Far East Journal of Mathematical Sciences (FJMS)
© 2014 Pushpa Publishing House, Allahabad, India
Available online at http://pphmj.com/journals/fjms.htm
Volume 87, Number 2, 2014, Pages 133-147

ON THE COMPLEXITY OF FINITE SEQUENCES OVER A FINITE SET

Dae-Seok Lee and Soonja Kang

CNU Science Education Center for the Gifted
Chonnam National University
Gwangju 500-757, Korea
e-mail: apig5942@naver.com

Department of Mathematics Education
Chonnam National University
Gwangju 500-757, Korea
e-mail: kangsj@jnu.ac.kr

Abstract

There are various factors by which we recognize a sequence is complicated. Thus, it is very difficult to assign a real number representing the degree of complexity to a sequence by considering all factors. J.-P. Allouche says that the way of measuring complexity of a sequence is classified into two main classes, algorithmic complexity and combinatorial complexity, according to view of looking at sequences. In this paper, we will give two definitions of complexity functions focused on the amount of the repetition in a sequence. Then we prove the equivalence of two definitions and obtain the approximations to the maximum and the average of complexities of sequences over a finite set of the length n.

Received: January 14, 2014; Revised: February 24, 2014; Accepted: March 8, 2014
2010 Mathematics Subject Classification: 11B85, 94A55.
Keywords and phrases: complexity, repetitive sequence, primitive sequence.
This work was supported by the Korea Foundation for the Advancement of Science and Creativity.

1. Introduction

It is natural to think that sequences 00000000 or 00001111 are simpler with a sequence 011010100 psychologically. How can we denote the degree of complexity about a given finite sequence mathematically? There are many studies devoted to complexity of finite or infinite sequences [1-3]. One way of measuring the complexity of a sequence is to evaluate how difficult it is to generate that sequence, called *algorithmic complexity*. The other one, called *combinatorial complexity*, is to look for its combinatorial properties, e.g., counting subwords, palindromes, squares, etc., occurring in the sequence. In this paper, we will focus on the combinatorial complexity in which the more repeated the terms of a sequence is, the less complicated it is. We will define a new complexity of a sequence as a kind of combinatorial complexity concentrated on the amounts of the repetitions in a sequence though there are many definitions in view of combinatorial complexity of a sequence [1-3]. Then we will obtain the approximations to the maximum and the average of complexities of sequences over a finite set of the length n.

2. Measuring the Complexity of Finite Sequence

We will consider only finite sequences and define a complexity of a finite sequence (called *word* in [3]) generated by elements in a finite set. The symbols and terms used in this paper are partially taken from the paper [1] on Lempel-Ziv's complexity and from Allouche and Shallit's book [3] on Automatic sequences.

Let $\mathcal{A}^*$ be the set of all finite length sequences (except null sequence) over a finite set $\mathcal{A}$. Let $l(S)$ denote the length of $S \in \mathcal{A}^*$ and let

$$\mathcal{A}^n = \{S \in \mathcal{A}^* : l(S) = n\}, \quad n \geq 1.$$

When a sequence $S \in \mathcal{A}^*$ is denoted by $S = s_1 s_2 s_3 \cdots s_n$, $S(i, j) = s_i s_{i+1} \cdots s_j$, $i \leq j$ is called *substring* (called *subword* in [3]) of a sequence S, consisting of the adjacent terms of the sequence S. In particular, if $S = aaa \cdots a \in \mathcal{A}^*$, then we denote $S = a^n$.

If $Q = q_1 q_2 \cdots q_m \in A^m$ and $R = r_1 r_2 \cdots r_n \in \mathcal{A}^n$, the *concatenation* of Q and R is a new sequence $S = QR = q_1 q_2 \cdots q_m r_1 r_2 \cdots r_n \in \mathcal{A}^{m+n}$. In this case, we call Q the *prefix* of S and S the *extension* of Q. If $S_i \in \mathcal{A}^*$, $(i = 1, 2, ..., n)$, then the concatenation of $S_i's$ means $S = S_1 S_2 \cdots S_{n-1} S_n = (\cdots(((S_1 S_2) S_3) S_4) S_5 \cdots)$ and $S^n = SSS \cdots SS$ if $S_i = S$, $i = 1, 2, ..., n$.

