12. Some open problems

As a conclusion, let us emphasize the underlying arithmetic structure and the interaction between the geometric and symbolic nature of the dynamical systems we have considered throughout this book. We are mainly interested in the two following problems: first, finding geometric interpretations of various symbolic dynamical systems including those generated by substitutions, and secondly, developing multidimensional continued fraction algorithms reflecting the dynamics of the systems.

12.1 The S-adic conjecture

Let us start with an algorithmic approach. For more details on the subject, the reader is referred to [140, 142].

A sequence is said to have *sub-linear complexity* if there exists a constant C such that for every positive integer n, $p(n) \leq Cn$.

Numerous combinatorial, ergodic or arithmetic properties can be deduced from this indication on the growth-order of the complexity function.

Let us start with a purely combinatorial result due to Cassaigne [90].

Theorem 12.1.1 (Cassaigne). *If the complexity function $p(n)$ of a sequence is sub-linear, then the first difference of the complexity $p(n+1)-p(n)$ is bounded.*

Furthermore, an upper bound on the first difference can explicitly be given.

Let us note that the equivalent result does not hold in the case of sub-quadratic complexity sequences: in [140], one can find an example of sequence such that its complexity function is sub-quadratic and has unbounded second differences $p(n+2)+p(n)-2p(n+1)$.

12.1.1 S-adic expansions

The following result [140] can be deduced from Theorem 12.1.1 and considerations on the graph of words. The complexity function of a symbolic dynamical system is defined as the function which counts the number of distinct factors of the *language* of this system, i.e., the union of the sets of factors of the sequences of the system.

Theorem 12.1.2 (Ferenczi). *Let X be a minimal symbolic system on a finite alphabet $\mathcal{A}$ such that its complexity function $p_X(n)$ is a sub-linear function, or, equivalently (following Theorem 12.1.1), such that $p_X(n+1)-p_X(n)$ is bounded for every $n \geq 1$; then there exists a finite set of substitutions $\{\sigma_i,\ 1 \leq i \leq c\}$, over an alphabet $\mathcal{D} = \{0,...,d-1\}$, an application φ from $\mathcal{D}$ to $\mathcal{A}$, and an infinite sequence $(i_n)_{n\geq 1}$ with values in $\{1,\ldots,c\}$ such that*

$$\inf_{0\leq r\leq d-1} |\sigma_{i_1}\sigma_{i_2}...\sigma_{i_n}(r)| \to +\infty$$

when $n \to +\infty$, and any word of the language of the system is a subword of $\varphi(\sigma_{i_1}\sigma_{i_2}...\sigma_{i_n})(0)$ for some n.

The above proposition can be read as follows: *minimal systems with sub-linear complexity are generated by a finite number of substitutions.* Using a variation of the Vershik terminology, we propose to call such systems *S-adic systems.* We similarly call a sequence *S-adic* if the symbolic dynamical system generated by the sequence is itself S-adic.

It will be noted that the fact that the lengths of the words tend to infinity, which generalizes the notion of *primitivity* of a substitution, is necessary to make Theorem 12.1.2 nonempty. It can be seen in the proof of [140] that this prevents us from getting a universal bound on the number d of letters of the alphabet $\mathcal{D}$. However, in an important particular case generalizing the Arnoux-Rauzy sequences [42], we do have a universal upper bound:

Proposition 12.1.1. *For minimal systems over a three-letter alphabet such that*

$$p_X(n+1) - p_X(n) \leq 2$$

for every n large enough, Theorem 12.1.2 is satisfied with $d \leq 3$ and $c \leq 3^{27}$.

A measure-theoretic consequence of Theorem 12.1.2 is that a minimal and uniquely ergodic system of sub-linear complexity cannot be strongly mixing (see Chap. 3 for definitions).

