand, gives us symbolic dynamics for the map Φ . A fundamental difference between these two maps is that the rotation is of entropy 0, while Φ is of positive entropy. In this section we compute the natural extension (in the sense of Rokhlin (cf. [41]) for this map, and recover in this way symbolic dynamics for the scenery flow.

Of course one could easily build abstractly the natural extension of the subshift of finite type, by simply taking the allowed biinfinite sequences. In this way we recover the 2-sided subshift defined in §3. However we prefer to build the natural extension in a more concrete way. There is some freedom in this construction of the natural extension; the choices we make may seem arbitrary, but in fact it will be seen that they come directly from the geometric construction sketched in §4.1.

We use for this purpose a slight modification of the above coding: instead of working with σ_0, σ_1 , consider the maps τ_0, τ_1 defined from $\{\mathbf{0}, \mathbf{1}\}$ to $\{\mathbf{0}, \mathbf{1}\}^*$ by

$$\begin{aligned} \tau_0 : \quad \mathbf{0} &\mapsto \mathbf{0}, \\ &\mathbf{1} \mapsto \mathbf{10}; \\ \tau_1 : \quad \mathbf{0} &\mapsto \mathbf{10}, \\ &\mathbf{1} \mapsto \mathbf{1}. \end{aligned}$$

so that $\tau_0 = \sigma_0$, but τ_1 and σ_1 differ by their order.

We can use these two substitutions also to recode Sturmian sequences, which leads to the following proposition.

Proposition 4.6. *Let u be a Sturmian word. If u is of type 0, then*

- (i) *If u is not a special word, then either $u = \tau_0(v)$, where v is a Sturmian word, or $u = S\tau_0(v)$, where v is a Sturmian word that starts with $\mathbf{1}$ (but not both).*
- (ii) *If u is a special word, we can write both $u = \tau_0(v_0)$ and $u = S\tau_0(v_1)$, where $Sv_0 = Sv_1$ is a special Sturmian word.*

If u is of type 1 then, if $u_0 = \mathbf{1}$, we have $u = \tau_1(v)$, and if $u_0 = \mathbf{0}$, we have $u = S\tau_1(v)$, where v is a Sturmian word.

The proof is similar to that of Proposition 4.4 in §4.3, but slightly easier since for τ_1 , unlike for σ_1 , we do not need to treat special cases. We can then define a variant of the map Φ .

Definition 4.13. We denote by Ψ the map $\Psi : \Sigma \rightarrow \Sigma$ that takes u to the unique Sturmian word v such that $u = \tau_0(v)$ or $u = \tau_1(v)$, or, if it does not exist, to the unique Sturmian word v such that $u = S\tau_0(v)$ or $u = S\tau_1(v)$.

Using the map Ψ , we can again code all Sturmian sequences using the same alphabet as before, now with the graph given in Fig.4.3.

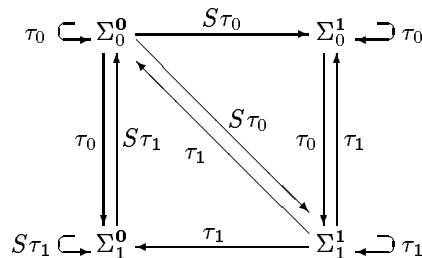


Fig.4.3. The dual graph for additive coding

The restrictions posed on coding sequences are the same as before. Consider now a biinfinite sequence $(\epsilon_n, \mathbf{a}_n)_{n \in \mathbb{Z}}$. We associate to this biinfinite sequence the pair of infinite sequences $(\epsilon_n, \mathbf{a}_n)_{n \in \mathbb{N}}$ and $(\delta_n, \mathbf{b}_n)_{n \in \mathbb{N}}$, where $\mathbf{b}_n = \mathbf{a}_{-n}$ and $\delta_n = 1 - \epsilon_{-n-1}$.

It is easy to check that this is a one-to-one and onto correspondence between admissible biinfinite sequences, and pairs of admissible infinite sequences; for example, suppose that $(\delta_n, \mathbf{b}_n)_{n \in \mathbb{N}}$ is not admissible, and that for some n , $(\delta_n, \mathbf{b}_n) = (0, \mathbf{1})$, $\mathbf{b}_{n+1} = \mathbf{0}$. Then we compute $(\epsilon_{-n-1}, \mathbf{a}_{-n-1}) = (1, \mathbf{0})$ and $\mathbf{a}_{-n} = \mathbf{1}$; hence the biinfinite sequence $(\epsilon_n, \mathbf{a}_n)_{n \in \mathbb{Z}}$ is not admissible.

This gives us a concrete combinatorial model for the natural extension as follows. To any biinfinite admissible sequence $(\epsilon_n, \mathbf{a}_n)_{n \in \mathbb{Z}}$, first associate a pair of infinite admissible sequences $(\epsilon_n, \mathbf{a}_n)_{n \in \mathbb{N}}$ and $(\delta_n, \mathbf{b}_n)_{n \in \mathbb{N}}$. To the first infinite sequence, we associate a Sturmian word u_+ , using substitutions σ_0, σ_1 ; to the second, we associate a Sturmian word u_- , using τ_0, τ_1 . We get two Sturmian words with the same initial letter, and the process is now clearly one-to-one: u_- tells us how we must decode u_+ , and gives us the past.

Formally, we define $\tilde{\Sigma} = \{(u, v) | u, v \in \Sigma, u_0 = v_0\}$; this is the space of pairs of Sturmian words with the same initial letter; we write $\tilde{\phi} : \tilde{\Sigma} \rightarrow \tilde{\Sigma}$ for the map which recodes the first word and decodes the second. This map is the natural extension of Φ . We obtain

Proposition 4.7. The map $\tilde{\Phi} : \tilde{\Sigma} \rightarrow \tilde{\Sigma}$ is conjugate to the subshift of finite type $S : \Lambda \rightarrow \Lambda$ by associating to each biinfinite coding sequence a pair of Sturmian words.

It is not an accident that we recover the same symbolic system as in §3. In fact we have the following

Theorem 4.2. The transformation $\Phi_a : \Delta_a \rightarrow \Delta_a$ is conjugate to $\tilde{\Phi} : \tilde{\Sigma} \rightarrow \tilde{\Sigma}$.

Proof. We first define a map from Δ_a to $\tilde{\Sigma}$. Consider some point in the cross-section Δ_a for the scenery flow, with coordinates $(l_0, l_1, h_0, h_1, x, y)$. To this point we associate two Sturmian words in the following way. We have a fundamental domain for a lattice consisting of two rectangles, the left one labelled $\mathbf{0}$ and the right one labelled $\mathbf{1}$, and a point (x, y) in this fundamental domain. We consider the Sturmian word which is the itinerary of the orbit of this point for the vertical flow; it is a Sturmian word associated to the rotation by angle l_1 on an interval of length $l_0 + l_1$. Hence it determines l_0, l_1 up to a factor. But because we are in Δ_a , we have $\sup(l_0, l_1) = 1$, so these are completely determined. In fact, to compute l_0, l_1 , we do not need to know here the precise word we get, only the system to which it belongs,

that is, the sequence a_n above. Knowing this word itself determines the number x . (We give the precise formulas in §6.) In the same way, the Sturmian word given by the horizontal flow determines the vector (h_0, h_1) up to a factor; but since we have $h_0 l_0 + h_1 l_1 = 1$, it is in fact completely determined. Moreover, the word also defines y . In this way we have associated to each point of Δ_a an element of $\tilde{\Sigma}$ (which is uniquely determined, except for degenerate cases when the point (x, y) happens to be in the horizontal or vertical trajectory of the origin; in this case, we have two or four possible sequences; it is easy to check that this corresponds to the exceptional sequences we have seen before).