$S \in \mathcal{A}^*$ is called the *repetitive sequence* if there exists a sequence $S' \in \mathcal{A}^*$ such that $S = (S')^n$ and S' is called the *unit of repetition* in S. In particular, if S is denoted by $S = (S')^2$, then S is called a *square*. A sequence $S \in \mathcal{A}^*$ is said to be *primitive* if there has no solution to $S = (S')^k$ for k an integer ≥ 2 and $S' \in \mathcal{A}^*$. In other words, a primitive sequence is not repetitive.

The multiple power notation of a sequence S means the arrangement of S denoted by powers of substrings. There are various multiple power notations of a sequence S. For example, when $S = aababaababbbb$, the multiple power notations of S are $(a(ab)^2)^2 b^3$, $a(ab)^2 a^2 bab^4$, $a^2 (ba)^2 (ab)^2 b^3 \cdots$. The number (considering the repeated occurrence) of letters showed in the multiple power notation M of a sequence S is denoted by $\nu(M)$. So we can see $\nu((a(ab)^2)^2 b^3) = 4$ because a, a, b, b, are occurred in $(a(ab)^2)^2 b^3$ and $\nu(a(ab)^2 a^2 bab^4) = 7$ because a, a, b, a, b, a, b are occurred in $a(ab)^2 a^2 bab^4$.

We define the *complexity* of a sequence $S \in \mathcal{A}^*$ by

$$C_*(S) = \min\{\nu(M) : M \text{ is a multiple power notation of } S\}. \tag{2.1}$$

This definition of a complexity of S is more or less complicated to compute by computer or to prove some properties of complexity. Thus, we suggest another definition of complexity which is equivalent to definition as above. It is convenient to compute and prove some properties about complexity using the following definition of complexity. We will prove the equivalence of two definitions of complexity and then obtain some properties using the following definition of complexity.

Definition 2.1. The function $C : \mathcal{A}^* \to \mathbb{N}$ is defined inductively as follows:

(1) $C(S) = 1$ if $l(S) = 1$.

(2) Suppose that $C(S)$ is defined for $S \in \mathcal{A}^*$ with $l(S) < n$. For a sequence $S \in \mathcal{A}^*$ with $l(S) = n$, we define

$$C(S) = \min\{\{C(S_1) + C(S_2) : S = S_1S_2\} \cup \{C(S') : S = S'^k, k \geq 2\}\}. \quad (2.2)$$

This $C(S)$ as above is called the *complexity* of a sequence $S \in \mathcal{A}^*$.

Remark 2.2. For $A, B \in \mathcal{A}^*$, we have $C(AB) \leq C(A) + C(B)$.

We will show that two definitions $C(S)$ and $C_*(S)$ defined as above are equivalent.

Theorem 2.3. $C(S) = C_*(S)$ *for every sequence* $S \in \mathcal{A}^*$.

Proof. It is clear that $C(S) = C_*(S) = 1$ for S with $l(S) = 1$. Suppose that $C(S) = C_*(S)$ holds for every $S \in \mathcal{A}^*$ with $l(S) < n$. For a sequence $S \in \mathcal{A}^*$ with $l(S) = n$.

(1) We show that $C_*(S) \leq C(S)$.

