SEQUENCES OF LOW COMPLEXITY: AUTOMATIC AND STURMIAN SEQUENCES

Valérie BERTHÉ
Institut de Mathématiques de Luminy
CNRS-UPR 9016
Case 907, 163 avenue de Luminy
F-13288 Marseille Cedex 9
France

Contents

Abstract

The complexity function is a classical measure of disorder for sequences with values in a finite alphabet: this function counts the number of factors of given length. We introduce here two characteristic families of sequences of low complexity function: automatic sequences and Sturmian sequences. We discuss their topological and measure-theoretic properties, by introducing some classical tools in combinatorics on words and in the study of symbolic dynamical systems.

1 Introduction

The aim of this course is to introduce two characteristic families of sequences of low "complexity": automatic sequences and Sturmian sequences (complexity is defined here as the combinatorial function which counts the number of factors of given length of a sequence over a finite alphabet). These sequences not only occur in many mathematical fields but also in various domains as theoretical computer science, biology, physics, cristallography...

We first define some classical tools in combinatorics on words and in the study of symbolic dynamical systems: the complexity function and frequencies of factors in connection with the notions of topological and measure-theoretic entropy (Sections 2 and 3), the graphs of words (Section 4), special and bispecial factors (Section 5). Then we study Sturmian sequences in Section 6: these sequences are defined as the sequences of minimal complexity among non-ultimately periodic sequences. 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. Then we introduce automatic sequences in Section 7: an automatic sequence is defined as the image by a letter-to-letter projection of a fixed point of a substitution of constant length or equivalently as a finite-state function of the representation of the index in a given basis. We emphasize on the connections with transcendence of formal power series with coefficients in a finite field. In particular, we will try to answer the following question: how to recognize if a sequence is automatic or not? We conclude this course by discussing the connections between sequences with a linear growth order for the complexity function, and substitutions.

2 Complexity Function

2.1 Definition

Let us introduce a combinatorial measure of disorder for sequences over a finite alphabet: this notion is called *(symbolic) complexity.* For more information on the subject, we refer the reader to the surveys [8, 43] and to the course [59].

In all that follows we restrict ourselves to sequences over a finite alphabet indexed by the set $\mathbb{N}$ of non-negative integers. A *factor* of the infinite sequence $u = (u_n)_{n\in\mathbb{N}}$ is a finite block w of consecutive letters of u, say $w = u_{n+1}\cdots u_{n+l}$; l is called the *length* of w, denoted by $|w|$. Let $p(n)$ denote the *complexity function* of sequence u with values in a finite alphabet: it counts the number of distinct factors of length n of the sequence u. The complexity function is obviously non-decreasing and for any integer n, one has $1 \le p(n) \le d^n$, where d denotes the cardinality of the alphabet.

This function can be considered to *measure the predictability* of a sequence. The first difference of the complexity function counts the number of possible extensions in the sequence of factors of given length. We call *right extension* (respectively *left extension*) of a factor w a letter x such that wx (respectively xw) is a factor of the sequence. Let w^+ (respectively w^-) denote the number of right (respectively left) extensions of w. (One may have

$w^- = 0$ but always $w^+ \geq 1$.) We have

$$p(n+1) = \sum_{|w|=n} w^+ = \sum_{|w|=n} w^-,$$

and thus

$$p(n+1) - p(n) = \sum_{|w|=n} (w^+ - 1) = \sum_{|w|=n} (w^- - 1).$$

Exercise 2.1 (see [31, 54]) Prove that a sequence is *ultimately periodic* (i.e., periodic from a certain index on) if and only if its complexity function satisfies

$$\exists n, \quad p(n) \leq n \Longleftrightarrow \exists C, \ \forall n \quad p(n) \leq C.$$

What happens in the case of a sequence defined over $\mathbb{Z}$?

The complexity function is a measure of disorder connected to the topological entropy: the *topological entropy* [1] is defined as the exponential growth rate of the complexity as the length increases

$$H_{top}(u) = \lim_{n \to +\infty} \frac{\log_d(p(n))}{n}.$$

It is easy to check that this limit exists because of the subadditivity of the function $n \mapsto \log(p(n))$. Note that the word *entropy* is used here as a measure of randomness or disorder. For a survey on the connections between entropy and sequences, see [13].

The study of the complexity is mainly concerned with the following three questions.

- How to compute the complexity of a sequence?
- Which functions can be obtained as the complexity function of some sequence?
- Can one deduce from the complexity a geometrical representation of sequences?

We will see how to answer the first question by introducing special and bispecial factors, in some particular cases of substitutive sequences (Section 5). The second question is still very much in progress and far from being solved (in particular in the case of positive entropy): for a survey on the question, see [24, 43]. Although the complexity function is in general not sufficient to describe a sequence, we will see in Section 6 that much can be said on the geometrical properties in the case of lowest complexity, i.e., in the case of *Sturmian sequences*: these sequences are defined to have exactly $n+1$ factors of length n, for any integer n. By Exercise 2.1 a sequence with complexity satisfying $p(n) \leq n$ for some n is ultimately periodic. Sturmian sequences have thus the minimal complexity among all sequences that are not ultimately periodic.

Exercise 2.2 Deduce from Exercise 2.1 that every prefix of a Sturmian sequence appears at least two times in the sequence. Deduce that the factors of every Sturmian sequence appear infinitely often (such a sequence is called *recurrent*).

2.2 Frequencies and Measure-Theoretic Entropy

The purpose of this section is to introduce a more "precise" (in a sense that we will see in Section 2.3) measure of disorder of sequences, connected with frequencies of factors. The *frequency* $f(B)$ of a factor B of a sequence (called *density* in Host's course) is defined as the limit, if it exists, of the number of occurrences of this block in the first k terms of the sequence divided by k.

Exercise 2.3 Construct a sequence for which the frequencies of letters do not exist.

Let us first introduce *the block entropies* for sequences with values in a finite alphabet in order to define the notion of *measure-theoretic entropy*. These sequences of block entropies were first introduced by Shannon in information theory, to measure the entropy of the English language (see [65]).

Let u be a sequence with values in the alphabet $\mathcal{A} = \{1, \cdots, d\}$. We suppose that all the block frequencies exist for u. Let

$$P(x|x_1 \cdots x_n) = \frac{f(x_1 \cdots x_n x)}{f(x_1 \cdots x_n)},$$

where $x_1 \cdots x_n$ is a block of non-zero frequency and x a letter. Intuitively $P(x|x_1 \cdots x_n)$ is the conditional probability that the letter x follows the block $x_1 \cdots x_n$ in the sequence u. We are going to associate with the sequence u two sequences of block entropies $(H_n)_{n\in\mathbb{N}}$ and $(V_n)_{n\in\mathbb{N}}$.

For all $n \geq 1$, let

$$V_n = \sum L(f(x_1 \cdots x_n)),$$

where the sum is over all the factors of length n and $L(x) = -x \log_d(x)$, for all $x \neq 0$ and $L(0) = 0$. We put $V_0 = 0$.

For all $n \geq 1$, let

$$H_n = \sum{}' f(x_1 \cdots x_n) H(x_1 \cdots x_n), \tag{1}$$

where the sum is over all the blocks of length n of non-zero frequency and

$$H(x_1 \cdots x_n) = \sum_{x \in \mathcal{A}} L(P(x/x_1 \cdots x_n)).$$

We put $H_0 = V_1$. The sequence $(H_n)_{n\in\mathbb{N}}$ measures in some way the properties of predictability of the initial sequence u.

Exercise 2.4 Prove that: $\forall n \in \mathbb{N},\ H_n = V_{n+1} - V_n$. (This classical property in information theory is called the *chain-rule*.)

Thus, $(H_n)_{n\in\mathbb{N}}$ is the discrete derivative of $(V_n)_{n\in\mathbb{N}}$. Note that $(V_n)_{n\in\mathbb{N}}$ is a non-decreasing sequence, since $H_n \geq 0$ for all n.

It can be shown that $(H_n)_{n\in\mathbb{N}}$ is a monotonic non-increasing sequence of n (see, for instance [16]). The intuitive meaning of this is that the uncertainty about the choice of the next symbol decreases when the number of known preceding symbols increases. From the non-increasing behaviour of the positive sequence $(H_n)_{n\in\mathbb{N}}$, we deduce the existence of the limit $\lim\limits_{n\to+\infty} H_n$. We have: $\forall n,\ H_n = V_{n+1} - V_n$ and $\sum\limits_{k=0}^{n-1} H_k = V_n$. By taking Cesàro means, we obtain:

$$\lim_{n\to+\infty} H_n = \lim_{n\to+\infty} \frac{V_n}{n}.$$

This limit is called the *measure-theoretic entropy* of the sequence u, it is the limit of the entropy per symbol of the choice of a block of length n, when n tends to infinity.

2.3 Variational Principle

What is the relation between the sequences $(H_n)_{n\in\mathbb{N}}$ and $(V_n)_{n\in\mathbb{N}}$? We have:

$$\forall n,\ nH_n \leq \sum_{k=0}^{n-1} H_k = V_n = \sum L(P(x_1 \cdots x_n)).$$

By concavity of the function L we get: $\forall n \geq 1$, $V_n \leq \log_d p(n)$. Hence the following proposition:

Proposition 2.5 *We have $H_n \leq \frac{\log_d(p(n))}{n}$, for all $n \geq 1$.*

We hence get:

$$\lim_{n\to+\infty} H_n = \lim_{n\to+\infty} \frac{V_n}{n} = H(u) \leq H_{top}(u) = \lim_{n\to+\infty} \frac{\log_d(p(n))}{n}.$$

This inequality is a particular case of a basic relationship between topological entropy and measure-theoretic entropy called the *variational principle* (for a proof see [53]).