The S-adic expansion is known in an explicit way for the symbolic dynamical systems generated by Sturmian sequences, by the Arnoux-Rauzy sequences [42], for the systems generated by some binary codings of rotations [124] (i.e., for codings with respect to a partition of the unit circle into intervals the lengths of which are larger than or equal to the angle of the rotation), and for systems generated by some interval exchanges [227], [144]. Let us note that in the Sturmian case, one gets a more precise result: every Sturmian sequence is S-adic in the sense of Chap. 6, the iteration following the Ostrowski expansion (with respect to the angle), of the initial point, the orbit of which is coded (for more details, see [35]). In the other cases (Arnoux-Rauzy sequences, binary codings, interval exchanges) one uses multidimensional continued fraction expansions.

12.1.2 The S-adic conjecture

The converse of Theorem 12.1.2 is clearly untrue. It is sufficient to consider a non-primitive substitution of quadratic complexity function to produce a counter-example, as for instance, the fixed point beginning with a of the substitution $a \mapsto ab$, $b \mapsto bc$, $c \mapsto c$ [269]. Nevertheless, in the case where the initial sequence u is minimal, the fact that it is a fixed point of a substitution guarantees that the complexity is sub-linear. This is probably not true any more if the sequence u is generated by two substitutions, even if u is minimal. Such an example was proposed by Boshernitzan: one uses a primitive substitution which appears infinitely often in the iteration, and a substitution of quadratic complexity which appears in long ranges in the iteration.

Question. Let u be a sequence generated by the iteration of a finite number of substitutions; which restrictions should one add to these substitutions, so that the sequence u has sub-linear complexity?

We still have to find a *stronger* form of S-adicity which would be equivalent to sub-linear complexity. This is the *S-adic conjecture*, due to Host, stating that minimal systems are of sub-linear complexity if and only if they are *strongly* S-adic. Theorem 12.1.2 means that one half of the conjecture is proved: *minimal systems of sub-linear complexity are S-adic.*

In other words, Theorem 12.1.2 implies that systems with a "reasonably low" language complexity can be built with a "reasonably low" algorithmic complexity. The *Kolmogorov-Chaitin complexity* of an infinite sequence associates to n the length of the shortest program (for example on a universal Turing machine) able to produce the first n terms of the sequence (see [222]). Now, Theorem 12.1.2 states that, if the symbolic complexity of a minimal sequence is small enough, there is an explicit algorithm which produces the system, with bounds on its length in some sub-cases.

Question. Can Theorem 12.1.2 be improved to get an equivalence between a "reasonably low" symbolic complexity and a "reasonably low" Kolmogorov-Chaitin complexity?

A result in the same direction has been proved recently in [68]: a Sturmian sequence for which the angle α is an algebraic number of degree two or the number e can be recognized by a Turing machine in linear time.

12.1.3 Linear recurrence

Durand gives a sufficient condition in [128] for a sequences to have sub-linear complexity. Namely, let u be a given recurrent sequence and let W be a factor of the sequence u. Let us recall that a *return word* over W is a word V such that VW is a factor of the sequence u, W is a prefix of VW and W has exactly two occurrences in VW. A sequence is *linearly recurrent* if there exists a constant $C > 0$ such that for every factor W, the length of every return word V of W satisfies $|V| \leq C|W|$. Let us note that such a sequence has sub-linear complexity [132]. Unfortunately, this condition is strictly stronger than having sub-linear complexity. Durand shows that a sequence is linearly recurrent if and only if it is S-adic with bounded partial quotients, i.e., every substitution comes back with bounded gaps in the S-adic expansion [128]. In particular, a Sturmian sequence is linearly recurrent if and only if the partial quotients in the continued fraction expansion of its angle are bounded, which means in particular, that not every S-adic sequence is linearly recurrent.

12.1.4 Periodic case and substitutive sequences

Given an S-adic sequence, one can ask whether this sequence is substitutive. In the Sturmian case, Sturmian sequences which are substitutive correspond to quadratic angles (for more details, see [105, 383]). This can be deduced from Durand's characterization on substitutive sequences based on return words (to be more precise, the notion of derived sequence can be geometrically seen using the induction). Hence, this result can be considered as a generalization of Galois' theorem. See also [73] for a connected result: if all the parameters of an exchange of intervals belong to the same quadratic extension, the sequence of induced (by performing always the same induction process) is ultimately periodic.