(i) If there are $S_1, S_2 \in \mathcal{A}^*$ such that $S = S_1S_2$ and $C(S) = C(S_1) + C(S_2)$. Since $l(S_1) < n$, $l(S_2) < n$,

$$C(S_1) + C(S_2) = C_*(S_1) + C_*(S_2). \quad (2.3)$$

Let M_1 and M_2 be the multiple power notations of S_1 and S_2 with $C_*(S_1) = \nu(M_1)$ and $C_*(S_2) = \nu(M_2)$, respectively. Then M_1M_2 is a multiple power notation of the sequence $S = S_1S_2$. Since $\nu(M_1M_2) = \nu(M_1)$

$+ \nu(M_2) = C_*(S_1) + C_*(S_2)$ and $\nu(M_1M_2) \geq C_*(S_1S_2) = C_*(S)$, from the definition (2.1) and (2.3) we obtain

$$C(S) = C(S_1) + C(S_2) = C_*(S_1) + C_*(S_2) \geq C_*(S).$$

Hence, the inequality $C_*(S) \leq C(S)$ holds.

(ii) If there are $S' \in \mathcal{A}^*$ and $k \geq 2$ such that $S = S'^k$ and $C(S) = C(S')$, since $l(S') < n$,

$$C(S) = C(S') = C_*(S'). \tag{2.4}$$

Let M' be a multiple power notation of S' with $C_*(S') = \nu(M')$. Then $(M')^k$ is a multiple power notation of S such that $\nu((M')^k) = \nu(M') = C_*(S')$. Hence by definition (2.1) and (2.4),

$$C(S) = C(S') = C_*(S') = \nu((M')^k) \geq C_*(S).$$

(2) We show that $C_*(S) \geq C(S)$. We need to show that for any multiple power notation M of a sequence S, $\nu(M) \geq C(S)$. Note that any multiple power notation M of a sequence S is denoted by either $(M')^k$, $k \geq 2$ for some multiple power notation M' or $M = M_1M_2$ for some multiple power notations M_1, M_2 of substrings S_1, S_2, respectively.

(i) Let $M = (M')^k$ and M' be a multiple power notation of some substring S' of S. Since $l(S') < n$, by definition

$$\nu(M) = \nu(M') \geq C_*(S') = C(S') \geq C(S'^k) = C(S).$$

(ii) Let any multiple power notation M of a sequence S be denoted by $M = M_1M_2$ for some multiple power notations M_1, M_2 of substrings $S_1, S_2 \in A^*$, respectively. Then

$$\begin{aligned} \nu(M) &= \nu(M_1) + \nu(M_2) \geq C_*(S_1) + C_*(S_2) \\ &= C(S_1) + C(S_2) \geq C(S_1S_2) = C(S). \end{aligned}$$

□

Let R_S be the set of (i, j) such that the substring $S(i, j)$ of S is a square of a primitive sequence (they call this S the square in [3]).

Remark 2.4. If $Q \in \mathcal{A}^*$ is a prefix of $S \in \mathcal{A}^*$, then $R_Q \subset R_S$.

We denote $n + R_s$ by $n + R_S = \{(n+i, n+j) : (i, j) \in R_S\}$. We define a function $u : R_S \to \mathbb{N}$ by $u(r) = u(i, j) = \frac{j-i+1}{2}$ for $r = (i, j) \in R_S$. We denote the number of elements r in R_S such that $u(r) = u$ by $\#(S, u)$, i.e.

$$\#(S, u) = |\{r \in R_S : u(r) = u\}|,$$

where $|\{\cdot\}|$ means the number of elements in a set. For example, if $S = aabcbcbcacbcaacaabab$, then $R_S = \{(1, 2), (3, 6), (4, 7), (5, 8), (13, 14), (16, 17), (17, 20)\}$, $u(3, 6) = 2$, $u(13, 14) = 1 \cdots$ and $\#(S, 2) = |\{(3, 6), (4, 7), (5, 8), (17, 20)\}| = 4$.

3. The Approximations to the Maximum and the Average of Complexities of Sequences in $\mathcal{A}^n$

Let $M[C(A^n)]$ be the average of complexities of sequences in $\mathcal{A}^n$, i.e.

$$M[C(\mathcal{A}^n)] = \frac{\sum_{S \in \mathcal{A}^n} C(S)}{|\mathcal{A}|^n}.$$

First, we will explore the lower bound of $M[C(\mathcal{A}^n)]$.

