2.1 COMPLEXITY AND FREQUENCIES OF CODINGS OF ROTATIONS

Objekttyp: **Chapter**

Zeitschrift: **L'Enseignement Mathématique**

Band (Jahr): **44 (1998)**

Heft 1-2: **L'ENSEIGNEMENT MATHÉMATIQUE**

PDF erstellt am: **21.07.2024**

2.1 COMPLEXITY AND FREQUENCIES OF CODINGS OF ROTATIONS

A *factor of the infinite sequence* u is a finite block w of consecutive letters of u, say $w = u_{n+1} \cdots u_{n+d}$; d is called the *length* of w, denoted by $|w|$. Let $p(n)$ denote the *complexity function* of the sequence u with values in a finite alphabet: it counts the number of distinct factors of given length of the sequence u. For more information on the subject, we refer the reader to the survey [2].

With the above notation, consider a coding u of the orbit of a point x under the rotation by angle α with respect to the partition $\{[\beta_0, \beta_1[, [\beta_1, \beta_2[, \ldots, [\beta_{p-1}, \beta_p[\}$. Let $I_k = [\beta_k, \beta_{k+1}[$ and let R denote the rotation by angle α. A finite word $w_1 \cdots w_n$ defined on the alphabet $\Sigma = \{0, 1, \ldots, p-1\}$ 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 $(\{x+n\alpha\})_{n \in \mathbf{N}}$ is dense in the unit circle, which implies that $w_1 w_2 \cdots w_n$ is a factor of u if and only if $I(w_1, \ldots, w_n) \neq \varnothing$. In particular, the set of factors of a coding does not depend on the initial point x of this coding. Furthermore, the connected components of these sets are bounded by the points

$$\{k(1-\alpha) + \beta_i\}, \text{ for } 0 \leq k \leq n-1, \ 0 \leq i \leq p-1.$$

Let us recall that the *frequency* $f(B)$ of a factor B of a sequence is the limit, if it exists, of the number of occurrences of this block in the first k terms of the sequence divided by k. Thus the frequency of the factor $w_1 \cdots w_n$ exists and is equal to the density of the set

$$\left\{k \,\middle|\, \{x + k\alpha\} \in I(w_1, \ldots, w_n)\right\},$$

which is equal to the length of $I(w_1, \ldots, w_n)$, by uniform distribution of the sequence $(\{x + n\alpha\})_{n \in \mathbf{N}}$. These sets consist of finite unions of intervals. More precisely, if for every k, $\beta_{k+1} - \beta_k \leq \sup(\alpha, 1-\alpha)$, then these sets are connected; if there exists K such that $\beta_{K+1} - \beta_K > \sup(\alpha, 1-\alpha)$, then the sets are connected except for $w_1 \cdots w_n$ of the form a_K^n (see [1]) (the notation a_K^n denotes the word of length n obtained by successive concatenations of the letter a_K). Let us note that there exists at most one integer K satisfying $\beta_{K+1} - \beta_K > \sup(\alpha, 1-\alpha)$. We thus have the following lemma, which links the three distance theorem and related results to the frequencies of codings of rotations.

LEMMA 1. *Let u be a coding of a rotation by irrational angle α on the unit circle with respect to the partition*

$$\{[\beta_0, \beta_1[,\ [\beta_1, \beta_2[, \ldots, [\beta_{p-1}, \beta_p[\} ,$$

such that the lengths of the intervals of the partition are less than or equal to $\sup(\alpha, 1-\alpha)$. *Then the frequencies of factors of length n of the sequence u are equal to the lengths of the intervals bounded by the points*

$$\{k(1-\alpha) + \beta_i\} , \ \textit{for}\ 0 \le k \le n-1,\ 0 \le i \le p-1 .$$

In particular, if the partition is equal to $\{[0, 1-\alpha[, [1-\alpha, 1[\}$, i.e., if u is a Sturmian sequence, the intervals $I(w_1, \ldots, w_n)$ are exactly the $(n+1)$ intervals bounded by the points

$$0,\ \{(1-\alpha)\}, \ldots, \{(n(1-\alpha)\} .$$

Therefore there are exactly $n+1$ factors of length n and the complexity of a Sturmian sequence satisfies $p(n) = n+1$, for every n. Furthermore, the lengths of these intervals are equal to the frequencies of factors of length n.

In fact, this complexity function characterizes Sturmian sequences. Indeed, any sequence of complexity $p(n) = n+1$, for every n, is a Sturmian sequence, i.e., there exists α irrational in $]0, 1[$ and x such that this sequence is the coding of the orbit of x under the rotation by angle α with respect either to the partition $\{[0, 1-\alpha[, [1-\alpha, 1[\}$ or $\{]0, 1-\alpha],]1-\alpha, 1]\}$ (see [40]) (the coding of the orbit of α is called the *characteristic sequence* of α). Note that a sequence whose complexity satisfies $p(n) \le n$, for some n, is ultimately periodic (see [19] and [39]). Sturmian sequences thus have the minimal complexity among sequences not ultimately periodic. 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.

In the general case of a coding of an irrational rotation, the complexity has the form $p(n) = an + b$, for n large enough (see Theorem 10 below and [1], for the whole proof). The converse is not true: every sequence of ultimately affine complexity is not necessarily obtained as a coding of rotation. Didier gives in [23] a characterization of codings of rotations. See also [46], 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 [22] or [1]).

2.2 The graph of words

The aim of this section is to introduce the Rauzy graph of words of a sequence, in order to obtain results concerning the frequencies of factors of this sequence. This follows an idea of Dekking, who expressed the block frequencies for the Fibonacci sequence by using the graph of words (see [20] and also [8]). Note that Boshernitzan also introduces in [8] a graph for interval exchange maps (homeomorphic to the Rauzy graph of words) in order to prove Theorem 3, which can be seen as a result on frequencies.

Let us note that precise knowledge of the frequencies of a sequence with values in a finite alphabet $\mathcal{A}$ allows a precise description of the measure associated with the dynamical system $(\overline{\mathcal{O}(u)}, T)$: here T denotes the one-sided shift which associates to a sequence $(u_n)_{n\in\mathbf{N}}$ the sequence $(u_{n+1})_{n\in\mathbf{N}}$ and $\overline{\mathcal{O}(u)}$ is the orbit closure under the shift T of the sequence u in $\mathcal{A}^{\mathbf{N}}$, equipped with the product of the discrete topologies (it is easily seen that $\overline{\mathcal{O}(u)}$ is the set of sequences of factors belonging to the set of factors of u). Indeed, we define a probability measure μ on the Borel sets of $\overline{\mathcal{O}(u)}$ as follows: the measure μ is the unique T-invariant measure defined by assigning to each cylinder $[w]$ corresponding to the sequences of $\overline{\mathcal{O}(u)}$ of prefix w, the frequency of w, for any finite block w with letters from $\mathcal{A}$. Let us note that a dynamical system obtained via a coding of irrational rotation is *uniquely ergodic*, i.e., there exists a unique T-invariant probability measure on this dynamical system, which is thus determined by the block frequencies.

The Rauzy graph Γ_n of words of length n of a sequence with values in a finite alphabet is an oriented graph (see, for instance, [41]), which is a subgraph of the de Bruijn graph of words. 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., more precisely, 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 (such an edge is labelled by xWy). Thus there are $p(n+1)$ edges and $p(n)$ vertices, where $p(n)$ denotes the complexity function. A sequence is said to be *recurrent* if every factor appears at least