12.2 Multidimensional continued fraction expansions

12.2.1 Arithmetics and S-adicity

One of the main interests of the S-adic expansion lies in the fact that it provides an arithmetic description of the sequences we consider. For instance, the sequence (i_n) (in Theorem 12.1.2) is governed by the continued fraction expansion of the angle in the Sturmian case. More generally, in numerous cases, the "partial quotients" (i.e., the successive powers which appear in the iteration in Theorem 12.1.2), provide a generalized continued fraction expansion which describes the combinatorial properties of the sequence: one can find such examples of continued fraction expansions in [143, 144, 145, 124, 227, 291, 300, 385]. The techniques which are usually used in these problems are on the one hand, the use of the graphs of words (see for instance [61]), and on the other hand, the induction process (see for instance Chap. 3 and Chap. 6).

The usual continued fraction algorithm provides the best rational approximations of a real number α; it is fundamentally connected to the toral rotation of $\mathbb{T}^1$, $x \mapsto x + \alpha$. It describes in a natural way the combinatorial properties of Sturmian sequences, as well as the properties on the one hand, of their geometric representation as a coding of a rotation, and on the other hand, of their associated symbolic dynamical system. What can be said now when considering the two-dimensional torus $\mathbb{T}^2$, or when trying to get simultaneous approximation results?

There exist several unidimensional or multidimensional symbolic objects which generalize Sturmian sequences in a natural way and provide multidimensional continued fraction algorithms. One can for instance code a rotation over $\mathbb{T}^1$ with respect to a partition into two intervals or more, the lengths of which do not depend on the angle α of the rotation. Such sequences have been studied for instance in [5, 124, 302, 63].

Three particular generalizations have been the subject of recent studies: they have been dubbed "The good, the bad and the ugly" by a Texan author.

12.2.2 The good: interval exchanges

Three-interval exchanges are fundamentally connected (via an induction process) to these sequences. Ferenczi, Holton and Zamboni develop in [143, 144, 145] a multidimensional algorithm which generates the orbits of the discontinuity points and opens new ways in the study of the ergodic and spectral properties of interval exchanges. Let us note that this algorithm is produced in a combinatorial way by catenative rules of production of bispecial factors, or in a geometric way, by the observation of the evolution of the intervals corresponding to the bispecial factors. It can also be expressed in terms of a new induction process which does not correspond to the usual ones. See also [227] for an S-adic expansion.

12.2.3 The bad: Arnoux-Rauzy sequences

Let us consider now the *Arnoux-Rauzy sequences*. It will be recalled that these are recurrent sequences defined over a three-letter alphabet with the following extra combinatorial property [42]: for every n, there is exactly one right special factor and one left special factor of length n. It will be recalled that a factor W of the sequence u is called right special (respectively left special) if W is a prefix (respectively suffix) of at least two words of length $|W|+1$ which are factors of the sequence u (see also Chap. 6). Let us note that they can be similarly defined over any alphabet of larger size, say d; one thus obtains sequences of complexity $(d-1)n+1$. Arnoux-Rauzy sequences are a natural generalization of Sturmian sequences, since they share with Sturmian sequences the fundamental combinatorial property of the unicity of the right (and left) special factors of given length. Contrary to the Sturmian case, these sequences are not characterized by their complexity function any more.

One knows perfectly well the S-adic expansion of the Arnoux-Rauzy sequences [42], and more precisely, the S-adic expansion of the associated dynamical system and the S-adic expansion of the sequence which has as prefixes the left special factors; following the terminology in the Sturmian case, this sequence is called the *left special sequence* or the *characteristic* *sequence*. The combinatorial properties of the Arnoux-Rauzy sequences are well-understood and are perfectly described by a two-dimensional continued fraction algorithm defined over a subset of zero measure of the simplex introduced in [385, 300] and in [96]. By using this algorithm, one can express in an explicit way the frequencies of factors of given length [381], one can count the number of all the factors of the Arnoux-Rauzy sequences [250], study their balance properties [93], or prove that the associated dynamical system has always simple spectrum [96].