Note that if S is not repetitive and does not include any substrings which are repetitive, then $C(S) = l(S)$.

Theorem 3.1. *For* $S \in \mathcal{A}^n$, $n \in \mathbb{N}$,

$$C(S) \geq n - \sum_{r \in R_S} u(r). \tag{3.1}$$

Proof. We prove the inequality (3.1) inductively. When $l(S) = 1$, it is easily proved that the equality in (3.1) holds. Suppose that the inequality (3.1) holds for all sequence $S \in \mathcal{A}^*$ with $l(S) < n$. For sequence S with $l(S) = n$,

(i) If there are $S_1, S_2 \in \mathcal{A}^*$ such that $S = S_1S_2$ and $C(S) = C(S_1) + C(S_2)$. Let $l(S_1) = n_1$, $l(S_2) = n_2$. The inequality (3.1) holds for substrings S_1, S_2 since $l(S_1) < n$, $l(S_2) < n$. Since $n = n_1 + n_2$, $R_{S_1} \cap n_1 + R_{S_2} = \varnothing$ and $R_{S_1} \cup n_1 + R_{S_2} \subset R_S$,

$$C(S) = C(S_1) + C(S_2) \geq \left\{n_1 - \sum_{r \in R_{S_1}} u(r)\right\} + \left\{n_2 - \sum_{r \in R_{S_2}} u(r)\right\}$$

$$= (n_1 + n_2) - \sum_{r \in R_{S_1} \cup n_1 + R_{S_2}} u(r)$$

$$\geq n - \sum_{r \in R_S} u(r).$$

(ii) If there are $S' \in \mathcal{A}^*$ and $d \geq 2$ such that $S = S'^d$ and $C(S) = C(S')$, let $S' = S''^k$, where S'' is primitive. Then since $l(S'') = m < n$, $n = mdk$ and $\{((k-1)m + 1, (k+1)m), (km + 1, (k+2)m), \ldots, ((dk-2)m + 1, dkm)\} \subset R_S$,

$$C(S) = C(S') \geq km - \sum_{r \in R_{S'}} u(r)$$

$$= n - m(dk - k) - \sum_{r \in R_{S'}} u(r)$$

$$\geq n - \sum_{i=k-1}^{dk-2} u(1 + im, (i+2)m) - \sum_{r \in R_{S'}} u(r)$$

$$\geq n - \sum_{r \in R_S} u(r).$$

The second inequality from below holds since $u(1 + im, (i + 2)m) = m$ for $k - 1 \leq i \leq dk - 2$. $\square$

Corollary 3.2. *For* $n \geq 1$,

$$M[C(\mathcal{A}^n)] = \frac{\sum_{S \in \mathcal{A}^n} C(S)}{|\mathcal{A}|^n} \geq n - \sum_{u=1}^{\left[\frac{n}{2}\right]} u \times \left(\frac{\sum_{S \in \mathcal{A}^n} \#(S, u)}{|\mathcal{A}|^n} \right).$$

Proof. Since $C(S) \geq n - \sum_{r \in R_S} u(r)$ by Theorem 3.1,

$$M[C(\mathcal{A}^n)] = \frac{\sum_{S \in \mathcal{A}^n} C(S)}{|\mathcal{A}|^n} \geq \frac{\sum_{S \in \mathcal{A}^n} \left(n - \sum_{r \in R_S} u(r) \right)}{|\mathcal{A}|^n}$$

$$\geq n - \frac{\sum_{S \in \mathcal{A}^n} \sum_{r \in R_S} u(r)}{|\mathcal{A}|^n}$$

$$= n - \frac{\sum_{S \in \mathcal{A}^n} \sum_{u=1}^{\left[\frac{n}{2}\right]} u \times \#(S, u)}{|\mathcal{A}|^n}$$

$$= n - \sum_{u=1}^{\left[\frac{n}{2}\right]} u \times \left(\frac{\sum_{S \in \mathcal{A}^n} \#(S, u)}{|\mathcal{A}|^n} \right). \qquad \square$$

