Global complexities for infinite sequences

Mira-Cristiana Anisiu*
"T. Popoviciu" Institute of Numerical Analysis
P. O. Box 68, 3400 Cluj-Napoca

1 Introduction

The language complexity of a finite word or infinite sequence is aimed to give a measure of the number of different factors in the given word or sequence. In fact, the definitions in the finite or infinite case coincide.

Let A be a finite nonvoid alphabet and $U = u_0 u_1 ...$ an infinite sequence with $u_i \in A$, $i \in \mathbb{N}$. The ***complexity of the sequence*** will be a function $p_U : \mathbb{N}^* \to \mathbb{N}$ given by

$$p_U(n) = \sharp L_n(U), \quad n \in \mathbb{N}^*, \tag{1}$$

$\sharp$ denoting the cardinal of the set $L_n(U)$ of the factors of length n in U.

For the finite word w of length $q \in \mathbb{N}^*$, the ***complexity*** p_w may be conceptually defined in the same way. But, because $\sharp L_n(w) = 0$ for $n > q$, we can consider only the restriction on the set $\{1, ..., q\}$, so $p_w : \{1, ..., q\} \to \mathbb{N}$,

$$p_w(n) = \sharp L_n(w), \quad n \in \{1, ...q\}. \tag{2}$$

The complexity functions p_U, respectively p_w, estimate the richness in factors of length n of a sequence or of a word, for every $n \in \mathbb{N}^*$ or $n \in \{0, ..., q\}$. It would be of great help to have a global indicator of the complexity.

In the case of finite words there are two known definitions, the first one of total complexity given independently by Iványi [5] and Shallit [6], the second of maximal complexity, suggested by Rauzy.

*E-mail: mira@math.ubbcluj.ro

Definition 1 *For the word w of length $q \in \mathbb{N}^*$, the* ***total complexity*** *is given by*

$$K_w = \sum_{j=1}^{q} p_w(j), \tag{3}$$

and the ***maximal complexity*** *is*

$$C_w = \max_{j=1}^{q} p_w(j). \tag{4}$$

The aim of this paper is to define similar global complexities (not depending on the specific length n of the factors) for infinite sequences.

2 The significance of total and maximal complexity for words

Given a word of length q, the complexity function p_w may be considered as a vector in the space $\mathbb{R}^q$. On this space we have the well-known Minkowski norm $\|x\|_1 = \sum_{j=1}^{q} |x_j|$, Chebyshev norm $\|x\|_\infty = \max_{j=1}^{q} |x_j|$, or Euclid norm $\|x\| = \left(\sum_{j=1}^{q} x_j^2 \right)^{1/2}$, which are of course equivalent. It is obvious that if we denote

$$p_w = (p_w(1), ..., p_w(q)),$$

we shall have for the total complexity

$$K_w = \|p_w\|_1,$$

and for the maximal complexity

$$C_w = \|p_w\|_\infty.$$

We can also consider a Euclidian complexity given by $\|p_w\|$, which has the disadvantage that its values are not in general integer numbers.

The total and maximal complexities were used by Ferenczi and Kasa [4] to define upper and lower complexities for infinite sequences, in the following way:

– the ***upper*** and ***lower total finite-word complexity function*** by

$$\begin{aligned} K_U^{+}(n) &= \max_i K(u_i u_{i+1} ... u_{i+n-1}), \\ K_U^{-}(n) &= \min_i K(u_i u_{i+1} ... u_{i+n-1}); \end{aligned} \tag{5}$$

– the ***upper*** and ***lower maximal finite-word complexity function*** by

$$\begin{aligned} C_U^{+}(n) &= \max_i C(u_i u_{i+1} ... u_{i+n-1}), \\ C_U^{-}(n) &= \min_i C(u_i u_{i+1} ... u_{i+n-1}). \end{aligned} \tag{6}$$

These two notions, having many interesting properties, are extensively studied in [4]. Being functions of $n \in \mathbb{N}^*$, the upper and lower total (maximal) finite-word complexity functions are similar to the initial complexity function p_U. What we intend is to define global complexities for the case of infinite sequences too.

3 Global complexities for infinite sequences

There already exist some notions to estimate the global complexity for infinite sequences, as for example the *topological entropy* [1]

$$h_U = \lim_{n\to\infty} \frac{\log p_U(n)}{n \log \sharp A}, \tag{7}$$

which satisfies

$$0 \le h_U \le 1$$

for every sequence U.

The definitions we give extend those for finite words to infinite sequences; to this aim it is necessary to "normalize" the values of the complexity function $p_U(n)$ by dividing them with $(\sharp A)^n$.

Definition 2 *For the infinite sequence U, the* ***total complexity*** *is given by*

$$K_U = \sum_{n=1}^{\infty} \frac{p_U(n)}{(\sharp A)^n}, \tag{8}$$

and the ***maximal complexity*** *by*

$$C_U = \sup_{n=1}^{\infty} \frac{p_U(n)}{(\sharp A)^n}. \tag{9}$$

Remark 1 *The total complexity in definition 2 is not necessarily finite; for example, for the Champernowne word containing successively all the binary written numbers* 011011100101111... *we have* $p_U(n) = 2^n$ *and* $K_U = \infty$, *while* $C_U = 1$.

