6. Sturmian Sequences

In this chapter, we will study symbolic sequences generated by an irrational rotation. Such sequences appear each time a dynamical system has two rationally independent periods; this is a very typical situation, arising for example in astronomy (with the rotation of the moon around the earth, and of the earth around the sun), or in music (with the building of musical scales, related to the properties of $\log 3/\log 2$), and such sequences have been studied for a long time. These sequences, or related objects, appear in the mathematical literature under many different names: rotation sequences, cutting sequences, Christoffel words, Beatty sequences, characteristic sequences, balanced sequences, Sturmian sequences etc...

The latter term was coined by Hedlund and Morse in 1943; they showed that, in a sense, they are the simplest non-trivial sequences, and gave two simple combinatorial characterizations of these sequences. The name comes from a relation with the Sturm comparison theorem (see for example [341], p. 104). In elementary terms, for an irrational number $\alpha \in [0,1]$, the positions of the sequence of zeroes of $\sin(\alpha\pi x + \beta\pi)$ with respect to the integers (zeroes of $\sin(\pi x)$) gives a sequence u, defining u_n as the number of zeroes in the interval $[n, n+1[$; every Sturmian sequence is of this type.

In Sec. 6.1, we define 3 types of sequences: Sturmian sequences, balanced sequences and rotation sequences. We prove that Sturmian sequences are exactly the non-eventually periodic balanced sequences, and that all rotations sequences are Sturmian sequences. We also discuss dynamical systems generated by Sturmian sequences, in the sense of Chap. 3, and prove that each such system contains a sequence with peculiar properties, the so-called "special sequence".

In Sec. 6.2, we generalize these notions to biinfinite sequences; this is technically easier for the following discussion.

In Sec. 6.3, we discuss another way to present Sturmian sequences, using a special encoding process related to an expanding map on the set of all Sturmian sequences, with a simple Markov partition, and derive various consequences; we give several variations of the encoding process, and establish their properties.

In Sec. 6.4, we show how these coding sequences translates to arithmetic; in particular, we use this to prove that all Sturmian sequences are rotation sequences, and give explicit expression for the angle and initial point of the rotation sequence in term of the coding sequence.

In Sec. 6.5, we study the case of periodic coding sequences. We obtain various consequences, algebraic and geometric, among them an explicit expression of the Markov coding for the toral automorphism canonically associated with a periodic coding sequence.

The main theme of this paper is the use of the combinatorial properties of Sturmian sequences and their coding sequences; hence, we tried, as much as possible, to rely on these combinatorial properties to derive from them arithmetic and geometric properties. Other viewpoints are clearly possible, and we provide in the exercises alternative proofs for a number of results. In particular, the exercises of Sec. 6.1 give the original proof that all Sturmian sequences are rotation sequences, a result which is obtained in Sec. 6.4. The combinatorial properties of Sturmian sequences are often easier to understand if one remembers that these are rotation sequences.

For more references on the subject, the reader is referred to [57] and to its impressive bibliography; see also [5, 36, 53, 81, 60, 81, 246, 247].

6.1 Sturmian sequences. Basic properties

We begin this section by giving two equivalent definitions of Sturmian sequences: as sequences of minimal unbounded complexity, and as balanced non-ultimately periodic sequences. In the second part, we give a typical example of Sturmian sequence: the so-called rotation sequences; we will later prove that all Sturmian sequences are of this type (an alternative proof is found in the exercises of the present section). In the last part, we investigate properties of the dynamical system defined by a Sturmian sequence.

6.1.1 Two definitions of Sturmian sequences

Let u be a sequence, with values in a finite set (alphabet) $\mathcal{A}$; we recall from Chap. 1 the following definition and proposition:

Definition 6.1.1. *The language of u is the set $\mathcal{L}(u)$ of finite words that occur in u; we denote by $\mathcal{L}_n(u)$ the set of words of length n that occur in u. The complexity function of u is the function p_u which, to each integer n, associates the number Card $\mathcal{L}_n(u)$ of distinct words of length n that occur in u.*

Proposition 6.1.1. *The complexity p_u is an increasing function. If u is eventually periodic, then p_u is bounded. If there is an n such that $p_u(n+1) = p_u(n)$, then u is eventually periodic.*

From this proposition, we deduce that, if u is not eventually periodic, we must have $p_u(n) \geq n+1$, because p_u has strictly increasing integral values, and $p_u(1)$ must be at least 2, otherwise there would be only one letter and u would be constant. In this chapter, we will be interested in non-periodic sequences with the smallest complexity:

Definition 6.1.2. *A sequence u is called* Sturmian *if it has complexity $p_u(n) = n+1$. We will denote by Σ' the set of Sturmian sequences.*

We will define, in Sec. 6.2, the related set Σ of biinfinite Sturmian sequences; as we will see, this set is slightly more complicated to define, but it is technically easier to use in many proofs.

These sequences are also called *sequences of minimal complexity*, or *sequences with minimal block growth*. The words that occur in a Sturmian sequences cannot disappear:

Proposition 6.1.2. *A Sturmian sequence is* recurrent*, that is, every word that occurs in the sequence occurs an infinite number of times.*

Proof. Suppose that a word U, of length n, occurs in a Sturmian sequence u a finite number of times, and does not occur after rank N. Let v be the sequence defined by $v_k = u_{k+N}$. It is clear that the language of v is contained in that of u, and does not contain U. Hence we must have $p_v(n) \leq n$. This implies that v is eventually periodic, and hence so is u; this is a contradiction. ∎

Since we have $p_u(1) = 2$, Sturmian sequences are sequences over two letters; in this chapter, we fix the alphabet $\mathcal{A} = \{\mathbf{0}, \mathbf{1}\}$. The definition of Sturmian sequences implies a very easy consequence, which we will use several times:

Lemma 6.1.1. *If u is Sturmian, then exactly one of the words* **00**, **11** *does not occur in u.*

Proof. We have $p_u(2) = 3$, so there are exactly three words of length 2 occuring in u. By the previous proposition, **0** and **1** each occur an infinite number of times in u, which implies that **01** and **10** both occur in u. But **00** and **11** are the two other words of length 2, and exactly one of them must occur. ∎

Definition 6.1.3. *We say that a Sturmian sequence is of* type 0 *if* **1** *is* isolated*, that is, if* **11** *does not occur in the sequence. This is equivalent to saying that* **0** *occurs more frequently than* **1**. *We say that the sequence is of type* 1 *if* **00** *does not occur. We denote by Σ'_0 (respectively Σ'_1) the set of Sturmian sequences of type 0 (respectively of type 1).*

Every element of Σ' is either of type 0 or of type 1; it is not immediately clear that Sturmian sequences exist at all, Chap. 3 gives the simplest example (see Exercise 3.3.1):

Example 6.1.1. The Fibonacci sequence $u = 010010100\ldots$, fixed point of the Fibonacci substitution

$$\begin{aligned}\sigma : \mathbf{0} &\mapsto \mathbf{01}\\ \mathbf{1} &\mapsto \mathbf{0}\end{aligned}$$

is a Sturmian sequence.

There is another useful characterization of Sturmian words, for which we shall need the following notation and definition:

Notation 6.1.1 *If U is a finite word over the alphabet $\mathcal{A}$, we denote by $|U|$ the length of U, and $|U|_{\mathbf{a}}$ the number of occurrences of the letter $\mathbf{a}$ in U.*

Definition 6.1.4. *A sequence u over the alphabet $\{\mathbf{0},\mathbf{1}\}$ is* balanced *if, for any pair of words U, V of the same length occuring in u, we have $||U|_{\mathbf{1}} - |V|_{\mathbf{1}}| \leq 1$.*

This means that, for words of $\mathcal{L}_n(u)$, the number of occurrences of $\mathbf{1}$ can take at most two consecutive values; we note that this property is very strong: this is the smallest possible number of values for non-periodic sequences. As we show below, in that case, this property is equivalent to Sturmian. We will often use the following technical lemma:

Lemma 6.1.2. *If the sequence u is not balanced, there is a (possibly empty) word W such that $\mathbf{0}W\mathbf{0}$ and $\mathbf{1}W\mathbf{1}$ occur in u.*

Proof. If u is not balanced, we can find two words A and B of length n such that $|A|_{\mathbf{1}} - |B|_{\mathbf{1}} > 1$. We first prove that we can suppose $|A|_{\mathbf{1}} - |B|_{\mathbf{1}} = 2$. Call A_k (resp. B_k) the suffix of length k of A (resp. B), and $d_k = |A_k|_{\mathbf{1}} - |B_k|_{\mathbf{1}}$. We have $d_n > 1$, $d_0 = 0$, and a short case study proves that $|d_{k+1} - d_k|$ is 0 or 1. But then, an intermediate value argument for integer valued functions shows that there is k such that $d_k = 2$. Suppose now that A and B are words of minimal length with this property. Write $A = \mathbf{a}_0\mathbf{a}_1\ldots\mathbf{a}_{n-1}$ and $B = \mathbf{b}_0\mathbf{b}_1\ldots\mathbf{b}_{n-1}$. We must have $\mathbf{a}_0 = \mathbf{a}_{n-1} = \mathbf{1}$ and $\mathbf{b}_0 = \mathbf{b}_{n-1} = \mathbf{0}$, otherwise we could find a shorter pair by removing some prefix (the first letter if the two initial letters are the same, a longer prefix if $\mathbf{a}_0 = \mathbf{0}$ and $\mathbf{b}_0 = \mathbf{1}$). It is then easy to prove by induction that for all the remaining letters, we have $\mathbf{a}_k = \mathbf{b}_k$, which proves the lemma. ■

Theorem 6.1.1. *A sequence u is Sturmian if and only if it is a non eventually periodic balanced sequence over two letters.*

Proof. We first prove the "if" part, by contradiction. Suppose that u is not Sturmian, we will show it is not balanced. Let n_0 be the smallest integer such that $p_u(n_0+1) \geq n_0+3$. We have $n_0 \geq 1$, since $p_u(1) = 2$. Because $p_u(n_0) = n_0 + 1$, there are at least two words U and V of length n_0 that can be extended on the right in two ways. Since n_0 is the smallest integer with this property, U and V differ only in the first letter; hence there is a word W such that $U = \mathbf{0}W$ and $V = \mathbf{1}W$. We have proved that $\mathbf{0}W\mathbf{0}$ and $\mathbf{1}W\mathbf{1}$ occur in u, hence this sequence is not balanced.

We now prove the converse: suppose that u belongs to Σ'; this implies, by definition, that it is a non eventually periodic sequence on two letters. We do a proof by contradiction, and suppose that it is not balanced. By the lemma 6.1.2, there is a word W such that $\mathbf{1}W\mathbf{1}$ and $\mathbf{0}W\mathbf{0}$ occur in u. We consider such a word of minimal length $n+1$, and we write $W = w_0w_1\dots w_n$.

Remark that, under this assumption, if we have a pair of words U, V of same the length such that $||U|_\mathbf{1} - |V|_\mathbf{1}| \geq 2$, then their length is at least $n+3$, that is, $\mathbf{0}W\mathbf{0}$ and $\mathbf{1}W\mathbf{1}$ is a pair of non balanced words of minimal length; for we can apply the same argument as in the proof of Lemma 6.1.2 to show that such a pair contains a pair $\mathbf{0}W'\mathbf{0}$ and $\mathbf{1}W'\mathbf{1}$.

The word W cannot be empty, otherwise $\mathbf{00}$ and $\mathbf{11}$ occur in u, which is impossible by lemma 6.1.1. For the same reason, we must have $w_0 = w_n$; more generally, we must have $w_k = w_{n-k}$, that is, W is a *palindrome*, otherwise, letting $w_k = \mathbf{0}$, $w_{n-k} = \mathbf{1}$, we see that $\mathbf{0}w_1\dots w_{k-1}\mathbf{0}$ and $\mathbf{1}w_{n-k+1}\dots w_n\mathbf{1}$ is a non balanced pair of smaller length, which is impossible. Recall that a palindrome is a word that stays the same when it is read backwards; for example, *in girum imus nocte et consumimur igni* is a classical palindrome in Latin, *a man, a plan, a canal: Panama!* a classical palindrome in English.

But now, we know that there are $n+2$ words of length $n+1$; W can be extended in two ways on the right and on the left, and all the others can be extended in only one way. Exactly one of $\mathbf{0}W$ and $\mathbf{1}W$ (suppose it is $\mathbf{0}W$) can be extended in two ways on the right (if a word can be extended in two ways on the right, all its suffixes can also). So, the words $\mathbf{0}W\mathbf{0}$, $\mathbf{0}W\mathbf{1}$ and $\mathbf{1}W\mathbf{1}$ occur in u, but not $\mathbf{1}W\mathbf{0}$. Let i be the rank of an occurence of $\mathbf{1}W\mathbf{1}$ in u. We prove:

Lemma 6.1.3. *The word* $\mathbf{0}W$ *cannot occur in* $u_iu_{i+1}\dots u_{i+2n+3}$.

Proof of Lemma 6.1.3. The length of $u_iu_{i+1}\dots u_{i+2n+3}$ is $2n+4$, the length of $\mathbf{1}W\mathbf{1}$ is $n+3$, and the length of $\mathbf{0}W$ is $n+2$. Hence this lemma exactly means that the first letter of an occurence of $\mathbf{0}W$ cannot occur in an occurence of $\mathbf{1}W\mathbf{1}$ in u. Suppose the beginning of $\mathbf{0}W$ overlaps $\mathbf{1}W\mathbf{1}$: $\mathbf{0}$ can obviously not be the first or last letter in $\mathbf{1}W\mathbf{1}$; if it is w_k, the overlap means that $\mathbf{0}w_0\dots w_{n-k} = w_kw_{k+1}\dots w_n\mathbf{1}$; but this implies that $w_k = \mathbf{0}$ and $w_{n-k} = 1$, and we get a contradiction, since W is a palindrome. ∎

End of the proof of Theorem 6.1.1. It is immediate that there are exactly $n+3$ words of length $n+2$ occuring in $u_iu_{i+1}\dots u_{i+2n+3}$; but there are $n+3$ words of length $n+2$, and $\mathbf{0}W$ does not occur, as we just proved, so at least one word occurs twice. But all these words can be extended in a unique way on the right (only $\mathbf{0}W$ can be extended in two ways on the right), hence u is eventually periodic: this is a contradiction. ■

From this characterization of Sturmian sequences as balanced sequences, we get the following proposition.

Proposition 6.1.3. *The* frequency *of* $\mathbf{1}$ *in a Sturmian sequence* u*, defined as the limit of* $\frac{|u_0u_1\dots u_{n-1}|_{\mathbf{1}}}{n}$ *when* n *tends to* ∞*, is well defined, and is irrational.*

Proof. Let a_n be the minimum number of $\mathbf{1}$ that occur in a word of length n occuring in u. Since $|u_0u_1\dots u_{n-1}|_{\mathbf{1}}$ is either a_n or a_n+1, it is enough to prove that the limit of a_n/n exists and is irrational.

A word of length $kq+r$ can be split in k words of length q and 1 word of length r, and we get the inequality $ka_q \leq a_{kq+r} \leq k(a_q+1)+r$. If we consider $n > q^2$, we can write $n = kq+r$, with $k \geq q$ and $0 \leq r < q$; since $r < k$, we check that $\frac{r}{n} < \frac{k}{n} \leq \frac{1}{q}$, hence the second part of the inequality gives $\frac{a_n}{n} \leq \frac{a_q}{q} + \frac{2}{q}$. One checks also immediately that the quantity $ra_q - n$ is negative, since $n \geq a_n \geq ka_q > ra_q$; hence the first part of the inequality above implies that $\frac{n}{q}(a_q-1) = ka_q + \frac{1}{q}(ra_q - n) \leq a_n$, and, after dividing by n, we get:

$$\frac{a_q}{q} - \frac{1}{q} \leq \frac{a_n}{n} \leq \frac{a_q}{q} + \frac{2}{q}$$

It follows that the sequence $(a_n/n)_{n\in\mathbb{N}}$ is a Cauchy sequence, and hence converges to some limit α.

Suppose that this limit is rational, equal to $\frac{p}{q}$. From the inequality $ka_n \leq a_{kn} < a_{kn}+1 \leq k(a_n+1)$, we deduce that, if n divides n', then $\frac{a_n}{n} \leq \frac{a_{n'}}{n'} < \frac{a_{n'}+1}{n'} \leq \frac{a_n+1}{n}$. In particular, the sequence $\left(\frac{a_{2^nq}}{2^nq}\right)_{n\in\mathbb{N}}$ is increasing, and the sequence $\left(\frac{a_{2^nq}+1}{2^nq}\right)_{n\in\mathbb{N}}$ is decreasing; we get the inequality:

$$\frac{a_q}{q} \leq \frac{a_{2^nq}}{2^nq} < \frac{a_{2^nq}+1}{2^nq} \leq \frac{a_q+1}{q}.$$

But this sequence must converge to $\frac{p}{q}$; this is only possible if $a_q = p$ and $a_{2^nq} = 2^np$ for all n, or $a_q+1 = p$, and $a_{2^nq}+1 = 2^np$ for all n.