These sequences can also be described as an exchange of six intervals of the unit circle [42]. Let us recall that the *Tribonacci sequence* (i.e., the fixed point beginning by a of the *Rauzy substitution*: $a \mapsto ab,\ b \mapsto ac,\ c \mapsto a$), is an Arnoux-Rauzy sequence. It will be noted that in this case, one obtains good approximation results [97]. The dynamical system generated by this sequence is isomorphic to a rotation of the torus $\mathbb{T}^2$ (for more details, see Chap. 7 and 8). More generally, in the periodic case, i.e., in the case where the S-adic expansion of the special sequence is purely periodic, this sequence is hence a fixed point of a Pisot type, primitive, unimodular substitution (and conversely). All the results of Chap. 7 and Chap. 8 apply. In particular, the dynamical systems generated by such sequences are obtained as toral rotations (the sufficient condition of Theorem 8.4.1 holds).

It was believed that all Arnoux-Rauzy sequences were issued from toral rotations, and more precisely, that they were natural codings of rotations over $\mathbb{T}^2$. We say that a sequence u is a *natural coding* of a rotation if there exists a measurable dynamical system (X, T, μ), which is itself measure-theoretically isomorphic to a rotation, a generating partition $\mathcal{P} = \{P_1, P_2, P_3\}$ of X, and a point $x \in X$ such that

$$\forall n \in \mathbb{N},\ u_n = i \text{ if and only if } T^n(x) \in P_i;$$

furthermore, the map T is supposed to be a piecewise translation, the translation vector taking exactly one value on each P_i. Namely, in all the examples, we know that the rotation is constructed as an exchange of domains in $\mathbb{R}^2$, the pieces of the exchange being isomorphic with the cylinders (see Chap. 7 and Chap. 8).

This conjecture was disproved in [93], where an example of totally unbalanced Arnoux-Rauzy sequence is constructed (a sequence is said to be *totally unbalanced* if there exists a letter a such that for each positive integer n, there exist two factors of the sequence with equal length, with one having at least n more occurrences of the letter a than the other). Namely, a natural coding of a rotation cannot be totally unbalanced following [138]. Let us note that the partial quotients (in the sense of the S-adic expansion or in the sense of the associated continued fraction) are unbounded in the above example of totally unbalanced Arnoux-Rauzy sequence.

The following questions are thus natural.

Questions. How to characterize the Arnoux-Rauzy sequences which are natural codings of rotations? In a weaker form, how to characterize the Arnoux-Rauzy sequences which are measure-theoretically isomorphic to rotations? Is the linear recurrence a sufficient condition for either one or the other question?

Are there Arnoux-Rauzy sequences which are weakly mixing? What are the eigenvalues of the Arnoux-Rauzy sequences? Do they admit a rotation as a factor?

12.2.4 The ugly: codings of two rotations

A second way of generalization of the Sturmian case consists in introducing a second parameter. One thus gets two dual approaches: one can either code a rotation of angle (α, β) in the two-dimensional torus $\mathbb{T}^2$, or a $\mathbb{Z}^2$-action by two irrational rotations of angle α and β over $\mathbb{T}^1$. In the first case, one gets a unidimensional sequence, in the second, a two-dimensional sequence. We discussed this first case by considering Arnoux-Rauzy sequences and their connections with rotations over $\mathbb{T}^2$ in the previous section. Consider now the second approach.

