Free group automorphisms and tilings

Valérie Berthé

1 Introduction

Free group morphisms are classical objects which prove to be very useful in many mathematical fields (algebraic and topological geometry, combinatorial and geometric group theory, hyperbolic geometry, and so on), as well as in theoretical computer science and physics. Among them, nonnegative morphisms play an important role. A morphism of a free group is said nonnegative if the reduced words representing the images of the letters contain no inverses of the letters; nonnegative morphisms can be considered as morphisms of the underlying free monoid and are often called *substitutions*. Substitutions are hence very simple combinatorial objects (roughly speaking, these are rules to replace a letter by a word) which produce infinite sequences by iteration, the main simplification being here that we have no problem of cancellations. Substitutive sequences generate symbolic dynamical systems which have a rich structure as shown by the natural interactions with combinatorics on words, ergodic theory, linear algebra, spectral theory, geometry of tilings, theoretical computer science, Diophantine approximation, transcendence, graph theory, and so on. See for instance the references in [7].

Notice that the notion of substitution we consider here differs from that used for self-similar tilings (see for instance [9]); in this framework, substitutions produce matching rules acting on a finite set of prototiles and determining the ways in which the tiles are allowed to fit together locally; the best known example is the Penrose tiling.

The aim of this survey is to focus on invertible substitutions (that is, substitutions that are free group automorphisms) in connection with some topological properties of tilings that are associated with substitutions with some prescribed algebraic properties (we consider here the Pisot case). We will see that if invertible substitutions over a two-letter alphabet are completely characterized, much less is known in the higher-dimensional case.

2 Definitions

Let us first recall some basic terminology and notions. For a detailed introduction to the subject, see [8] and [7].

Let $\mathcal{A}$ denote a finite alphabet. The set of all finite words over $\mathcal{A}$ is denoted by $\mathcal{A}^\star$. We denote by ε the empty word. The set $\mathcal{A}^\star$ endowed wit the concatenation is called the *free monoid* generated by $\mathcal{A}$.

A *substitution* σ is a nonerasing morphism of the free monoid $\mathcal{A}^\star - \{\varepsilon\}$; it extends to a morphism of $\mathcal{A}^\star$ and to a map over $\mathcal{A}^{\mathbb{N}}$ by concatenation. A *fixed point* of the substitution σ is an infinite sequence u with $\sigma(u) = u$.

Substitutions are very efficient tools for producing sequences. Let σ be a substitution over the alphabet $\mathcal{A}$, and a be a letter such that $\sigma(a)$ begins with a and $|\sigma(a)| \geq 2$. Then there exists a unique fixed point u of σ beginning with a. This sequence is obtained as the limit in $\mathcal{A}^\star$ (when n tends toward infinity) of the sequence $\sigma^n(a)\ldots$, which is easily seen to converge.

For instance, the *Fibonacci sequence* is the fixed point v beginning with a of the the *Fibonacci substitution* σ defined over the two-letter alphabet $\{a, b\}$ by $\sigma(a) = ab$ and $\sigma(b) = a$.

The Fibonacci sequence provides the more classical example of a Sturmian sequence. *Sturmian sequences* are defined as the one-sided sequences having exactly $n + 1$ factors of length n. A factor of a sequence is a finite string of letters which appear consecutively in this sequence. This combinatorial definition has the particularity of being equivalent to the following simple geometrical representation: a Sturmian sequence codes the orbit of a point of the unit circle under a rotation by irrational angle α with respect to a partition of the unit circle into two intervals of lengths α and $1 - \alpha$. Sturmian sequences have thus remarkable combinatorial and arithmetical properties and have been extensively studied (see for instance [5, 7]).

3 Invertible substitutions

Let Γ_d denote the free group over d letters. A substitution over a d-letter alphabet can be naturally extended to Γ_d by defining $\sigma(s^{-1}) = (\sigma(s))^{-1}$. It is said to be *invertible* if it is an automorphism of the free group, that is, if there exists a map $\eta : \Gamma_d \to \Gamma_d$ such that $\sigma\eta(a) = \eta\sigma(a) = a$ for every $a \in \Gamma_d$.

The Fibonacci substitution is invertible: its inverse is $a \mapsto b$, $b \mapsto b^{-1}a$. For a nice introduction to the free groups and invertible substitutions, see [5]; see also [6].

Invertible substitutions over a two-letter alphabet are completely characterized (see the references in [5]). In particular they are Sturmian, i.e., they preserve Sturmian words. Furthermore, the monoid of invertible substitutions is finitely generated. A set of generators is given by: ($a \mapsto b$, $b \mapsto a$), ($a \mapsto ab$ $b \mapsto a$), and ($a \mapsto ba$, $b \mapsto a$). Note that similarly the set $SL(2, \mathbb{N})$ of matrices with determinant 1 is also finitely generated by elementary matrices (it is even a free monoid, which can be considered as an algebraic version of the classical continued fraction expansion). These two results are closely related. Indeed, one can associate with a substitution its matrix of incidence M defined as follows: the entry $M_{i,j}$ counts the number of occurrences of the letter i in the word $\sigma(j)$. Every matrix in $SL(2, \mathbb{N})$ is the matrix of an invertible substitution. Note furthermore that one can generate Sturmian sequences as a limit of the composition of the two following substitutions (with matrix in $SL(2, \mathbb{N})$): $a \mapsto a$, $b \mapsto ab$, and $a \mapsto ba$, $b \mapsto b$.