The two limits $\lim_{n\to+\infty} H_n$ and $\lim_{n\to+\infty} \frac{\log_d(p(n))}{n}$ are distinct in general and the notion of measure-theoretic entropy for a sequence is more precise. But the sequences we are mostly dealing with here are *deterministic*, i.e., sequences with zero entropy. Therefore neither the metrical nor the topological entropy can distinguish between these sequences.

3 Symbolic Dynamical Systems

Recall some basic notions on symbolic dynamical systems. For a detailed introduction to the subject, see [57]. Let $\mathcal{A}$ denote a finite alphabet; here we work with the space $\mathcal{A}^{\mathbb{N}}$, whereas in Host's course it is $\mathcal{A}^{\mathbb{Z}}$.

Endow the set $\mathcal{A}^{\mathbb{N}}$ of all sequences with values in the finite set $\mathcal{A}$ with the product of discrete topologies on $\mathcal{A}$. This set is thus a compact space. The topology defined on $\mathcal{A}^{\mathbb{N}}$ is equivalent to the topology defined by the following metrics: for $x, y \in \mathcal{A}^{\mathbb{N}}$

$$d(x,y) = (1 + \inf\{k \geq 0;\ x_k \neq y_k\})^{-1}.$$

Two sequences are thus close to each other if their first terms coincide. The *cylinder* $[w]$, where $w = w_1 \ldots w_n$ belongs to $\mathcal{A}^n$, is the set of sequences of the form

$$[w] = \{x \in \mathcal{A}^{\mathbb{N}} |\ x_0 = w_1,\ x_1 = w_2, \ldots, x_{n-1} = w_n\}.$$

Cylinders are closed and open sets and span the topology.

The space $\mathcal{A}^{\mathbb{N}}$ is complete as a metric compact space. Let us deduce from this the existence of fixed points of substitutions. A *substitution* defined on the finite alphabet $\mathcal{A}$ is a map from $\mathcal{A}$ to the set of words defined on $\mathcal{A}$, denoted by $\mathcal{A}^*$, extended to $\mathcal{A}^*$ by concatenation, or in other words, a homomorphism of the free monoid $\mathcal{A}^*$ (see also [49] for a precise study of substitution dynamical systems).

Exercise 3.1 Let σ be a substitution and a be a letter such that $\sigma(a)$ begins by a and $|\sigma(a)| \geq 2$. Prove that there exists a unique sequence beginning with a satisfying $\sigma(u) = u$. This sequence is called a *fixed point* of the substitution.

For instance, the Fibonacci sequence is defined as the fixed point beginning with 1 of the following substitution

$$\sigma(1) = 10,\ \sigma(0) = 1.$$

Let T denote the following map defined on $\mathcal{A}^{\mathbb{N}}$, called the *one-sided shift*:

$$T((u_n)_{n \in \mathbb{N}}) = (u_{n+1})_{n \in \mathbb{N}}.$$

The map T is uniformly continuous, onto but not necessarily one-to-one on $\mathcal{A}^{\mathbb{N}}$.

Exercise 3.2 Recall that a sequence is said to be *recurrent* if every factor appears at least two times, or equivalently if every factor appears an infinite number of times in this sequence.

Prove that a sequence u is recurrent if and only if there exists a strictly increasing sequence $(n_k)_{k \in \mathbb{N}}$ such that

$$u = \lim_{k \to +\infty} T^{n_k} u.$$

Let u be a sequence with values in $\mathcal{A}$. Define $\overline{\mathcal{O}}(u)$ as the positive orbit closure of the sequence u under the action of the shift T, i.e., the closure of the set $\mathcal{O}(u) = \{T^n(u),\ n \geq 0\}$. The set $\overline{\mathcal{O}(u)}$ is a compact metric space, and thus complete. It is also T-*invariant*: $T(\overline{\mathcal{O}(u)}) \subset \overline{\mathcal{O}(u)}$. In other words T may be considered as acting on $\overline{\mathcal{O}}(u)$.

Exercise 3.3 1. Prove that

$$\overline{\mathcal{O}}(u) = \{x \in \mathcal{A}^{\mathbb{N}},\ L(x) \subset L(u)\},$$

where $L(x)$ denotes the set of factors of the sequence x.

2. Prove that u is recurrent if and only if T is onto on $\overline{\mathcal{O}}(u)$.

Let X be a non-empty compact metric space and T be a continuous map from X to X. The system (X, T) is called a *topological dynamical system.* For instance, $(\overline{\mathcal{O}(u)}, T)$ is a topological dynamical system. A topological dynamical system is called *minimal* if every closed T-invariant set E is either equal to the full set X or to the empty set.

Exercise 3.4 • Prove that (X, T) is minimal if and only if $X = \overline{\mathcal{O}(x)}$, for every element x of X.

- A sequence is said to be *uniformly recurrent* if every factor appears infinitely often and with bounded gaps (or, equivalently, if for every integer n, there exists an integer m such that every factor of u of length m contains every factor of length n). Prove that a sequence u is uniformly recurrent if and only if $(\overline{\mathcal{O}(u)}, T)$ is minimal. (If w is a factor of u, write

$$\overline{\mathcal{O}(u)} = \bigcup_{n \in \mathbb{N}} T^{-n}[w],$$

and conclude by a compactness argument.)

The following special case of the Daniell-Kolmogorov consistency theorem (see for instance [73]) establishes the existence of a certain probability measure on $(\overline{\mathcal{O}(u)}, T)$. A Borel probability measure μ defined on $(\overline{\mathcal{O}(u)}, T)$ is called *T-invariant* if $\mu(T^{-1}(B)) = \mu(B)$, for any Borel set B.

Theorem 3.5 *Let u be a sequence on $\mathcal{A} = \{1, \ldots, d\}$. Consider a family of maps $(p_n)_{n \geq 1}$, where p_n is a map from $\mathcal{A}^n$ to $\mathbb{R}$, such that*

- *for any word w in $\mathcal{A}^n$, $p_n(w) \geq 0$,*
- $\displaystyle\sum_{i=1}^{d} p_1(i) = 1,$

- *for any word $w = w_1 \dots w_n$ in $\mathcal{A}^n$, $p_n(w) = \sum_{i=1}^{d} p_{n+1}(w_1 \dots w_n i)$.*

Then there exists a unique probability measure μ on $\mathcal{A}^{\mathbb{N}}$ defined on the cylinders by $\mu([w_1 \dots w_n]) = p_n(w_1 \dots w_n)$.

Furthermore, if for any n and for any word $w = w_1 \dots w_n$ in $\mathcal{A}^n$,

$$p_n(w) = \sum_{i=1}^{d} p_{n+1}(iw_1 \dots w_n),$$

then this measure is T-invariant.

In particular, if all frequencies exist, then there exists a unique T-invariant probability measure which assigns to each cylinder the frequency of the corresponding factor. Moreover suppose the symbolic dynamical system $(\overline{\mathcal{O}(u)}, T)$ *uniquely ergodic*, i.e., there exists a unique T-invariant probability measure μ on this dynamical system. Thus a precise knowledge of the frequencies allows a complete description of the measure μ. For instance, a symbolic dynamical system obtained via the fixed point of a primitive substitution [49, 57], or via a Sturmian sequence is uniquely ergodic.

4 The Graph of Words

The Rauzy graph Γ_n of words of length n of a sequence on a finite alphabet $\mathcal{A}$ (of cardinality d) is an oriented graph (see, for instance, [58]), which is a subgraph of the de Bruijn graph of words[1] (see [32]). Its vertices are the factors of length n of the sequence and the edges are defined as follows: there is an edge from U to V if V follows U in the sequence, i.e., if there exists a word W and two letters x and y such that $U = xW$, $V = Wy$ and xWy is a factor of the sequence. There are $p(n+1)$ edges and $p(n)$ vertices, where $p(n)$ denotes the complexity function.

Exercise 4.1 Prove that the graphs of words of a sequence are always connected. Prove the following equivalence (see [61]):

[1]The de Bruijn graph of words corresponds to the graph of words of a sequence of maximal complexity ($\forall n, p(n) = d^n$) and was introduced by de Bruijn in order to construct circular finite sequences of length d^n with values in $\{0, 1, \dots, d-1\}$ such that every factor of length n appears once and only once: such a sequence corresponds to a Hamiltonian closed path in de Bruijn graph.

- the sequence u is recurrent,
- every factor of u appears at least twice,
- the graphs of words are strongly connected.

Let U be a vertex of the graph. Denote by U^+ the number of edges of Γ_n with origin U and by U^- the number of edges of Γ_n with end vertex U. In other words, U^+ (respectively U^-) counts the number of right (respectively left) extensions of U. Recall that

$$p(n+1) = \sum_{|U|=n} U^+ = \sum_{|U|=n} U^-,$$

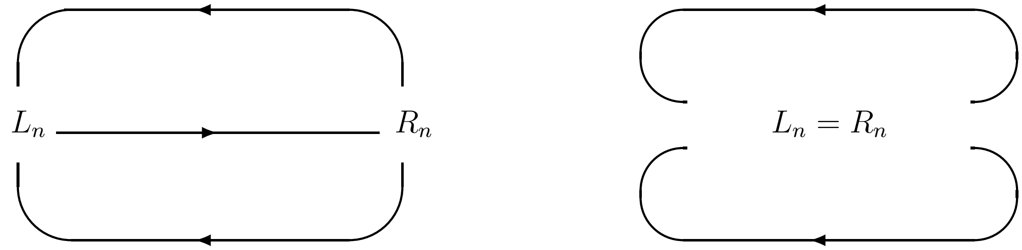
and thus

$$p(n+1) - p(n) = \sum_{|U|=n} (U^+ - 1) = \sum_{|U|=n} (U^- - 1).$$

Exercise 4.2 Recall that a Sturmian sequence is defined as a sequence of complexity function $p(n) = n+1$, for every positive integer n, and that it is recurrent (Exercise 2.2).