Next we will describe the effect of the map Φ_a on this pair of Sturmian words. We have two cases to consider, depending on whether l_0 is larger or smaller than l_1 ; we shall consider only the first case. There are three situations, summarized in Fig. 4.4.

Fig. 4.4. The Recoding of Sturmian Words.

In each case, we are stacking the left part, or rectangle **0**, over rectangle **1**; we are given two words u, v for the first domain, and we want to find the corresponding words for the restacked domain. If (x, y) is in rectangle **1**, then u' and v' also begin with **1**, and we can write $u = \sigma_0(u')$, $v' = \tau_1(v)$. If (x, y) is in the left part of rectangle **0**, then we have $u = S\sigma_0(u')$ and $v' = \tau_1(v)$; if (x, y) is in the right part of rectangle **0**, then we have $u = \sigma_0(u')$ and $v' = S\tau_1(v)$.

We see that the return map for the scenery flow is exactly the map $\tilde{\Phi}$ we have described before: take the σ -coding for the vertical orbit, the τ -coding of the horizontal orbit, remove the first term $(\epsilon_0, \mathbf{a}_0)$ of the vertical orbit and add $(1 - \epsilon_0, \mathbf{a}_1)$ to the coding of the horizontal orbit.

From the point of view of the geometry, the reason for using σ in one direction and τ in the other is that the process we have described is not symmetric: we stack always on the upper part, but we unstack alternatively on the right and on the left.

4.5. Shift Dynamical Systems on Sturmian Sequences: the Adic Viewpoint

Recall that the system (Ω, S) denotes the shift map S on Ω , the orbit closure of some Sturmian word u . In fact one can check, just by looking at the coding sequences, whether or not u belongs to Ω , as follows. The coding sequence for u is written using an alphabet \mathcal{A} of four letters. At the end of §4.2, we defined the coding sequence for a system, using the alphabet of two letters $\{0, 1\}$. There is a natural projection from the first alphabet onto the second, taking $(\epsilon_n, \mathbf{a}_n)$ to ϵ_n . We note that u belongs to the system defined by the projection of the coding sequence.

Fixing a Sturmian system and its coding sequence, we would like to express in terms of coding sequences the shift on Sturmian words corresponding to the rotation. There is a simple answer to this question:

Proposition 4.8. *Let u be a Sturmian word, and let $(\epsilon_n, \mathbf{a}_n)$ be its coding sequence. Let*

N be the smallest integer (if it exists) such that $(\epsilon_N, \mathbf{a}_N)$ is different from $(0, \mathbf{0})$ and $(1, \mathbf{1})$. The coding sequence for Su is the sequence $(\epsilon_n, \mathbf{b}_n)$ with $\mathbf{b}_n = \mathbf{a}_n$ if $n > N$, and $\mathbf{b}_n = 1 - \mathbf{a}_N$ for $n \leq N$. If N is not defined, then u is one of the two preimages of the fixed point, and the coding sequence for Su is either $(\epsilon_n, \mathbf{0})$ or $(\epsilon_n, \mathbf{1})$.

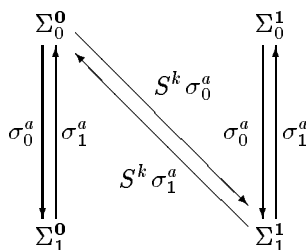


Fig.4.5. The Transition Graph for Multiplicative Coding

We can consider this map as a sort of odometer; the operation consists in finding the first term of the sequence which is not $(0, \mathbf{0})$ or $(1, \mathbf{1})$, changing the corresponding letter \mathbf{a}_N to its opposite, and carrying this letter for the first N ranks.

However, for this purpose it is more natural and interesting to use the multiplicative notation alluded to at the end of §4.3. We will see this below, showing that it is, in a slightly different notation, an adic system, as defined by Vershik (cf. [49, 50]). The precise formulas will be given in §6.

4.6 Symbolic Dynamics for the First Return map of the Scenery Flow to the Multiplicative Cross-Section

We will here only give the main results.

When we recode a given Sturmian word, instead of recording each step we can continue until we get a word of a different type. If the initial word u is of type 0, the new word v will be of type 1. We can write $u = \sigma_0^a(v)$ if the symbol $S\sigma_0$ was not used, or $u = \sigma_0^b S\sigma_0^c(v)$; in the latter case, the first letter v_0 of v must be 1, and it is easy to check that one can write $u = S^{b+1}\sigma_0^{b+c}(v)$. If we write k for $b+1$ and a for $b+c$, since $c \geq 1$, we see that the admissible sequences correspond to the paths in the graph represented in Fig.4.5 with the condition $0 < k \leq a$ (this graph corresponds to the graph given in Fig. 4.2, after grouping all the arrows that stay on the same level of the graph).

Each admissible sequence is uniquely defined by a sequence of pairs (a_n, k_n) of integers, with $0 \leq k_n \leq a_n$. It is clear from the graph that the admissible sequences are exactly those for which the maximal runs with $k_n = 0$, (except perhaps for the initial one) are of even length.

This is a sofic condition. One would like to in fact achieve a Markov condition; for that we will give a new related coding. The idea behind the first one was to approximate the given sequence u by shifting one of the fixed points of the sequence of substitutions, writing $u = S^{k_0}\sigma_0^{a_0}(S^{k_1}\sigma_1^{a_1}(\dots))$. We can instead try to approximate u by preimages, writing: $u = \mathbf{0}^{b_0}\sigma_0^{a_0}(\mathbf{1}^{b_1}\sigma_1^{a_1}(\dots))$.

We could start again as in §4, by proving that each Sturmian sequence u of type 0 can be written in a unique way $u = \sigma_0(v)$ or $u = \mathbf{0}\sigma_0(v)$, and defining in this way additive and multiplicative coding sequences. We leave this to the reader, giving here only the result:

Proposition 4.9. *Let u be a Sturmian word of type 0. Then one can write in a unique way $u = \mathbf{0}^{b_0}\sigma_0^{a_0}(\mathbf{1}^{b_1}\sigma_1^{a_1}(\dots))$, where (a_n) is a sequence of strictly positive integers, where*

(b_n) is a sequence of integers such that $0 \leq b_n \leq a_n$; and where, if $b_n = a_n$, then $b_{n-1} = 0$. Moreover the sequence (a_n, b_n) uniquely defines u unless the sequence (b_n) is 0 except for a finite set. In that case there are two possible values for u , each a preimage of the same order of the special word of the Sturmian system of u .

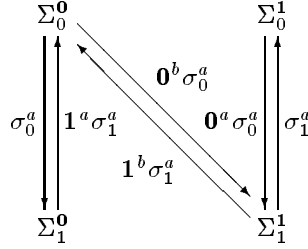


Fig. 4.6. The Transition Graph for Prefix Multiplicative Coding

The relation between b_n and k_n is simple: let N be the smallest integer such that $k_n > 0$. Then we have

- for $n < N$, $b_n = k_n = 0$;
- $b_N = a_N - k_N + 1$;
- for $n > N$ and $k_n > 0$, then $b_n = a_n - k_n$;
- for $n > N$ and $k_n = 0$, if $b_{n-1} = 0$, then $b_n = a_n$; if $b_{n-1} > 0$, then $b_n = 0$.

The admissible sequences are read from the graph shown in Fig.4.6. We have adopted the convention that $0 < b < a$. Each state is given by the type of the word u , together with the initial letter of the preimage of u (which is well-defined except when u is the special word).