Let $(\alpha, \beta) \in \mathbb{R}^2$, such that $1, \alpha, \beta$ are rationally independent, and let $\rho \in \mathbb{R}$. Let u be the two-dimensional sequence defined over $\mathbb{Z}^2$ with values in $\{0, 1\}$ by

$$\forall (m,n) \in \mathbb{Z}^2,\ (u(m,n) = 0 \Longleftrightarrow m\alpha + n\beta + \rho \in [0, \alpha[\text{ modulo } 1).$$

Such a two-dimensional sequence has $mn+n$ rectangular factors of size (m, n) and is *uniformly recurrent* (i.e., for every positive integer n, there exists an integer N such that every square factor of size (N, N) contains every square factor of size (n, n)). The function which counts the number of rectangular factors of given size is called the *rectangle complexity function* (for more details, see Sec. 12.3). Conversely, every two-dimensional uniformly recurrent sequence with complexity $mn + n$ admits such a geometric description [65].

Let us see why these sequences can be considered as a generalization of Sturmian sequences.

They have the smallest complexity function known among two-dimensional sequences which are uniformly recurrent and not periodic. They are obtained as a letter-to-letter projection of two-dimensional sequences defined over a three-letter alphabet which code a plane approximation (let us us recall that Sturmian sequences code and describe discrete lines [54, 57]): one can approximate a plane with irrational normal by square faces oriented along the three coordinates planes; this approximation is called a *discrete plane*; after projection on the plane $x + y + z = 0$, along $(1, 1, 1)$, one obtains a tiling of the plane with three kinds of diamonds, namely the projections of the three possible faces

(for more details, see [363, 66], see also [298, 102]). One can code this projection over $\mathbb{Z}^2$ by associating with each diamond the name of the projected face. One thus gets a sequence with values in a three-letter alphabet. These sequences are shown to code a $\mathbb{Z}^2$-action over the unit circle $\mathbb{T}^1$ [66].

These later sequences are generated by two-dimensional substitutions governed by the Jacobi–Perron algorithm. Namely, a geometric interpretation of the Jacobi–Perron algorithm is given in [34]: in the same way as classical continued fractions can be interpreted in terms of induction of rotations, this algorithm can be considered as an induction algorithm for a group of rotations operating on the unidimensional torus; it is not trivial that this group can be induced on a subinterval to obtain a new group of rotations, but by inducing on a suitable interval, the induction of the $\mathbb{Z}^2$-action we consider is again generated by a $\mathbb{Z}^2$-action by a pair of rotations. However, unlike the classical $\mathbb{Z}$-action, the generators of the $\mathbb{Z}^2$-action are not canonically defined (since we can find an infinite number of bases for the lattice $\mathbb{Z}^2$). This is to relate to the fact that there seems to be no way of defining a "best" two-dimensional continued fraction algorithm. Using this induction process, we define in [34] a sequence of two-dimensional substitutions associated with the Jacobi–Perron algorithm (see also Chap. 8) which generate the two-dimensional Sturmian sequences mentioned above over a three-letter alphabet. We shall allude again to this notion of two-dimensional substitution in Sec. 12.3.

12.3 Combinatorics on words for two-dimensional sequences

When one works with multidimensional sequences, fundamental problems in the definition of the objects appear; for instance, how can one define a multidimensional complexity function? Consider two-dimensional sequences, i.e., sequences defined over $\mathbb{Z}^2$ with values in a finite alphabet. A natural notion of complexity consists in counting the rectangular factors of given size: we thus define the *rectangle complexity function*: $(m, n) \mapsto P(m, n)$. This notion is not completely satisfactory since we choose a basis of the lattice $\mathbb{Z}^2$. For a more general definition of complexity, see [310]. The following questions are thus natural:

Questions. Can one characterize a two-dimensional sequence with respect to its rectangle complexity function? Which functions do exist as rectangle complexity functions?

A two-dimensional sequence is *periodic* if it is invariant under translation, i.e., if it admits a nonzero vector of periodicity with integer coordinates. Note that the fact that the lattice of vector periods has rank 2 is characterized by a bounded rectangle complexity function. There is no characterization of periodic sequences by means of the complexity function: one can construct two-dimensional sequences with a nonzero periodicity vector of arbitrarily large complexity function; namely, consider the sequence $(u_{m+n,n})_{(m,n)\in\mathbb{Z}^2}$, where the unidimensional sequence $(u_n)_{n\in\mathbb{Z}}$ has maximal complexity. Conversely Nivat has conjectured the following:

Conjecture 12.3.1. If there exists (m_0, n_0) such that $P(m_0, n_0) \leq m_0 n_0$, then the two-dimensional sequence u is periodic.