We prove that it is impossible to have $a_{2^nq} = 2^na_q$ for all n: since the sequence is not periodic, there is at least one word U of length q in the language of u such $|U|_1 = a_q+1$. Since u is recurrent, this word occurs an infinite number of times, so that it must occur in two positions congruent mod q. Hence, we can find a word of length 2^nq that can be split in words of length q, with at least two occurences of U; the number of $\mathbf{1}$ in this word is at least 2^na_q+2, so that $a_{2^nq} > 2^na_q$. A similar proof shows that $a_{2^nq}+1$ cannot be equal to $2^n(a_q+1)$ for all n, and we get a contradiction. ■

From this theorem, it follows that the language of u (see Definition 1.1.2) is not so easy to compute:

Corollary 6.1.1. *If u is a Sturmian sequence, the language $\mathcal{L}(u)$ is not* rational *(or equivalently, it cannot be recognized by a finite automaton)*

Proof. Suppose that $\mathcal{L}(u)$ is rational; we can apply the pumping lemma (see also Chap. 4): there is an N such that any word U of length more than N can be split as AWB, where W is a nonempty word with the following property: for any n, AW^nB belongs to $\mathcal{L}(u)$. Hence W^n also belongs to $\mathcal{L}(u)$; the frequency of $\mathbf{1}$ in W^n is equal to $|W|_1/|W|$, which is rational, and it tends to the frequency of $\mathbf{1}$ in the sequence u. By the preceding proposition, we get a contradiction. ∎

It is possible to define the balance property for sets of finite words; the following easy result, proved in [256] (see also Chap. 4), will be used in later exercises:

Exercise 6.1.1. We will say that a set E of finite words is balanced if, for any pair of words U, V in E, and for any words U', V' of same length that occur in U, V, we have $||U|_1 - |V_1|| \leq 1$. Prove that a balanced set of words of length n contains at most $n+1$ distinct words. (*(Hint: recurrence.)*

Exercise 6.1.2. As a consequence, prove that if u and v are two Sturmian sequences with the same frequency, they have the same language.(*Hint: Show that the minimum number of* $\mathbf{1}$ *in a word of length n is the same for sequences u and v. Deduce that $\mathcal{L}_n(u) \cup \mathcal{L}_n(v)$ is a balanced set.*)

Exercise 6.1.3. Prove the converse: the frequency of $\mathbf{1}$ in a Sturmian sequence depends only on the language, not on the sequence itself, and explain how one can compute the frequency, knowing only the language. Deduce that every Sturmian sequence is minimal, that is, every word that occurs in the sequence occurs with bounded gaps (see Chap. 3). (*Hint: Suppose that U occurs in u with arbitrarily large gaps. Prove that there exists in the orbit closure of u a sequence v where U does not occur. Deduce that v is eventually periodic, and get a contradiction.*)

6.1.2 Rotation sequences

We now generalize the example of the Fibonacci sequence. We first need some definitions.

Notation 6.1.2 *Let x be a real number. The* integral part *of x is the integer $[x] = \sup\{n \in \mathbb{Z} | n \leq x\}$. The* fractional part *of x is the real number $\{x\} = x - [x]$. The* ceiling *of x is the integer $\lceil x \rceil = \inf\{n \in \mathbb{Z} | n \geq x\}$.*

Definition 6.1.5. *A* rotation sequence *is a sequence u such that there is an irrational number $\alpha \in [0,1]$ and a real number β such that:*

$$(\forall n \in \mathbb{N})\,(u_n = [(n+1)\alpha + \beta] - [n\alpha + \beta]) \tag{6.1}$$

$$\textit{or } (\forall n \in \mathbb{N})\,(u_n = \lceil (n+1)\alpha + \beta \rceil - \lceil n\alpha + \beta \rceil) \tag{6.2}$$

The number α will be called the angle *associated with the rotation sequence, and β the* initial point.

Rotation sequences occur in a number of classical situations. The simplest one is the natural coding associated with a rotation. Consider the rotation:

$$R : \mathbb{T}^1 \to \mathbb{T}^1 \qquad x \mapsto x + \alpha \mod 1$$

where $\mathbb{T}^1$ is the circle, identified to $\mathbb{R}/\mathbb{Z}$, or to the interval $[0, 1[$, by the function e^{ix}. We consider the two intervals I_0, I_1 on $\mathbb{T}^1$ delimited by 0 and $1 - \alpha$. We denote by ν the *coding function* defined by $\nu(x) = \mathbf{0}$ if $x \in I_0$, $\nu(x) = \mathbf{1}$ otherwise; then the rotation sequence defined by α and β is just the sequence $\nu(R^n(x))$, that is the coding of the positive orbit of β under the rotation R.

Remark. Of course, this definition is not precise enough: to properly define ν, we need to state what happens at the endpoints of the two intervals. We will always take the intervals closed on one side and open on the other. If we take $I_0 = [0, 1 - \alpha[$, $I_1 = [1 - \alpha, 1[$, this corresponds to chosing the integral part in the definition of the rotation sequence. If we take the other choice $I_0 =]0, 1 - \alpha]$ and $I_1 =]1 - \alpha, 1]$, it corresponds to the choice of the ceiling function. This choice is irrelevant, except if there is a positive integer n such that $\beta + n\alpha$ is an integer, that is β is congruent to $-n\alpha$ mod. 1. In that case, α and β define two possible rotation sequences, depending on this choice. We will see later that the rotation sequence defined by the angle α and the initial point α, and the two rotation sequences defined by the angle α and the initial point $1 - \alpha$, play an important role.

In most proofs, we will only consider rotation sequences defined using $[x]$; it is usually immediate to check each time that the proof works also for $\lceil x \rceil$.

Exercise 6.1.4. Prove the following alternative characterization: consider the line $y = \alpha x + \beta$ on the plane, with α irrational. This line cuts all the horizontal lines $y = n$, and all the vertical lines $x = n$. This defines a so-called "*cutting sequence*": write $\mathbf{0}$ each time the line cuts a vertical line, and $\mathbf{1}$ each time it cuts a horizontal line. Prove that the cutting sequence is a rotation sequence associated with the angle $\frac{\alpha}{1+\alpha}$ and the initial point $\frac{\beta}{1+\alpha}$ (we need to be more precise in the case where the line goes through an integral point, but this happens at most once; in this case, we write $\mathbf{01}$ or $\mathbf{10}$ as we like).

Remark. In the case $\alpha < 1$, we can also define a Sturmian sequence by defining $u_n = \mathbf{1}$ if the line crosses a horizontal $y = n$ for x between n and $n+1$, $u_n = \mathbf{0}$ otherwise. We recover in this way the sequence of the definition. In this setting, the number α is also sometimes called the *slope* of the sequence, and β the *intercept*.

An other classical example is the *square billiard*: we shoot a ball in a square billiard, with initial irrational slope a. The ball bounces on the sides of the billiard according to the laws of elastic shock, and we write **0** for a vertical side, and **1** for a horizontal side. This can be reduced in a straightforward way to a cutting sequence.

Since we will prove later that all Sturmian sequences are of this type, it is important to keep these examples in mind, because they can give intuition on the properties of Sturmian sequences.

Proposition 6.1.4. *Every rotation sequence is a Sturmian sequence.*

Proof. Because they are simple and instructive, we will give two proofs.

We first prove that a rotation sequence is balanced. From the definition of the sequence, we can compute the number of **1** in $u_k \dots u_k + n$:

$$|u_k \dots u_{k+n}|_{\mathbf{1}} = [(n+k+1)\alpha + \beta] - [k\alpha + \beta]$$

And it is easy to check that, for a fixed n, this number can take only two values, $[n\alpha]$ or $[n\alpha] + 1$, therefore the sequence is balanced.

It is a little more subtle to prove that the sequence is Sturmian. Consider the sequence as defined by the coding of the orbit of β for the rotation R with respect to the partition $\mathcal{P}$ in two intervals. The letter u_n tells us in which element of the partition $\mathcal{P}$ is $R^n\beta$, or equivalently in which element of $R^{-n}\mathcal{P}$ is β. The first n letters will give the position of β with respect to the partition intersection of $\mathcal{P}, \dots R^{1-n}\mathcal{P}$. In the same way, words of length n in the sequence will give the position of iterates of β with respect to this partition. But an easy computation shows that this partition is defined on the circle by the $n+1$ points $0, -\alpha, \dots, -n\alpha$, so that it consists of $n+1$ intervals (since α is irrational, the $n+1$ points are distinct). Hence, there can be no more than $n+1$ different words of length n. This proof is less straightforward than the first, but it can be generalized to more complicated cases (cf. [31, 40, 39]). ■

We can build in this way a large number of Sturmian sequences. In fact, we get in this way **all** sturmian sequences: the converse of the proposition is true, and it is one of the main points of Sec.s 6.3 and 6.4 to prove this result and to determine explicitly the angle and initial point of a given Sturmian sequence.

Exercise 6.1.5. It is possible to prove that Sturmian sequences are rotation sequences in an elementary way, as it is done in the original paper of Hedlund and Morse [256], by using the results of exercises 6.1.1 and 6.1.2, along the following lines.

1) The sequence v of angle α and initial point 0 is Sturmian, with frequency α.

2) If u is an arbitrary Sturmian sequence of frequency α, there is a sequence n_k such that $S^{n_k}v$ tends to u. (*Hint: use Exercise 6.1.2.*)

3) There is a real number β such u is the rotation sequence (or one of the two rotation sequences) with angle α and initial point β.

Our approach will give us a more complete information on α and β.

Exercise 6.1.6. Prove that the language associated with a Sturmian sequence is *palindromic*, that is, the reverse word of any word in $\mathcal{L}_u$ also belongs to $\mathcal{L}_u$; this property is very easy for a rotation sequence (for example, one can use the obvious symmetry on the cutting sequence starting from the origin), but there is a purely combinatorial proof using Exercise 6.1.1. (*Hint: if we consider a balanced set of words, and add to it the one of these words read backwards, it stays balanced; but a balanced set of words of length n contains at most $n+1$ words.*)

6.1.3 Dynamical systems associated with Sturmian sequences

We have defined in Chap. 1 the dynamical system associated with a sequence. Let us recall the main points: we consider the set of all sequences on $[0,1]$, with the product topology (this is a compact set, as product of finite sets), and the shift S on this set: the image $v = Su$ of a sequence u is the sequence v defined by $v_i = u_{i+1}$. In other words, v is u, with first letter u_0 deleted. To a sequence u, we associate the set X_u if there is no risk of confusion, which is the closure of $\{S^n u | n \in \mathbb{N}\}$, the orbit of u for the shift. The set X_u is by construction shift-invariant, and (X_u, S) is the dynamical system associated with the shift. It is easy to show that the language of any sequence in X_u is included in the language of u, but in general it could be strictly smaller. The dynamical system associated with a Sturmian sequence has special properties:

Theorem 6.1.2. *The system associated with a Sturmian sequence is one-to-one, except one point that has two preimages.*

Proof. We first prove that the system is onto. Let v be a point in X_u; if v is of the form $S^n u$, with $n > 0$, then it is the image of $S^{n-1}u$. If it is not in the orbit of u, then there is a sequence n_k, with $\lim_{k\to\infty} n_k = \infty$, such that $S^{n_k}u$ tends to v; from $S^{n_k-1}u$, we can, by compacity, extract a sequence that has a limit w, and this sequence w satisfies $Sw = v$. The only difficult point is to prove that u has a preimage. But we have seen that the sequence u is recurrent, that is, every word that occurs in u occurs an infinite number of times. Hence, the initial word of length k of u occurs again at some position n_k of u, with $n_k > 0$, and the sequence $S^{n_k}u$ tends to u. For the same reason as before, u also has a preimage in X_u.

We now prove that there is exactly one point with two preimages. The proof we just gave shows that any word in $\mathcal{L}_n(u)$ can be extended on the left, since it appears past the initial position. Since there are $n+1$ words of length n, exactly one of them can be extended in two different ways; we call this word G_n. It is clear that any prefix of G_{n+1} can be extended on the left in two ways, so G_n is the prefix of length n of G_{n+1}. If a sequence u has two preimages, all its prefixes are the G_n; but there is exactly one such sequence in X_u. ∎

Definition 6.1.6. *We call the G_n the* left special words *of $\mathcal{L}_n(u)$, and the sequence g which has the G_n as prefixes is called the* left special sequence *of X_u.*

Exercise 6.1.7. Define the *right special word* R_n, and prove that R_n is the reverse word of L_n. Prove that, if a_n is the minimum number of $\mathbf{1}$ in a word of length n occuring in the Sturmian sequence u, then $|\mathbf{0}L_n|_1 = a_n$.

We can find explicitly all the preimages of g, and this will be useful later:

Proposition 6.1.5. *The sequence g has two preimages $\mathbf{0}g$ and $\mathbf{1}g$, and two preimages of order 2, $\mathbf{10}g$ and $\mathbf{01}g$.*

In other words, the preimages of g differ only on two letters.

Proof. The preimages of g are of course $\mathbf{0}g$ and $\mathbf{1}g$. Suppose now that $\mathbf{1}$ is isolated in g (that is, $\mathbf{11}$ does not occur). Then the only preimage of $\mathbf{1}g$ has to be $\mathbf{01}g$. Because $\mathbf{1}$ is isolated, we can suppose that g starts with a string of k $\mathbf{0}$ followed by a $\mathbf{1}$. If the preimage of $\mathbf{0}g$ is $\mathbf{00}g$, then the language of u contains a string of $n+2$ $\mathbf{0}$, in $\mathbf{00}g$, and a string of n $\mathbf{0}$ surrounded by $\mathbf{1}$ in $\mathbf{1}g$, which is impossible since u is balanced. ∎

Exercise 6.1.8. Prove by recurrence that, for each n, m, the words $D_n\mathbf{01}G_m$ and $D_n\mathbf{10}G_m$ belong to the language of u. (*Hint: prove that if a_n is the minimum number of $\mathbf{1}$ in a word of length n in u, we have $a_{i+j} = a_i + a_j$ or $a_{i+j} = a_i + a_j + 1$; then use Exercise 6.1.2.*)

Prove that, for each $n \geq 3$, the special sequence has two preimages of order n, $D_{n-2}\mathbf{10}g$ and $D_{n-2}\mathbf{01}g$.

When we consider cutting sequences, the meaning of the proposition becomes clear: the special sequence corresponds to the line through $(0,0)$, which can be coded in two ways at this point; in all the other points, the coding is unique, and when we establish that all Sturmian sequences are cutting sequences, we have an immediate proof of the proposition and the exercise.

Remark. The special words have many other combinatorial properties. Some of them, the bi-special words, are at the same time left special and right special, hence they are palindromes; one can prove that any word of $\mathcal{L}(u)$ is contained in a bi-special word, and give rules to generate them (the so-called Rauzy rules): this is another way to obtain the results of Sec. 6.3.

Exercise 6.1.9. Find the special word of the Fibonacci substitution. Consider now the substitution σ defined by $\sigma(\mathbf{0}) = \mathbf{010}$ and $\sigma(\mathbf{1}) = \mathbf{10}$. Find its special word, in terms of the fixed words of the substitution. How is σ related to the Fibonacci substitution? (*Hint: show that, for any sequence u, the image by σ of $\mathbf{01}u$ and $\mathbf{10}u$ differ only on the first two letters; show that σ has 2 fixed points that differ only in the first 2 letters.*)

Exercise 6.1.10. Prove that the set Σ' is not closed, and that its closure is the set of balanced sequences. Prove that the complement of Σ' in its closure is countable. (*Hint: the set balanced sequences is closed, and the set of periodic balanced sequences is countable.*)

6.2 Biinfinite Sturmian sequences

In this section, we consider biinfinite sequences; this will allow some simplifications of Sec. 6.1. We will in a first part define balanced sequences, and Sturmian biinfinite sequences. In the second part, we will show that, for a definition in terms of complexity, we need to take some precautions in the case of biinfinite sequences. In a third part, we will consider dynamical systems generated by biinfinite sequences, and their relations with the systems considered in the previous section.

6.2.1 Balanced biinfinite sequences

We consider now sequences $u = (u_n)_{n\in\mathbb{Z}} \in \{\mathbf{0}, \mathbf{1}\}^{\mathbb{Z}}$. For such a sequence, we need to define a special position to differentiate from shifted sequences:

Definition 6.2.1. *The* initial letter *of a biinfinite sequence u is u_0.*

We can define, in the same way as before, the words occuring in the sequence, and the definition of balanced words still makes sense:

Definition 6.2.2. *A biinfinite sequence u over the alphabet $\{\mathbf{0}, \mathbf{1}\}$ is* balanced *if and only if, for any pair of words U, V of the same length occuring in u, we have $||U|_{\mathbf{1}} - |V|_{\mathbf{1}}| \leq 1$.*

The definition of periodicity is the same for infinite or biinfinite sequences; however, there is a difference for eventual periodicity. We say that a sequence u is *positively eventually periodic* if there exist an integer $p > 0$ and an integer $N \in \mathbb{Z}$ such that, for all $i > N$, $u_{i+p} = u_i$. The sequence is *negatively eventually periodic* if, for all $i < N$, $u_{i-p} = u_i$.