A slightly different system will be produced if in place of the substitutions σ_0, σ_1 we use τ_0, τ_1 . In this case, it turns out to be best to use a third multiplicative system given by

$$\begin{aligned} \tau_0^a : \quad & \mathbf{0} \mapsto \mathbf{0}, \\ & \mathbf{1} \mapsto \mathbf{10}^a; \\ \tau_1^a : \quad & \mathbf{0} \mapsto \mathbf{1}^a \mathbf{0}, \\ & \mathbf{1} \mapsto \mathbf{1}. \end{aligned}$$

It is now easy to prove that any sequence u of type 0 can be written $u = S^k \tau_0^a(v)$, where v is a sequence of type 1, and any sequence u of type 1 can be written $u = S^k \tau_1^a(v)$, where v is a sequence of type 0. We can in this way associate to any sequence u a pair of sequences (a_n, k_n) . We will define a multiplicative system (a_n, b_n) , explaining how, at each step, one gets (a, b) from (a, k) .

If u is of type 0, then, if $u = \tau_0^a(v)$, with $v_0 = \mathbf{0}$, we define $b = a$; otherwise, if $u = S^k \tau_0^a(v)$, with $v_0 = \mathbf{1}$, we define $b = \max(0, k - 1)$.

If u is of type 1, then, if $u = \tau_1^a(v)$, with $v_0 = \mathbf{1}$, we define $b = a$; otherwise, if $u = S^k \tau_1^a(v)$, with $v_1 = \mathbf{1}$, we define $b = \max(0, a - k - 1)$.

Fig. 4.7. The Horizontal Tiling

It is now possible to prove that this sequence is well-defined, and that it completely defines the Sturmian word, except for the orbit of the special words. What is more interesting is that the set of admissible sequences is defined by the Markov condition $0 \leq b_n \leq a_n$, and $b_n = a_n$ implies $b_{n+1} = 0$. This is dual to the condition given above for the substitutions σ_0, σ_1 .

We will not give proofs for these facts here, since we will see later an arithmetic interpretation which will make them clear.

§5. A Line Tiling Model and the Natural Extension of Rotation Renormalisation

We gave in §1 a completely algebraic expression for the scenery flow, and in §3,4 two distinct completely symbolic expressions. Here we shall give a partly symbolic expression, using tilings of the real line (cf. [8] for more detailed explanations from the view point of quasi-crystals).

We begin with a tiling of the plane associated to a lattice and its L -shaped fundamental domain. For a point (x, y) in this fundamental domain, we can consider the non-periodic tiling induced on the horizontal line by the intersection with the translates of the two basic rectangles (cf. Fig. 4.7). This is a tiling of the line by two kinds of intervals, each of length l_0, l_1 , with a marked origin point x . We assume from now on that l_0 and l_1 are incommensurable. Such a marked tiling is completely defined by the following:

- a biinfinite rotation sequence $(u_n)_{n \in \mathbb{Z}}$. (We note that, since the one-sided sequence $(u_n)_{n \in \mathbb{N}}$ is Sturmian, we will be able to apply the results of the previous section.)
- the lengths l_0, l_1 of the two intervals.
- the position x of the marked point, counted positive (from the left endpoint) if x is in the interval labelled **1**, and negative (from the right endpoint) in the interval **0**.

The action of the scenery flow is geometrically that of zooming in toward smaller scales at a constant rate. That is, we are multiplying by $e^{t/2}$ the three numbers l_0, l_1, x . At a certain time, we will exit the fundamental domain (when $\inf(l_0, l_1) = 1$); we have then to subdivide the largest tile. We subdivide either the tiles **0** in two pieces, labelled by **1** and **0** on the left and right respectively, while not changing tiles labelled **1**, or we subdivide the tile **1** by taking **1** and **0** on the left and right, while leaving the tiles **0** unchanged. Thus there are two cases:

If $l_0 > l_1$, we replace (l_0, l_1) by $(l_0 - l_1, l_1)$; in this case, if $x > 0$, we keep x and replace u by $\tau_1(u)$. If $l_1 - l_0 < x < 0$, we keep x and replace u by $S\tau_1(u)$; if $-l_0 < x < l_1 - l_0$, we replace x by $x + l_0$ and u by $\tau_1(u)$. We have the analogous formulas if $l_0 < l_1$, replacing τ_1 by τ_0 .

We mention that these different cases do not depend on the sequence u , but rather on the values l_0, l_1, x . On the other hand, if one wishes to consider negative time for the scenery flow, one needs to take into account the sequence u . This sequence can be written in a unique way as $u = \sigma_a(u')$ or $u = S\sigma_a(u')$, and this determines the formulas used to compute the new lengths. One can view this as follows: l_0 and l_1 determine the small-scale structure of the tiling, while u determines the large-scale structure.

There is also a dual tiling, on the vertical orbit of (x, y) , with similar properties. However here we always take coordinates from the left of the interval, and therefore change formulas according to the permutations σ_0, σ_1 rather than τ_0 and τ_1 .

We can now make completely precise the relation between these tilings and the set E^* we defined in §3.1. By definition a *biinfinite* Sturmian sequence is a sequence $(v_n)_{n \in \mathbb{Z}}$ such that for each k , the sequence $(u_n)_{n \in \mathbb{N}}$ defined by $u_n = v_{k+n}$ is Sturmian.

Definition 5.1. *An irrational Sturmian quasi-crystal is a line tiling by two types of intervals, of rationally independent lengths, with combinatorics defined by a biinfinite Sturmian sequence, and with a chosen origin. Two irrational Sturmian quasi-crystals are equivalent if they differ only on a finite number of tiles.*

Two distinct equivalent quasi-crystals differ in exactly two tiles; they correspond to a horizontal line through the origin of the L -shaped fundamental domain.

The subdivision map defined above acts on the set of Sturmian quasicrystals, and respects equivalence.

Definition 5.2. *The set of equivalence classes of irrational quasi-crystals up to equivalence and subdivision is called the set of quasi-crystal hierarchies, denoted by \mathcal{T} .*

It is immediate that the scenery flow acts on \mathcal{T} , by multiplying all real coordinates by $e^{\frac{1}{2}}$. It is now straightforward to check

Theorem 5.1. *The scenery flow on \mathcal{T} is topologically conjugate to the scenery flow on E^* .*

This was the first model the authors found for the scenery flow, and it explains the origin of the name: combining the ideas of [7] and [20], we wanted to model how the “scenery” one sees changes as one zooms down to a point in a nested tiling of the line given by renormalization. Thus, what one sees as time progresses in the flow defined by this magnification is that the hierarchy in the nested tiling will change, and smaller details will come into view, while the structure visible at the beginning remains recorded in the large-scale combinatorial structure of the tiling (cf. also [3]).

Remark 5.1. A particular interesting case is the one where the sequence of subdivisions, written as $\tau_0^{a_0} \tau_1^{a_1} \dots$, is periodic; this is the case when l_0/l_1 is a Galois integer (a quadratic number with an immediately periodic continued fraction expansion). In that case, there is a natural way to get a backward orbit when only l_0, l_1 are given. This consists of extending backwards the coding sequence as a periodic orbit, and then taking the sequence u this determines. This sequence appears as fixed point for some substitution; hence we get the (much studied) case of self-similar Sturmian tilings, and in particular the simplest case related to the golden number, the Fibonacci tiling (cf. [17]). These are closely related to the study of toral automorphisms. Indeed, in that case the fundamental domain for the plane lattice appears as Markov partition for a toral automorphism (cf. [10]); we will give below the explicit arithmetic formulas.