- For any positive integer n, prove that there exists a unique factor of length n having two right (respectively left) extensions: such a factor is called a *right (*respectively *left) special factor* (or also *expansive* factor) and is denoted from now on by R_n (respectively L_n).
- Prove that the graph of words Γ_n of a Sturmian sequence has the two following possible forms.

- Deduce from the morphology of the graph of words Γ_n that every Sturmian sequence is uniformly recurrent. One can first prove that every factor of a Sturmian sequence is a subfactor of a factor of the form R_n and then deduce from the morphology of the graph Γ_n that R_n appears with bounded gaps.



Exercise 4.3

- Prove that if the sequence u is uniformly recurrent and non-constant, then the graph Γ_n has no edge of the form $U \to U$, for n large enough.
- Suppose that the sequence u is uniformly recurrent. Prove that if the graph of words Γ_{n+1} is Hamiltonian (i.e., there exists a closed oriented path passing exactly once through every vertex), then the graph Γ_n is Eulerian (there exists a closed path passing exactly once through every edge) and that $U^+ = U^-$, for every vertex of Γ_n. Is the converse true?

4.1 The Line Graph

The *line graph* $D(\Gamma_n)$ of the graph of words Γ_n is defined as follows: its vertices are the edges of Γ_n (i.e., the factors of length $n+1$); given two vertices u and v in $D(\Gamma_n)$, there is an edge from u to v if the enpoint of the edge labelled u in Γ_n is the origin of the edge labelled v. It is easily seen that the edges of the line graph correspond to words of length $n+2$ such that their prefix and their suffix of length $n+1$ are factors of the sequence u. The line graph of Γ_n is thus a subgraph of Γ_{n+1}.

Exercise 4.4 Study the evolution of the graph of words from Γ_n to Γ_{n+1} for a Sturmian sequence by using the line graph. (Distinguish between the two possible forms of the graph).

Remark 4.5 In [61] Rote uses the graph of words and the line graph for the study of sequences of complexity $p(n) = 2n$, for every n (see also [41]). The study of the evolution of the graph of words for any Sturmian sequence is a very powerful method and contains all the information concerning the sequence: Arnoux and Rauzy have thus proved that every Sturmian sequence is generated by the composition of two substitutions (see [11]); one can also study the frequencies of factors of given length (see Section 8.2 and [14]) or covering numbers for rotations (see [15, 26]).

4.2 Graph and Frequencies

Let us see how to deduce from the morphology of the graphs of words results concerning the frequencies of factors. This follows an idea of Dekking who expressed the block frequencies for the Fibonacci sequence, by using the graph of words (see [34]).

In this section we restrict ourselves to sequences for which the frequencies exist. Observe that the function which associates to an edge labelled by xWy the frequency of the factor xWy is a *flow*. Indeed, it satisfies Kirchhoff's current law: the total current flowing into each vertex is equal to the total current leaving the vertex. This common value is equal to the frequency of the word corresponding to this vertex.

Lemma 4.6 *Let U and V be two vertices linked by an edge such that $U^+ = 1$ and $V^- = 1$. Then the two factors U and V have the same frequency.*

Proof. Write $U = xW$ and $V = Wy$, where x and y are letters. As $U^+ = 1$, U has a unique right extension y. Similarly, V has a unique left extension x. Thus $f(U) = f(Uy) = f(xWy) = f(xV) = f(V)$, where f denotes the frequency.

A *branch* of the graph Γ_n is a longest sequence of maximal length $(U_1, \ldots, U_m)$ of connected edges of Γ_n, possibly empty, satisfying

$$U_i^+ = 1, \quad \text{for} \quad i < m, \quad U_i^- = 1, \quad \text{for} \quad i > 1.$$

Therefore, the edges of a branch have the same frequency and the number of frequencies of factors of given length is bounded by the number of branches of the corresponding graph, as expressed below (see [18]).

Theorem 4.7 *For a recurrent sequence of complexity function $p(n)$, the frequencies of factors of given length, say n, take at most $3(p(n+1) - p(n))$ values.*

Proof. Let V_1 denote the set of factors of length n having more than one extension. In other words V_1 is the subset of vertices of the graph Γ_n defined as follows: $U \in V_1$ if and only if $U^+ \geq 2$. The cardinality of V_1 satisfies

$$\operatorname{card}(V_1) = \sum_{|U|=n,\ U^+ \geq 2} 1 \leq \sum_{|U|=n} (U^+ - 1) = p(n+1) - p(n).$$

Let V_2 denote the subset of vertices of the graph Γ_n defined as follows: $U \in V_2$ if and only if $U^+ = 1$ and if V denotes the unique vertex such that there is an edge from U to V in Γ_n, then $V^- \geq 2$. In other words, U belongs to V_2 if and only if $U = xW$, where x is a letter and where the factor W of the sequence u has a unique right extension but at least two left extensions. The cardinality of V_2 satisfies:

$$\mathrm{card}(V_2) \leq \sum_{V^- \geq 2} V^- = \sum_{V^- \geq 2} (V^- - 1) + \sum_{V^- \geq 2} 1 \leq 2(p(n+1) - p(n)).$$

Thus there are at most $3(p(n+1) - p(n))$ factors in $V_1 \cup V_2$.

Let U be a factor of length n belonging neither to V_1 nor to V_2: $U^+ = 1$ and the unique word V such that there is an edge from U to V in Γ_n satisfies $V^- = 1$. The two factors U and V thus have the same frequency. Now consider the path of the graph beginning at U and consisting of vertices which do not belong to V_1 nor to V_2. The last vertex of this path belongs to either V_1 or to V_2, and has the same frequency as U.

Remark 4.8 In fact we have proved that the frequencies of factors of length n take at most $p(n+1) - p(n) + r_n + l_n$ values, where r_n (respectively l_n) denotes the number of factors having more than one right (respectively left) extension.

We deduce from this result that if $p(n+1) - p(n)$ is uniformly bounded with n, the frequencies of factors of given length take a finite number of values. Indeed, using a theorem of Cassaigne quoted below (see [23]), we can easily state the following corollary.

Theorem 4.9 *If the complexity $p(n)$ of a sequence on a finite alphabet is sub-affine, i.e.,*

$$\exists (a, b),\ \forall n,\ p(n) \leq an + b,$$

then $p(n+1) - p(n)$ is bounded.

Corollary 4.10 *If a sequence has a sub-affine complexity then the frequencies of its factors of given length take a finite number of values.*

We will see in Section 7.6 (Theorem 7.26) that fixed points of uniform substitutions (i.e., substitutions such that the images of the letters have the same length) or fixed point of primitive substitutions have sub-affine complexities.

In particular, in the Sturmian case ($\forall n$, $p(n) = n + 1$), Theorem 4.7 implies that the frequencies of factors of given length of a Sturmian sequence take at most three values. We will come back to this in Section 6.2.

Note that Theorem 4.9 above does not hold anymore for the second-difference of the complexity $p(n + 2) + p(n) - 2p(n + 1)$ in the case of a sub-quadratic complexity (see the counterexample in [42]).

5 Special factors

The aim of this section is to introduce the notions of special and bispecial factors in order to evaluate the second-difference of the complexity, which is often easier to compute than the complexity itself. Indeed, in the case of a low complexity, the number of special factors and bispecial factors, which is low, is quite easy to evaluate. For a more detailed exposition, see [24].

Let u be a sequence with values in a two-letter alphabet $\mathcal{A}$. Let w^+ (respectively w^-) denote the number of right (respectively left) extensions of w. Recall that

$$p(n+1) - p(n) = \sum_{|w|=n} (w^+ - 1) = \sum_{|w|=n} (w^- - 1).$$

A factor is said to be a *left special factor* (respectively *right special factor*) if it has more than one left (respectively *right*) extension. We use the notation of [24], in which the case of a bigger-sized alphabet is also considered. A factor is said to be *bispecial* if it is both a right and a left special factor. More precisely, we distinguish three cases according to the cardinality $c(w)$ of $L(u) \cup \mathcal{A}w\mathcal{A}$, where $L(u)$ denotes the set of factors of the sequence u and w is a bispecial factor, the operation considered here being the concatenation. We have obviously $2 \leq c(w) \leq 4$.

- If $c(w) = 2$, then w is called *a weak bispecial factor*,
- if $c(w) = 3$, then w is called *an ordinary bispecial factor*,
- if $c(w) = 4$, then w is called *a strict bispecial factor.*

Exercise 5.1 Let $b_w(n)$ (respectively $b_s(n)$) denote the number of weak (respectively strict) bispecial factors of the sequence u of size n. Prove that the second-difference of the complexity is given by

$$p(n+2) + p(n) - 2p(n+1) = s(n+1) - s(n) = b_s(n) - b_w(n),$$

where $s(n) = p(n+1) - p(n)$.

Exercise 5.2 Consider the Fibonacci sequence. Prove that every factor w can be uniquely written as follows: $w = r_1\sigma(x)r_2$, where x is a factor, $r_1 \in \{\varepsilon, 0\}$, and $r_2 = 1$ if the last letter of w is 1, and $r_2 = 0$, otherwise. Prove by induction that the bispecial factors of the Fibonacci sequence are all ordinary. Deduce that this sequence is Sturmian.

Exercise 5.3 Let u be the Thue-Morse sequence defined as the fixed point beginning by 0 of the following substitution: $\sigma(0) = 01$ and $\sigma(1) = 10$. We will compute the complexity function of the Thue-Morse sequence in two ways.

1. Prove that every factor w can be written as follows: $w = r_1\sigma(x)r_2$, where x is a factor and $r_i \in \{\varepsilon, a, b\}$. If $|w| \geq 5$, then this decomposition is unique.