Lemma 6.2.1. *A balanced sequence is positively eventually periodic if and only if it is negatively eventually periodic.*

Proof. As we saw in the preceding section, if we consider a sequence U_n of words occuring in u, such that $|U_n| = n$, then the frequency $\frac{|U_n|_{\mathbf{1}}}{n}$ of $\mathbf{1}$ in U_n converges to some limit α. If we take another sequence of words V_n occuring in u, since the sequence u is balanced, the limit is the same; in particular, in that case, $\lim_{n\to\infty} \frac{|u_0 u_1 \ldots u_{n-1}|}{n} = \lim_{n\to\infty} \frac{|u_{-n+1} \ldots u_{-1} u_0|}{n} = \alpha$.

As we proved in the preceding section, α is rational if and only if the sequence u is positively eventually periodic (in that case, if W is a word corresponding to the period, the sequence $\frac{|u_0 u_1 \dots u_{n-1}|}{n}$ converges to $\frac{|W|_\mathbf{1}}{|W|}$). But the symmetric proof, using the sequence $(u_{-n})_{n\in\mathbb{N}}$, shows that α is rational if and only if the sequence u is negatively eventually periodic. This proves the result. ■

As a consequence, it suffices to say that a balanced sequence is eventually periodic, in which case, it must be simultaneously positively and negatively eventually periodic.

Exercise 6.2.1. Prove that, if u is balanced and eventually periodic, the period, and the words that occur infinitely often, are the same in the positive and the negative direction.

Definition 6.2.3. *A biinfinite sequence over the alphabet* $\{\mathbf{0}, \mathbf{1}\}$ *is* Sturmian *if and only if it is balanced and not eventually periodic.*

Remark. In the original paper [256], Sturmian bi-infinite sequences are defined as balanced sequences; however, we do not need two different words for the same concept, and eventually periodic balanced sequences turn out to give some problems in the framework of the next section; this is the reason of the present definition.

6.2.2 Complexity of biinfinite sequences

We can as before define the complexity $p_u(n)$ of a biinfinite sequence u as the number of distinct words of length n that occur in u. The function p_u is still increasing; however, the basic property is no more true: there are eventually periodic biinfinite sequences of unbounded complexity. The simplest such sequence is $\mathbf{0}^\infty\mathbf{1}\mathbf{0}^\infty$, that is, the biinfinite sequence u defined by $u_0 = \mathbf{1}$ and $u_n = \mathbf{0}$ if $n \neq 0$.

It is nevertheless possible to define Sturmian biinfinite sequences in terms of complexity.

Proposition 6.2.1. *A biinfinite sequence u is Sturmian if and only if it is a sequence of complexity $n + 1$ that is not eventually periodic.*

Remark. The statement can seem ambiguous; when we say that the biinfinite sequence u is not eventually periodic, this can have two different meanings for a general sequence: we can require, either that the sequence is not eventually periodic in one direction, or that it is not eventually periodic in both directions. In this case however, there is no ambiguity, since we proved that, for a balanced sequence, or for a sequence of complexity $n + 1$, positive eventual periodicity is equivalent to negative eventual periodicity.

Proof. Suppose that u is not positively eventually periodic and of complexity $n+1$; then the restricted sequence $(u_n)_{n\in\mathbb{N}}$ is an infinite sequence of complexity at most $n+1$ (some words could have disappeared) that is not eventually periodic. Hence it is of complexity exactly $n+1$, so it is Sturmian, and balanced, by the results in Sec. 6.1. But then, since all the words that occur in the biinfinite sequence also occur in the positive infinite sequence, the original sequence u is balanced, and not eventually periodic, so it is Sturmian. The symmetric proof works if u is not negatively eventually periodic.

Conversely, suppose that the biinfinite sequence u is Sturmian. Then, the positive infinite sequence is balanced and not eventually periodic, so that it is of complexity $n+1$. The same is true for all sequences $(u_{N+n})_{n\in\mathbb{N}}$, for all $N\in\mathbb{Z}$; hence all these sequences have the same language, and the union of all these languages is the language of the biinfinite sequence u. Hence this biinfinite sequence has complexity $n+1$. ∎

Exercise 6.2.2. Prove that there are infinitely many eventually periodic biinfinite sequences of complexity $n+1$, and that this set is countable.

Exercise 6.2.3. Prove that there are sequences of complexity $n+1$ that are not balanced. (*Hint:* $\mathbf{0}^\infty\mathbf{1}^\infty$.)

The methods of the next section will allow us to describe completely the set of biinfinite sequences of complexity $n+1$; we can then show that, if they are not Sturmian, then they are ultimately periodic in both directions, and that the frequencies of $\mathbf{1}$ in the future and the past are either equal (and the sequence is balanced), or Farey neighbors, that is, rational numbers $\frac{p}{q}, \frac{p'}{q'}$ such that $|pq'-p'q|=1$ (and the sequence is not balanced); we have given above examples of both types.

6.2.3 Dynamical systems generated by Sturmian biinfinite sequences

We consider the shift S on $\{\mathbf{0},\mathbf{1}\}^{\mathbb{Z}}$, defined by $Su=v$, where $v_n=u_{n+1}$ for all $n\in\mathbb{Z}$. The difference with the preceding section is that this shift is invertible, since $S^{-1}u=w$, where $w_n=u_{n-1}$.

As in the preceding section, we can associate to a Sturmian sequence u its orbit closure X_u, that is, $X_u=\overline{\{S^n u | n\in\mathbb{Z}\}}$. The set X_u is invariant by S, and the shift on X_u is one-to-one, because it is on the total space $\{\mathbf{0},\mathbf{1}\}^{\mathbb{Z}}$. But X_u is also S^{-1}-invariant, and this implies that the shift on X_u is onto.

The set $\{\mathbf{0},\mathbf{1}\}^{\mathbb{Z}}$ has a natural product topology, given by the metric defined $d(u,v)=1/2^k$, where $k=\inf\{n\in\mathbb{N} | u_n\neq v_n \text{ or } u_{-n}\neq v_{-n}\}$. Hence we can consider (X_u,S) as a topological dynamical system for the induced topology.

Definition 6.2.4. *A* Sturmian system *is a dynamical system* (X_u,S)*, where* X_u *is the orbit closure of a Sturmian biinfinite sequence.*

Exercise 6.2.4. Prove that a Sturmian dynamical system is recurrent and minimal.

There is a natural projection Π from $\{\mathbf{0},\mathbf{1}\}^{\mathbb{Z}}$ to $\{\mathbf{0},\mathbf{1}\}^{\mathbb{N}}$. This projection sends a biinfinite Sturmian system X_u to an infinite Sturmian system $(X_u)'$. While the projection from the total space is highly non-injective, its restriction to X_u is almost one-to-one, as we show in the next exercise.

Exercise 6.2.5. Prove that the projection $\Pi : X_u \to (X_u)'$ is at most 2-to-one; prove that it is one-to-one except on a countable set, and describe this set.

We can define a different pseudo-metric on X_u in the following way:

Definition 6.2.5. *We set* $\delta(u,v) = \limsup_{n\to\infty} \frac{\#\{i\in\mathbb{Z} \mid -n\le i\le n, u_i\ne v_i\}}{2n+1}$

This pseudo-metric turns out to have nice properties.

Exercise 6.2.6. 1) Prove that the superior limit in the preceding definition is in fact a limit.

2) Prove that, if $\delta(u,v)=0$, then either $u=v$, or u and v differ in exactly two places.

3) Prove that, in the last case, u and v project to the orbit of the special point.

4) Prove that δ is shift invariant.

5) Prove that the quotient space of X_u by the relation $\delta(u,v) = 0$ is isometric to a circle, and the shift is semi conjugate to a rotation on that circle.

In the language of ergodic theory, this shift is a 2-to-1 extension of a circle rotation, which is 1-to-1 except on a countable set. In particular, it is measurably isomorphic to a rotation.

Remark. We know that the properties in Exercise 6.2.6 are true, because every Sturmian sequence is a rotation sequence, and it is easy in that case to compute δ; however, I do not know a direct combinatorial proof of these properties (that is: the combinatorial proof asked for in this exercise!), and I should be interested to know one; finding a purely combinatorial proof of the above exercise is an open problem. Such a proof could be interesting, since it could extend to other symbolic systems, such as the systems generated by some substitution.

We see that pairs of biinfinite sequences that differ only in two points play a particular role; they are pre-images of the orbit of the special sequence used in the last section. It is not difficult to show that, for each $n\in\mathbb{Z}$, there are exactly two sequences that differ exactly in n and $n+1$.

Definition 6.2.6. *The* fixed sequences *of a Sturmian system X_u are the two sequences that differ exactly at indices 0 and 1.*

These two sequences play an important role in the next section; they project to the preimages of order two of the special sequence.

In the next section, we will define a renormalization procedure for Sturmian sequences, and show that fixed sequences renormalize to fixed sequences; we will then be able to produce these sequences as fixed points for an infinite product of substitutions, hence the name.

6.3 Coding sequences for Sturmian sequences and Sturmian systems

A Sturmian sequence contains very little information: among the 2^n possible words of length n, it uses only $n+1$ words. For example, if the sequence is of type 0, every **1** is isolated, so the next letter must be a **0**, and gives no new information. Hence, it must be possible to recode this sequence in a more compact way. The basic idea is to group, in a sequence of type 0, any **1** with an adjacent **0**, to obtain a new sequence. It turns out that a small miracle occurs: this new sequence is again Sturmian, and the procedure can be iterated; to the given Sturmian sequence, we can associate an infinite coding sequence, and the initial word of length n of the Sturmian sequence is, for a typical Sturmian sequence, defined by an initial word of the coding sequence of length approximately $\log n$.

The procedure is simple, but there are difficulties: first of all, the coding procedure is not canonical: for a sequence of type 0, we can group every **1** with the preceding or the next **0**, that is, we can use two different substitutions, and similarly for sequences of type 1. We will see below that there is also a problem for the coding of the initial letter, and that we can recode using a suffix or a prefix; this gives a total of 16 possible coding procedures. The choice of the procedure is arbitrary, and several of them can be convenient for different purposes, as we shall see.

A second problem is that, if we consider one-sided infinite sequences, the coding is not well defined for all sequences. We will see in exercises that the sequences in the positive orbit of the special sequence admit two possible codings. However, this is not really a serious problem: for a given Sturmian system, it arises only on a countable set; but, this makes exact proofs and statement quite cumbersome. This difficulty disappears for biinfinite sequences; for that reason, from now on, we will only consider biinfinite sequences, except in exercises.

A third problem is that, even for biinfinite sequences, there are pairs of sequences (the fixed sequences of the system and their negative orbit) that admit the same coding. The underlying reason is that these pairs of sequences correspond to the same point for the associated circle rotation.

An important question is then to identify all possible coding sequences for Sturmian sequences. It turns out that the admissible coding sequences form a sofic system, defined by a simple automaton, and even a subshift of finite type for a well-chosen coding procedure (see the definitions below).

A last remark: it is possible to code a given sequence, or the language associated with this sequence; coding the language is the same thing as associating a coding sequence to the dynamical system generated by the given sequence. We will first, since it is easier, recode Sturmian systems; a system is completely defined by its frequency α, and in Sec. 6.4, we will see that the coding we get is closely related to the continued fraction expansion of α. The coding sequence for a rotation sequence defines the angle α and the initial point β, and we will give arithmetical interpretations of these codings in Sec. 6.4.

6.3.1 Recoding a Sturmian sequence

From now on, unless specifically stated, we consider only biinfinite Sturmian sequences.

As we remarked at the beginning, in a Sturmian sequence, one of the letters is always isolated. We recall that the sequence is of type 0 if $\mathbf{1}$ is isolated, of type 1 otherwise. It is clear that any Sturmian sequence of type 0 can be written using only the words $\mathbf{0}$ and $\mathbf{10}$. We will make this formal:

Definition 6.3.1. *We note σ_0 the substitution defined by $\sigma_0(\mathbf{0}) = \mathbf{0}$, $\sigma_0(\mathbf{1}) = \mathbf{10}$, and by σ_1 the substitution defined by $\sigma_1(\mathbf{0}) = \mathbf{01}$, $\sigma_1(\mathbf{1}) = \mathbf{1}$.*

These substitutions extend in a natural way to finite words (as morphism of the free monoid), to infinite sequences, and to biinfinite sequences (with the convention that, if $u = \sigma(v)$, then u_0 is the first letter of the word $\sigma(v_0)$.

Lemma 6.3.1. *If u is Sturmian of type 0 (resp. type 1), then there exists a unique v such that either $u = \sigma_0(v)$, or $u = S\sigma_0(v)$(resp. $u = \sigma_1(v)$ or $u = S\sigma_1(v)$)*

Exercise 6.3.1. Give a formal proof of this lemma; remark that an image $u = \sigma_0(v)$ can never satisfy $u_{-1}u_0 = \mathbf{10}$, hence the shift $S\sigma_0(v)$ is necessary in that case.

The remarkable fact is that v is Sturmian:

Lemma 6.3.2. *A biinfinite sequence v is Sturmian, if and only if $\sigma_0(v)$ is Sturmian.*

Proof. It is easy to prove that v is positively eventually periodic if and only if $\sigma_0(v)$ is positively eventually periodic. Hence we can restrict to the case where both v and $\sigma_0(v)$ are not eventually periodic.

Suppose that v is Sturmian. If $\sigma_0(v)$ is not Sturmian, by Lemma 6.1.2, there is a (possibly empty) word W such that $\mathbf{1}W\mathbf{1}$ and $\mathbf{0}W\mathbf{0}$ occur in $\sigma_0(v)$. But $\sigma_0(v)$, by construction, contains only isolated $\mathbf{1}$. Hence we can write

$W = \mathbf{0}V\mathbf{0}$, and $V\mathbf{0} = \sigma_0(V')$. Then, taking inverse images, v must contain $\mathbf{1}V'\mathbf{1}$ and $\mathbf{0}V'\mathbf{0}$, which is impossible.

Suppose that v is not Sturmian; again, there is a (possibly empty) word W such that $\mathbf{0}W\mathbf{0}$ and $\mathbf{1}W\mathbf{1}$ occur in v. Since the sequence v is biinfinite, it contains the word $\mathbf{a0}W\mathbf{0}$, where $\mathbf{a}$ is an arbitrary letter. Then $\sigma_0(v)$ contains the words $\sigma_0(\mathbf{1})\sigma_0(W)\sigma_0(\mathbf{1}) = \mathbf{10}\sigma_0(W)\mathbf{10}$ and $\sigma_0(\mathbf{a})\sigma_0(\mathbf{0})\sigma_0(W)\sigma_0(\mathbf{0}) = \sigma_0(\mathbf{a})\mathbf{0}\sigma_0(W)\mathbf{0}$. Because, in any case, $\sigma_0(\mathbf{a})$ ends with $\mathbf{0}$, we see that $\sigma_0(v)$ contains the words $\mathbf{10}\sigma_0(W)\mathbf{1}$ and $\mathbf{00}\sigma_0(W)\mathbf{0}$: therefore, it is not balanced. ■

We can define in this way a "recoding map" on sturmian sequences:

Definition 6.3.2. *We denote by $\Phi : \Sigma \to \Sigma$ the map defined by $\Phi(u) = v$, where v is the unique sequence such that $u = S^\epsilon \sigma_i(V)$, with $\epsilon = 0$ or 1, $i = 0$ or 1.*

Definition 6.3.3. *We denote by Σ_0 (resp. Σ_1) the set of Sturmian sequences of type 0 (resp. 1).*

We denote by $\Sigma_0^{\mathbf{0}}$ (resp. $\Sigma_0^{\mathbf{1}}$, $\Sigma_1^{\mathbf{0}}$, $\Sigma_1^{\mathbf{1}}$) the set of Sturmian sequences of type 0 such that the initial letter is $\mathbf{0}$ (resp. of sequences of type 0 with initial letter $\mathbf{1}$, of type 1 with initial letter $\mathbf{0}$, of type 1 with initial letter $\mathbf{1}$).

This partition is nicely related to the map Φ:

Exercise 6.3.2. Prove the following properties:

1) $u = \sigma_0(\Phi(u))$ if and only if $u \in \Sigma_0$ and u and v have the same initial letter.

2) $u = S\sigma_0(\Phi(u))$ if and only if $u \in \Sigma_0$ and u and v have a different initial letter, in which case $u_0 = \mathbf{0}$ and $\Phi(u)_0 = \mathbf{1}$.

3) $\Phi(\Sigma_0^{\mathbf{0}}) = \Phi(\Sigma_1^{\mathbf{1}}) = \Sigma$, and Φ restricted to $\Sigma_0^{\mathbf{0}}$ (resp. $\Sigma_1^{\mathbf{1}}$) is one-to-one.

4) $\Phi(\Sigma_0^{\mathbf{1}}) = \Sigma_0^{\mathbf{1}} \cup \Sigma_1^{\mathbf{1}}$, and Φ restricted to $\Sigma_0^{\mathbf{1}}$ is one-to-one.