Remark 5.2. If we restrict ourselves to the central two intervals **0** and **1**, we see a sequence of rotations of smaller and smaller domain. This is what we get from the usual continued fraction; we can view this process as a renormalization of circle rotations. However, there is no well-defined way to invert this process, because for any rotation, there are always two rotations from which it can come via an elementary renormalization. If we renormalize for some time, we get a tiling of the circle by two intervals by looking at the intervals given by the orbit of 0. We can then come back in a unique way to the initial rotation, by considering this finite tiling (but no further). One can see the scenery flow we describe in this section as the natural extension of the renormalization of circle rotations: each Sturmian tiling describes a possible past for the renormalization of the rotation defined by the two basic intervals. We will elaborate on this remark in the next section.

Remark 5.3. We have shown, in particular, that the space of Sturmian tilings corresponding to given tiles and a given Sturmian system can be parametrized by a torus. However as before this parameterization is only almost one-to-one: to recover the torus, we have to identify equivalent pairs of Sturmian tilings, corresponding to the special word of the Sturmian system, which differ only on two tiles and are parametrized by the same point of the torus. Modulo this identification, the space of tilings becomes homeomorphic to a torus. More complicated cases, for example Penrose tilings, can be studied in the same way (cf. [40]).

§6. An Arithmetic Interpretation

We have shown above that the symbolic dynamics completely defines a point in the cross-section Δ_m defined in §2, so that the two Sturmian words completely determine the geometrical coordinates of the point. We will give in this section precise arithmetic formulas, and recover the connection with continued fractions.

The results given here were worked out in a seminar at Luminy (cf. [4] for more details). Vershik and Sidorov also (independently) found these results in the framework of adic systems (cf. [50]).

6.1. The Ostrowsky Number System

In this section we will define two number systems that correspond to the coding sequence above. Such systems have been studied by a number of authors previously (cf. [37, 43, 28, 29, 50]).

To discuss the first system, it will be best to change our convention for circle rotations. Now we take a rotation of angle α on the circle written as the interval $[-1, \alpha[$, together with the two natural continuity intervals $[-1, 0[$ and $[0, \alpha[$. We will give an explicit formula for recovering the geometric location of a point x in the circle from its combinatorial coordinates given by the first coding sequence of the Sturmian word u corresponding to x , in terms of the sequence of pairs of integers (a_n, b_n) defined in §4.6.

Fig. 6.1. The Ostrowsky Induction Process

The main theorem is the following:

Theorem 6.1. *Let u be the Sturmian word determined by the itinerary of a point x with respect to the rotation of angle α on $[-1, \alpha[$. Let (a_n, b_n) be the sequence determined by the coding of u in sequences of prefixes for σ_0, σ_1 . Then (a_n) is the continued fraction expansion of α , and if we define $\alpha_0 = \alpha$, $\alpha_{n+1} = \{1/\alpha_n\}$ and $\beta_n = (-1)^{n+1} \alpha_0 \alpha_1 \cdots \alpha_n$, we have*

$$x = \sum_{n=0}^{\infty} b_n \beta_n.$$

Proof. The main point to notice is that the substitution σ_0 appears naturally when one

looks at the first return map on the complement of the image of $[0, \alpha[$. Indeed, consider the initial rotation R and the induced map R' (which is also a rotation) and a point x in the domain of R' . To this point we associate a Sturmian word u defined by the action of R , and another Sturmian word v , defined by the action of R' . The relation between these words is simple, since the orbit of x under the action of R' is a subset of the orbit under the action of R , and since, by definition, we have removed all points in the orbit that follow immediately a point of the orbit in interval $\mathbf{0}$ (the removed points must be in the interval $\mathbf{1}$). Thus, to obtain u from v , we simply replace all symbols $\mathbf{0}$ in v by $\mathbf{01}$ while keeping all symbols $\mathbf{1}$ unchanged; this tells us that $u = \sigma_0(v)$.

After repeating this operation at most $[1/\alpha]$ times, we get a rotation defined on the interval $[\alpha[1/\alpha - 1], \alpha[$, with associated substitution $\sigma_0^{[1/\alpha]}$. If the point x is in this interval, we can proceed; otherwise, we must take the first point in the positive orbit of x which is in the inducing interval. This point is easily seen to be $x' = R^b(x)$ if the associated Sturmian word u can be written as $\mathbf{0}^b \sigma_0^a(v)$.

We now renormalize the induced rotation, as shown in Fig.6.1, to get a biggest interval of length 1 on the left; all that is needed is to multiply by $-1/\alpha$; this gives a rotation of angle $\{1/\alpha\}$, and a new point $x_1 = -x'/\alpha$. We have clearly

$$x = -b_0\alpha - x_1\alpha.$$

Since we have now returned to the initial situation (except that now the interval $\mathbf{1}$ is on the left), we can repeat the process, using alternatively σ_0 and σ_1 . We have proved that $a_n = [1/\alpha_n]$, where (α_n) is the orbit of α under the continued fraction map. That is, $[a_0, a_1, \dots]$ is the continued fraction expansion of α .

A simple recurrence shows that one can define a sequence of points x_n , all bounded by 1, such that

$$x = \sum_{i=0}^n b_i \beta_i + \beta_n x_n.$$

Since the β_n converge exponentially fast to 0, the series converges, completing the proof.

Remark 6.1. It is a classically known fact that if p_n/q_n are the best rational approximations for α , given by the first n terms for the continued fraction expansion, then $p_n - q_n\alpha = \beta_n$. Here the numbers β_n satisfy the recurrence relation $\beta_{n+1} = \beta_{n-1} - a_n\beta_n$.

We mention that it is possible, although not very helpful to one's understanding, to give a proof of the Markov condition on the coefficients b_n by means of this relation.

6.2. The Dual Number System

To make use of the dual coding given by the substitutions τ_0, τ_1 defined in §4.4, it will be more convenient to consider a rotation of angle α on $[0, 1[$. We will induce on the smallest interval if is on the left, and on the image of the smallest interval otherwise; all of the properties of the coding can be deduced from this algorithm.

Theorem 6.2. *Let u be the Sturmian word determined by the itinerary of a point x with respect to the rotation of angle α on $[0, 1[$. Let (a_n, b_n) be the sequence determined by the third multiplicative coding of u from §4.6, using τ_0 and τ_1 . Then (a_n) is the continued fraction expansion of α , and if we define $\alpha_0 = \alpha$, $\alpha_{n+1} = \{1/\alpha_n\}$ and $\gamma_n = \alpha_0\alpha_1 \cdots \alpha_n$, we have*

$$x = \sum_{n=0}^{\infty} b_n \gamma_n.$$

Proof. We reproduce the preceding proof almost exactly, except that one needs here to make a distinction between words of type 0 and of type 1. For words of type 0, we consider the first point in the forward orbit that is in the inducing interval, while for words of type 1, we take the first such point in the *backward* orbit.

There is in this case an easier interpretation: we think of the sequence γ_n as a basis for a number system, and try to express x in this basis, using the greedy algorithm. The digits we then get will be the b_n , with the Markov condition an immediate consequence of the recurrence relation $\gamma_{n-1} = a_n \gamma_n + \gamma_{n+1}$.

6.3. An Explicit Conjugacy Between the Symbolic and Algebraic Models for the Scenery Flow

In §4.1, given a point in E^* , we defined an associated pair of Sturmian sequences given by the symbolic dynamics for the horizontal and vertical flows.

We can now compute the inverse of this map.