2. Prove that $p(2n) = p(n) + p(n+1)$ and that $p(2n+1) = 2p(n+1)$, for $n \geq 1$. Give an expression for the complexity function (see for instance [20]).

3. Find by induction the expressions of $b_s(n)$, $b_o(n)$ and $b_w(n)$ by studying the small length cases. Deduce an expression for the complexity.

6 Sturmian sequences

Sturmian sequences have received considerable attention in the literature. We refer the reader to the impressive bibliography of [21]. A recent account on the subject can also be found in [12].

6.1 A Particular Coding of Rotations

We introduce a large family of Sturmian sequences obtained by coding the orbit of a point of the unit circle under an irrational rotation.

Let $\{x\}$ denote as usual the fractional part of x (i.e., if $\lfloor x \rfloor$ denotes the largest integer not exceeding x, then $\{x\} = x - \lfloor x \rfloor$). Let α be an irrational number in $]0, 1[$ and consider the rotation R_α of angle α defined on the unit circle (identified with $[0, 1[$ or with the unidimensional torus $\mathbb{R}/\mathbb{Z}$): we have $R_\alpha(x) = x + \alpha \ (mod\ 1)$. The *positive orbit* of a point x of the unit circle under the rotation by angle α is the set of points $\{\{\alpha n + x\},\ n \geq 0\}$. We code the information concerning the orbit of x by a binary sequence. The *coding* of the orbit of x under the rotation by angle α with respect to the partition $\mathcal{P} = \{[0, 1-\alpha[, [1-\alpha, 1[\}$ is the sequence $u = (u_n)_{n\in\mathbb{N}}$ defined on the alphabet $\{0, 1\}$ as follows:

$$u_n = 1 \Leftrightarrow \{x + n\alpha\} \in [0, 1-\alpha[.$$

We could also choose to code the orbit of the rotation with respect to the partition $\mathcal{P}' = \{]0, 1-\alpha],]1-\alpha, 1]\}$. As α is irrational, the two sequences obtained by coding with respect to $\mathcal{P}$ or to $\mathcal{P}'$ are ultimately equal. The results stated below on codings with respect to $\mathcal{P}$ are obviously true for $\mathcal{P}'$.

Let $I_0 = [0, 1-\alpha[$ and $I_1 = [1-\alpha, 1[$. A finite word $w_1 \cdots w_n$ defined on the alphabet $\{0, 1\}$ is a factor of the sequence u if and only if there exists an integer k such that

$$\{x + k\alpha\} \in I(w_1, \ldots, w_n) = \bigcap_{j=0}^{n-1} R^{-j}(I_{w_{j+1}}).$$

As α is irrational, the sequence $(\{x+n\alpha\})_{n\in\mathbb{N}}$ is dense in the unit circle, which implies that $w_1 w_2 \ldots w_n$ is a factor of u if and only if $I(w_1, \ldots, w_n) \neq \emptyset$. In particular, the set of factors does not depend on the initial point x of this coding. Furthermore, one can check that the sets $I(w_1, \ldots, w_n)$ are connected and are bounded by the points $\{k(1-\alpha)\}$, for $0 \leq k \leq n-1$. There are $n+1$ such intervals (α is irrational) and thus $n+1$ factors of length n: the sequence u is therefore Sturmian.

Remark 6.1 In the general case of a coding of an irrational rotation with respect to a partition of the unit circle in l intervals, the complexity has the form $p(n) = an + b$, for n large enough (see [2]). Conversely, every sequence of ultimately affine complexity is not necessarily obtained as a coding of rotation. See for instance [61], where Rote studies the case of sequences of complexity $p(n) = 2n$, for every n. However, if the complexity of a sequence u has the form $p(n) = n + k$, for n large enough, then u is the image of a Sturmian sequence by a morphism, up to a prefix of finite length (see for instance [2, 25, 36]).

Exercise 6.2 Let u be a coding of the rotation by angle α defined as above. Prove that the orbit closure $\overline{\mathcal{O}(u)}$ is the set of sequences obtained by coding every point of the unit circle with respect to the partition $\mathcal{P}$ or $\mathcal{P}'$ under the rotation by angle α. (Use the fact that the set of factors of a coding depends neither on the initial point x nor on the choice of the partition but only on the angle of the rotation). Deduce the minimality of $(\overline{\mathcal{O}(u)}, T)$.

Now a natural question is whether all Sturmian sequences are obtained by coding a rotation as defined above. The answer is yes and is due to Morse and Hedlund (see [55]): this shows that in this case of low disorder, one can give a geometrical description of sequences defined up to their complexity function. When the complexity grows, this becomes much more difficult (see for instance [11] for a geometrical representation of a particular class of sequences of complexity $2n + 1$).

Theorem 6.3 (Hedlund and Morse) *A sequence u is Sturmian if and only if there exists an irrational α in $]0, 1[$ and x on the unit circle such that u is the coding of the orbit of x under the rotation by angle α with respect to one of the partitions* $\{[0, 1-\alpha[, [1-\alpha, 1[\}$ *or* $\{]0, 1-\alpha],]1-\alpha, 1]\}$.

Sturmian sequences are also characterized by the following properties.

- Sturmian sequences are exactly the non-ultimately periodic balanced sequences over a two-letter alphabet. A sequence is balanced if the difference between the number of occurrences of a letter in any two factors of the same length is bounded by one in absolute value.

- Sturmian sequences are codings of trajectories of irrational initial slope in a square billiard obtained by coding horizontal sides by the letter 0 and vertical sides by the letter 1.

- One can also consider Sturmian sequences as approximations of a line of irrational slope in the upper half-plane.

The last three properties can be easily deduced from the representation by a rotation: they are just geometrical reformulations; the first characterization of Sturmian sequences in terms of the balance property is much more difficult to establish and is an important step in the proof of Theorem 6.3.

6.2 Frequencies of Factors of Sturmian Sequences

We now consider properties of frequencies of factors of Sturmian sequences.

The frequency of the factor $w_1 \dots w_n$ exists and is equal to the density of the set

$$\{k \mid \{x + k\alpha\} \in I(w_1, \dots, w_n)\},$$

which is equal to the length of $I(w_1, \dots, w_n)$, by uniform distribution of the sequence $(\{x + n\alpha\})_{n\in\mathbb{N}}$. The lengths of these intervals are equal to the frequencies of factors of length n.

But we deduce from Theorem 4.7 the following result.

Theorem 6.4 *The frequencies of factors of given length of a Sturmian sequence take at most three values.*

Theorem 6.4 implies that the lengths of the intervals $I(w_1, \dots, w_n)$, and thus the lengths of the intervals obtained by placing the points $0, \{1 - \alpha\}, \dots, \{n(1-\alpha)\}$ on the unit circle, take at most three values. We thus have proved the following classical result in Diophantine approximation, called the *three-distance theorem* (see the survey [3]). In fact, this point of view and more precisely, the study of the evolution of the graphs of words with respect to the length n of the factors, allows us to give a proof of the most complete version of the three distance theorem, i.e., to express the exact number of factors having each of the three frequencies and the frequencies themselves (for more details, the reader is referred to [14]).

The three distance theorem was initially conjectured by Steinhaus and proved by V. T. Sós (see [67, 68, 69, 70]).

Theorem 6.5 *Let $0 < \alpha < 1$ be an irrational number and n a positive integer. The points $\{i\alpha\}$, for $0 \le i \le n$, partition the unit circle into $n+1$ intervals, the lengths of which take at most three values, one being the sum of the other two.*

More precisely, let $(\frac{p_k}{q_k})_{k\in\mathbb{N}}$ and $(c_k)_{k\in\mathbb{N}}$ be the sequences of the convergents and partial quotients associated to α in its continued fraction expansion (if $\alpha = [0, c_1, c_2, \dots]$, then $\frac{p_n}{q_n} = [0, c_1, \dots, c_n]$). Let $\eta_k = (-1)^k(q_k\alpha - p_k)$. Let n be a positive integer. There exists a unique expression for n of the form

$$n = mq_k + q_{k-1} + r,$$

with $1 \le m \le c_{k+1}$ and $0 \le r < q_k$. Then, the circle is divided by the points $0, \{\alpha\}, \{2\alpha\}, \dots, \{n\alpha\}$ into $n+1$ intervals which satisfy:

- *$n+1-q_k$ of them have length η_k (which is the largest of the three lengths),*

- *$r+1$ have length $\eta_{k-1}-m\eta_k$,*

- *$q_k-(r+1)$ have length $\eta_{k-1}-(m-1)\eta_k$.*

7 Automatic Sequences

The aim of this section is to introduce automatic sequences and to study them in connection with properties of algebraicity of formal power series over a finite field.

Let k be an integer greater than or equal to 2. Recall the definition of a finite complete deterministic k-automaton (also called 2-*tape automaton or transducer*). For more details, see the courses of Frougny, see [30] or see the surveys [5, 9]. A k-automaton is represented by a directed graph defined by:

- a finite set of *states* $\mathcal{S}=\{i=a_1,a_2,\cdots,a_d\}$, one of these states i is called the *initial state*;

- k *transition maps* (or "edges") from the set of states $\mathcal{S}$ into itself, denoted by $0,1,\cdots,k-1$;

- a set Y and a map φ from $\mathcal{S}$ into a set Y, called *output function* or *exit map*.

A sequence $(u(n))_{n\in\mathbb{N}}$ with values in Y is called k-*automatic* (it is also called a k-*uniform tag sequence* or a k-*recognizable* sequence) if it is generated by a k-automaton as follows: let $\omega(n)$ be the base k expansion of the integer n; starting form the initial state one feeds the automaton the sequence $\omega(n)$, the digits being read in growing order of powers; after doing this the automaton is in the state $a(n)$ (the automaton is said to be fed in *reverse* reading.) Then put $u(n)=\varphi(a_f)$. One can similarly give another definition of k-automaticity by reading the digits in the reverse order, i.e., by starting with the most significant digit, but these two notions are easily seen to be equivalent. (The automaton is said to be fed in *direct* reading.)