Now we are going to simplify $\frac{\sum_{S \in \mathcal{A}^n} \#(S, u)}{|\mathcal{A}|^n}$. Let $f_k(n)$ be the number of all primitive sequences in $\mathcal{A}^n$ when $|\mathcal{A}| = k$. For example, when $\mathcal{A} = \{a, b\}$, the number $f_2(4)$ of all primitive sequences in $\mathcal{A}^4$ is $f_2(4) = 2^4 - 4 = 12$ since the repetitive sequences are *aaaa*, *bbbb*, *abab*, *baba*. When $\mathcal{A} = \{a, b, c\}$, the number of all primitive sequences in $\mathcal{A}^6$ is $f_3(6) = 3^6$

$-33 = 696$ since the repetitive sequences are of the form $(S_1)^2$, $(S_2)^3$, $(S_3)^6$ for some primitive substrings S_1, S_2, S_3 of S and so the number of repetitive sequences is 33.

We find a formula evaluating the value of $f_k(n)$ in the following lemma.

Lemma 3.3 [3]**.** *Let* $f_k(n)$ *be the number of all primitive sequences in* $\mathcal{A}^n$ *when* $|\mathcal{A}| = k$. *We obtain*

$$f_k(n) = \sum_{d \mid n} \mu(d) k^{\frac{n}{d}}, \tag{3.2}$$

where μ *is a Möbius function defined by*

$$\mu(n) = \begin{cases} 1, & \text{for } n = 1, \\ 0, & n = cp^2,\ p \text{ is some prime number}, \\ (-1)^r, & n = p_1 p_2 p_3 \cdots p_r,\ p_i\text{'s are prime numbers}. \end{cases}$$

Proof. Möbius inverse formula says that if $g(n) = \sum_{d \mid n} f(d)$ for function *f*, *g*, then $f(n) = \sum_{d \mid n} \mu(d) g\left(\frac{n}{d}\right)$, $n \geq 1$. Now let $f(d) = f_k(d)$. Then $g(n) = \sum_{d \mid n} f_k(d) = k^n$. For let A_d be the set of elements $S \in A^n$ which are the form of $S = (S')^d$ for some primitive substring S' of S. Then $|\mathcal{A}_d| = f_k\left(\frac{n}{d}\right)$ and $\bigcup_{d \mid n} \mathcal{A}_d = \mathcal{A}^n$ and so we have

$$\begin{aligned} k^n &= |S^n| = |\textstyle\bigcup_{d \mid n} \mathcal{A}_d| \\ &= \sum_{d \mid n} |\mathcal{A}_d| = \sum_{d \mid n} f_k\left(\frac{n}{d}\right) = \sum_{d \mid n} f_k(d). \end{aligned}$$

Thus by Möbius inverse formula, we obtain

$$f_k(n) = \sum_{d \mid n} g\left(\frac{n}{d}\right) \mu(d) = \sum_{d \mid n} \mu(d) k^{\frac{n}{d}}. \qquad \square$$

Lemma 3.4. *When* $|\mathcal{A}| = k,$ *we have*

$$\sum_{S \in \mathcal{A}^n} \frac{\#(S, u)}{|\mathcal{A}|^n} = (n - 2u + 1) f_k(u) k^{-2u},$$

where $f_k(u)$ *can be obtained by* (3.2).

Proof. It is sufficient to prove that

$$\sum_{S \in \mathcal{A}^n} \#(S, u) = (n - 2u + 1) k^{n-2u} f_k(u).$$

Note that for $r = (i, j) \in R_S$ with $u(r) = u,$ we know that $j = i + 2u - 1$ and the substring $S(i, j)$ of S is denoted by $S(i, j) = (S')^2$ for primitive substring S' of S.