Theorem 6.3. *Let (u, v) be two Sturmian sequences with the same initial letter. Let $(a_n, b_n)_{n \in \mathbb{N}}$ be the multiplicative coding of u , and $(a'_n, b'_n)_{n \in \mathbb{N}}$ be the dual multiplicative coding of v , as explained in §4.6. Let α (resp. α') be the number whose continued fraction expansion is (a_n) (resp. (a'_n)), and let x (resp. y) be the number whose Ostrowski expansion is $(a_n, b_n)_{n \in \mathbb{N}}$ (resp. whose dual Ostrowski expansion is $(a_n, b_n)_{n \in \mathbb{N}}$). Let $k = 1 + \alpha\alpha'$. Then, the two Sturmian sequences correspond to a pair of boxes of total area 1, with respective dimensions $(1, 1/k)$ and $(\alpha, \alpha'/k)$, and a point with the coordinates $(x, y/k)$. Whether the largest box is on the left or on the right is given by the type of the sequence u .*

Remarks 6.4. The role of the constant k above is simply to normalize the total area.

We mention that there are integer number systems which are closely related to these real number systems. Consider the sequence q_n defined by $q_{n+1} = a_n q_n + q_{n-1}$, $q_0 = 1$, $q_{-1} = 0$ where these are the denominators of the best approximations associated with the continued fraction expansion $[a_0, a_1, \dots]$; then every positive integer N can be written in a unique way as $N = \sum_{i=0}^n b_i q_i$, where the integers b_i satisfy the conditions that $0 \leq b_i \leq a_i$, and that if $b_i = a_i$ then $b_{i-1} = 0$. This fact is clear from the greedy algorithm and the above recurrence relation.

This number system is very close to the first one we considered, the main difference being that here we have finite sums. It turns out that it corresponds in fact to those points in the positive orbit of 0 which have a finite coding. The point 0 itself gives two Sturmian words, the two fixed words of the sequence of substitutions; its second image corresponds to the special word of the system.

It is in fact possible to recover everything just using the integers; this is carried out in [50]. One can define a distance on the integers in the following way: if $N = \sum_{i=0}^n b_i q_i$ and

$N' = \sum_{i=0}^n b'_i q_i$, define $d(N, N') = 2^{-k}$, where k is the smallest number such that $b_k \neq b'_k$. If we take the completion of \mathbb{N} for this distance, and identify a countable number of pairs of elements in the completion (the two maximal elements for the lexicographic ordering, and their backwards orbit), we obtain the circle; the addition of 1 on the integers extends to a map on this completion which is the rotation by angle α .

One can proceed similarly with the dual system, writing each integer (this time positive or negative) as $N = \sum_{i=0}^n b_i (-1)^i q_i$. It is remarkable that the first number system allows us to

write positive or negative reals, but only positive integers, while with the the dual system we can write positive or negative integers, but only positive real numbers.

One would like to try to extend these results to other recurrent sequences, in particular to recurrent sequences of higher order. This has already been carried out for certain recurrent sequences, in particular for the recurrent sequence of degree three (the “Tribonacci” numbers) defined by $T_{n+3} = T_{n+2} + T_{n+1} + T_n$. In that case one gets a translation on a 2-torus, coded by a fractal domain (cf. [39, 33]).

We remark that with this system, one can develop real numbers only in a compact set. However, if instead of an infinite sequence $(a_n)_{n \in \mathbb{N}}$ we take a biinfinite sequence $(a_n)_{n \in \mathbb{Z}}$, we can give developments for any real number, in the form of sequences $(b_n)_{n > -k}$. The same type of development allows one, for example, to give arithmetic formulas for special Sturmian tilings of the line. Here again, we can define an adic topology; the completion, which consists of biinfinite sequences (b_n) , corresponds to the space of all Sturmian tilings.

A special case arises when the continued fraction sequence is periodic. (The simplest case of this is the well-known Zeckendorf expansion, which corresponds to the golden number.) It is then easy to build the natural extension of the coding process, by taking the binfinite extension of the stationary Markov process associated to the periodic continued fraction. (This has no meaning in the general case, since there is no single natural way to extend on the left a given continued fraction if it is not periodic.) In that way, we come to the theory of hyperbolic toral automorphisms, and the formulas above give an explicit description of certain Markov partitions for these maps of the torus. We can then obtain explicit coordinates for a Markov partition of the type $(x, y) = \left(\sum_{n=0}^{\infty} \epsilon_n \lambda^n, \sum_{n=-\infty}^{-1} \eta_n \lambda'^n \right)$, where λ' is the conjugate of a Galois integer λ , ϵ_n, η_n belong to a finite set of elements of $\mathbb{Q}[\lambda]$, and where the sequence is subject to some explicit Markov condition.

§7. Rauzy Induction on the Exchange of Three Intervals

We present here a further model of the scenery flow in terms of exchanges of three intervals.

Recall that an interval exchange transformation (IET) is a one-to-one map of an interval to itself, which is everywhere continuous on the right, and is continuous with derivative 1 except on a finite set. Thus geometrically, an interval exchange transformation consists in cutting the given interval in a finite number of (left-closed, right-open) subintervals and permuting them by translations. Therefore, an interval exchange transformation on k intervals is completely defined by a vector $\lambda = (\lambda_1, \dots, \lambda_k)$ of positive reals, giving the length of the intervals, together with a permutation π of k symbols.

A fundamental property is that the first return map of an IET with k intervals on a subinterval is an IET with at most $k + 2$ intervals. Moreover, if the induction subinterval is carefully chosen, then the first return map will again be an IET on k intervals. (For this, the induction interval must be an **admissible** subinterval, that is, both extremities must be of the type $T^k(a_i)$, for $k \in \mathbb{Z}$ and a_i a discontinuity point; and the interval must contain no point of the orbit between a_i and $T^k(a_i)$ (cf. [44]).)

A basic tool for studying dynamics of interval exchanges is the Rauzy induction procedure: given an interval exchange (λ, π) , we then induce on the largest admissible subinterval containing 0; one checks from the definition that this interval is $[0, \min(\lambda_n, \lambda_{\pi^{-1}n})]$.

The first non-trivial example of an interval exchange is the exchange of two intervals,

that is, a rotation. In the preceding section, the coding using τ_0 and τ_1 is the coding one obtains from the Rauzy induction for this rotation. If one instead, however, induces on a non-admissible interval, then one will have produced not a rotation but an exchange of three intervals.

The next case to consider is the exchange of three intervals. In order to get something non-trivial, one needs to use the permutation (321). The reason is that three of the other permutations yield interval exchanges with an interval of fixed points, while the remaining two permutations on three letters yield a rotation. However, as is well-known (cf. [31]), the Rauzy induction applied to an exchange of three intervals will give a rotation.

Now this induced map is indeed a rotation, but in fact it contains more information than just the rotation: there is also a distinguished point in one of the two continuity intervals, coming from the removable discontinuity. In particular, the Rauzy induction for this interval exchange will not be the same as Rauzy induction applied to the corresponding rotation, since it will take into account that distinguished point.

Fig. 6.2. Plane Tilings Given by Exchanges of 3 Intervals

One can check that, considering the sequence of permutations obtained by the Rauzy induction procedure for three intervals, the sequences obtained will consist of all of the non-eventually-constant infinite paths in the following graph:

$$\hookrightarrow (231) \longleftrightarrow (321) \longleftrightarrow (312) \leftarrow$$

In fact, these are equivalent: giving an exchange of three intervals, or giving a rotation together with a distinguished point in its domain. It follows that any Sturmian word determines (as above) a unique exchange of three intervals.

The Rauzy induction is a two-to-one map, corresponding to the additive continued fraction. Veech (in [44, 45]) explains how to build explicitly the natural extension of this map. The idea is to build rectangles of respective heights h_1, \dots, h_k over the intervals of length $\lambda_1, \dots, \lambda_k$, and to choose a point on each vertical boundary of these rectangles. These heights must be chosen in such a way that it is possible to identify the boundaries so as to get a flat surface, the point on the boundaries giving singularities (cone points) of this surface. This condition on the heights is given by a set of linear equations that depends only on the permutation. We will not go into detail here, remarking only that, for example for the permutation (321), the condition is $h_2 = h_1 + h_3$, while for the permutation (231), the condition is $h_2 = h_3$. The meaning of these conditions will be clear from Fig.6.2.