Let us note that the conjecture was proved for factors of size $(2, n)$ or $(n, 2)$ in [311]. A more general conjecture is given in [312, 310, 311]. Let us remark that in the higher-dimensional case, counterexamples to the conjecture can be produced [311].

It is interesting with respect to this conjecture, to consider the "limit" case of sequences of rectangle complexity function $mn+1$. Such sequences are fully described in [92] and are all proved to be non recurrent. The two-dimensional Sturmian sequences mentioned in Sec. 12.2.4 of complexity $mn + n$ are conjectured to be the uniformly recurrent sequences with smallest complexity function (it remains to give a more precise meaning to the term "smallest").

Let us end this section with the problem of the definition of two-dimensional substitutions. Unlike the classical one-dimensional case, the notion of two-dimensional substitutions is not trivial. Consider a map which associates with a letter a pointed two-dimensional pattern. We first need more information to know where to place the image of a letter. Two natural questions then arise. First, it is not immediate to prove the consistency if one wants to apply the substitution to a finite pattern or to a double sequence. Secondly, how can one iterate such a process to generate a double sequence? We introduce in [34] two notions of substitutions. We first define *pointed substitutions*: given the value of the initial sequence at the point x, we deduce the value of the image sequence on a pointed pattern situated at a point y that can be computed from x and its value. This is however inconvenient for explicit computation, since we need at each step global information. In particular we are not able to iterate it in order to generate a double sequence. It is much more convenient to be able to use a local information, i.e., *local rules* (this exactly what is done when one computes one-dimensional substitutions: one does not compute the exact position of a given pattern, but one only uses the fact that patterns follow each other). Roughly speaking, a local rule says how to place the image of a pointed letter with respect to the images of the letters belonging to a finite neighborhood. If we know the image of the initial point, we can compute the values of adjacent points by using a finite number of patterns, and in this way, compute the image of the complete sequence.

Question. Can one build other examples of substitutions endowed with local rules? Given a substitution with local rules, does the limit of the iterates cover $\mathbb{Z}^2$? Can one introduce a different notion of substitution which can be used for the generation of two-dimensional sequences?

12.4 Substitutions, translations and toral automorphisms

The general question of the isomorphism between a substitutive dynamical system and a rotation brings many open questions.

The dynamical system generated by the Tribonacci substitution is measure-theoretically isomorphic to a translation on the torus $\mathbb{T}^2$. The isomorphisms is a continuous onto map from the orbit closure of the fixed point of the substitution to the torus. The images of the three basic cylinders corresponding to the letters are connected, and even simply connected domains of the torus. The three basic sets are bases for the three cylinders of a Markov partition for the automorphism of the torus $\mathbb{T}^3$ associated with the substitution. Do these properties extend to some other symbolic dynamical systems?

Let us recall (Theorem 7.10.1) that the symbolic dynamical system associated with a primitive unitary substitution over a two-letter alphabet which satisfies the coincidence condition is measure-theoretically isomorphic to a rotation on $\mathbb{T}^1$ (for more details, see Chap. 7). The coincidence condition states the following for a substitution σ over the two-letter alphabet $\{0, 1\}$: there exist two integers k, n such that $\sigma^n(0)$ and $\sigma^n(1)$ have the same k-th letter, and the prefixes of length $k - 1$ of $\sigma^n(0)$ and $\sigma^n(1)$ have the same number of occurrences of the letter 0.

Question. Is the symbolic dynamical system associated with a primitive unimodular substitution over a two-letter alphabet measure-theoretically isomorphic with a rotation on the unidimensional torus $\mathbb{T}^1$? In other words, does every primitive substitution over two letters satisfy the coincidence condition?

Let us note that the Morse substitution does not satisfy the coincidence condition.