7.1 Automata and Transcendence

Let $\mathbb{F}_q$ be the finite field with q elements, and let p denote its characteristic. We thus have p prime and $q = p^s$, where s is a nonzero natural integer. The field $\mathbb{F}_q((1/X))$ of *Laurent formal power series* with coefficients in $\mathbb{F}_q$ is the field of formal power series series of the form

$$u(-d)X^d + \cdots + u(0) + u(1)X^{-1} + \cdots,$$

where the coefficients $u(i)$ belong to $\mathbb{F}_q$. Similarly, we denote by $\mathbb{F}_q((X))$ the field of formal power series of the form

$$u(-d)X^{-d} + \cdots + u(0) + u(1)X + \cdots$$

A formal power series F is called *algebraic over* $\mathbb{F}_q(X)$ if there exists a non-trivial polynomial P with coefficients in $\mathbb{F}_q(X)$ such that $P(F) = 0$. In the converse case, F is called *transcendental* over $\mathbb{F}_q(X)$.

The following theorem due to Christol, Kamae, Mendès France and Rauzy (see [27, 28] and also [5]) gives a necessary and sufficient condition of algebraicity for a formal power series with coefficients in a finite field.

Theorem 7.1 (Christol, Kamae, Mendès France and Rauzy) *Let $u = (u(n))_{n\in\mathbb{N}}$ be a sequence with values in $\mathbb{F}_q$. The following conditions are equivalent:*

1. *the formal power series $\sum_{n\geq 0} u(n)X^n$ is algebraic over the field $\mathbb{F}_q(X)$,*
2. *the q-kernel $N_q(u)$ of the sequence u is finite, where $N_q(u)$ is the set of subsequences of the sequence $(u(n))_{n\in\mathbb{N}}$ defined by*

   $$N_q(u) = \{(u(q^k n + r))_{n\in\mathbb{N}};\ k \geq 0;\ 0 \leq r \leq q^k - 1\},$$

3. *the sequence u is q-automatic,*
4. *the sequence u is the image by a letter-to-letter projection of a fixed point of a substitution of constant length.*

The last equivalence is due to Cobham (see [30]) and the equivalence between 2 and 3 dates back to Eilenberg in [39].

Remark 7.2

- The same theorem holds by considering a series $\sum u(n)X^{-n}$ in $\mathbb{F}_q((1/X))$ with the same definition for the q-kernel. Indeed, $\sum u(n)X^{-n}$ is transcendental over $\mathbb{F}_q(X)$ if and only if $\sum u(n)X^n$ is transcendental over $\mathbb{F}_q(X)$.
- The proof of the equivalence between 2 and 3, and between 2 and 4, is constructive.
- The equivalences between 2, 3 and 4 is true for any sequence taking its values in a set of cardinality q, where q is not necessarily a power of a prime.
- The notion of p-automaticity can also be expressed as follows in terms of first-order logic: a sequence is generated by a p-substitution if and only if it is p-definable (it can be defined in the theory $(\mathbb{N}, +, V_p)$, where V_p is the function "valuation" that associates to x the highest power of p that divides x (or 1 if $x = 0$)). For more details, the reader is referred to the survey [22].
- Durand has extended in [37] this characterization of automatic sequences to the case of uniformly recurrent sequences generated by substitutions of non-constant length, by introducing the notion of return words.

Exercise 7.3

1. Show that a sequence is p-automatic if and only if it is p^k-automatic for any non-zero power of the prime p.
2. Consider the Thue-Mose sequence $(S_2(n))_{n\in\mathbb{N}}$ defined over $\mathbb{F}_2$, where $S_2(n)$ is the sum modulo 2 of the coefficients of the digits in the base 2 expansion of the integer n. Build a 2-automaton generating the Thue-Morse sequence.
3. Build a d-automaton generating the characteristic sequence of the set of powers of a fixed positive integer d.
4. Build a d-automaton generating the characteristic sequence of the set of integers divisible by d.
5. Build a 2-automaton generating the characteristic sequence of the set of integers of base 2-expansion of the form 1^n0^m1, for $n, m > 0$ and $n + m$ odd.

6. Prove that the characteristic sequence of the set of integers of base 2-expansion of the form $1^n0^{n+1}1$, for $n > 0$, is not 2-automatic.

7.2 Applications

In this section we give some applications of Theorem 7.1. The first two are easy consequences of the theorem.

- Let $\sum u(n)X^n$ be an algebraic formal power series. Let a and b be two natural integers. The series $\sum_{n\geq 0} u(an+b)X^n$ is algebraic.

- Let $p \geq 2$ be an integer. Let $S_p(n)$ be the sum modulo p of the digits of n in base p. The series $\sum_{n\geq 0} S_p(n)X^n$ is algebraic.

Remark 7.4 The series $\sum_{n\geq 0} S_p(n^2)X^n$ is transcendental. More precisely, let R be a polynomial with coefficients in $\mathbb{Q}$ such that $R(\mathbb{N}) \subset \mathbb{N}$. The formal power series $\sum_{n\geq 0} S_p(R(n))X^n$ is algebraic over $\mathbb{F}_p$ if and only if the degree of R is less than or equal to 1 (see [4]).

The **Hadamard product** of two series $\displaystyle\sum u(n)X^n$ and $\displaystyle\sum v(n)X^n$ is defined as the series $\displaystyle\sum u(n)v(n)X^n$. By considering the the notion of q-kernel, we easily deduce the following.

Theorem 7.5 *The Hadamard product of two algebraic formal power series with coefficients in a finite field is algebraic.*

Note that the following theorem, due to Cobbham [29], produces more examples of transcendental series. We will not give here the proof of this theorem, which is rather difficult.

Theorem 7.6 *Let u be a sequence which is both k-automatic and k'-automatic. If k and k' are multiplicatively independent (i.e., if $\frac{\log(k)}{\log(k')}$ is irrational), then the sequence u is ultimately periodic.*

We deduce from this theorem the following result of transcendence, which answers, in the case of formal power series with values in a finite field, an analogous question attributed to Mahler and still open: given $(u(n))_{n\in\mathbb{N}}$ a binary sequence such that the series $\sum u(n)2^{-n}$ and $\sum u(n)3^{-n}$ are algebraic over $\mathbb{Q}$, is this sequence ultimately periodic?

Theorem 7.7 *Let $(u(n))_{n\in\mathbb{N}}$ be a binary sequence such that $\sum u(n)X^n$ considered as an element of $\mathbb{F}_2((X))$ and $\sum u(n)X^n$ considered as an element of $\mathbb{F}_3((X))$ are algebraic. Then, this sequence is ultimately periodic, i.e., both series are rational.*

Another application of Cobham's Theorem is the following (see for instance [7]).

Theorem 7.8 *Let r be an integer greater than or equal to 2. The series $\sum_{k=0}^{+\infty} X^{r^k}$ is algebraic over $\mathbb{F}_q(X)$ if and only if r is a power of p.*

Proof. Write
$$\sum_{k=0}^{+\infty} X^{r^k} = \sum_{n\geq 1} u(n)X^n,$$
where $u = (u(n))_{n\in\mathbb{N}}$ is the characteristic sequence of the set of powers of r. The series $\sum_{k=0}^{+\infty} X^{r^k}$ is algebraic over $\mathbb{F}_q(X)$ if and only if the sequence u is p-automatic. But it is easily seen that the sequence u is r-automatic (Exercise 7.3) and not ultimately periodic. Hence the series $\sum_{k=0}^{+\infty} X^{r^k}$ is algebraic over $\mathbb{F}_q(X)$ if and only if r is a power of p.

Remark 7.9 The formal power series $\sum u(n)X^n$ belongs to $\mathbb{F}_q(X)$ if and only if the sequence $(u(n))_{n\in\mathbb{N}}$ is ultimately periodic. Note that in the real case we just have the following implication: if the sequence $(u(n))_{n\in\mathbb{N}}$ is ultimately periodic, then the series $\sum u(n)X^n$ belongs to $\mathbb{Q}(X)$. The rational series $\sum nX^n$ shows that the converse is not true.

It is natural to consider the connections between transcendence in the real case and in positive characteristic. Indeed, a formal power series is algebraic in positive characteristic if the sequence of its coefficients has some kind of order, whereas irrational algebraic real numbers cannot have a too regular expansion. Loxton and van der Poorten [51] have conjectured the following (this conjecture is often quoted as a theorem, but there seems to be a gap in the proof).

Conjecture 7.10 *If the sequence of the coefficients in the base q-expansion of a real number is automatic, then this number is either rational or transcendental.*

This conjecture illustrates, like Cobhams's Theorem, the fact that transcendence deeply depends on the frame in which it is considered.

7.3 The Multidimensional Case

The Christol, Kamae, Mendès France and Rauzy theorem can be generalized to the multidimensional case. In particular, Salon has generalized this theorem to the case of a formal power series with a finite number of indeterminates and with coordinates in a finite field, say $\sum_{n_i \geq 0} u(n_1, n_2, \cdots, n_d) X_1^{n_1} \cdots X_d^{n_d}$ (see for instance [62] and [63]). The generalization of the q-kernel is given in this case by:

$$N_q(u(n_1, n_2, \cdots, n_d)) = \{u(q^k n_1 + r_1, q^k n_2 + r_2, \cdots, q^k n_d + r_d),$$

$$k \geq 0,\ 0 \leq n_i \leq q^k - 1,\ \text{for}\ \ i = 1\ \text{to}\ \ d\}.$$

Recall that a formal power series $F = \sum_{n_i \geq 0} u(n_1, n_2, \cdots, n_d) X_1^{n_1} \cdots X_d^{n_d}$ is said to be algebraic over $\mathbb{F}_q(X_1, X_2, \cdots, X_d)$ if there exists a nontrivial polynomial P with coefficients in $\mathbb{F}_q(X_1, X_2, \cdots, X_d)$ such that $P(F) = 0$.