One can then use the Rauzy induction to build a flow on the space of rectangles. Starting

with a set of rectangles with total base width equal to 1, we apply Rauzy induction, and then multiply the first coordinate by $e^{\frac{1}{2}}$ and the second by $e^{-\frac{1}{2}}$, to reduce the new base width to again be 1. We can show

Theorem 7.1. *The natural extension of Rauzy induction is the first-return map of the scenery flow to a cross-section.*

This now is a different cross-section from those we gave in §2; it is made of three rectangles instead of two. One can in this context find a symbolic dynamics and arithmetic; we leave the computation of the precise formulas to the reader.

It would seem at first sight that this version does not provide anything new. It is however interesting to note that, for Rauzy induction on three intervals, the lengths of intervals are given by 3×3 matrices. Therefore in the periodic case we should get for the matrix eigenvectors of order three. However, since this corresponds to a periodic continued fraction, we can get in this case only quadratic integers, and not cubic integers as might be expected. In fact one can, in a similar way, produce matrices of arbitrary size by introducing extra points and thus considering the rotation to be an exchange of k intervals. By applying the Rauzy induction, we will then get interval exchanges “of rotation class” (cf. [36]). This leads to a generalization of the scenery flow, where we now follow a finite set of points in the torus (cf. the section which follows). There is a similar notion for general interval exchanges: some interval exchanges can be reduced, by Rauzy induction, to interval exchanges on a smaller number of intervals, with marked points, while others are primitive in this respect. It would be interesting to have a combinatorial condition that would insure primitivity.

We mention that a deep study exchanges of 3 intervals has been made recently, by Ferenczi, Holton and Zamboni, using a different induction procedure; this yields many surprising results, notably an example of an interval exchange on 3 intervals that is measurably, but not topologically, conjugate to a rotation. It would be interesting to understand the link with the viewpoint of Veech and Rauzy.

Another point of interest is that the flow on the natural extension of the Rauzy induction has a well-defined geometric interpretation, as the Teichmüller flow on strata of quadratic differentials (cf. [46]). In the next section, we will show how to interpret the scenery flow as Teichmüller flow on the twice punctured torus.

§8. A Geometric Model: Pairs of Foliations and the Teichmüller Scenery Flow

It is well known that the diagonal flow on the quotient $SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$ is naturally isomorphic to the geodesic flow on the modular surface, and can be identified with the Teichmüller flow of conformal structures on the torus.

If we consider the once-punctured torus, the Teichmüller flow stays the same. The reason is that the torus is a homogeneous space, with a transitive group of automorphisms. So, if T and T' are two tori equipped with conformal structures, and if there is a conformal automorphism from T to T' , then, for any $x \in T$ and $x' \in T'$, there is a conformal automorphism that sends x to x' .

Another way to explain this is as follows. The Teichmüller flow acts on the space of quadratic differentials. (A quadratic differential q is locally defined by $q = f(z) dz^2$, where z is a local complex coordinate and f is a holomorphic function; a change of coordinates $z = \phi(w)$ transforms this formula in $q = g(w) dw^2$, where $g(w) = f \circ \phi(w) \cdot \phi'(w)^2$; the function f is allowed to have poles only at the punctures, if any, of the surface.) On the

compact torus, a quadratic differential has no pole, so, by the index formula, it has no zero and hence must be a constant. On the once-punctured torus, a quadratic differential has at most a simple pole, at the puncture; its quotient by a constant quadratic differential gives a conformal function on the torus with at most a simple pole, whence this conformal function must be constant, i.e. the puncture is a removable singularity. We see therefore that the space of quadratic differentials is the same for the torus and the once-punctured torus, so their Teichmüller flows are the same.

Things are different for the twice-punctured torus: we have here (cf. [35]) three possible cases, describing three strata of quadratic differentials:

—two simple poles and two zeroes of order 1, giving a stratum of non-orientable quadratic differentials (this is the “main stratum”, of maximal dimension: almost all (in the measure or topological sense) quadratic differentials on the twice punctured torus belong to this stratum);

—two simple poles and a zero of order 2, giving a stratum of non-orientable quadratic differentials;

—no pole and no zero, giving a stratum of constant (orientable) quadratic differentials.

For the same reason as above, it is impossible to have only one simple pole: either the two singularities are removable and the quadratic differential is constant, or neither is removable.

The first two strata are quite complicated; the last one seems almost trivial, and one could think that, since it consists in constant quadratic differentials, it is isomorphic to the case of the non-punctured torus. This however is not the case: the formal definition of this constant quadratic differential includes the two punctures, and since the group of conformal automorphisms is transitive, but not 2-transitive on the torus, a given quadratic differential with different choices for the punctures will give different elements of this stratum.

In fact, as one can check, the space of constant quadratic differentials on the twice punctured torus is a set of full measure in the space $SA(2, \mathbb{Z}) \backslash SA(2, \mathbb{R})$ we defined in §1 (since the two punctures must, by definition, be distinct, we cannot get the points $(M, \vec{0})$; the corresponding punctures would in that case be the same). Using the fact that a quadratic differential can be locally (away from the zeroes and the punctures) written as the square of a holomorphic one-form $q = \omega^2$, one gets two transverse foliations, given by the real and imaginary part of this holomorphic one-form. It is elementary to check that these two foliations possess a well-defined transverse invariant measure. In fact these two foliations together with their transverse invariant measures define a flat structure on the surface, hence a conformal structure, and moreover a quadratic differential.

In particular, if a quadratic differential on the twice punctured torus is non-singular, it is constant, and so it defines a pair of transverse linear measured foliations on the torus. We choose by convention one of the punctures to be 0; taking the universal cover of the torus, the two foliations lift to transverse measured foliations, which can be used as coordinates; the puncture 0 lifts to some lattice Γ , and the other puncture lifts to some distinct translate of Γ . The choice of which of these two punctures is 0 is irrelevant, since it amounts to a change by $-Id$, while we must in any case quotient by the group of automorphisms of the twice-punctured torus.

The Teichmüller flow consists in multiplying one of the transverse measures by $e^{t/2}$ and the other by $e^{-t/2}$. Restricting to the stratum of constant quadratic differentials, and coordinatizing by means of the space $SA(2, \mathbb{Z}) \backslash SL(2, \mathbb{R})$, we obtain

Theorem 8.1. *The Teichmüller flow on the stratum of constant quadratic differentials*

on the twice punctured torus is conjugate to the restriction of the scenery flow g_t defined in §1 to the subset $\{(M, \vec{v}) \mid \vec{v} \neq 0\}$ in E .

It is clear that the analogous procedure can be carried out in the general case, the modular space of a Riemann surface. One then gets a fiber bundle, with fiber the given surface; this contains a suspension of any pseudo-Anosov diffeomorphism (sitting over periodic orbits). We expect to come back to this subject in a future paper.

§9. Shannon-Parry Measures for Induction on Sturmian Sequences

If one considers the Sturmian system associated to a fixed periodic continued fraction expansion, by taking the coding sequence one gets a subshift of finite type. By taking the natural extension, one can code in this way all toral automorphisms. We get here three measures: the invariant measure for the automorphism, and the invariant transverse measures for the stable and unstable foliations.