Let $\chi(S, S', i)$ be 1 if $S(i, i + 2u - 1) = S'^2$ and 0 elsewhere. We get

$$\begin{aligned}
\sum_{S \in \mathcal{A}^n} \#(S, u) &= \sum_{S \in \mathcal{A}^n} \sum_{i=1}^{n-2u+1} \sum_{l(S')=u,\, S':primitive} \chi(S, S', i) \\
&= \sum_{i=1}^{n-2u+1} \sum_{l(S')=u,\, S':primitive} \sum_{S \in \mathcal{A}^n} \chi(S, S', i) \\
&= \sum_{i=1}^{n-2u+1} \sum_{l(S')=u,\, S':primitive} k^{n-2u} \\
&= \sum_{i=1}^{n-2u+1} f_k(u) k^{n-2u} \\
&= (n - 2u + 1) f_k(u) k^{n-2u}
\end{aligned}$$

and so the result is obtained. □

Theorem 3.5. *If* $|\mathcal{A}| = k \geq 3,$ *we have*

$$M[C(\mathcal{A}^n)] > Kn,$$

where

$$K = 1 - \sum_{u=1}^{\infty} u f_k(u) k^{-2u}$$

$$= 1 - \sum_{a=1}^{\infty} a\mu(a) \frac{k^{1-2a}}{(1-k^{1-2a})^2} \geq 1 - \frac{k}{(k-1)^2} > 0$$

and μ *is a Möbius function.*

Proof. Since $2u - 1 \geq 1$ for any $u \in \mathbb{N}$ *and so* $n - 2u + 1 > n$,

$$M[C(\mathcal{A}^n)] = \frac{\sum_{S \in \mathcal{A}^n} C(S)}{|\mathcal{A}|^n} \geq n - \sum_{u=1}^{\left[\frac{n}{2}\right]} u \times \left(\frac{\sum_{S \in \mathcal{A}^n} \#(S, u)}{|\mathcal{A}|^n} \right)$$

$$= n - \sum_{u=1}^{\left[\frac{n}{2}\right]} u \times (n - 2u + 1) \times f_k(u) k^{-2u}$$

$$> n - \sum_{u=1}^{\left[\frac{n}{2}\right]} u \times n \times f_k(u) k^{-2u}$$

$$\geq n\left(1 - \sum_{u=1}^{\infty} u \times f_k(u) k^{-2u} \right).$$

Now if we let $K = 1 - \sum_{u=1}^{\infty} u f_k(u) k^{-2u}$, then we obtain $M[C(\mathcal{A}^n)] > Kn.$

Furthermore, since $\sum_{n=1}^{\infty} n x^n = \dfrac{x}{(1-x)^2}$,

$$\sum_{u=1}^{\infty} u f_k(u) k^{-2u} = \sum_{u=1}^{\infty} \frac{u f_k(u)}{k^{2u}}$$

$$< \sum_{u=1}^{\infty} u \frac{k^u}{k^{2u}} = \sum_{u=1}^{\infty} \frac{u}{k^u}$$

$$= \frac{\frac{1}{k}}{\left(1 - \frac{1}{k}\right)^2} = \frac{k}{(k-1)^2},$$

we can see $K = 1 - \sum_{u=1}^{\infty} u f_k(u) k^{-2u} \geq 1 - \frac{k}{(k-1)^2} > 0$. Also since $f_k(n) = \sum_{d \mid n} \mu(d) k^{\frac{n}{d}}$ by (3.2), we have

$$K = 1 - \sum_{u=1}^{\infty} u \left(\sum_{d \mid u} \mu(d) k^{\frac{u}{d}} \right) k^{-2u}$$