The notions of automaton and substitution can also be generalized in two dimensions. A two-dimensional substitution of constant length l associates to each letter a square array of letters of size (l, l). A two-dimensional k-automaton is defined similarly as a one-dimensional k-automaton but in this case the edges are labelled by pairs of integers in $[0, k-1]^2$. A sequence $(u(m, n))_{(m,n) \in \mathbb{Z}^2}$ is generated by the automaton $\mathcal{A}$ by reading simultaneously the digits of the base k expansions of m and n, the shortest expansion being completed with leading zeroes to get two strings of symbols of the length of the longest expansion (without leading zeroes).

We thus have the following theorem due to Salon (see [62] and [63]).

Theorem 7.11 *The series $\sum u(n_1, n_2, \cdots, n_d) X_1^{n_1} \cdots X_d^{n_d}$ is algebraic over $\mathbb{F}_q(X_1, X_2, \cdots, X_d)$ if and only if the q-kernel of the sequence u is finite.*

The following results are easy applications of this theorem.

- Let u be an algebraic formal power series. The double formal power series $\sum u(m+n) X^m Y^n$ is algebraic.

- Let $\sum u(m,n)X^mY^n$ be algebraic. Let a,b,c,d be four integers. The series $\sum u(am+b,cn+d)X^mY^n$ is algebraic.

Exercise 7.12 Consider the substitution $\sigma : \{0,1\} \to \{0,1\} \times \{0,1\}$ defined by

$$\sigma(0) = \begin{matrix} 00 \\ 00 \end{matrix},$$

$$\sigma(1) = \begin{matrix} 11 \\ 10 \end{matrix}.$$

Prove that the double-sequence fixed point of this substitution generated by the successive images of 1 is equal to Pascal's triangle reduced modulo 2. Find the substitution generating Pascal's triangle modulo a prime p.

Remark 7.13 The double-sequence corresponding to Pascal's triangle modulo an integer d is automatic if and only if d is a power of a prime (see [10]).

7.4 Application to Diagonals

Another interesting consequence of this generalization to the multidimensional case is given by the following results. The *diagonal* of a double formal power series $\sum u(m,n)X^mY^n$ is defined as the series $\sum u(n,n)X^n$.

Theorem 7.14 *The diagonal of an algebraic formal power series with coefficients in a finite field is algebraic.*

Proof. Consider either the notion of q-kernel or the one-dimensional substitution defined by associating to each letter the "diagonal" of the square array of letters associated by the initial substitution.

Theorem 7.14 was first proved by Furstenberg in [46] and can be compared to the following theorem, also due to Furstenberg.

Theorem 7.15 *A series with coefficients in a finite field is algebraic if and only if there exists a rational double formal power series such that the initial series is the diagonal of this double series.*

Observe that this result still holds on $\mathbb{C}$ when considering two-indeterminate series but is false for series involving more indeterminates. For a survey on the subject, the reader is referred to [6].

Exercise 7.16 1. Conisder the Thue-Morse sequence $(S_2(n))_{n\in\mathbb{N}}$. Prove that the series $\sum S_2(n)X^n$ is the diagonal of the rational function in $\mathbb{F}_2(X,Y)$ defined by $Y(1+Y(1+XY)+X(1+XY)^{-2})^{-1}$.

2. Let $(u(n))_{n\in\mathbb{N}}$ be a sequence with values in the finite set X. Prove that if $\sum u(n)X^n$ is algebraic, then, for any $x \in X$, $\sum\limits_{v(n)=x} X^n$ is algebraic.

3. Let $(v(n))_{n\in\mathbb{N}}$ be the characteristic sequence of the set of powers of the prime p. Prove that the series $\sum v(n)X^n$ is the diagonal of the rational fraction of $\mathbb{F}_p(X,Y)$ defined by $X/(1-(X^{p-1}+Y))$.

4. Let $(w(n))_{n\in\mathbb{N}}$ be the characteristic sequence of the set of integers of base 2-expansion of the form 1^n0^m1, for $n,m>0$ and $n+m$ odd. Let $(x(n))_{n\in\mathbb{N}}$ be the characteristic sequence of the set of integers of base 2-expansion of the form $1^n0^{n+1}1$, for $n>0$. Let $(y(n))_{n\in\mathbb{N}}$ be the characteristic sequence of the set of squares. Prove that the Hadamard product of the series $\sum_{n\geq 0} w(n)X^n$ and $\sum_{n\geq 0} y(n)X^n$ is equal to $\sum_{n\geq 0} x(n)X^n$. Deduce from Exercice 7.3 that the sequence $(y(n))_{n\in\mathbb{N}}$ is not 2-automatic (see [60] and the survey [71]).

Remark 7.17 Christol, Kamae, Mendès France and Rauzy's theorem can also be extended to a general field of positive characteristic, which is not necessarily finite. Such a generalization is due to Sharif and Woodcock (see [66]) and Harase (see [48]). The results on the Hadamard poduct and on the diagonal still hold in this context. We can deduce namely the following corollary proved first by Deligne in [35].

Corollary 7.18 *The Hadamard product of two algebraic formal power series with coefficients in a field of positive characteristic is algebraic. The diagonal of an algebraic formal power series with coefficients in a field of positive characteristic is algebraic.*

Remark 7.19 Fresnel, Koskas and de Mathan have also generalized effectively Christol, Kamae, Mendès France and Rauzy's Theorem to the case of an infinite ground field [44].

7.5 Transcendence of the Bracket Series

The purpose of this section is to prove the following result, which gives us an example of application of the Christol, Kamae, Mendès France and Rauzy theorem. This result was first proved by Wade in [72]; the proof below is due to Allouche (see [7] and also [52]).

Theorem 7.20 *The series* $\sum_{k=1}^{+\infty} \frac{1}{[k]}$ *is transcendental over* $\mathbb{F}_q(X)$.

This proof makes use of the following consequence of the Christol, Kamae, Mendès France and Rauzy theorem.

Proposition 7.21 *Let* $(u(n))_{n\in\mathbb{N}}$ *be a sequence with values in* $\mathbb{F}_q$. *If the series* $\sum_{n\geq 0} u(n)X^{-n}$ *is algebraic over* $\mathbb{F}_q(X)$, *then the sequence* $(u(q^n - 1))_{n\in\mathbb{N}}$ *is ultimately periodic.*

Proof. Suppose that the series $\sum_{n\geq 0} u_n X^{-n}$ is algebraic; the sequence $u = (u(n))_{n\in\mathbb{N}}$ is thus q-automatic. Let $\mathcal{A}$ denote a finite q-automaton which generates the sequence u. The subsequence $(u(q^n - 1))_{n\in\mathbb{N}}$ is obtained by reading in the automaton $\mathcal{A}$ strings of ones. As the number of states of $\mathcal{A}$ is finite, a sufficiently long string of ones meets twice the same state. The sequence of states met is thus ultimately periodic, which implies that $(u(q^n - 1))_{n\in\mathbb{N}}$ is also ultimately periodic.

Remark 7.22 Let $\overline{UT^nV}$ denote, for any natural integer n, the integer of base-q expansion UT^nV, where U, T, V are words defined over $\{0, 1, \cdots, q-1\}$. We can similarly prove that if the series $\sum_{n\geq 0} u_n X^{-n}$ is algebraic, then the sequence $(u(\overline{UT^nV}))_{n\in\mathbb{N}}$ is ultimately periodic. This result corresponds to the classical *pumping lemma* in automata theory.

Exercise 7.23 Give another proof of Proposition 7.21, by using the notion of q-kernel.

Proof. Let us prove Theorem 7.20. We have:

$$\begin{aligned}\sum_{k\geq 1}\frac{1}{[k]} &= \sum_{k\geq 1}\frac{1}{(X^{q^k}-X)} = \sum_{k\geq 1}\frac{1}{X^{q^k}(1-(\frac{1}{X})^{q^k-1})} \\ &= \sum_{k\geq 1}\frac{1}{X^{q^k}}\sum_{j\geq 0}(\frac{1}{X})^{j(q^k-1)} = \frac{1}{X}\sum_{k\geq 1,\ j\geq 0}(\frac{1}{X})^{(j+1)(q^k-1)} \\ &= \frac{1}{X}\sum_{k\geq 1,\ j\geq 1}\frac{1}{X^{j(q^k-1)}} = \frac{1}{X}\sum_{n\geq 1}a(n)X^{-n},\end{aligned}$$

where $a(n)$ is the number (modulo the characteristic p) of decompositions of the integer n as $n = j(q^k - 1)$, with $k \geq 1$ and $j \geq 1$, i.e.,

$$a(n) = \sum_{k\geq 1,\ (q^k-1)|n} 1.$$

Clearly the series $\displaystyle\sum_{k\geq 1}\frac{1}{[k]}$ is transcendental over $\mathbb{F}_q(X)$ if and only if the series $X\sum_{k\geq 1}\frac{1}{[k]}$ is transcendental. Suppose that the series $\sum_{n\geq 1}a(n)X^{-n}$ is algebraic over $\mathbb{F}_q(X)$. This implies that the sequence $(a(n))_{n\in\mathbb{N}}$ is q-automatic and in particular that the subsequence $a((q^n-1))_{n\in\mathbb{N}}$ is ultimately periodic. This assertion leads to a contradiction.