This exercise proves, among other things, that the partition by the four sets $\Sigma_\epsilon^{\mathbf{a}}$ is a Markov partition for the map Φ.

Associated with these partitions, we can define maps that will be useful in the next subsections:

Definition 6.3.4. *We denote by τ (for "type") the map $\tau : \Sigma \to \{0, 1\}$ defined by $\tau(u) = 0$ if u is of type 0, $\tau(u) = 1$ if u is of type 1.*

We denote by η the map $\eta : \Sigma \to \{\mathbf{0}, \mathbf{1}\}$ $\quad u \mapsto u_0$ that sends u to its initial letter.

We denote by γ the map defined by $\gamma(u) = (\tau(u), \eta(u))$. This map tells us in which of the sets $\Sigma_i^{\mathbf{j}}$ is the sequence u.

It is easy to prove that the fixed sequences, as defined in the previous section, behave nicely under the map Φ.

Exercise 6.3.3. Prove that, if u and u' are a pair of fixed sequences that differ only at position 0 and 1, the same is true for the recoded sequences $\Phi(u)$ and $\Phi(u')$.

It is also possible to define a recoding map for one-sided infinite sequences, but there are some technical problems, as shown in the following exercise.

Exercise 6.3.4. 1) Prove that, for any sequence u of type 0, there exists a unique sequence v such that $u = \sigma_0(v)$, and that Sv is Sturmian.

2) Prove that v is not always Sturmian.

3) Prove that, if u, of type 0, is not a special sequence, there exists a unique Sturmian sequence v such either $u = \sigma_0(v)$ or $u = S\sigma_0(v)$.

4) Prove that if u is a special sequence of type 0, there exists a special sequences v such that $u = \sigma_0(\mathbf{0}v) = S\sigma_0(\mathbf{1}v)$.

It is now possible, by iterating the map Φ and coding with respect to a partition (that is, by using maps τ or γ), to associate a coding sequence with any biinfinite Sturmian sequence. We will do so in the next subsection for fixed sequences, giving a very simple rule for their generation, and for Sturmian systems; we consider general Sturmian sequences in the following subsection.

6.3.2 Coding sequences for Sturmian systems

We can recode Sturmian systems, by using the following lemma:

Lemma 6.3.3. *Let u and u' two Sturmian sequences in the same Sturmian system X; if u and u' are recoded respectively to v and v', v and v' belong to the same Sturmian system. In particular, v and v' have the same type.*

Proof. We can suppose that u and u' are of type 0. It is enough to prove that, if V is the initial word of v of length n, it occurs in v'. But V must occur infinitely often in v, otherwise $S^k v$ should be of complexity at most $P(n) = n$ for some k, and should be eventually periodic. Hence we can suppose that V is included in some word W, occuring in v, that starts with $\mathbf{1}$; the word $\sigma_0(W)$ occurs in u, so it occurs in u', and, because it begins with $\mathbf{1}$ and ends with $\mathbf{0}$, it can be recoded in only one way by σ_0, as W. This proves that W, and V, occur in v'. ■

This lemma shows that, for any Sturmian sequence, u, the coding sequence $(\tau(\Phi^n(u)))_{n\in\mathbb{N}}$ only depends on the Sturmian system X_u.

Definition 6.3.5. *The* (additive) coding sequence *of a Sturmian system X, is the sequence $\tau(\Phi^n(u)))_{n\in\mathbb{N}}$, for any sequence $u \in X$.*

This definition raises two natural questions: what are the *admissible sequences*, that is, the sequences that can occur as coding sequence for a Sturmian system? Is a Sturmian system completely defined by its additive coding sequence? We have a complete and constructive answer for the latter question:

Proposition 6.3.1. *A Sturmian system is completely defined by its additive coding sequence*

Proof. We will prove a little more: the fixed points of the system are completely defined by the additive coding sequence.

Let $(i_n)_{n\in\mathbb{N}}$ be the coding sequence, with $i_n = 0$ or 1. Using the exercise 6.3.3, we see that the two fixed points of the system can be written $u = \sigma_{I_0}\sigma_{i_1}\ldots\sigma_{i_n}(v)$, where v is the fixed point of the recoded system that has same initial letter as u. It is then enough to prove that this determines an arbitrarily long initial word $u_0\ldots u_k$ of u, independently of v.

Indeed, this shows that the fixed points of the system have as initial word $\sigma_{I_0}\sigma_{i_1}\ldots\sigma_{i_n}(\mathbf{0})$ and $\sigma_{I_0}\sigma_{i_1}\ldots\sigma_{i_n}(\mathbf{1})$, and it is easy to prove that the length of at least one of these words (and in fact both) tends to the infinite, hence the language is completely determined. ■

Exercise 6.3.5. *(Rauzy rules)* There is a very simple way to compute these initial words: let U_n and V_n be the sequence of words defined by:

- $U_0 = \mathbf{0}$, $V_0 = \mathbf{1}$
- $U_{n+1} = U_n$, $V_{n+1} = V_nU_n$ if $i_n = 0$
- $U_{n+1} = U_nV_n$, $V_{n+1} = V_n$ if $i_n = 1$

1) Prove that U_n (resp. V_n) are initial words of the fixed point that has $\mathbf{0}$ (resp. $\mathbf{1}$) as initial letter in the Sturmian system with coding sequence (i_n).

2) Prove that, with the exception of a finite set of indices, all the words U_n and V_n are suffixes of the same infinite negative sequence, and that this sequence is a Sturmian sequence with two right extensions.

3) Prove that if we remove from U_n or V_n a prefix of length 2, the remaining part is a palindrome.

A first application of this coding is the following:

Proposition 6.3.2. *The system generated by a Sturmian sequence is minimal.*

Proof. Recall that a system is minimal if it does not contain a nonempty closed invariant subset, or equivalently, if the orbit of any element is dense.

If the system generated by u is not minimal, it contains a sequence v whose orbit is not dense. Then, there is a word U that occurs in u, but not in v. This implies that $p_v(n) < n+1$ for some n, so that v is eventually periodic. Because the language of v is contained in that of u, there is a finite nonempty word W such that W^n occurs in u for all n. We will prove that this is impossible.

Since the sequence u is balanced, it cannot contain sequences of $\mathbf{0}$ or $\mathbf{1}$ of arbitrary length. Hence, W cannot be constant. Suppose that u is of type $\mathbf{0}$; after a cyclic permutation, we can suppose that the word W begins with $\mathbf{1}$ and ends with $\mathbf{0}$, and W can be recoded in a unique way in a word W' which is strictly shorter. Consider the sequence u' obtained by recoding u: its language contains all the powers of W'. We can iterate the procedure, and at each step we reduce the length of the word whose powers all belong to the language. After a finite number of steps, we get a word of length 1, so that the corresponding Sturmian sequence contains arbitrarily long sequences of $\mathbf{0}$ or $\mathbf{1}$: we just saw that this is impossible. ∎

Exercise 6.3.6. Prove directly, using Exercise 6.1.1, that a Sturmian dynamical system is minimal, and deduce that every point in this system is a Sturmian sequence, with the same frequency.

It is now easy to find the admissibility condition for additive coding sequences, thus solving the question above:

Proposition 6.3.3. *A sequence with value in* $\{0,1\}$ *is an admissible coding sequence for a Sturmian system if and only if* 0 *and* 1 *occur an infinite number of times, or equivalently if and only if it is not eventually constant.*

Proof. Consider a Sturmian sequence of type 0; let k be the minimum number of $\mathbf{0}$ between two $\mathbf{1}$ occuring in the sequence. Then the sequence can be recoded exactly k times using σ_0, to a sequence of type 1. By the same argument, any string of 0 or 1 in the additive coding sequence must be of finite length, and the sequence cannot be eventually constant.

Conversely, given a not eventually constant sequence, consider the set of words $U_n = \sigma_{i_0}(\sigma_{i_1}(\cdots(\sigma_{i_n}(\mathbf{0}))\cdots))$ of Exercise 6.3.5; it is easy to prove that these words are prefix of each other, and define a Sturmian sequence. The associated system has the given sequence $(i_n)_{n\in\mathbb{N}}$ as coding sequence. ∎

We can rewrite in another way the additive coding of X. If we group strings of 0 and 1, we can write the sequence as $0^{a_0}1^{a_1}\ldots0^{a_{2n}}1^{a_{2n+1}}\ldots$.

Definition 6.3.6. *This sequence* (a_n) *of integers, all strictly positive except maybe* a_0*, is called the* multiplicative coding sequence *for the system.*

We will see in the next part that this sequence is closely related to the frequency α: it is in fact the continued fraction expansion of $\frac{1-\alpha}{\alpha}$. The first sequence is related to the additive algorithm for the continued fraction expansion of the same number, this is the reason for the name.

The following is an immediate consequence of the previous proposition:

Proposition 6.3.4. *A sequence* (a_n) *of integers is an admissible multiplicative coding sequences if and only if* $a_0 \geq 0$ *and* $a_n > 0$ *for* $n \geq 1$.

The first integer can be 0, because of our convention that a_0 is the number of leading 0's, and the sequence could be of type 1.

Exercise 6.3.7. It is also possible to consider balanced eventually periodic sequences. Prove that the associated dynamical system is countable, with one finite orbit and one or two countable orbits asymptotic to the finite orbit in both directions, and that, to this system, on can associate an additive coding sequence that is eventually constant; prove that the frequency of such a system is rational. (*Hint: you can start by considering the sequences defined in a self-evident notation by:* $\mathbf{0}^\infty$, $\mathbf{0}^\infty\mathbf{1}\mathbf{0}^\infty$, *and* $(\mathbf{01})^\infty$, $(\mathbf{01})^\infty(\mathbf{10})^\infty$, $(\mathbf{10})^\infty(\mathbf{01})^\infty$.)

6.3.3 Coding sequences for Sturmian sequences

We saw in the preceding subsection that any Sturmian system can be defined in a unique way by its additive coding sequence, related to a sequence of substitutions. We would like to do the same thing with a Sturmian sequence, that is, to associate with it a coding sequence that defines it in a unique way.

Definition 6.3.7. *Let u be a Sturmian sequence. The* additive coding sequence *of u is the sequence $(\gamma(\Phi^n(u)))_{n\in\mathbb{N}}$.*

The two usual questions arise: does the additive coding sequence completely define the Sturmian sequence? What are the admissible coding sequences?

Proposition 6.3.5. *The additive coding sequence completely defines the Sturmian sequence, except if it is in the negative orbit of the fixed points; pair of preimages of same rank of the two fixed points of a system have the same coding.*

The admissible sequences are the non eventually constant sequences of vertices for infinite paths in the graph of Fig. 6.1, where we have represented all the possible transitions, labelling each edge by the corresponding transformation (we have given each vertex the name of the set $\Sigma_\epsilon^{\mathbf{a}}$ instead of $(\epsilon, \mathbf{a})$).

Proof. Let us first examine the admissibility condition.

It is clear that if a sequence u is equal to $S\sigma_0\Phi(u)$, then $\gamma(u) = (0, \mathbf{0})$ (it is of type 0, because we use substitution σ_0, and it begins with $S\sigma_0(\mathbf{1}) = \mathbf{0}$), and we have $\gamma(\Phi(u)) = (0, \mathbf{1})$ or $(1, \mathbf{1})$. A similar study for the other cases proves that the possible arrows must be as indicated on Fig. 6.1. Hence, all admissible coding sequences must correspond to paths in the graph.

As we saw above, there must be an infinite number of changes of type in the sequence $(\Phi^n(u))$; this is equivalent to saying that the path must not be eventually constant, since any path staying on the upper or lower level is eventually constant.

$$
\begin{tikzcd}[row sep=huge, column sep=huge]
\Sigma_0^{\mathbf{0}} \arrow[loop left, "\sigma_0"] \arrow[r, "S\sigma_0"] \arrow[d, shift right, "\sigma_0"'] \arrow[dr, shift left, "S\sigma_0"] & \Sigma_0^{\mathbf{1}} \arrow[loop right, "\sigma_0"] \arrow[d, shift right, "\sigma_0"'] \\
\Sigma_1^{\mathbf{0}} \arrow[loop left, "\sigma_1"] \arrow[u, shift right, "\sigma_1"'] & \Sigma_1^{\mathbf{1}} \arrow[loop right, "\sigma_1"] \arrow[l, "S\sigma_1"'] \arrow[u, shift right, "\sigma_1"'] \arrow[ul, shift left, "S\sigma_1"]
\end{tikzcd}
$$

Fig. 6.1. The transition graph for additive coding

It remains to prove that any non eventually constant path in the graph is admissible.

Consider now an infinite path in the graph, and the corresponding sequence $(\varepsilon_n, \mathbf{a}_n)$. From this sequence, we deduce a sequence c_n, with $c_n = \sigma_{\varepsilon_n}$ or $S\sigma_{\varepsilon_n}$ (if there exists a sequence u with coding sequence $(\varepsilon_n, \mathbf{a}_n)$, we must have $\Phi^n(u) = c_n \Phi^{n+1}(u)$).

The sequence of words $c_0 c_1 \dots c_n(\mathbf{a}_{n+1})$ defines an increasingly large part $u_{i_n} \dots u_0 \dots u_{j_n}$ of u. If both i_n and j_n are unbounded, the Sturmian sequence u is completely defined by $(\varepsilon_n, \mathbf{a}_n)$; hence it exists, and the sequence is admissible.

If i_n is bounded, then after a finite time there are no more shifts, and c_n is eventually of the form σ_0 or σ_1. Without loss of generality, we can suppose that all the c_n are of this form, or equivalently, that all the $\Phi^n(u)$ have same initial letter, and $(\mathbf{a}_n)$ is constant. But then, it is clear that u must be one of the two fixed points (the choice of the fixed point is given by the letter). Remark that, formally, only the positive part of the sequence is defined by the coding; however, the fixed point has a unique left extension, since it is in the negative orbit of the special sequence, and so it cannot be also in its positive orbit. Hence in that case also, the sequence is admissible.

There remains to check the case of a bounded sequence (j_n). It is then constant after a finite time, and we can suppose, replacing u by a suitable $\Phi^n(u)$, that $j_n = 0$ for all n. But then, the path on the graph must be restricted to the diagonal, since $\sigma_0(\mathbf{1})$ and $\sigma_1(\mathbf{0})$ cannot occur. It is easy to check that this diagonal path is the additive coding sequence of the preimages of the two special points. Hence, this sequence also is admissible, but it does not completely define the sturmian sequence.

We have proved that all non eventually constant path are admissible, and that they define a unique Sturmian sequence, except in the case when they restrict eventually to the diagonal; such paths define a pair of preimages of same order of the two fixed points of some Sturmian system. ∎

Remark. The basic idea of the theorem is that, in a Sturmian system, there is a one-to-one correspondence between Sturmian sequences and coding sequences, except on a countable set (negative orbit of the special sequence), but the precise details can seem puzzling. We will make the picture more clear in the next section: to a Sturmian system, we will associate a rotation on the circle. To any point, we associate a Sturmian sequence and a coding sequence, except for the orbit of 0; each point of the positive orbit of 0 corresponds to a pair of Sturmian sequences, and to a pair of coding sequences that are eventually, one of the form $(\epsilon_n, \mathbf{0})$ and the other one of the form $\epsilon_n, \mathbf{1})$, and each point of the negative orbit of 0 corresponds to a pair of Sturmian sequences and a unique coding sequence.

Exercise 6.3.8. We had to exclude from consideration eventually constant coding sequences; it would be nicer to allow also these sequences, so that the set of admissible sequences should be defined by an automaton. Explain how this could be done, by allowing all balanced sequences, including eventually periodic sequences.

Remark. Let $(\epsilon_n, \mathbf{a}_n)$ be an additive coding sequence for a Sturmian sequence u, and consider the natural projection of this sequence to (ϵ_n); the latter sequence is an additive coding sequence for a Sturmian system X, and it is immediate to check that u belongs to X.

It is also possible to characterize an orbit of the Sturmian system in this way:

Lemma 6.3.4. *Let u be a Sturmian sequence, and $v = S^k u$. Then there exists $k' \leq k$ such that $\Phi(v) = S^{k'}\Phi(u)$ is a recoded sequence for u. Moreover, if u is of type 0 (resp.1), we have $k' = k$ if and only if $u_0 \ldots u_{k-1} = \mathbf{0} \ldots \mathbf{0}$ (resp. $u_0 \ldots u_{k-1} = \mathbf{1} \ldots \mathbf{1}$*

Proof. Suppose that u is of type 0. Then there is a unique k' such that $\sigma_0(S^{k'}\Phi(u)) = v$ or $\sigma_0(S^{k'}\Phi(u)) = 1v$, so that $S^{k'}\Phi(u)$ is $\Phi(v)$. A simple case study shows that $k' < k$, except if all the first letters of u are $\mathbf{0}$. ∎

Proposition 6.3.6. *Let u be a Sturmian sequence, and $v = S^k u$. Then $\Phi^n(u)$ coincides eventually with $\Phi^n(v)$, except if u is in the negative orbit of a fixed point, and v in the positive orbit of this fixed point.*

Proof. By the preceding lemma, we have $\Phi^n(v) = S^{k_n}\Phi^n(u)$, where k_n is a decreasing sequence of integers. If it tends to 0, the proof is finished. Otherwise, it is constant, equal to k after a rank N. This means that, for $n > N$, u^n begins with k $\mathbf{0}$ if it is of type 0, and with k $\mathbf{1}$ if it is of type 1. It suffices to consider the case of a sequence of type 0 that recodes to a sequence of type 1 (this must occur infinitely often in the family) to check that we must have $k = 1$. We must then have $\phi^n(v) = \sigma_i \Phi^{n+1}(v)$ for $n \geq N$, hence $\phi^N(v)$ is a fixed point of its system. We conclude that v must be in the positive orbit of a fixed point, and u in the negative orbit of the same fixed point. ∎

6.3.4 Multiplicative coding sequences for Sturmian sequences

As in subsection 6.3.2, we can rearrange the additive coding sequence in blocks of symbols of same type 0 or 1. The idea is to "accelerate" the coding map Φ.