It is conjectured that every primitive unimodular substitution of Pisot type over two or three letters satisfies the coincidence condition. It is not known even in the case of two letters whether there exists a unimodular primitive Pisot substitution that does not satisfy the coincidence condition.

Conjecture 12.4.1. If the Perron–Frobenius eigenvalue of the incidence matrix of a primitive unimodular substitution is a Pisot number, then the dynamics of the substitution is measure-theoretically isomorphic to a rotation on the torus.

A connected problem is the following, addressed by Liardet in [158]: *is the odometer associated with the numeration scale related to a real number $\beta > 1$ isomorphic in measure with a translation on a compact group if β is a canonical Pisot number (i.e., β has a finite β-expansion and the associated companion polynomial is irreducible over $\mathbb{Q}$)?*

Other natural questions arise concerning the topological properties of the fractal domain X_σ associated with a substitution, such as the question of connectedness and simply connectedness.

Canterini gives a sufficient condition for the connectivity of the exchange of domains associated with primitive unimodular substitutions of Pisot type.

Question. Can one give a characterization of those substitutions for which the associated domains are connected or simply connected?

12.5 Arithmetics in $SL(3, \mathbb{N})$ and $SL(3, \mathbb{Z})$

We have considered here products of substitutions via the S-adic conjecture. This can also be expressed in terms of product of matrices. It is thus natural from an algebraic point of view to consider the properties of the monoid $SL(3, \mathbb{N})$ and of the group $SL(3, \mathbb{Z})$. This is the point of view we develop in this section.

12.5.1 Definitions

For $k \geq 2$, let $SL(k, \mathbb{N})$ (respectively $SL(k, \mathbb{Z})$) denote the set of matrices of determinant 1 with nonnegative (respectively integer) coefficients. This set endowed with the multiplication is a monoid (respectively a group).

The monoid $SL(2, \mathbb{N})$ is a free monoid with two generators (for more details, see Chap. 6):

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \; B = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

This expansion is a matricial translation of the Euclidean algorithm and it corresponds to the continued fraction expansion.

Consider now what happens in a higher dimension. We are interested in the study of the algebraic properties of $SL(k, \mathbb{N})$ and $SL(k, \mathbb{Z})$ (for $k \geq 3$), that is, in finding generators and “prime” elements.

A matrix A in $SL(3, \mathbb{N})$ is said to be a *unit matrix* if it admits an inverse in $SL(3, \mathbb{N})$, i.e., if there exists a matrix $B \in SL(3, \mathbb{N})$, such that $AB = Id$. Unit matrices are exactly the matrices of even permutation.

A matrix M in $SL(3, \mathbb{Z})$ is said to be a *prime matrix* if any product of the form $M = AB$, implies that A or B is a unit matrix. Note that in $SL(2, \mathbb{N})$ there are only four prime matrices, whereas in $SL(3, \mathbb{N})$, there are infinitely many prime matrices.

Question. Can one characterize the prime matrices in $SL(3, \mathbb{N})$? Can one give an algorithm which produces the decomposition of any element in $SL(3, \mathbb{N})$ into prime matrices?

Some motivations for these question are the following: first, such a decomposition can be seen as a multidimensional continued fraction algorithm; secondly, it can also provide a better understanding of the automorphisms of the torus.

12.5.2 Invertible substitutions

A substitution over the alphabet $\mathcal{A}$ is said to be *invertible* if it extends to a morphism of the free group generated by the alphabet $\mathcal{A}$.

Invertible substitutions over a two-letter alphabet are completely characterized (see Sec. 9.2 and Chap. 6). In particular they are Sturmian, i.e., they preserve Sturmian words. However, one is far from understanding the structure of the monoid of unimodular substitutions over a larger size alphabet (see [374]).

Question. Can one characterize invertible substitutions over a larger size alphabet?

Another connected problem is the following asked by Cassaigne in [158]. Is it possible to find an algorithm which decides whether the monoid generated by two square matrices of size two is free? The answer is negative in the three-dimensional case [94]; namely, this problem is undecidable.