Indeed, it is easily seen that $q^k - 1$ divides $q^n - 1$ if and only if k divides n. We thus have

$$a(q^n-1) = \sum_{k\geq 1,\ (q^k-1)|(q^n-1)} 1 = \sum_{k\geq 1,\ k|n} 1.$$

The subsequence $a((q^n-1))_{n\in\mathbb{N}}$ is supposed to be ultimately periodic. Thus there exist $n_0 \geq 1$ and $T \geq 1$ such that:

$$\forall n\geq n_0,\ \sum_{k\geq 1,\ k|n} 1 = \sum_{k\geq 1,\ k|n+T} 1 \mod p.$$

This implies

$$\forall n\geq n_0,\ \forall \mu\in\mathbb{N},\ \sum_{k\geq 1,\ k|n} 1 = \sum_{k\geq 1,\ k|n(1+\mu T)} 1 \mod p.$$

By the primes in arithmetic progression theorem, there exists a prime number $\omega > n_0$ and such that $\omega = 1 + \mu T$ for some integer μ. Then for $n = \omega$

$$\sum_{k \geq 1,\ k|\omega} 1 = \sum_{k \geq 1,\ k|\omega^2} 1 \mod p,$$

i.e.,

$$2 = 3 \mod p,$$

which is the desired contradiction.

Exercise 7.24 Let us give another proof of this result which does not involve the primes in arithmetic progression theorem. This very nice proof is due to Mendès France and Yao ([52]).

1. Prove that for any positive integers u, v, w, the number $q^w - 1$ divides $q^u(q^v - 2) + 1$ if and only if w divides the greatest common divisor of u and v.
2. Define the sequence $a_u = (a(q^u n + 1))_{n \in \mathbb{N}}$, for a fixed positive integer u. Let u, v be two distinct positive integers. Let h be the smallest integer such that h divides u and h does not divide v. Prove that $a_u(q^h - 2) - a_v(q^h - 2) \equiv 1$.
3. Deduce from this the transcendence of $\sum_{n \geq 1} a(n) X^{-n}$.
4. Prove similarly the following theorem [52].

Theorem 7.25 *Let $(n_k)_{k \in \mathbb{N}}$ be a sequence of elements of $\mathbb{F}_q$ which is not ultimately equal to zero. The formal power series $\sum_{k \geq 1} \frac{n_k}{[k]}$ is transcendental over $\mathbb{F}_q(X)$.*

7.6 Complexity and Frequencies

Recall that automatic sequences (and more generally substitutive sequences) have a strong underlying structure with respect to the complexity, as expressed by the following properties.

Theorem 7.26 *1. The complexity of a fixed point of a primitive substitution [57] or of a fixed point of a substitution of constant length [30] satisfies*

$$\forall n,\ p(n) \leq Cn.$$

2. The complexity of a substitution fixed point satisfies [40, 56]

$$\forall n,\ p(n) \leq Cn^2.$$

This disproves the automaticity of a high complexity sequence. Note that automatic sequences have rather similar complexity functions but very dissimilar spectral properties (see for instance [57]).

Let us review some results on the frequencies of factors of a fixed point of a substitution of constant length. For general results in the case of primitive substitutions, see [49, 57]. Recall in particular that for such sequences the frequencies exist and are strictly positive.

The following equirepartition result holds for frequencies of factors of primitive substitutions of constant length (see [35]).

Theorem 7.27 *Let σ be a primitive substitution of constant length. There exist $C_2 > C_1 > 0$ such that*

$$C_1 \leq nf(w) \leq C_2,$$

for all $n \geq 1$ and all factors w of length n.

One proves the following result by applying the properties of stochastic matrices to the n-th power of the matrix of the substitution divided by l^n, where l denotes the length of the substitution (see [30, 49, 57]).

Theorem 7.28 *If the frequencies of the factors of an automatic sequence exist, they are rational. Furthermore, if the corresponding substitution is primitive, then the frequencies exist.*

In particular, a Sturmian sequence cannot be automatic.

Cobham gives more precise results for automatic sequences with a letter of 0 frequency [30]. In particular we have the following result which can be considered as a criterion for testing the automaticity.

Theorem 7.29 *Let u be an automatic sequence and let a be a letter which occurs in u infinitely often with 0 frequency. Then the gaps beween successive occurrences of a in u satisfy*

$$lim\ sup_{n\to+\infty} \frac{\alpha_{n+1}}{\alpha_n} > 1,$$

where α_n denotes the index of the n-th occurrence of a in the sequence u

Exercise 7.30 Apply Theorem 7.29 to the characteristic sequence of the set of squares (see also Exercise 7.16).

Remark 7.31 Yao produces in [74] non-automaticity criteria motivated by a transcendence criterion due to de Mathan (see [33]). Note that Koskas gives a proof using automata of this last criterion in [50].

8 Conclusion

Let us conclude this lecture by discussing the connections between Sturmian sequences and automatic sequences (Section 8.1), and more generally between sequences of sub-affine complexity and substitutions (Section 8.2).

8.1 Automaticity and Sturmian sequences

Shallit introduces in [64] a *measure of automaticity* of a sequence u over a finite alphabet: the k-automaticity of u, $A_u^k(n)$, is defined as the smallest possible number of states in any deterministic finite automaton which generates the prefix of size n of this sequence. (By Christol, Kamae, Mendès France and Rauzy's theorem a sequence has a finite measure of automaticity if and only if this sequence is automatic.) This measure tells quantitatively how "close" a sequence is to being k-automatic.

Remark 8.1 The automaton is fed with the digits i, starting from the *least significant digit*: there are languages of low automaticity whose mirror image has high automaticity (see [47]).

A sequence can fail to be k-automatic if all the sequences in the k-kernel are distinct. A sequence is said to be *maximally diverse* if the subsequences $\{(u(kn+r))_{n\in\mathbb{N}} : \ k \geq 1, \ 0 \leq r \leq k-1\}$ are all distinct. Shallit proves in [64] that Sturmian sequences are maximally diverse, which shows that they are very far from being automatic, even when they are fixed points of substitution. More precisely, he deduces from the three distance theorem a measure of automaticity for some Sturmian sequences [64].

Theorem 8.2 *Let $0 < \alpha < 1$ be an irrational number with bounded partial quotients. Let $u_n = \lfloor (n+1)\alpha \rfloor - \lfloor n\alpha \rfloor$, for $n \geq 1$. The automaticity of the sequence $(u_n)_{n\geq 1}$ has the same order of magnitude as $n^{1/5}$.*

8.2 Sub-affine Complexity

Numerous combinatorial, ergodic or arithmetical properties hold in the case of a sub-affine complexity function. Consider a dynamical system generated by a minimal sequence with sub-affine complexity. Ferenczi proves in [42] the absence of strong mixing. Boshernitzan produces in [17, 19] an explicit upper bound for the number of ergodic measures. He proves furthermore that the following conditions imply the unique ergodicity [17, 19]:

$$\liminf_{n\to+\infty} \frac{p(n)}{n} < 2, \text{ or } \limsup_{n\to+\infty} \frac{p(n)}{n} < 3.$$

Furthermore Ferenczi deduces from Theorem 4.9 the following result [42]: a symbolic dynamical system generated by a minimal sequence of sub-affine complexity is generated by a finite number of substitutions (such a system is called *S-adic* following Vershik's terminology). One thus explicitely knows S-adic expansions of Sturmian sequences or of Arnoux-Rauzy sequences [11]. The reciprocal of this result is false. Consider indeed a substitution of quadratic complexity: it thus provides a counter-example.

The question of finding a characterization of sequences of sub-affine complexity in terms of S-adic expansions remains open. Note that Durand gives in [38] a sufficient (but not necessary) condition for a S-adic sequence to have sub-affine complexity.

References

[1] R. L. Adler, A. G. Konheim, M. H. Mcandrew, *Topological Entropy*, Trans. Amer. Math. Soc. **114**, pp. 309–319 (1965).

[2] P. Alessandri, *Codages de Rotations et Basses Complexités*, Université Aix-Marseille II, Thèse (1996).

[3] P. Alessandri, V. Berthé, *Three Distance Theorems and Combinatorics on Words*, Enseig. Math. **44**, pp. 103–132 (1998).

[4] J.-P. Allouche, *Somme des Chiffres et Transcendence*, Bull. Soc. math. France **110**, pp. 279–285 (1982).

[5] J.-P. Allouche, *Automates Finis en Théorie des Nombres*, Expo. Math., **5**, pp. 239–266 (1987).

[6] J.-P. Allouche, *Note sur un Article de Sharif et Woodcock*, Séminaire de théorie des Nombres de Bordeaux, Série II, **1**, pp. 631–633 (1989).

[7] J.-P. Allouche, *Sur la Transcendance de la Série Formelle* Π, Séminaire de Théorie des Nombres de Bordeaux **2**, pp. 103-117 (1990).

[8] J.-P. Allouche, *Sur la Complexité des Suites Infinies*, Bull. Belg. Math. Soc. **1**, pp. 133–143 (1994).

[9] J.-P. Allouche, M. Mendès France, *Automata and Automatic Sequences*, Actes de l'École de Physique Théorique des Houches: "Beyond quasicrystals", Les Éditions de Physique, Springer, pp. 293–367 (1995).

[10] J.-P. Allouche, F. von Haeseler, H.-O. Peitgen, G. Skordev, *Linear Cellular Automata, Finite Automata and Pascal's Triangle*, Discrete Applied Mathematics **66**, pp. 1–22 (1996).

[11] P. Arnoux, G. Rauzy, *Représentation Géométrique de Suites de Complexité* $2n+1$, Bull. Soc. Math. France **199**, pp. 199–215 (1991).

[12] J. Berstel, *Recent Results in Sturmian Words*, Developments in Language Theory II (DLT'95) (Dassow, Rozenberg, Salomaa eds) World Scientific, pp. 13–24 (1996).

[13] V. Berthé, *Entropy in Deterministic and Random Systems*, Actes de l'École de Physique Théorique des Houches: "Beyond quasicrystals", Les Éditions de Physique, Springer, pp. 441–463 (1995).