It is possible to recover these measures directly at the symbolic level. The recoding of Sturmian sequences gives a subshift of finite type, and we construct the Parry-Shannon measure (which is the measure of maximal entropy for the subshift) as in [10]. In other words it is the Gibbs measure for the potential function which is identically zero; so we call it the Gibbs state. Via the isomorphism to the flat torus, this is exactly Lebesgue measure projected from \mathbb{R}^2 .

The transverse measures are the following symbolically: they are the conditional Gibbs measures. These are dual, one conditioned on the past, and one on the future, and there are two interpretations for each: as transverse measure to one foliation, and as a measurement of length along the other. This measurement of length defines a sequence of scaling functions, when evaluated on 0-cylinders at successive levels k ; the scaling function assigns lengths to nested sub-tiles. In this way one builds up the scenery (a nested tiling of the real line), labelled by an infinite past or future in the symbol space, with lengths given by the sequence of scaling functions or equivalently by the conditional Gibbs measure for that tail.

These ideas are just as valid for the non-constant case. Consider a Sturmian system Ω_0 ; to each element of this system, we associate a coding sequence, and by shifting, some element in a Sturmian system Ω_1 . In this way we define a sequence of Sturmian systems Ω_n , together with shift maps $\sigma_n : \Omega_n \rightarrow \Omega_{n+1}$. Taking the natural extension, let us suppose that this sequence is biinfinite. We then define an operator $\mathcal{L} : \mathcal{C}(\Omega_n) \rightarrow \mathcal{C}(\Omega_{n+1})$ by $\mathcal{L}(f)(x) = \sum_{y: \sigma_n(y)=x} f(y)$. We also define the dual operator \mathcal{L}^* . Here, the notion of eigenvector is replaced by a sequence of positive vectors invariant by the operator. We have the following Proposition.

Theorem 9.1. *Let Ω be a Sturmian system of type 0, defined by the coding sequence $(a_n)_{n \in \mathbb{N}}$. Let Ω_n be the sequence of Sturmian systems obtained by recoding, and let (ν_n) be a sequence of measures invariant by the associated dual operator. If α is the number with $[a_0, a_1, \dots]$ as continued fraction expansion, and if we denote by $[0]$ and $[1]$ the basic cylinders in Ω (in the usual notation, and not with the recoding), we have*

$$\frac{\nu_0([1])}{\nu_0([0])} = \alpha.$$

Proof. The essential property is that it is easy to compute $\nu_0([0])$ and $\nu_0([1])$ if we know the same quantity for ν_1 . Indeed, we have $\nu_0([0]) = \nu_1([0]) + \nu_1([1])$ and $\nu_0([1]) = \nu_1([1])$.

More generally, if we denote by V_n the vector $\begin{pmatrix} \nu_n([0]) \\ \nu_n([1]) \end{pmatrix}$, we compute immediately $V_n = M_n V_{n+1}$, where $M_n = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ if Ω_n is of type 0, and $M_n = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ if it is of type 1.

This means that, if we denote by Λ the positive cone of \mathbb{R}^2 , V_0 must belong to the positive cone that is the intersection of the $M_0 M_1 \cdots M_n \Lambda$. But it is classical that this intersection reduces to a line whose slope is given by the continued fraction.

We deduce from this that the measure ν_0 (and hence the eigenmeasure sequence ν_n) is uniquely defined up to a multiplicative constant; the theorem is true for all ν_n , and knowing the measure ν_n on cylinders of length 1 allows one to compute the measure ν_0 on cylinders of length n .

The same turns out to be true for the eigenfunction: by the same proof, it is constant on each of the two cylinders, and the ratio of the two values is defined by the negative part of the biinfinite sequence (a_n) . The product of the eigenfunction by the eigenmeasure is a sequence of invariant measures for the shift. We think of this in the following way: instead of finding an invariant measure for a toral automorphism, we are now given a sequence of toral automorphisms on a sequence of tori, and we can define a natural invariant measure. This brings us to the notion of *mapping families* (cf. [3]).

Remark 9.1. What we have found here are special cases of invariant measures for adic systems (cf. [49]). However, it is interesting to consider not only invariant measures for a given dynamical system, but sequences of related measures for sequences of dynamical systems linked by induction. Examples are provided by a generalization of the Ruelle-Perron-Frobenius Theorem to nonstationary subshifts of finite type. This is related to random dynamics on the one hand and to the study of invariant differentiable structures on the other (cf. [3, 11] and the references given there).

§10. Some Generalizations

There are two main possible generalizations. The first one is to extend these ideas to other surfaces than the torus. A large part of this has already been done in the work of Veech (cf. [44–46]); in fact, one of the starting points for this paper was the study of [46], and in particular the role of the punctures that turn out to be removable singularities.

The symbolic dynamics associated to this situation is well understood: these are some types of sequences of complexity $(k-1)n+1$, generated by a finite number of substitutions, the admissible sequences of substitutions being defined by a sofic system (given by the so-called Rauzy graph).

There are still some unclear points: in particular, there should be a dual algorithm to the Rauzy algorithm; some other algorithms have been proposed (cf. [14, 16, 32]), and they are probably the dual algorithm we are looking for. What kind of arithmetical formulas one should get is not yet clear.

Another possible generalization is to go to a higher dimension; the paradigm here should be the Markov partitions for hyperbolic automorphisms of the torus \mathbb{T}^n . The situation here seems to be much more difficult, due to the fact that, by a theorem of Bowen (cf. [12]), the boundaries of such Markov partition, in dimension larger than 2, and for irreducible automorphisms, cannot be smooth. In fact, some explicit examples (with nice fractal boundaries) are known (cf. [39, 33]); in these cases, it has been possible to give explicitly symbolic dynamics and arithmetic formulas.

However, the only cases that have been explicitly studied as of this writing are “periodic” cases, and in particular those for which the eigenvalue is a Pisot number. The work of Dumont, Kamae and Takahashi (cf. [19]) could open some possibilities for the study of non-Pisot examples. We hope that it will be possible to extend the methods of the present paper to cover some non-periodic cases.

It is worth remarking that these two different generalizations could be linked: it is known (cf. [6]) that some particular symbolic dynamical systems can represent both an interval exchange and a translation on a torus of dimension three or more.

One can also try to extend to multidimensional continued fraction algorithms, such as the well-known Jacobi-Perron algorithm. A step in this direction has been made in [1], where this algorithm has been linked to \mathbb{Z}^2 -action by rotations on the circle. In that case, one can obtain a generalized Ostrowski expansion for elements of \mathbb{R}^2 , but as yet it is unclear whether there is in this case any structure which corresponds to the scenery flow.