Definition 6.3.8. *We define the map $\Psi : \Sigma \to \Sigma$ by $\Psi(u) = \Phi^{n_u}(u)$, where n_u is the smallest strictly positive integer such that $\Phi^n(u)$ is not of the same type as u.*

It is clear from the graph in Fig. 6.1 that, if u is of type 0, we have $u = \sigma_0^n \Psi(u)$ if u and $\Psi(u)$ have same initial letter, and $u = \sigma_0^a S \sigma_0 \sigma_0^b$ if they have different initial letter; in the latter case, the initial letter of u is **0**, and the initial letter of $\Psi(u)$ is **1**.

We can simplify this expression, using the following exercise:

Exercise 6.3.9. Prove that, if v is a sequence with initial letter **1**, we have the equality $\sigma_0^n(Su) = S^{n+1}\sigma_0^n(u)$

Hence, if u is of type 0, we can always write $u = S^k \sigma_0^a \Psi(u)$, and a similar property for Sturmian sequences of type 1.

Definition 6.3.9. *We will call first multiplicative coding sequence of a Sturmian sequence u the unique sequence (a_n, k_n) such that:*

- *If u is of type 0,$\Psi^n(u) = S^{k_n}\sigma_{\epsilon_n}^{a_n}\Psi^{n+1}(u)$, with $\epsilon_n = n \mod 2$.*
- *If u is of type 1, $(a_0, k_0) = (0, 0)$ and $\Psi^n(u) = S^{k_{n+1}}\sigma_{\epsilon_n}^{a_{n+1}}\Psi^{n+1}(u)$, with $\epsilon_n = n + 1 \mod 2$.*

The properties of this multiplicative coding sequences are summarized in the following theorem:

Theorem 6.3.1. *The admissible multiplicative coding sequences are the double sequences of integer (a_n, k_n), with $a_0 \geq 0$, $a_n > 0$ if $n > 0$, and $0 \leq k_n \leq a_n$, that are label of paths in the infinite graph of Fig. 6.2*

Each Sturmian sequence admits a unique multiplicative coding sequence.

Each coding sequence defines a unique Sturmian sequence, except the sequences that satisfy $k_n = 0$ for all n, which define the two fixed points of a system, and the sequences that satisfy eventually $k_n = a_n$, which define two preimages of the same order of the two fixed points.

Proof. The admissibility condition is immediate by looking at the graph for additive coding sequence: it suffices to consider all finite paths that stay on one level, except for the last state. The unicity condition is then just a reformulation of the additive case. ∎

Exercise 6.3.10. Reformulate the definition of the coding sequence by using defining a Markov partition for the map Ψ (it may be convenient to change slightly the definition to make it more symmetric with respect to type)

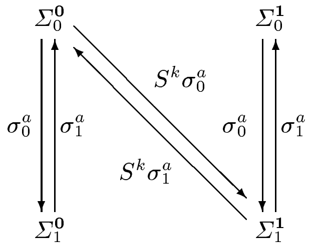

Fig. 6.2. The transition graph for multiplicative coding

We can reformulate the admissibility condition:

Proposition 6.3.7. *A sequence (a_n, k_n) is an admissible multiplicative coding sequence if it satisfies $a_0 \geq 0$, $a_n > 0$ if $n > 0$, $0 \leq k_n \leq a_n$, and the additional condition: in the sequence (k_n), any maximal block of 0, except maybe the initial block, is of even length.*

Proof. After coding by $S^k\sigma_0^a$, with $k > 0$, we must be in state $(1, 1)$. If we recode by σ_1^b, we get to the state $(0, 1)$, and it is clear on the graph that the next coding symbol can only be of the type σ_0^c; the proof is then clear by recurrence. ∎

The set of admissible coding sequences form what we call a sofic system (on an infinite alphabet), since it is defined by the set of admissible paths on a graph with a finite number of vertices; it is however not a finite type system (that is, a symbolic system where we can check whether a sequence is admissible by looking at subwords of bounded lengths), since we might need to consider arbitrarily long words to check whether a given sequence is admissible. We will make a small modification to improve this.

As we remarked at the beginning of this section, when we recode a sequence u of type 0, it is not always possible to write $u = \sigma_0(v)$; we have then the choice to write $u = S\sigma_0(v)$, or $u = \mathbf{0}\sigma_0(w)$, where v begins with $\mathbf{1}$ and $w = Sv$. If we recode n times, and if we call σ the product of the n substitutions involved, the first view point leads to $u = S^k\sigma(v)$, and the second to $u = X\sigma(w)$ (X is a finite word, prefix of u) , where (if $k > 0$) $w = Sv$. If v starts with the letter $\mathbf{a}$, we check that X is a suffix of $\sigma(\mathbf{a})$, and $k + |X| = |\sigma(\mathbf{a})|$.

We can build another coding, the second multiplicative coding sequence, using the idea of suffix; we just state the theorem, and explain the relation to the first coding sequence:

Theorem 6.3.2. *For any Sturmian sequence u, we can define a sequence (a_n, b_n), with $b_n \leq a_n$, and a family of recoded Sturmian sequences $w^{(n)}$,* *such that* $w^{(2n)} = \mathbf{0}^{b_{2n}}\sigma_0(w^{(2n+1)})$ *and* $w^{(2n+1)} = \mathbf{1}^{b_{2n+1}}\sigma_1(w^{(2n+2)})$*. The coding sequence* (a_n, b_n) *is uniquely defined. The sequence* (a_n, b_n) *uniquely defines the sequence* u*, except if* b_n *is eventually 0, in which case it defines two preimages of the same order of the two fixed points.*

The set of admissible sequences is completely characterized by the following properties:

- *all the* a_n *are strictly positive integers, except maybe the first one.*
- *the integers* b_n *satisfy* $b_n \leq a_n$*.*
- *if* $b_{n+1} = a_{n+1}$*, then* $b_n = 0$*.*

These admissibility conditions are summarized in the graph of Fig. 6.3

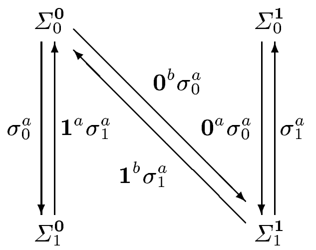

Fig. 6.3. The transition graph for prefix multiplicative coding.

We leave the proof as an exercise, since it is completely similar to the preceding one, the basic idea being to study precisely the first step of the coding process.

Definition 6.3.10. *The sequence* (a_n, b_n) *is called the second multiplicative coding of* u*.*

The interest of this second coding is that the admissibility condition is now of finite type; we see that, as for the first multiplicative coding, there is a one-to-one relation between Sturmian sequences and special sequences, except on the orbit of the special sequence. It is in fact easy to compute one of the coding sequences, knowing the other, by the next proposition:

Proposition 6.3.8. *Let* u *be a Sturmian sequence not in the positive orbit of the special sequence, and let* (a_n, k_n) *and* (a_n, b_n) *be the first and second multiplicative coding sequence of* u*. Let* $N = \inf\{n \in \mathbb{N} | k_n \neq 0\}$*.*

- *For* $n < N$*, we have* $b_n = k_n = 0$*.*
- $b_N = 1 + a_N - k_N \neq 0$*.*

- *For* $n > N$ *and* $k_n \neq 0$*, we have* $b_n = a_n - k_n$*.*
- *For* $n > N$ *and* $k_n = 0$*, if* $k_{n-1} = b_{n-1} = 0$*, then* $b_n = a_n$*; otherwise,* $b_n = 0$*.*

Proof. The idea is to consider the family of recoded sequences $(v^{(n)}$ and $(w^{(n)}$ associated with the two coding processes; it is easy to prove that $w^{(n)} = v^{(n)}$ for $n < N$, and $w^{(n)} = Sv^{(n)}$ for $n \geq N$. The condition then follows, by studying the first letter of v_n. ∎

Exercise 6.3.11. Define a map and a partition associated with this coding; prove that the map is infinite-to-one.

Exercise 6.3.12. *Review Exercise* The reader may feel confused by this variety of recoding sequences; we can organize them in a more systematic way.

We saw before that any coding map will use, for sequences of type 0, one of the two substitutions σ_0, τ_0, and the shift S or its inverse, and similarly for sequences of type 1. This gives 16 elementary coding maps.

We will denote $\Phi_{\epsilon_0,\delta_0,\epsilon_1,\delta_1}$, with all subscripts equal $+1$ or -1, the coding map given, on sequences of type 0, by the use of σ_0 if $\epsilon_0 = 1$ and τ_0 if $\epsilon_0 = -1$, and $S\sigma_0$ or $S\tau_0$ if $\delta_0 = 1$, or $S^{-1}\sigma_0$ or $S^{-1}\tau_0$. If $\delta_0 = -1$, and similarly for sequences of type 1, so that, for example, the map Φ defined above is equal to $\Phi_{1,1,1,1}$.

It is quite obvious that these coding are not all completely different, and we can make this more precise.

Let E be the flip that exchanges **0** and **1** (it is the trivial substitution $\mathbf{0} \mapsto \mathbf{1}$, $\mathbf{1} \mapsto \mathbf{0}$).

1) Prove that E is a one-to-one map on the set of all Sturmian sequences, and that it conjugates $\Phi_{\epsilon_0,\delta_0,\epsilon_1,\delta_1}$ to $\Phi_{\epsilon_1,\delta_1,\epsilon_0,\delta_0}$.

Let R be the *retrogression* map, that is, the map that takes u the the *retrograde* sequence defined by $v_n = u_{-1-n}$ (this map reads u backwards, and exchange the positive and the strictly negative part of the sequence; for many reasons, it turns out to be useful to have no fixed letter, as one would get by the natural definition $v_n = u_{-n}$).

2) Prove that R is a one-to-one map on the set of Sturmian sequences, and that it preserves all Sturmian systems.

3) Prove that R conjugates $\Phi_{\epsilon_0,\delta_0,\epsilon_1,\delta_1}$ and $\Phi_{-\epsilon_0,-\delta_0,-\epsilon_1,-\delta_1}$.

4) Prove that we obtain in this way 6 classes of coding maps globally invariant by flip and exchange, 2 of them with 4 members, and 4 of them with 2 members.

We will see in Sec. 6.6 another map (duality) on the set of all coding maps.

5) Prove that the coding sequence of a Sturmian system does not depend on the choice of the coding map.

Remark. See [173], where Sturmian sequences are expanded under one of the above form to get an explicit formula which computes the supremum of all real numbers $p > 0$ for which there exist arbitrarily long prefixes which are pth-powers.

6.3.5 The dual additive coding for Sturmian sequences

In this section, we will define another coding that will be useful in Sec. 6.4 and 6.6. We leave the proofs to the reader.

As we stated at the beginning of this part, the choice of the two substitutions σ_0, σ_1 is somewhat arbitrary. We could for example replace the substitution σ_1 by τ_1, such that $\tau_1(\mathbf{0}) = \mathbf{10}$, and $\tau_1(\mathbf{1}) = \mathbf{1}$ (this amounts to grouping any isolated $\mathbf{0}$ in a sequence of type 1 with the preceding $\mathbf{1}$, instead of the next $\mathbf{1}$). We can then recode any sequence of type 1 as $\tau_1(v)$ or $S\tau_1(v)$.

It is then possible to associate to each Sturmian sequence a *dual coding sequence*, using now $(\sigma_0, S\sigma_0, \tau_1, S\tau_1)$, as we did before.

It is now possible to associate to the pair (σ_0, τ_1) a curious multiplicative coding: first group the substitutions in blocks $\sigma_0\tau_1^{a_n-1}$ and $\tau_1\sigma_0^{a_n-1}$; then recode, using alternatively suffix and prefix; that is, write $u = \mathbf{0}^b\tau_1\sigma_0^{a_n-1}(v)$ or $u = S^b\sigma_0\tau_1^{a_n-1}(v)$, depending on the case.

Exercise 6.3.13. Prove that, in this way, one can associate to any Sturmian sequence a double sequence of integers (a_n, b_n), where (a_n) is the coding sequence for the Sturmian system, and b_n satisfies $0 \leq b_n \leq a_n$, and $b_n = a_n$ implies $b_{n+1} = 0$.

Remark. Remark that, for this coding, the admissibility condition is a condition of finite type, which is exactly similar to the condition we obtained in the last section, except that the direction is reversed; this is why we call this the dual coding. The next section gives an arithmetic and dynamic natural interpretation to this strange coding.

6.3.6 Dynamical systems on coding sequences

The set of additive or multiplicative coding sequences is invariant by the shift; in this way, we can define a dynamical system, by shifting coding sequences. It is important to note that this new dynamical system is completely different from the Sturmian dynamical systems we considered in previous sections.

Let us note by Γ the set of additive coding sequences; there is an almost one-to-one map from the set Σ of biinfinite sturmian sequences to γ. This map conjugates the map Φ on Σ to the shift on Γ. It is easy to check that Φ is at least 2-to-1, and one can prove that it is of strictly positive topological entropy, and ergodic for a suitable measure(this is clear on Γ, since it is an irreducible subshift of finite type).

One can also consider the shift on the set of all Sturmian sequences; we saw above that it is a union of disjoint Sturmian systems. To each Sturmian system, we can associate the coding sequences with a given sequence of substitutions; it is then possible to write explicitly the map corresponding on coding sequences to the Sturmian shift. This is what is called an adic system, and it is somewhat similar to an odometer.

Exercise 6.3.14. Let (a_n, k_n) be the first multiplicative coding sequence of a Sturmian sequence u; write down explicitly the multiplicative coding sequence for Su. Do the same for the second multiplicative coding sequence, and for the dual multiplicative coding.

6.4 Sturmian sequences. Arithmetic properties and continued fractions

The main purpose of this section is to make effective the coding sequence we defined above.

More precisely, if we are given a Sturmian sequence, we have explained how to obtain, using the map Φ, its coding sequence. We will now associate with a given multiplicative coding sequence (a_n, b_n), a rotation sequence whose angle and convergent are defined by explicit arithmetic formulas in term of a_n and b_n, using continued fraction expansion and Ostrowski expansion.

This proves in a constructive way that every Sturmian sequence is a rotation sequence.

In the first subsection, we study the induction of a rotation on a subinterval, and show that it is directly related to the recoding of Sturmian sequences. We show that there are several possibilities for induction, in one-to-one relation with the different recoding processes we already explained. We give explicit arithmetic formulas for a particular induction, obtained by iteration of the induction on the image of the largest continuity interval.

In Sec. 6.4.2 and 6.4.3, we give without proofs the basic facts on continued fractions and Ostrowski expansion; we apply these formulas in Sec. 6.4.4 to recover the angle and incident of a rotation sequence from its coding sequence. We pay particular attention there to the possibility of multiple coding, and its geometric and arithmetic meaning.

We survey in Sec. 6.4.5, for later use, the other possibilities of induction algorithm, and in particular Rauzy induction and its link with the dual multiplicative algorithm.

6.4.1 Induction of rotations and recoding of rotation sequences

To make the following computations easier, we make a slight change of notation; we consider a rotation as an exchange of two intervals, and we renormalize these intervals so that the larger interval has length one. We can suppose without loss of generality that the largest interval is on the left (that is, the sequence is of type 0), and we denote by θ the length of the smallest one.

To be more formal, we are now studying the map R:

$$
\begin{aligned}
R: \quad & [-1,\theta[\to [-1,\theta[\\
& x \mapsto x+\theta \text{ if } x<0 \\
& x \mapsto x-1 \text{ if } x \geq 0.
\end{aligned}
$$

For the initial rotation, with a domain of length one, the angle α is given by $\alpha = \frac{\theta}{1+\theta}$.

The map R has two continuity intervals, which we will denote by $I_0 = [-1,0[$ and $I_1 = [0,\theta[$; we also give names to their images $J_0 = R(I_0) = [\theta-1,\theta[$ and $J_1 = R(I_1) = [-1,\theta-1[$.

Definition 6.4.1. *The* induced map *of R on J_0, denoted by $R_{|J_0}$, is the map defined on J_0 by $R_{|J_0}(x) = R^{n_x}(x)$, where n_x is the first return time of x to J_0, $n_x = \inf\{n > 0 | R^n(x) \in J_0\}$.*

This map is very related to the previous section; to make this relation explicit, we first define itineraries with respect to a partition.