One is far from understanding the structure of the monoid of invertible substitutions over a larger size alphabet: indeed the monoid of invertible substitutions over a d-letter alphabet is no more finitely generated, for $d > 2$; neither does $SL(3, \mathbb{N})$ (for more details, see [7]). However, these two results, which could be thought nearly equivalent, are in fact completely unrelated. Indeed there are matrices with nonegative entries of determinant 1 that are not matrices of invertible substitutions for $d = 3$.

4 Tilings associated with substitutions

The aim of this section is to describe how to associate in a natural way with a substitution (with some prescribed algebraic properties) some periodic and autosimilar tilings of the space with fractal boundary.

4.1 One-dimensional case

The simplest example of a self-similar non-periodic tiling is the Fibonacci tiling of the line; it is a tiling by two intervals, one interval a, of length $\phi = \frac{1+\sqrt{5}}{2}$, and interval b of length 1, the structure of the tiling being the unique fixed point of the Fibonacci substitution $a \mapsto ab$, $b \mapsto a$. If we divide each interval a in an interval a' of length 1, and an interval b' of length $1/\phi$, and rename the interval b as a', we obtain a new tiling (called *inflation* of the initial tiling) which is homothetic to the initial one by an homothety of ratio $1/\phi$; on the other hand, it is easy to check that each interval b is preceded by an interval a; if we merge each pair of consecutive intervals a, b as one interval a'', of length ϕ^2, and rename the remaining intervals as b'', we obtain another tiling (*deflation* of the initial one), homothetic to the initial

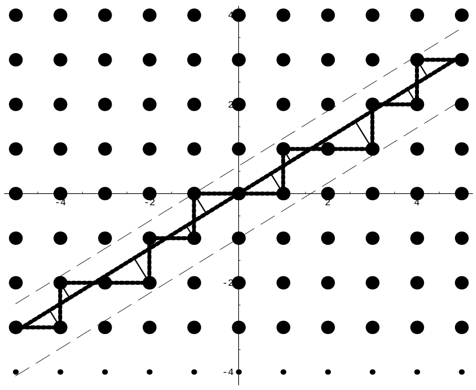

Figure 1: The Fibonacci tiling by the "cut and project" method

one by an homothety of ratio ϕ.

This remarkable tiling has been much studied, starting with the pioneering work of de Bruijn [3, 4]; it can be obtained in a number of different ways, notably by the "cut and project" method (see for instance [11]), as projection of a discrete approximation of the line through 0 with slope $1/\phi$ (see Figure 1).

Let us note that in the general case the symbolic sequences associated with such tilings of the lines (with an irrational slope) are the Sturmian sequences.

4.2 Higher-dimensional case

Let us try to generalize this approach in a higher-dimensional space. Let σ be a substitution defined over the alphabet $\{1, 2, 3\}$. Let M_σ denote the *matrix of incidence* of the substitution σ. We will suppose that this matrix is *primitive* (that is, there exists a power of the matrix that is strictly positive). We can then suppose, replacing if necessary σ by a power σ^n and permuting letters, that $\sigma(1)$ begins with 1, and that there exists an infinite word u starting with 1 such that $\sigma(u) = u$. We will assume that the substitution σ is of *Pisot type*, that is, M_σ has a real eigenvalue λ_1, and all other eigenvalues of M_σ have a nonzero modulus strictly less than one (the behaviour of non-Pisot substitutions is much more complicated, and largely unknown). This implies that the characteristic polynomial of the matrix is irreducible. In that case, there is one *expanding direction* and a *contracting plane* $\mathcal{P}$ which is the kernel of the linear form given by the left eigenvector of M_σ corresponding to λ_1. Consider for instance the Tribonacci substitution $1 \mapsto 12$,

$2 \mapsto 12, 3 \mapsto 1$.

One associates with the infinite word $u = (u_n)_{n\in\mathbb{N}}$ a broken line starting from 0 in $\mathbb{Z}^3$ and approximating the expanding direction as follows. Let $f : \{1,2,3\}^* \to \mathbb{Z}^3$ be defined by $f(w) = |w|_1 e_1 + |w|_2 e_2 + |w|_3 e_3$, where (e_1, e_2, e_3) denotes the canonical basis of $\mathbb{R}^3$ and $|w|_i$ denotes the number of occurrences of the letter i in the word w. Let us consider the following set of points in $\mathbb{Z}^3$: $\{f(u_0 u_1 \dots u_n);\ n \in \mathbb{N}\}$. The broken line is defined by joining these points with segments. (Intuitively, we start from the origin, and read the fixed point, moving one step in direction e_i if we read letter i.) Let π denote the projection over the contracting plane $\mathcal{P}$ of the matrix along the expanding direction. We project these points by π, and define the set X_σ as the closure of the projection:

$$X_\sigma := \overline{\{\pi(f(u_0 \dots u_n));\ n \in \mathbb{N}\}},$$

and for $i \in \{1,2,3\}$

$$X_\sigma^i := \overline{\{\pi(f(u_0 \dots u_n));\ n \in \mathbb{N} \text{ and } u_n = i\}}.$$

The set X_σ is called the *atomic surface* or *Rauzy fractal* associated with σ, following [10]. See also [13] or [1] for similar tilings in connection with beta-numeration. For a survey on results connected with these sets, see [7].