For this reason, instead of definition 2 we propose one for normal complexity, which will have values less than 1.

Definition 3 *For the infinite sequence* U, *the* ***normal complexity*** *is given by*

$$\mathcal{K}_U = \sum_{n=1}^{\infty} \frac{1}{(\sharp A)^n} \frac{p_U(n)}{1 + p_U(n)}. \quad (10)$$

Remark 2 *In the definitions above, there appears the sequence*

$$c_n = \frac{p_U(n)}{(\#A)^n}, \quad n \in \mathbb{N}^*.$$

Because of the inequality

$$p_U(n+1) \leq (\#A)\ p_U(n), \quad n \in \mathbb{N}^*$$

holding for each complexity function, the sequence c_n *is non-increasing, so* $C_U = c_1 = 1$. *For the Champernowne sequence we have* $c_n = 1, \quad n \in \mathbb{N}^*$.

4 Global complexities for sequences with known complexity functions

In the following we consider sequences for which the complexity function is known and try to determine (at least approximately) their global complexities. The alphabet has three symbols in example 4.1, $a + b$ in 4.2 and two symbols for Sturmian sequences and those in 4.3 and 4.4. The topological entropy is $h_U = 0$, excepting example 4.4.

4.1 Sequences defined by billiards in the cube

A sequence generated by the structure of billiard trajectories in the cube, associating to a trajectory starting with totally irrational direction the sequence with values in {1, 2, 3} given by coding 1 (respectively 2, 3) any time the particle rebounds on a frontal (lateral, horizontal) side of the cube, was shown in [2] to have the complexity $p_U(n) = n^2 + n + 1$.

Proposition 1 *For a sequence defined by the cubic billiard with totally irrational direction, the total, maximal and normal complexity will be*

$$K_U = 2.75,\ C_U = 1 \text{ and } \mathcal{K}_U \approx 0.3994006256.$$

4.2 Sequences with $p_U(n) = an + b$

For an important class of sequences the complexity function is linear.

Proposition 2 *Let U be a sequence having the complexity function given by $p_U(n) = an + b$, $n \in \mathbb{N}^*$ $(a \in \mathbb{N}^*,\ b \in \mathbb{N}\ ,\ a + b \geq 2)$. Its total, maximal and normal complexity will be*

$$K_U = {}_2F_1(1, \tfrac{2a+b}{a}; \tfrac{a+b}{a}; \tfrac{1}{a+b}),\quad C_U = 1,$$

$$\mathcal{K}_U = \tfrac{1}{a+b+1}\, {}_3F_2\left(1, \tfrac{2a+b}{a}, \tfrac{a+b+1}{a}; \tfrac{a+b}{a}, \tfrac{2a+b+1}{a}; \tfrac{1}{a+b}\right),$$

where ${}_\alpha F_\beta$ denotes the hypergeometric function.

Remark 3 *Well-known sequences of this type are Sturmian sequences (which are not ultimately periodic, but are recurrent), for which $a = b = 1$; they have the global complexities given by*

$$K_U = 3,\ C_U = 1 \text{ and } \mathcal{K}_U = 7/2 - 4\ln 2 \approx 0.727411278.$$

4.3 The power sequence

Let us consider the words $v_i = 0^i$, $w_i = 1^i$ and $u_i = v_i w_i$, $i \in \mathbb{N}^*$, u_i being obtained by concatenation of v_i and w_i. The power sequence U is given by

$$U = u_1 u_2 u_3 ... = 010011000111... \tag{11}$$

and has the complexity $p_U = n(n+1)/2 + 1$.

Proposition 3 *For the power sequence (11), the global complexities are given by*

$$K_U = 5,\ C_U = 1,\ \mathcal{K}_U \approx 0.7595479501.$$

4.4 Champernowne sequence

This sequence was mentioned in Section 3 and it is obtained by writing successively all the binary written numbers. Its complexity function is $p_U(n) = 2^n$ and the topological entropy is $h_U = 1$.

Proposition 4 *For the Champernowne sequence, the global complexities are given by*

$$K_U = \infty, \; C_U = 1, \; \mathcal{K}_U \approx 0.7644997803.$$

Acknowledgements. The author thanks the Ministry of Science and Technology for supporting partially this work (3004GR/1997).

References

[1] J.-P. Allouche, Sur la complexité des suites infinies, *Bull. Belg. Math. Soc.* **1**(1994), 133-143

[2] P. Arnoux, Ch. Mauduit, I. Shiokawa, Jun-ichi Tamura, Complexity of sequences defined by billiards in the cube, *Bull. Soc. Math. France* **122**(1)(1994), 1-12

[3] S. Ferenczi, Complexity of sequences and dynamical systems, to appear in *Discrete Math.* **206**(1999)

[4] S. Ferenczi, Z. Kása, Complexities for finite factors of infinite sequences, to appear in *Theor. Comput. Sci.* **218**(1)(1999)*, 177-195*

[5] A. Iványi, On the d-complexity of words, *Ann. Univ. Sci. Budapest Sect. Comput.* **8**(1987), 69-90

[6] J. Shallit, On the maximum number of distinct factors in a binary string, *Graphs and Combinatorics* **9**(1993), 197-200