Definition 6.4.2. *Let $T: X \to X$ be a map, and $(X_i)_{i\in\mathcal{A}}$ be a partition of X.*

For any point $x \in X$, the itinerary *of x under T with respect to the partition $(X_i)_{i\in\mathcal{A}}$ is the sequence u, with values in the set $\mathcal{A}$ of indices of the partition, defined by $u_n = i$ if $T^n(x) \in X_i$.*

Remark. It is immediate to show that the itinerary of any point under R with respect to the partition (I_0, I_1) (or J_0, J_1) is a *rotation sequence*, in the sense of Section 6.1.2.

Exercise 6.4.1. Explain how the angle and initial point are changed when we apply to the sequence the flip E and the retrogression R defined in the previous section.

Use this to prove that, in any Sturmian system generated by a rotation sequence, there are exactly two sequences that are invariant under retrogression (these are the biinfinite equivalent of the palindromes of even length; there is also one sequence that is invariant by the map that takes u to v defined by $v_n = u_{-n}$, it is the biinfinite equivalent of palindromes of odd length).

Lemma 6.4.1. *The map $R_{|J_0}$ is again a rotation (or an exchange of 2 intervals).*

If x is a point of J_0, let u be the itinerary of x under the map R with respect to the partition I_0, I_1, and v be the itinerary of x under the induced map $R_{|J_0}$ with respect to the partition $(I_0 \cap J_0, I_1 \cap J_0)$.

We have $u = \sigma_0(v)$.

If x is in J_1, then both $R(x)$ and $R^{-1}(x)$ are in J_0; in that case, let u be the itinerary of x with respect to the partition (I_0, I_1), and v (resp. w) be the itinerary of $R^{-1}(x)$ (resp. $R(x)$) under the induced map $R_{|J_0}$ with respect to the partition $(I_0 \cap J_0, I_1 \cap J_0)$.

We have $u = S\sigma_0(v) = \mathbf{0}\sigma_0(w)$.

Exercise 6.4.2. Prove the lemma.

Remark. We see that the induction is the exact analogue of the recoding process defined in Subsection 6.3.1. Sequences of type 0 correspond to rotations where the interval on the left (I_0) is the longest (this is the case we considered above), and sequences of type 1 to rotations for which the interval on the right is the largest. When we iterate the induction, we alternate between sequences of type 0 and sequences of type 1.

We have the choice for the induction interval; we described above the induction on J_0, but we could also induce on I_0, giving rise to a different substitution.

We see also that we have the choice of images or preimages for points out of the induction interval; this corresponds to recoding either by shifting, or by using prefixes.

Exercise 6.4.3. Give a geometric model for the coding maps $\Phi_{\epsilon_0,\delta_0,\epsilon_1,\delta_1}$ defined in the previous section, in terms of induction; characterize in each case the unique point that is in the intersection of all the induction domains. In the case under study, it is, in the convention of this section, the point 0, and this is the reason for chosing this convention. In other cases, like $\Phi_{1,1,-1,1}$, it is the left extremity of the interval, and it would be more convenient to keep the initial convention of defining the rotation on $[0, 1[$.

It is now natural to iterate this induction operation as many times as possible.

Theorem 6.4.1. *let R be, as above, the rotation of parameter θ on $[-1, \theta[$. Let $a = \left[\frac{1}{\theta}\right]$. Let R' be the map defined by*

$$\begin{aligned} R' : \quad & [-\{\tfrac{1}{\theta}\}, 1[\to [-\{\tfrac{1}{\theta}\}, 1[\\ & x \mapsto x + 1 \text{ if } x < 0 \\ & x \mapsto x - \{\tfrac{1}{\theta}\} \text{ if } x \geq 0 \end{aligned}$$

For any $y \in [-\{\frac{1}{\theta}\}, 1[$, *and any integer* b *such that* $0 \leq b \leq a$, *with* $b = 0$ *if* $y > 1 - \{\frac{1}{\theta}\}$, *define* $x = \theta(y - b)$.

If u *is the itinerary of* x *under* R *with respect to partition* I_0, I_1, *and if* v *is the itinerary of* y *under* R' *with respect to the natural partition* $[-\{\frac{1}{\theta}\}, 0[\,, [0, 1[$, *we have:*

$$u = \mathbf{0}^b \sigma_0^a(v).$$

Proof. The proof is an immediate consequence of the previous proposition, once the proposition is rewritten in the good order: x should be given first, and y computed from x.

First remark that we can induce on the image of the left continuity interval a times; each time, the left interval decreases by θ, until it becomes of length $1 - a\theta < \theta$; we can then no more use the same induction (the return time would be larger than 2). If x is in the induction interval, and if u, v are its itineraries with respect to R and its induced map, iteration of the previous proposition shows that $u = \sigma_0^a(v)$. Otherwise, let b be the return time of x to the induction interval, and $z = x + b\theta$ its first image; it is easy to prove that $z < (a+1)\theta - 1$ (make a picture!), and that, for the itinerary u, v of x, y, we have $u = \mathbf{0}^b \sigma_0^a(v)$.

It remains to remark that, to renormalize, we must divide by θ, and we recover the formulas of the theorem.

■

Exercise 6.4.4. Explain what happens in the variants, when we use preimages instead of images for those points that are out of the induction interval, or when we iterate induction on I_0.

Exercise 6.4.5. Explain how we can iterate the construction of the previous proposition. (*Hint: we can either change the type of induction, inducing now as many times as possible on* J_1, *or conjugate the induced map by* $x \mapsto -x$, *so that we recover the initial situation. This translates to symbolic dynamics by exchanging* $\mathbf{0}$ *and* $\mathbf{1}$.)

6.4.2 Continued fractions

We are interested here only in the combinatorics of the usual continued fraction, not in its approximation and diophantine property. We will give a very brief exposition, without proofs, to recall the arithmetic expressions we need. For more details, see for instance [167, 205].

First of all, we recall that, for any sequence $(a_n)_{n>0}$ of strictly positive integers, the sequence $(x_n)_{n>0}$ defined by

$$x_1 = \frac{1}{a_1} \qquad x_2 = \frac{1}{a_1 + \cfrac{1}{a_2}} \qquad x_n = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ldots + \cfrac{1}{a_n}}}}$$

converges to a real number x.

Reciprocally, to any irrational number $x \in [0,1]$, there exists a unique sequence $(a_n)_{n>0}$ of strictly positive integers such that x is the limit of the sequence defined above. We denote this by:

$$x = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\dots + \cfrac{1}{a_n + \dots}}}}$$

.

It is also customary to note: $x = [0; a_1, \dots, a_n, \dots]$. We can generalize this expansion to all irrational numbers by replacing the initial 0 by $a_0 = [x]$, hence the presence of the initial 0; as we will only consider numbers in the interval $[0,1]$, we always have $a_0 = 0$.

Definition 6.4.3. *Let $x = [0; a_1, \dots, a_n, \dots]$. The integers a_n are called the* partial quotients *of x.*

The rational numbers $\frac{p_n}{q_n} = \cfrac{1}{\mathbf{a}_1 + \cfrac{1}{a_2 + \cfrac{1}{\dots + \cfrac{1}{a_n}}}}$ *are called the* convergents *of x.*

Exercise 6.4.6. Prove that the convergents satisfy the following relation: $p_n = a_n p_{n-1} + p_{n-2}$, $q_n = a_n q_{n-1} + q_{n-2}$ for $n > 2$.

Exercise 6.4.7. Prove that any rational number can be also written as a continued fraction, but now the sequence is finite, and it is not uniquely defined: to each rational number correspond exactly two continued fraction expansions, of the form $[0; a_1, \dots, a_n]$ and $[0; a_1, \dots, a_n - 1, 1]$

We will use the following facts:

Proposition 6.4.1. *Let $T :]0,1[\to [0,1[$, $x \mapsto \{\frac{1}{x}\}$, and $F :]0,1[\to \mathbb{N}$, $x \to [\frac{1}{x}]$.*

For any irrational number $x \in]0,1[$, the partial quotients of x are given by $a_n = F(T^{n-1}(x))$.

Define $\beta_n = (-1)^n T(x) T^2(x) \dots T^n(x)$. We have:

$$\forall n \in \mathbb{N},\ q_n \alpha - p_n = \beta_n.$$

Remark. There is a nice way to explain the continued fraction algorithm, which is directly related to the previous subsection. Consider two intervals, of respective length α and β; we suppose that $\alpha < \beta$, and that the ratio is $x = \frac{\alpha}{\beta}$. Now substract α to β as many times (that is, $a_1 = [\frac{\beta}{\alpha}]$ times) as you can. The remaining interval is now smaller than α; substract it from α, obtaining a new integer a_2, and iterate. It is not difficult to check that we obtain in this way the continued fraction expansion of x. If we start with a normalized interval $\beta = 1$, we can check that the successive lengths are the absolute values of the numbers $q_n\alpha - p_n = \beta_n$ defined above.

In the rational case, the algorithms terminates after a finite number of steps; we can suppose that both intervals are of integer length, and in that case this is exactly Euclid's algorithm.

This algorithm, however, gives information only on the lengths of the intervals; the induction process gives also information on orbits of particular points, and we will see that we can account for that by a small change in this algorithm: instead of deleting some intervals, we stack them over other intervals, thus building what is called "Rokhlin towers" (see also Chap. 3).

6.4.3 The Ostrowski numeration system

We first define the Ostrowski numeration system on the integers. For more details, see [268, 340, 198, 199, 260, 184, 165, 360] and see also the survey [62].

Let $(a_n)_{n>0}$ be a sequence of strictly positive integers, and let q_n be the sequence recursively defined by $q_0 = 1$, $q_1 = a_1$, $q_n = a_n q_{n-1} + q_{n-2}$ for $n \geq 2$ (the q_n are the denominators of the convergents of the number $\theta = [0; a_1, \ldots, a_n, \ldots])$. We have the following proposition:

Proposition 6.4.2. (The Ostrowski numeration system on the integers) *Any integer $k \geq 0$ can be written as*

$$k = \sum_{n=1}^{N} b_n q_{n-1},$$

where the coefficients b_n are integers such that $0 \leq b_n \leq a_n$, and $b_n = a_n$ implies $b_{n-1} = 0$.

Proof. It is clear from definition that the sequence q_n is strictly increasing. We apply the greedy algorithm: let N be the largest integer such that $q_{N-1} \leq k$. Let $b_N = [\frac{k}{q_{N-1}}]$. It is clear that $b_N \leq a_N$, otherwise we should have $q_N < (a_N + 1)q_N \leq k$. The integer $k - a_N q_{N-1}$ is strictly smaller than q_{N-1}, and if $b_N = a_N$, it must be strictly smaller than q_{N-2}, ot herwise we should have $q_N = a_N q_{N-1} + q_{N-2} \leq k$. The result is then clear by induction. ■

Exercise 6.4.8. Prove that the expansion is unique, and that the possible expansions are completely characterized by the nonstationary Markov condition $0 \leq b_n \leq a_n$, and $b_n = a_n$ implies $b_{n-1} = 0$.

This numeration system translates to a numeration system on the real numbers in the following way. We fix as above a sequence $(a_n)_{n>0}$ of partial quotients corresponding to a real irrational number θ. We define $\theta_0 = \theta$, $\theta_n = T^n(\theta)$, where T is the Gauss map $T(x) = \{\frac{1}{x}\}$, and $\beta_n = (-1)^{n+1}\theta_0\theta_1 \ldots \theta_n$.

Theorem 6.4.2. (The Ostrowski numeration system on the reals)*Any real number $x \in [-1, \theta]$ can be written as*

$$x = \sum_{n=1}^{\infty} b_n \beta_n,$$

where the coefficients b_n are integers such that $0 \leq b_n \leq a_n$, and $b_n = a_n$ implies $b_{n-1} = 0$.

Exercise 6.4.9. Prove the theorem; examine the unicity of the expansion, and the characterization of admissible expansions. (*Hint: it can be useful to prove first the recurrence relation $\beta_n = a_n\beta_{n-1} + \beta_{n-2}$.*)

Remark. Note that there are infinitely many Ostrowski systems, one for each sequence of partial quotients.

It is also possible to define a dual Ostrowski numeration system:

Theorem 6.4.3. (The dual Ostrowski numeration system on the reals)*Any real number $x \in [0, 1+\theta]$ can be written*

$$x = \sum_{n=1}^{\infty} b_n |\beta_n|,$$

where the coefficients b_n are integers such that $0 \leq b_n \leq a_n$, and $b_n = a_n$ implies $b_{n+1} = 0$.

Proof. The proof here is simpler than in the preceding theorem; the sequence $|\beta_n|$ is a strictly decreasing sequence of positive numbers. Hence, we can apply a kind of greedy algorithm: we consider the smallest n such that $|\beta_n| < x$, and we substract $|\beta_n|$ from x as many times as possible. The numbers $|\beta_n|$ satisfy a recurrence relation that immediately implies the Markov condition. ∎

This dual numeration system, as the preceding one, is associated with a numeration system on the integers, but this time on negative and positive integers:

Proposition 6.4.3. *Any integer $k \in \mathbb{Z}$ can be written in a unique way*

$$k = \sum_{n=1}^{N} b_n (-1)^n q_{n-1},$$

where the coefficients b_n are integers such that $0 \leq b_n \leq a_n$, and $b_n = a_n$ implies $b_{n+1} = 0$.

Exercise 6.4.10. Prove this proposition.

Remark. It is important for the sequel, specially for Sec. 6.6, to remark that, in the dual numeration system, the Markov condition, compared to that of the original numeration system, has changed direction: when $b_n = a_n$, we obtain a condition on the next coefficient, and not the previous coefficient.

6.4.4 Sturmian sequences and arithmetic

Consider now, in the notation defined at the beginning of this section (rotation on the interval $[-1, \theta[$), a rotation sequence u of unknown angle θ and incident x; that is, the sequence is given as the itinerary of the point x under the rotation with respect to the partition (I_0, I_1). We want to recover the angle and incident from the symbolic sequence u.

We associate to the sequence u its second multiplicative coding (a_n, b_n) as defined in Sec. 6.3.4. We can then recover θ and x as follows:

Theorem 6.4.4. *Let u be a rotation sequence which is the itinerary of a point x under the rotation of angle θ on $[-1, \theta[$. Let (a_n, b_n) be the second multiplicative coding of u. We have:*

$$\theta = [0; a_1, a_2, \ldots, a_n, \ldots]$$

that is, (a_n) is the continued fraction expansion of θ.

Moreover, if we define as above $\theta_0 = \theta$, $\theta_n = T^n(\theta)$, where T is the Gauss map $T(x) = \{\frac{1}{x}\}$, and $\beta_n = (-1)^{n+1}\theta_0\theta_1 \ldots \theta_n$, we have:

$$x = \sum_{n=1}^{\infty} b_n \beta_n$$

that is, the sequence (b_n) is the Ostrowski expansion of x with respect to the sequence (a_n).

Proof. We know by hypothesis that u is a rotation sequence. The second multiplicative coding process gives a family $u^{(n)}$ of sequences, such that $u^{(n)} = \mathbf{0}^{b_{n+1}} \sigma_0^{a_{n+1}}(u^{(n+1)})$ or $u^{(n)} = \mathbf{1}^{b_{n+1}} \sigma_1^{a_{n+1}}(u^{(n+1)})$, depending on the parity of n. With each of these sequences are associated an angle θ_n and an incident x_n; but Theorem 6.4.1 gives us the relation between these quantities, namely $\theta_n = \frac{1}{a_{n+1}+\theta_{n+1}}$ and $x_n = \theta_n(x_{n+1} - b_{n+1})$; the proof of the theorem follows. ∎

We obtain now a constructive proof that any Sturmian sequence is a rotation sequence (for more details, see [35], see also [265, 360]:

Theorem 6.4.5. *Every Sturmian sequence is a rotation sequence, whose angle and incident are explicitly determined by the second multiplicative coding.*

Proof. Let u be a Sturmian sequence, and let (a_n, b_n) be its second multiplicative coding.

Consider the numbers θ whose continued fraction is (a_n), and x whose Ostrowski expansion with respect to (a_n) is (b_n). These define a rotation sequence, whose second multiplicative coding is, by construction, the same as that of u. Hence this rotation sequence is equal to u. ∎

Remark. Strictly speaking, we should have discussed the orbit of the fixed point, since for these sequences, the additive or multiplicative coding does not completely define the sequence; however, the two sequences obtained differ only in two places, and if one is a rotation sequence, so does the other, the only difference being whether one uses right-closed or left closed intervals.

6.4.5 Arithmetic of the dual coding

We defined in the end of the previous section a dual multiplicative coding; this can be interpreted in term of the dual Ostrowski expansion.

Theorem 6.4.6. *Let u be a Sturmian sequence of type 0, and let (a_n, b_n) be its dual arithmetic coding. The sequence u can be obtained as the coding of x under the rotation of angle θ on $[0, 1+\theta[$, where $\theta = [0; a_1, \ldots, a_n, \ldots]$, and x is the positive real number whose dual Ostrowski expansion, with respect to the sequence (a_n), is (b_n).*

Exercise 6.4.11. Prove this theorem; we must use here a different kind of induction, called Rauzy induction, where we always induce on the longest of I_0 and J_1.