REFERENCES

- [1] Arnoux, P., Berthé, V. & Ito, S., Discrete planes, \mathbb{Z}^2 -actions, Jacobi-Perron algorithms and substitutions [J], To appear in *Annales de l'Institut Fourier*.
- [2] Adler, R. L. & Flatto, L., Geodesic flows, interval maps and symbolic dynamics [J], *Bulletin of the American Mathematical Society*, **25**(1991), 229–334.
- [3] Arnoux, P. & Fisher, A. M., Anosov families, renormalization and the scenery flow [M], work in progress, 2000.
- [4] Arnoux, P., Ferenczi, S. & Hubert, P., Trajectories of rotations [J], *Acta Arithmetica*, **87**(1999), 209–217.
- [5] Arnoux, P. & Hubert, P., Fractions continues sur les surfaces de Veech [J], *Journal d'Analyse Mathématique* **81** (2000) 35–64.
- [6] Arnoux, P., Un exemple de semi-conjugaison entre un échange d'intervalles et une translation sur le tore [J], *Bulletin de la Société Mathématique de France*, **116**(1988), 489–500.
- [7] Arnoux, P., Le codage du flot géodésique sur la surface modulaire [J], *l'enseignement Mathématique*, **40**(1994), 29–48.
- [8] Arnoux, P., Recoding sturmian sequences on a subshift of finite type: Chaos from order, a worked out example [A], In *Complex Systems*, E. Goles and S. Martinez (eds) [C], Santiago de Chile, (2001), 1–67.
- [9] Artin, E., Ein mechanisches System mit quasiergodischen Bahnen [J], *Abh. Math. Sem. Univ. v. Hamburg*, **3**(1924), 170–175.
- [10] Adler, R. L. & Weiss, B., Similarity of automorphisms of the torus [J], *Memoirs of the American Mathematical Society*, **98**(1970), 1–43.
- [11] Bedford, T. & Fisher, A. M., Ratio geometry, rigidity and the scenery process for hyperbolic Cantor sets [J], *Ergodic Theory and Dynamical Systems*, **17**(1997), 531–564.
- [12] Bowen, R., Markov partitions are not smooth [J], *Proceedings of the American Mathematical Society*, **71**(1978), 130–132.
- [13] Berstel, J. & Séébold, P., Morphismes de Sturm [J], *Bull. Belg. Math. Soc.*, **1**(1994), 175–189.
- [14] Carroll, C., A cutting and stacking approach to pseudo-Anosov automorphisms [R], 1994, Preprint.
- [15] Coven, E. M. & Hedlund, G. A., Sequences with minimal block growth [J], *Math. Systems Theory*, **7**(1973), 138–153.
- [16] daRocha, L. F. C., Another induction for interval exchange maps [R], 1994, Preprint.
- [17] deBruijn, N. G., Sequences of zeros and ones generated by special production rules [J], *Kon. Nederl. Akad. Wetensch. Proc. Ser. A*, **84**(1981), 38–52, (=Indagationes Mathematicae 43).
- [18] deBruijn, N. G., Updown generation of Beatty sequences [J], *Kon. Nederl. Akad. Wetensch. Proc. Ser. A*, **92**(1989), 385–407, (=Indagationes Mathematicae 51).
- [19] Dumont, J. M., Kamae, T. & Takahashi, S., Minimal cocycles with scaling properties and substitutions [J], *Israel Journal of Mathematics*, **95**(1996), 393–410.
- [20] Fisher, A. M., Integer Cantor sets and an order-two ergodic theorem [J], *Ergodic Theory and Dynamical Systems*, **13**(1992), 45–64.
- [21] Grabner, P. J., Liardet, & Tichy, R. F., Odometers and systems of numeration [J], *Acta Arithmetica*, **70**(1995), 103–123.
- [22] Hedlund, G. A., Sturmian minimal sets [J], *American Journal of Mathematics*, **66**(1944), 605–620.

- [23] Hedlund, G. A. & Morse, M., Symbolic dynamics II: Sturmian trajectories [J], *American Journal of Mathematics*, **62**(1940), 1–42.
- [24] Hopf, E., Statistik der geodätischen linien in mannigfaltigkeiten negativer krümmung [J], *Ber. Verh. Sächs. Akad. Wiss. Leipzig*, **91**(1939), 261–304.
- [25] Hopf, E., Ergodic theory and the geodesic flow on surfaces of constant negative curvature [J], *Bulletin of the American Mathematical Society*, **77**(1971), 863–877.
- [26] Ito, S., Number theoretic expansions, algorithms and metrical observations [A], Séminaire de théorie des nombres de Bordeaux, pages 3.01–3.27, 1984.
- [27] Ito, S., Some skew-product transformations associated with continued fractions and their invariant measures [J], *Tokyo Journal of Mathematics*, **9**(1986), 115–133.
- [28] Keane, M., Irrational rotations and quasi-ergodic measures [A], Publications des séminaires de Mathématiques de l'Université de Rennes [C], 1979, 17–26.
- [29] Keane, M., Sur les mesures quasi-ergodiques des translations irrationnelles [J], *Comptes Rendus de l'Académie des Sciences*, **272**(1971), 54–55.
- [30] Kraaikamp, C., Maximal s -expansions are Bernoulli shifts, *Bulletin de la société Mathématique de France*, **121**(1993), 117–131.
- [31] Katok, A. B. & Stepin, A. M., Approximations in ergodic theory [J], *Russian Mathematical Surveys*, **22**(1967), 76–102.
- [32] Lopes, A. O. & daRocha, L. F. C., Invariant measures for Gauss maps associated with interval exchanges maps [J], *Indiana University Mathematical Journal*, **43**(1994), 1399–1438.
- [33] Messaoudi, A., Autour du fractal de Rauzy [D], PhD thesis, Université d'Aix-Marseille 2, 1996.
- [34] Mignosi, F. & Séebold, P., Morphismes sturmiens et règles de Rauzy [R], Technical Report 91-74, Laboratoire d'Informatique Théorique et de Programmation, 1991.
- [35] Masur, H. & Smillie, J., Quadratic differentials with prescribed singularities and pseudo-Anosov diffeomorphisms [J], *Commentarii Math. Helv.*, **68**(1993), 289–307.
- [36] Nogueira, A. & Rudolph, D., Topological weak-mixing of interval exchange maps [J], *Ergodic Theory and Dynamical Systems*, **17**(1997), 1183–1209.
- [37] Ostrowsky, A., Bemerkungen zur Theorie der Diophantischen Approximationen i, ii [J], *Abh. Math. Sem. Hamburg*, **1**(1922), 77–98 & 250–251.
- [38] Paul, M., Minimal symbolic flows having minimal block growth [J], *Mathematical System Theory*, **8**(1974), 309–315.
- [39] Rauzy, G., Nombres algébriques et substitutions [J], *Bulletin de la Société Mathématique de France*, **110**(1982), 147–178.
- [40] Robinson Jr, E. A., The dynamical properties of Penrose tilings [J], *Transactions of the American Mathematical Society*, **348**(1996), 4447–4464.
- [41] Rohlin, V. A., Exact endomorphisms of Lebesgue spaces [J], *American Mathematical Society Translations*, **39**(1964), 1–36.
- [42] Series, C., The modular surface and continued fractions [J], *Journal of the London Mathematical Society*, **31**(1985), 69–80.
- [43] Sós, V., On the distribution mod. 1 of the sequences $n\alpha$ [J], *Ann. Univ. Sci. Budapest Eötvös Sect. Math.*, **1**(1958), 127–134.
- [44] Veech, W. A., Gauss measures for transformations on the space of interval exchange maps [J], *Annals of Mathematics*, **115**(1982), 201–242.
- [45] Veech, W. A., The metric theory of interval exchange transformations i,ii,iii [J], *American Journal of Mathematics*, **106**(1984), 1331–1422.
- [46] Veech, W. A., The Teichmüller geodesic flow [J], *Annals of Mathematics*, **124**(1986), 441–530.
- [47] Vershik, A. M., Arithmetic isomorphisms of hyperbolic automorphisms of the torus and of sofic shifts [J], *Functional analysis and its applications*, **26**(1992), 170–173.
- [48] Vershik, A. M., Locally transversal symbolic dynamics [J], *Algebra i Analiz*, **6**(1994), 94–106 (in Russian).
- [49] Vershik, A. M., The adic realizations of the ergodic actions with the homeomorphisms of the Markov compact and the ordered Bratteli diagrams [J], *Teor. Predstav. Din. Systemy. Kombin. i Algoritmy. Metody.*, **1**(1995), 120–126.
- [50] Vershik, A. & Sidorov, N., Arithmetic expansions associated with the rotation of a circle and continued fractions [J], *Algebra i Analiz*, **5**(1993), 97–115 (in Russian).