It is not difficult to see that the vectors $f(u_0 u_1 \dots u_n)$ stay within bounded distance of the expanding direction, which is exactly the direction given by the vector of frequencies of the letters $1, 2, 3$ in the infinite word u. Furthermore, the set X_σ is bounded, and hence compact. The translates of the atomic surface by the vectors of the lattice $\mathcal{L}_0 := \mathbb{Z}\pi(e_1 - e_2) + \mathbb{Z}\pi(e_1 - e_3)$ cover the contracting plane $\mathcal{P}$:

$$\cup_{\gamma \in \mathcal{L}_0} (X_\sigma + \gamma) = \mathcal{P}.$$

Note that one can also prove that X_σ is the closure of its interior.

The proof is a direct consequence of the Perron-Frobenius theorem.

4.3 Some open questions

We do not know in the general case whether the sets X_i can overlap, although in all known examples it has been proved that their intersection has measure 0. The same question arises for X_σ and its tranlates by the lattice $\mathcal{L}_0$, that is, do the translates of X_σ tile peoriodically the contracting plane? These are probably the two more important questions in this field. If one assumes that the substitution is *unimodular* (the absolute value of the determinant of the matrix equals one), and if one adds an extra combinatorial assumption, that is, the *strong coincidence*

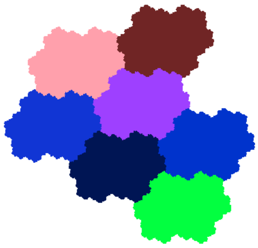

Figure 2: The autosimilar and periodic tiling generated by the Rauzy fractal.

condition, then it can be proved that the sets X_σ^i are disjoint of each other, up to a set of measure zero, and form a partition of X_σ. For more details, see [2]. Furthermore, the second question (the question of tiling) can be also answered by an algorithmic condition which only depends on the substitution [12]. Indeed, A. Siegel has produced in [12] an explicit and effective combinatorial condition (based on a clever graph study) for the atomic surface associated with a unimodular Pisot substitution over d letters to generate a regular tiling of $\mathbb{R}^{d-1}$. Many questions also arise concerning the topological properties of these Rauzy fractals. For instance, they appear to be non-connected or non-simply connected in some examples. Let us end with the following question inspired by the two-letter case, which is a further motivation for the study of invertible substitutions: does the fact that the substitution is invertible imply any kind of topological property?

References

[1] S. Akiyama *Self affine tiling and Pisot numeration system*, "Number Theory and its Applications", ed. by K. Györy and S. Kanemitsu, Kluwer (1999), 7–17.

[2] P. Arnoux, S. Ito *Pisot substitutions and Rauzy fractals*, Bull. Belgian Math. Soc. **8** (2001), 181–207.

[3] N.G. de Bruijn *Sequences of zeros and ones generated by special production rules*, Kon. Nederl. Akad. Wetensch. Proc. Ser. A, **84** (1981), 38–52.

[4] N.G. de Bruijn *Updown generation of Beatty sequences*, Kon. Nederl. Akad. Wetensch. Proc. Ser. A, **92** (1989), 385–407.

[5] M. Lothaire *Algebraic Combinatorics on Words*, Chapitre 2: Sturmian words, par J. Berstel et P. Séébold, to appear.

[6] R. C. Lyndon, P. E. Schupp *Combinatorial group theory*, Classics in Mathematics, Springer-Velag, Berlin, 2001.

[7] N. Pytheas Fogg *Substitutions in Dynamics, Aritmetics and Combinatrorics*, V. Berthé, S. Ferenczi, C. Mauduit, A. Siegel eds., Lecture Notes Math. **1794**, Springer Verlag, 2002.

[8] M. Queffélec, *Substitution Dynamical Systems. Spectral Analysis*, Lecture Notes in Math. **1294**, Springer-Verlag (1987).

[9] C. Radin *Space tilings and substitutions*, Geometriae Dedicata **55** (1995), 257–264.

[10] G. Rauzy *Nombres algébriques et substitutions*, Bull. Soc. math. France **110** (1982), 147–178.

[11] M. Senechal *Quasicrystals and geometry*, Cambridge University Press, Cambridge, 1995.

[12] A. Siegel *Pure discrete spectrum dynamical system and periodic tiling associated with a substitution*, preprint, 2002.

[13] W. P. Thurston *Groups, Tilings and Finite state automata*, AMS Colloquium lectures, 1989.