Exercise 6.4.12. Work out the arithmetic formulas associated with the various coding maps we defined in the previous section, and their multiplicative counterpart.

It might be useful to determine first the point whose expansion is 0: this is the intersection point of the successive induction domains that was the subject of Exercise 6.4.3

6.5 Sturmian substitutions. Dynamical interpretations

6.5.1 Some classical theorems on periodic continued fractions

The numbers whose continued fraction expansion is periodic are specially interesting, and we have two classical theorems:

Theorem 6.5.1. (Lagrange's Theorem) *The continued fraction expansion of x is eventually periodic if and only if x is a quadratic number.*

The "only if" part is easy, and analogues of it are easily proved for all known generalized continued fraction algorithms; one way to prove it is that, if the expansion is periodic, one can write $x = R(x)$, where R is a rational fraction with integer coefficients; one can also write a matrix form of the continued fraction expansion, and look at eigenvectors of a matrix in $SL(2, \mathbb{Z})$; the "if" part is more difficult, and does not generalize easily to multidimensional algorithms, although some cases are known.

Theorem 6.5.2. (Galois' Theorem) *The continued fraction of x is purely periodic if and only if x is a quadratic integer in $[0, 1]$ whose conjugate is smaller than -1.*

6.5.2 Sturmian systems with periodic coding

We are interested in Sturmian systems and sequences whose coding is periodic; these are linked with quadratic numbers, by the two previous theorems.

Let X be a Sturmian system with periodic coding; we can suppose that its multiplicative coding is given by a sequence $(a_n)_{n>0}$ of period d, that is $a_{n+d} = a_n$. We can suppose that d is even (otherwise, just consider the double of the period). In that case, it is easy to check that the map Ψ^d, defined in Sec. 6.3 preserves Ω. (If d were odd, the map Ψ^d would send X on the flipped system $E(X)$.)

Another way to explain that is to remark that, for any element of X, we can consider the second multiplicative coding $(a_n, b_n)_{n\in\mathbb{N}}$; the sequence (a_n) depends only on the system, and it is periodic of period d. The sequence (b_n) satisfies a Markov condition, and it is immediate to check that the set of coding sequences for elements of X is not invariant by S, but it is invariant by S^d; grouping the coding sequence in blocks of length d, we obtain a subshift of finite type.

We can be more explicit: let us denote by σ the composed substitution $\sigma = \sigma_0^{a_1}\sigma_1^{a_1}\sigma_0^{a_3}\dots\sigma_1^{a_d}$.

Then the fixed points of the system X are the fixed points of σ; the system X is the substitutive dynamical system associated with σ. Any element $u \in X$ can be written in a unique way $u = S^k\sigma(v)$, with $k < |\sigma(v_0)|$; we have of course $v = \Psi^d(u)$, and the map Ψ^d can be seen as some kind of inverse of σ on X.

Exercise 6.5.1. Prove that the sequence $(k_n)_{n\in\mathbb{N}}$ obtained by iteration completely defines the initial Sturmian sequence u, except if k_n is ultimately zero, in which case we obtain points in the positive orbits of the two fixed points.

Exercise 6.5.2. Instead of giving just numbers k_n, one could give the corresponding prefixes of $\sigma(v_0)$.
1) Explain what kind of prefixes can arise.
2) Characterize admissible sequences of prefixes.
3) Explain how one could code with suffixes.

6.5.3 One-dimensional dynamical systems associated with Sturmian substitutions

We just explained how, on a Sturmian system with periodic coding, one can define two dynamics: the shift, and the map Ψ^d, inverse of the substitution σ.

We have seen in previous sections, and proved in Sec. 6.4, that there is a nice geometric model for the shift on a Sturmian system: a rotation on an interval, coded by the two continuity intervals. It is natural to look for a geometric model for the map Ψ^d in this framework. In fact, Sec. 6.4.1 hints at the possible model: we showed in this section that a variant of the coding map can be given as a map on an interval, of the form $x \mapsto \frac{x}{\theta} + b$, where b is an integer. The problem here is that the domain and the image of this map are different intervals, the domain corresponding to the initial rotation, and the image corresponding to the induced rotation.

If however the system has periodic coding of even period d, after inducing d times we recover the initial rotation, and the composed map is now from an interval to itself. We see that this composed map is locally an homothety of constant factor $\lambda = \theta_0\theta_1 \ldots \theta_{d-1}$, where $\theta_0 = \theta$, the angle of the system, and $\theta_{n+1} = \left\{\frac{1}{\theta_n}\right\}$.

Exercise 6.5.3. Prove that λ is the smallest eigenvalue of the matrix associated to the substitution σ, and that it is a quadratic integer.

The map Ψ^d, reduced to the Sturmian system, is in fact associated to a generalized λ-shift, and we leave to the reader the proof of the following theorem, which is the essential result of this subsection:

Theorem 6.5.3. *Let X be a Sturmian system of type 0, with periodic coding of even period d. Let $R : [-1, \theta[\to [-1, \theta[$ be the corresponding rotation of angle θ, as defined in Sec. 6.4.1. Let $\sigma = \sigma_0^{a_1}\sigma_1^{a_1}\sigma_0^{a_3} \ldots \sigma_1^{a_d}$ be the corresponding substitution, and λ its smallest eigenvalue.*

There is a finite partition of the domain of R in intervals $K_1, \ldots, K_n$ a finite set of quadratic numbers $d_1, \ldots, d_n$ in $\mathbb{Q}[\lambda]$, and a map $F : [-1, \theta[\to [-1, \theta[\quad x \times \frac{x}{\lambda} + d_i$ if $x \in K_i$ such that:

The rotation sequence u (resp. v) associated to x (resp. $F(x)$) under R satisfy $v = \Psi^d(u)$.

Furthermore, the image of any of the intervals K_i under F is either I_0 or I_1.

Exercise 6.5.4. Characterize explicitly the set of digits d_i *(it is interesting to use here the prefixes of images of the letters under the substitution σ)*

Exercise 6.5.5. Prove in this setting that the Ostrowski system reduces to an expansion in powers of λ, with digits in a finite set of quadratic integers; characterize the set of admissible expansions (it satisfies a Markov condition). This can be another way to answer the preceding exercise.

Exercise 6.5.6. Explain what happens if we change the type of the induction, and how the corresponding map F is changed.

The most simple example is of course the golden number. This does not fall exactly under the framework of this section, since the minimal period is 1, so it is not even. It is however interesting to work out this case.

Exercise 6.5.7. We consider the Fibonacci substitution σ defined by $\sigma(\mathbf{0}) = \mathbf{01}$, $\sigma(\mathbf{1}) = \mathbf{0}$.

1) Show that the corresponding dynamical system is the rotation by ϕ, the golden number.

2) Show that it can be coded by the set of sequences $(\epsilon_n) \in \{0,1\}$ such that $\epsilon_i\epsilon_{i+1} = 0$ (that is, the word 11 is not allowed).

3) Show that the induction map admits as geometric model the map $[0,1[\to [0,1[\ x \mapsto \{\phi x\}$.

4) Show that this map admits a unique invariant measure absolutely continuous with respect to Lebesgue measure, and compute this invariant measure. (*Hint: the density is constant on intervals.*)

6.5.4 Two-dimensional dynamical systems related to Sturmian substitutions

The coding sequences are infinite sequences, with admissibility conditions of finite type. The shift on these sequences is not a one-to-one map, and the geometric model we just gave is far from being injective.

It is natural to try to extend the coding sequences to biinfinite sequences, and to look for a one-to-one system that projects on the given one. This is a very simple example of what is called *natural extension* to define the natural extension as an inductive limit, but we would like to obtain more concrete models.

For a system of finite type, it is natural to consider biinfinite sequences subject to the same finite type condition. This gives us a first model, as a two-sided subshift of finite type.

We will see in Sec. 6.6 how we can associate to this biinfinite coding sequence a pair of Sturmian sequences (u^+, u^-) with same initial letter, by considering the negative part of the coding as the dual coding for the sequence u^-. The shifted coding sequences is then associated to $(\Psi^d(u^+), S^k\sigma(u^-))$ for a suitable k.

Another way, more geometric, to understand two-sided coding sequences is the following: since the coding sequence describes what occurs under induction, extending this sequence on the left means that we are trying to "exduce" the given system, that is, to find a larger system of which it is induced on a suitable set. There is in general no clear way to do this; however, it is possible in the present case to find a nice representation of this natural extension as a toral automorphism.

We just give here the idea of the construction, since the details are quite involved. More details can be found in Sec. 6.6, where we work out the non-periodic case, and where the combinatorics is simpler, since we use the additive map Φ.

With a Sturmian system X, one can associate a rotation R on two intervals I_0, I_1; if X has periodic coding, we can find two intervals I'_0, I'_1 such that the induced map of R on these smaller intervals is conjugate by an homothety to the initial map. The map Ψ^d corresponds to stacking subintervals over I'_0, I'_1; if we know at which level in the stack is the corresponding point, we can recover the initial point, and "exduce" the transformation. However, if we have a finite stack, we can exduce only a finite number of times.

The idea is then to consider to rectangles, of respective basis I_0, I_1, to cut these rectangles in slices and to stack these slices over I'_0, I'_1. If we take for heights the coordinates of an appropriate eigenvector, the final pair of rectangles will be the image of the initial one by a measure-preserving dilation.(We show below the figure corresponding to the simplest example, period 2, continued fraction expansion $(1, 1)$, matrix $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, substitution $\sigma(\mathbf{0}) = \mathbf{010}$, $\sigma(\mathbf{1}) = \mathbf{10}$). If we compose with this dilation, we obtain a map from a pair of rectangles to itself, and one can prove that, after suitable identification on the boundary, this map becomes a toral automorphism corresponding to the matrix of the substitution σ.

6.5.5 Similarity of toral automorphisms: the Adler-Weiss theorem

We can recover in this way the Markov coding that was the basis of Adler-Weiss theorem on similarity of toral automorphism with the same entropy. For more details, see [4, 3].

It is based on a few simple lemmas:

Lemma 6.5.1. *Any hyperbolic matrix in $SL(2, \mathbb{Z})$ is conjugate, in $SL(2, \mathbb{Z})$, to a matrix with positive coefficients.*

Proof. Since the matrix is hyperbolic, it has a contracting and an expanding direction, which both have irrational slope.

But it is easy to prove that, for any two independent vectors U, V with irrational slope which form an oriented basis, it is possible to find a basis of $\mathbb{Z}^2$ such that U is in the positive cone, and V is in the second quadrant.

Let us do this for a basis formed of an expanding and a contracting vector; it is clear that, after the change of basis, the image of the positive cone by the map is included in the positive cone, hence the matrix in the new basis is positive. ∎

But positive matrix in $SL(2,\mathbb{Z})$ have quite special properties.

Definition 6.5.1. *We denote by $SL(2,\mathbb{N})$ the set of positive matrices in $SL(2,\mathbb{Z})$, that is, the set of matrices with positive integral coefficients and determinant 1.*

Lemma 6.5.2. *The set $SL(2,\mathbb{N})$, endowed with the multiplication on matrices, is a free monoid on the two matrices $M_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $M_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$*

Exercise 6.5.8. Prove this. (*Hint: prove that, unless it is the identity, for any element $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, one column is bigger than the other, i.e., $a \geq b$ and $c \geq d$, or $a \leq b$ and $c \leq d$; deduce that this matrix can be decomposed in a unique way as a product of the two given matrices.*)

We can now associate to that matrix a substitution: to the matrix M decomposed as

$$M = M_0^{a_1} M_1^{a_2} \dots M_1^{a_d},$$

we associate the substitution σ defined as $\sigma_0^{a_1} \sigma_1^{a_2} \dots \sigma_1^{a_d}$.

Exercise 6.5.9. Prove that we can now define a fundamental domain for the torus, made of a pair of rectangles with sides parallels to the eigendirections; the substitution dynamical system associated with σ corresponds to the rotation, first return map on the basis of the rectangles of the flow along the unstable direction; the natural extension of the shift on coding sequences is the initial toral automorphism.

Show how one can define a dual substitution, associated with the flow along the stable direction.

In this way, one can associate in a canonical way a Markov coding for any hyperbolic automorphism of the 2-torus. Adler and Weiss then proceeded to prove, using this coding, that any two automorphisms with same entropy are measurably isomorphic (see for instance [4, 3]). Note that this is a much deeper part: as proves the following exercise, there are automorphisms with same entropy that are not conjugate in $SL(2,\mathbb{Z})$, and one can prove that these cannot be topologically conjugate (one must know that, for an automorphism of the 2-torus, the entropy is the logarithm of the dominant eigenvalue, and remark that this only depends on the trace).

Exercise 6.5.10. Give an example of 2 elements of $SL(2,\mathbb{N})$ with same trace that are not conjugate in $SL(2,\mathbb{Z})$. (*Hint: if they are conjugate, they are conjugate mod 2; consider the matrices* $\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}$ *and* $\begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$.)

6.6 Natural extension for the recoding of Sturmian sequences

In the preceding section, we worked out a natural extension for the map Ψ^d on a Sturmian system with periodic coding of period d.

It is tempting to work out a similar result for the map Φ on the set Σ of all sturmian sequences.

From the formal viewpoint, there is no difficulty: in Sec. 6.3.3, we showed that the sequences $(\gamma(\Phi^n(u)))_{n\in\mathbb{N}}$ give a symbolic model of Φ as a one-sided subshift of finite type on a finite alphabet; to obtain the natural extension, it suffices to consider the 2-sided subshift satisfying the same condition.

This first version of the natural extension i not very satisfying; we will give below, first a geometric model, then a symbolic model as a map on the set of pairs of Sturmian sequences with same initial. For more details, see for instance [323, 32, 36].

6.6.1 Geometric model for the natural extension: the pairs of rectangles

We use here the idea of Sec. 6.4. We showed there that, to any Sturmian sequence, one can associate first a rotation, on a normalized interval $[-1,\alpha]$ or $[-\alpha,1]$ depending on the type of the sequence, and a point in this interval, such that the given Sturmian sequence is the orbit of the given point under the rotation of angle α or $-\alpha$.

We showed also that the map Φ amounts to an inducing and stacking operation; if one can keep the height of stacking, one can recover from the sequence $\Phi(u)$ the initial sequence u. However, if one consider only stacks of integer heights, one can only recover a finite number of preimages.

It works better if one consider real heights for the stacks.

Let us now consider pairs of rectangles, such that:

- one rectangle has width 1, and the other irrational width $\alpha < 1$.
- the total area is 1.
- Both rectangles have a common lower vertex in $(0,0)$, and empty intersection.
- there is a distinguished point (x,y).

We can now define a map on the set of all these pairs of marked rectangles, by cutting a slice of width α of the larger rectangle, stacking it on the narrower rectangle, and renormalizing by a measure-preserving dilation so that the size of the new larger rectangle becomes 1 (see Fig. 6.4).

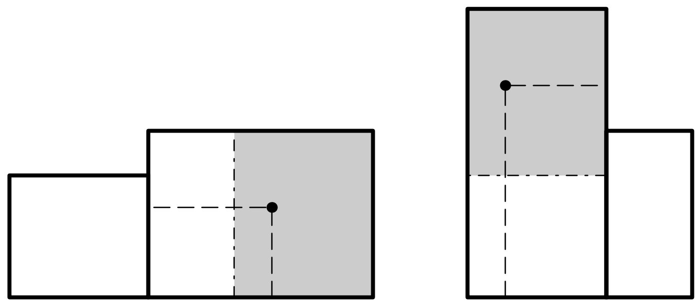

Fig. 6.4. The stacking map

It is clear, for geometric reasons, that this map is one-to-one (one can now "unstack", using the rectangle of larger height), and it obviously projects to Φ, just by forgetting the heights.

Definition 6.6.1. *We denote by Δ_a the set of* 6*-tuples* $(a, b, c, d, x, y) \in \mathbb{R}^6$ *such that: a, b, c, d are all positive,* $\sup(a, b) = 1$*, $ad + bc = 1$, $-a \le x \le b$, and $0 \le y < d$ if $x < 0$, $O \le y < c$ if $x \ge 0$.*

We denote by $\Delta_{a,0}$ the set of elements of Δ_a such that $a = 1 > b$, and $\Delta_{a,1}$ its complement (elements of Δ_a such that $a < b = 1$)

Of course, these equations just define the set of marked rectangles, with 4 variables for the shape of the rectangles, and 2 variables for the marked point; this set is really of dimension 4, because of the two relations $\sup(a, b) = 1$ and $ad + bc = 1$ (the total area is 1); remark that, to recover traditional notations, d is the height of the rectangle of width a, not of the rectangle of width b. Hence, in each of the two sets $\Delta_{a,0}$, $\Delta_{a,1}$, we can reduce to 4 coordinates, namely, the width and height of the narrower rectangle, and the coordinates of the marked points. We leave it as an exercise to prove that, in these coordinates,the analytic form of the map is given as follows:

Definition 6.6.2. *We denote by $\tilde{\Phi}$ the map defined on $\Delta_{a,1}$ by:*

- *if* $a < 1/2$

$$(a,d,x,y) \mapsto \left(\frac{a}{1-a}, (d+1-ad)(1-a), \frac{x}{1-a}, y(1-a)\right) \quad \textit{if } x < 1-a$$
$$(a,d,x,y) \mapsto \left(\frac{a}{1-a}, (d+1-ad)(1-a), \frac{x-1}{1-a}, (y+d)(1-a)\right) \textit{if } x \geq 1-a$$

- *if* $a > 1/2$

$$(a,d,x,y) \mapsto \left(\frac{1-a}{a}, (1-ad)a, \frac{x}{a}, ya\right) \quad \textit{if } x < 1-a$$
$$(a,d,x,y) \mapsto \left(\frac{1-a}{a}, (1-ad)a, \frac{x-1}{a}, (y+d)a\right) \quad \textit{if } x \geq 1-a;$$

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

The map $\tilde{\Phi}$ *is defined by similar formulas on* $\Delta_{a,0}$.