[14] V. Berthé, *Fréquences des Facteurs des Suites Sturmiennes*, Theoret. Comput. Sci. **165**, pp. 295–309 (1996).

[15] V. Berthé, N. Chekhova, S. Ferenczi, *Covering Numbers: Arithmetics and Dynamics for Rotations and Interval Exchanges*, J. Anal. Math., to appear.

[16] P. Billingsley, *Ergodic Theory and Information*, John Wiley and Sons, New York (1965).

[17] M. Boshernitzan, *A Unique Ergodicity of Minimal Symbolic Flows with Linear Block Growth*, J. Anal. Math. **44**, pp. 77–96 (1984).

[18] M. Boshernitzan, *A Condition for Minimal Interval Exchange Maps to be Uniquely Ergodic*, Duke Math. J. **52**, pp. 723–752 (1985).

[19] M. Boshernitzan, *A condition for unique ergodicity of minimal symbolic flows*, Ergodic Theory Dynam. Sys. **12**, pp. 425–428 (1992).

[20] S. Brlek, *Enumeration of Factors in the Thue-Morse Word*, Discr. Appl. Math. **24**, pp. 83–96 (1989).

[21] T. C. Brown, *Descriptions of the Characteristic Sequence of an Irrational*, Canad. Math. Bull **36**, pp. 15–21 (1993).

[22] V. Bruyère, G. Hansel, C. Michaux, *Logic and p-Recognizable Sets of Integers*, Bull. Belg. Math. Soc. **1**, pp. 191–238 (1994).

[23] J. Cassaigne, *Special Factors of Sequences with Linear Subword Complexity*, Developments in Language Theory II (DLT'95) (Dassow, Rozenberg, Salomaa eds) World Scientific, pp. 25–34 (1996).

[24] J. Cassaigne, *Complexité et Facteurs Spéciaux*, Bull. Belg. Math. Soc. **4**, pp. 67–88 (1997).

[25] J. Cassaigne, *Sequences with Grouped Factors*, Developments in Language Theory III (DLT'97), Publications of the Aristotle University of Thessaloniki, pp. 211–222 (1998).

[26] N. Chekhova, *Covering Numbers of Rotations*, Theoret. Comput. Sci., to appear.

[27] G. Christol, *Ensembles presque Périodiques k-Reconnaissables*, Theoret. Comp. Sci. **9**, pp. 141–145 (1979).

[28] G. Christol, T. Kamae, M. Mendès France, G. Rauzy, *Suites Algébriques, Automates et Substitutions*, Bull. Soc. Math. France **108**, pp. 401-419 (1980).

[29] A. Cobham, *On the Base Dependence of Sets of Numbers Recognisable by Finite Automata*, Math. Systems Theory **3**, pp. 186–192 (1969).

[30] A. Cobham, *Uniform Tag Sequences*, Math. Systems Theory **6**, pp. 164–192 (1972).

[31] E. M. Coven, G. A. Hedlund, *Sequences with Minimal Block Growth*, Math. Systems Theory **7**, pp. 138–153 (1973).

[32] N. G. de Bruijn, *A Combinatorial Problem*, Rominklijke Netherland Academic Van Wetenschepen Proc. **49** Part. 20, pp. 758–764 (1946).

[33] B. de Mathan, *Irrationality Measures and Transcendence in Positive Characteristic*, J. Number Theory **54**, pp. 93–112 (1995).

[34] F. M. Dekking, *On the Prouhet-Thue-Morse Measure*, Acta Universitatis Carolinae, Mathematica et Physica **33**, pp. 35–40 (1992).

[35] P. Deligne, *Intégration sur un Cycle Évanescent*, Invent. Math. **76**, pp. 129–143 (1984).

[36] G. Didier, *Caractérisation des N-Écritures et Applications à l'Étude des Suites de Complexité Ultimement* $n + c^{ste}$, Theoret. Comput. Sci. **215**, pp. 31–49 (1999).

[37] F. Durand, *A Characterization of Substitutive Sequences using Return Words*, Discrete Math. **179**, pp. 89–101 (1998).

[38] F. Durand, *Linearly Recurrent Subshifts*, Ergod. Theory Dynam. Sys., to appear.

[39] S. Eilenberg, *Automata, Languages, and Machines*, Vol. A Academic Press (1974).

[40] A. Ehrenfeucht, K. P. Lee, G. Rozenberg, *Subwords of various classes of deterministic developmental languages without intercations*, Theoret. Comput. Sci. **1** (1975), 59–75.

[41] S. Ferenczi, *Les Transformations de Chacon : Combinatoire, Structure Géométrique, Lien avec les Systèmes de Complexité* $2n + 1$, Bull. Soc. math. France **123**, pp. 271–292 (1995).

[42] S. Ferenczi, *Rank and Symbolic Complexity*, Erg. Theory Dynam. Sys. **16**, pp. 663–682 (1996).

[43] S. Ferenczi, *Complexity of Sequences and Dynamical Systems*, Discrete Math., to appear.

[44] J. Fresnel, M. Koskas, B. de Mathan, *Automata and Transcendence in Positive Characteristic*, J. Number Theory, to appear.

[45] C. Frougny, *Number Representation and Finite Automata*, this volume.

[46] H. Furstenberg, *Algebraic Functions over Finite Fields*, J. Algebra **7**, pp. 271–277 (1967).

[47] I. Glaister, J. Shallit, *Automaticity III: Polynomial Automaticity and Context-Free Languages (extended abstract)*, Mathematical foundations of Computer Science 1996 (Cracow), Lecture Notes in Comput. Sci. **1113**, pp. 382–393 (1996).

[48] T. Harase, *Algebraic Elements in Formal Power Series Rings*, Israel J. Math. **63**, pp. 281–288 (1988).

[49] B. Host, *Substitutions and Bratelli Diagrams*, this volume.

[50] M. Koskas, *Complexité de Suites, Fonctions de Carlitz*, University of Bordeaux I, Thesis (1995).

[51] J. H. Loxton, A. J. van der Poorten, *Arithmetic Properties of Automata: Regular Sequences*, J. Reine Angew. Math. **392**, pp. 57–69 (1988).

[52] M. Mendès France, J. Y. Yao, *Transcendence and the Carlitz-Goss Gamma Function*, J. Number Theory **63**, pp. 396–402 (1997).

[53] M. Misiurewicz, *A short proof of the variational principle for a* $\mathbb{Z}_+^n$ *action on a compact space*, Int. Conf. Dyn. Systems in Math. Physics, Société Mathématique de France, Astérisque **40**, pp. 147–158 (1976).

[54] M. Morse, G. A. Hedlund, *Symbolic Dynamics*, Amer. J. Math. **60**, pp. 815–866 (1938).

[55] M. Morse, G. A. Hedlund, *Symbolic Dynamics II: Sturmian Trajectories*, Amer. J. Math. **62**, pp. 1–42 (1940).

[56] J.-J. Pansiot, *Complexité des facteurs des mots infinis engendrés par morphismes itérés*, Lecture Notes in Comput. Sci. **172**, pp. 380–389 (1984).

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

[58] G. Rauzy, *Suites à Termes dans un Alphabet Fini*, Sém. de Théorie des Nombres de Bordeaux, 25-01–25-16 (1983).

[59] G. Rauzy, *Low Complexity and Geometry*, Dynamics of Complex Interacting Systems (Santiago, 1994), 147–177, Nonlinear Phenom. Complex Systems, 2, Kluwer Acad. Publ., Dordrecht (1996).

[60] R. Ritchie, *Finite automata and the set of squares*, J. Assoc. Comput. Mach. **10**, pp. 528–531 (1963).

[61] G. Rote, *Sequences with Subword Complexity* $2n$, J. Number Theory **46**, pp. 196–213 (1994).

[62] O. Salon, *Suites Automatiques à Multi-indices et Algébricité*, C. R. Acad. Sci. Paris, Série I **305**, pp. 501–504 (1987).

[63] O. Salon, *Suites Automatiques à Multi-indices*, Séminaire de Théorie des Nombres, Exposé 4, 1986-1987.

[64] J. Shallit, *Automaticity IV: Sequences, Sets, and Diversity*, J. Théor. Nombres Bordeaux **8**, pp. 347–367 (1996).

[65] C. E. Shannon, *A Mathematical Theory of Communication*, Bell System Tech. J. **27**, pp. 379–423, 623–656 (1948).

[66] H. Sharif, C. F. Woodcock, *Algebraic Functions over a Field of Positive Characteristic and Hadamard Products*, J. Lond. Math. Soc. **37**, pp. 395–403 (1988).

[67] V. T. Sós, *On the Theory of Diophantine Approximations I*, Acta Math. Acad. Sci. Hungar. **8**, pp. 461–472 (1957).

[68] V. T. Sós, *On the Distribution mod 1 of the Sequence* $n\alpha$, Ann. Univ. Sci. Budapest, Eötvös Sect. Math. **1**, pp. 127–134 (1958).

[69] J. Surányi, *Über die Anordnung der Vielfachen einer reellen Zahl mod 1*, Ann. Univ. Sci. Budapest, Eötvös Sect. Math. **1**, pp. 107–111 (1958).

[70] S. Świerczkowski, *On Successive Settings of an Arc on the Circumference of a Circle*, Fundamenta Math. **46**, pp. 187–189 (1958).

[71] D. S. Thakur, *Automata and Transcendence*, Contemporary Mathematics, **210**, pp. 387–399 (1998).

[72] L. J. Wade, *Certain Quantities Transcendental over $GF(p^n, x)$*, Duke Math. J. **8**, pp. 701-720 (1941).

[73] P. Walters, *An Introduction to Ergodic Theory*, Graduate Texts in Mathematics **79**.

[74] J.-y. Yao, *Critères de non-automaticité et leurs applications*, Acta Arith. **80** (1997), 237–248.