Remark that it is very easy to prove that the map preserves the Lebesgue measure (for the coordinates (a,d,x,y) on $\Delta_{a,1}$), since the Jacobian is 1; the total mass of this measure is infinite.

Exercise 6.6.1. Strictly speaking, this defines this is correct only on a set of measure 1; explain what should be done, in term of strict and large inequality and of irrational numbers, so that the map is defined everywhere.

We can of course iterate this map; we can define pair of rectangles of type 0 (if the widest rectangle is on the left), and of type 1 (if it is on the right), and we can define a "multiplicative map" $\tilde{\Psi}$ by iterating till we change type. It is easy to compute this iterated map:

Definition 6.6.3. *We denote by* Δ_m *the set of marked pairs of rectangles such that the widest rectangle is also the highest one (that is, in the above notations,* $a > b$ *is equivalent to* $d > c$*); we denote by* $\Delta_{m,0}$ *the subset such* $a = 1$, $\Delta_{m,1}$ *the subset such that* $b = 1$.

Definition 6.6.4. *The map* $\tilde{\Psi}$ *is defined on* $\Delta_{m,1}$ *by:*

$$(a,d,x,y) \mapsto \left(\left\{\frac{1}{a}\right\}, a - da^2, \frac{x}{a}, ya\right) \quad \textit{if } x < a\left\{\frac{1}{a}\right\}$$
$$(a,d,x,y) \mapsto \left(\left\{\frac{1}{a}\right\}, a - da^2, -\left\{\frac{1-x}{a}\right\}, ya + da + (a - da^2)\left\lfloor\frac{1-x}{a}\right\rfloor\right)$$
$$\textit{if } x \geq l\left\{\frac{1}{l}\right\},$$

and similarly on $\Delta_{m,0}$.

Exercise 6.6.2. Prove that this map preserves Lebesgue measure, and that this invariant measure in finite, of measure $2\ln 2$? (*Hint: by definition of* $\Delta_{m,1}$*, one has* $d < c$*, hence* $d(a+1) < 1$.)

Exercise 6.6.3. Prove that this maps factors over the classical continued fraction map, and also over a 2-dimensional continued fraction, and over a natural extension of the usual continued fraction.

Exercise 6.6.4. Prove that the maps $\tilde{\Phi}$ and $\tilde{\Psi}$ are one-to-one (except on a set of measure 0, as explained in Exercise 6.6.1 above), and compute the reciprocal maps.

Exercise 6.6.5. Explain how one can obtain, using map $\tilde{\Psi}$, a two-sided sequence (a_n, k_n) of positive integers, and how one can recover the initial coordinates from this two sided sequence. (*Hint: difficult; this sequence is related to, but not equal to, the Ostrowski expansion considered in Sec. 6.4.*)

6.6.2 Symbolic model for the natural extension: pairs of Sturmian sequences

We can at this point recover the initial sturmian sequence. The idea is to identify, by a rotation, the upper and lower sides of the pair of rectangles, and also the free left and right sides. The quotient space is easily seen to be a torus (a more formal way to see it is proposed in the next exercise).

Exercise 6.6.6. Prove that any pair of rectangles that share a lower vertex at $(0,0)$ and whose interior do not intersect tile the plane periodically, and express the group Γ of the tiling as a function of the height and width of the rectangles.

Deduce that there is a natural identification on the boundary such that, modulo this identification, the pair of rectangles becomes isomorphic to the torus $\mathbb{Z}^2/\Gamma$.

This torus comes with two natural flows: a vertical flow $((x,y) \mapsto (x, y+t))$ and a horizontal flow $((x,y) \mapsto (x+t, y)$; any vertical trajectory cuts the two rectangles, and we obtain in this way a sequence u_n, itinerary of this trajectory with respect to the two rectangles; it is an easy exercise to prove that this sequence is in fact a rotation sequence.

One can do the same with the horizontal flow, and obtain in this way another Sturmian sequence v; remark that we have, by definition, $u_0 = v_0$.

We can then give another symbolic model for the natural extension of the map Φ, as a map on the set of pairs of Sturmian sequences (u, v) with same initial, that takes (u, v) to (u', v'), where (u', v') are the Sturmian sequences that correspond to the marked point (x, y) in the new "stacked" domain; one has clearly $u' = \Phi(u)$.

Exercise 6.6.7. Explain how one can compute v' from u and v. (*Hint: use the pictures below.*)

It is obviously possible to iterate this map, and we obtain in this way a natural extension of the map Ψ.

Exercise 6.6.8. Explain how one can give Markov Symbolic dynamics for the maps $\tilde{\Phi}$ and $\tilde{\Psi}$, and their symbolic counterparts.

Show that u (resp. v) determines the future (resp. the past) of this symbolic dynamics.

Make explicit the dual algorithm that determines v as a function of the past symbolic dynamics.

6.6.3 The geodesic flow on the modular surface

It is clear that, if we do not know exactly the sequence u, but only the corresponding Sturmian system, we cannot determine the abscissa of the marked point x, but only the width of the rectangles, and similarly for the sequence v.

This amounts to forgetting (x, y) in the above formulas for maps Φ and Ψ; we thus obtain a map on a subset of the plane that is a natural extension of the continued fraction map.

There is a nice way to represent this: we consider the space of lattices of covolume one of the plane; this set is naturally isomorphic to $SL(2,\mathbb{Z})\backslash SL(2,\mathbb{R})$.

Exercise 6.6.9. Prove that $Sl(2,\mathbb{R})$ is isomorphic to the set of oriented basis of $\mathbb{R}^2$ such that the unit square has area 1.

Prove that $SL(2,\mathbb{Z})$ is the group of automorphisms of $\mathbb{Z}^2$.

Deduce that $SL(2,\mathbb{Z})\backslash SL(2,\mathbb{R})$ is the space of lattices of the plane of covolume 1.

There is a natural action on this set of lattices, by diagonal matrices $g_t = \begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{-\frac{t}{2}} \end{pmatrix}$.

It is possible to prove that each lattice admits a fundamental domain shaped as a pair of rectangle; those lattices for which one of the rectangles has area 1 form a section of the flow, and the map $\tilde{\Phi}$ is in fact, if we forget the marked point, a first return map to this section.

Exercise 6.6.10. Work out the section corresponding to the map Ψ, find the return time to this section and prove that the space $SL(2,\mathbb{Z})\backslash SL(2,\mathbb{R})$ has finite volume.

Exercise 6.6.11. The flow g_t is usually called the geodesic flow on the modular surface, for a reason explained in this exercise.

One consider the set $\mathbb{H} = \{x + iy \in \mathbb{C} | y > 0\}$ (Poincaré half plane), endowed with the metric $\frac{dx^2+dy^2}{y^2}$

1) Prove that the group $SL(2,\mathbb{R})$ acts on $\mathbb{H}$ by:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} .z = \frac{az+b}{cz+d}$$

2) Prove that it is an action by isometries (i.e., $z \mapsto M.z$ is an isometry).

3) Prove that $SL(2,\mathbb{R})$ acts transitively on $\mathbb{H}$ and on $T^1\mathbb{H}$, deduce that $T^1\mathbb{H}$ is isomorphic to $PSL(2,\mathbb{R}) = SL(2,\mathbb{R})/\{Id, -Id\}$ (remark that the action of $-Id$ is trivial).

We will fix the isomorphism by associating the identity matrix and the vertical unit vector at i.

4) Prove that the curve ie^t is a geodesic.

5) Prove that the right action of the group $\{g_t | t \in \mathbb{R}\}$ defined above on $SL(2,\mathbb{R})$ is the geodesic flow on the Poincaré half plane.

6) Prove that the quotient of $\mathbb{H}$ by the action of $SL(2,\mathbb{Z})$ is a surface with one cusp and two singular points (conic points of angle $\frac{\pi}{2}$ and $\frac{Pi}{3}$). This surface is called the modular surface.

Hence the action of the one-parameter group g_t turns out to be exactly the geodesic flow on the unit tangent bundle of the modular surface.

Another model of the map Φ could be given in this context.

6.6.4 The scenery flow

It is possible to find a similar model for the maps Φ and Ψ. The idea is to consider, not only lattices in the plane, but also cosets of these lattices.

This can be formalized by taking, not linear groups, but affine groups; we define the group $SA(2,\mathbb{R})$ of measure preserving affine maps of the plane, of the form $(x,y) \mapsto (x,y).M + v$, where M is a matrix of determinant 1, and v is an element of $\mathbb{R}^2$. It is easy to check that the subset $SA(2,\mathbb{Z})$ of elements with integral coefficients is a subgroup, and we can consider the left quotient $SA(2,\mathbb{Z})\backslash SA(2,\mathbb{R})$.

Exercise 6.6.12. Compute the product of the elements (M,v) and (M',v') of $SL(2,\mathbb{R})$ (we use the same convention as in the preceding paragraph).

Show that $SA(2,\mathbb{R})$ is isomorphic to a subgroup of $SL(3,\mathbb{R})$, by

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

The group $SA(2,\mathbb{R})$ acts on the right on the quotient $SA(2,\mathbb{Z})\backslash SA(2,\mathbb{R})$. In particular, one can study the action of the flow g_t, or *scenery flow* .

The maps $\tilde{\Phi}$ and $\tilde{\Psi}$ turn out to be the first return map of the scenery flow to a suitable section. For more details, see [33, 36].

We will just remark here that this flow is Anosov, and that it admits a number of different interpretation (Teichmüller flow on the twice punctured torus, renormalizations on Sturmian quasi-crystals, see [33]).

6.7 Miscellaneous remarks

6.7.1 Problems of orientation

A recurrent problem in the theory of continued fractions or Sturmian systems is whether one should use any invertible maps, or only orientation preserving invertible maps.

For example, the flip substitution E: $\mathbf{0} \mapsto \mathbf{1}$, $\mathbf{1} \mapsto \mathbf{0}$ reverses orientation, as does the Fibonacci substitution $\mathbf{0} \mapsto \mathbf{01}$, $\mathbf{1} \mapsto \mathbf{0}$.

In this paper, we have chosen to work only in the orientation preserving case, that is, with substitutions whose abelianization belongs to $SL(2, \mathbb{Z})$. This is the reason why, for example, we consider only continued fractions with even periods.

It is perfectly possible to work with orientation reversing maps, and this can simplify some results. For example, it is known that two numbers are equivalent modulo maps $z \mapsto \frac{az+b}{cz+d}$, with $ad - bc = \pm 1$, if and only if their continued fraction expansion coincide eventually, up to a power of the shift, and that they are equivalent modulo $SL(2, \mathbb{Z})$ if and only if their continued fraction expansion coincide eventually, up to an even power of the shift.

In the same way, one could use only the maps σ_0, τ_0 and E, and consider only Sturmian sequences of type 0; each time one obtains sequences of type 1, we just need to apply E to recover a sequence of type 0.

6.7.2 The circle as completion of $\mathbb{N}$ for an adic distance

In Sec. 6.4, we defined the Ostrowski expansion of integers and of real numbers; the relation between these two expansions might have seemed unclear to the reader, and we can give here a nice explanation.

Suppose that a sequence $(a_n)_{n \in \mathbb{N}}$ of strictly positive integers is given, and consider the associated Ostrowski expansion of the integers.

We define an adic distance on the set of integers in the following way: let $n = \sum \epsilon_i q_i$ and $m = \sum \delta_i q_i$ be two integers, we define $d(n, m) = 0$ if $n = m$, and otherwise $d(n, m) = 2^{-k}$, where k is the smallest integer such that $\epsilon_k = \delta_k$.

Exercise 6.7.1. Prove that, for this distance, we have $lim_{i \to \infty} q_i = 0$; prove that $\mathbb{N}$ is not a complete space for this distance.

Exercise 6.7.2. Prove that the map $n \mapsto n+1$ is continuous for the topology induced by this distance.

The space $\mathbb{N}$ is not complete, but it can be completed:

Exercise 6.7.3. Prove that the completion of $\mathbb{N}$ for the Ostrowski distance is a circle, and that the addition of 1 on $\mathbb{N}$ extends to a rotation on this circle.

In fact, this Ostrowski topology on $\mathbb{N}$ is none else than the topology induced by the embedding of $\mathbb{N}$ on the circle as the orbit of 0; the fact that the sequence q_n tends to 0 is related to the fact that, if the q_n are the denominators of the convergents of α, $|q_n\alpha - p_n|$ tends to 0, that is, $R_\alpha^{q_n}(x)$ tends to x for any point x on the circle.

6.7.3 Sturmian substitutions and automorphisms of free groups

In Sec. 6.5, we could have raised a more general question: what are the substitutions whose fixed points are Sturmian? These substitutions are completely known. For more details, see [56, 55, 57] and Chap. 9.

Definition 6.7.1. *A substitution σ is* Sturmian *if the image by σ of any Sturmian sequence is a Sturmian sequence.*

Exercise 6.7.4. Prove that the fixed point of a primitive substitution is Sturmian if and only if this substitution is Sturmian.

Exercise 6.7.5. Prove that the composition of two Sturmian substitutions is a Sturmian substitution.

Hence, the Sturmian substitutions form a monoid, the Sturm monoid, which is very well understood. One knows a presentation by generators and relations (see for instance [57]).

In particular, σ_0, τ_0 and the flip E generate the Sturm monoid.

The submonoid generated by τ_0 and E is called the *monoid of standard morphisms*; it preserves all infinite special words. The submonoid generated by σ_0 and E preserves all fixed words; it contains, as a submonoid of order 2, the monoid generated by σ_0 and $\sigma_1 = E\sigma_0 E$; these are the particular morphisms we considered in the preceding sections. They have the advantage to have determinant 1, and to preserve the fixed words.

Note that there are easy tests to check whether a substitution is Sturmian:

Proposition 6.7.1. *[56] A primitive substitution σ is Sturmian if and only if the word $\sigma(\mathbf{10010010100101})$ is a balanced word.*

Proposition 6.7.2. *[370]A substitution σ over a two-letter alphabet is Sturmian if and only if the words $\sigma(\mathbf{01})$ and $\sigma(\mathbf{10})$, which have the same length, differ only in 2 consecutive indices, where* **01** *is replaced by* **10**.

This last property has a nice geometric interpretation: to a word on two letters, one can associate a path in the plane, starting from 0, and going one step up for each **1** and one step to the right for each **0**. Then the morphism is Sturmian if and only if the paths corresponding to $\sigma(\mathbf{01})$ and $\sigma(\mathbf{10})$ differ exactly on the boundary of a unit square (see [134]).

Another remarkable fact is that a substitution is Sturmian if and only if it extends to an isomorphism of the free group on two generators (see [370] and Chap. 9).

It is not immediately clear whether a word is a fixed word of some substitution:

Exercise 6.7.6. Let u be one of the fixed points of $\mathbf{0} \mapsto \mathbf{010}$, $\mathbf{1} \mapsto \mathbf{10}$.
Let v be the fixed point of $\mathbf{0} \mapsto \mathbf{01}$, $\mathbf{1} \mapsto \mathbf{0}$.
Prove that $v = S^2 u$.

It is easy to prove that there can be only a countable number of fixed points of primitive substitutions, since there is only a countable number of substitutions, and a primitive substitution has a finite number of fixed points. One can also prove that the fixed point of a substitution belongs to a system corresponding to a periodic continued fraction expansion (i.e., a system with a particular quadratic slope).

The following question is thus natural: what are the Sturmian sequences that are fixed points of a substitution? For an answer, see [105, 383].

6.7.4 The problem of the Gauss measure

When we work on the set Σ of all biinfinite Sturmian sequences, we would like to speak of sets of measure 0 (for example, to say that almost all sequences are completely defined by their coding sequence, or that the union of orbits of the fixed points is of measure 0).

The problem is that there is no obvious natural measure on Σ.

One could use the map Φ, and look for an invariant measure for this map, using the fact that it can be coded as a subshift of finite type.

However, the most interesting measure should be the pullback of the Gauss measure, and this can probably be obtained as a Gibbs measure for the subshift of finite type associated to Φ.