$$= 1 - \sum_{u=1}^{\infty} u \left(\sum_{u=ab} \mu(a) k^{b} \right) k^{-2u}$$

$$= 1 - \sum_{u=1}^{\infty} \sum_{u=ab} u\mu(a) k^{b} k^{-2u}$$

$$= 1 - \sum_{ab=1}^{\infty} ab\mu(a) k^{b} k^{-2ab}$$

$$= 1 - \sum_{a=1}^{\infty} \sum_{b=1}^{\infty} ab\mu(a) k^{b} k^{-2ab}$$

$$= 1 - \sum_{a=1}^{\infty} a\mu(a) \sum_{b=1}^{\infty} b(k^{1-2a})^{b}$$

$$= 1 - \sum_{a=1}^{\infty} a\mu(a) \frac{k^{1-2a}}{(1 - k^{1-2a})^2}. \qquad \square$$

Lemma 3.6 (Fekete's Lemma)**.** *If for sequence* $\{a_n\}_{n \in \mathbb{N}}$ *in* $\mathbb{R}$, $a_{n+m} \leq a_n + a_m$ *for all* $n, m \in \mathbb{N}$, *then the following holds*:

$$\lim_{n \to \infty} \frac{a_n}{n} = \inf_{n \in \mathbb{N}} \frac{a_n}{n}.$$

Theorem 3.7. (1) *The average of complexities of sequences in* $\mathcal{A}^{m+n}$ *is equal or less than the sum of the average of complexities of sequences in* $\mathcal{A}^{m}$ *and the average of complexities of sequences in* $\mathcal{A}^{n}$, *i.e.*

$$M(C(\mathcal{A}^{m+n})) \leq M(C(\mathcal{A}^{m})) + M(C(\mathcal{A}^{n})).$$

(2) *The maximum of complexities of sequences in* A^{m+n} *is equal or less than the sum of the maximum of complexities of sequences in* $\mathcal{A}^{m}$ *and the maximum of complexities of sequences in* $\mathcal{A}^{n}$, *i.e.*

$$\max(C(\mathcal{A}^{m+n})) \leq \max(C(\mathcal{A}^{m})) + \max(C(\mathcal{A}^{n})),$$

where $\max(C(\mathcal{A}^{n})) = \max\{C(S) : S \in \mathcal{A}^{n}\}.$

Proof. (1) By Definition 2.1 of $C(S)$, we can get easily

$$\begin{aligned}
M(C(\mathcal{A}^{m+n})) &= \frac{\sum_{S \in \mathcal{A}^{m+n}} C(S)}{|\mathcal{A}|^{m+n}} \\
&\leq \frac{1}{|\mathcal{A}|^{m+n}} \sum_{S \in \mathcal{A}^{m+n}} \{C(S(1, m)) + C(S(m+1, m+n))\} \\
&= \frac{\sum_{S \in \mathcal{A}^{m+n}} C(S(1, m))}{|\mathcal{A}|^{m+n}} + \frac{\sum_{S \in \mathcal{A}^{m+n}} C(S(m+1, m+n))}{|\mathcal{A}|^{m+n}} \\
&= \frac{\sum_{S \in \mathcal{A}^{m}} |\mathcal{A}|^{n} C(S)}{|\mathcal{A}|^{m+n}} + \frac{\sum_{S \in \mathcal{A}^{n}} |\mathcal{A}|^{m} C(S)}{|\mathcal{A}|^{m+n}} \\
&= \frac{\sum_{S \in \mathcal{A}^{m}} C(S)}{|\mathcal{A}|^{m}} + \frac{\sum_{S \in \mathcal{A}^{n}} C(S)}{|\mathcal{A}|^{n}} \\
&= M(C(\mathcal{A}^{m})) + M(C(\mathcal{A}^{n})).
\end{aligned}$$

(2) Let S be the sequence with $\max(C(\mathcal{A}^{m+n})) = C(S)$. Then by 2.2 we have

$$\begin{aligned}\max(C(\mathcal{A}^{m+n})) &= C(S) = C(S(1,\ m)S(m+1,\ m+n))\\ &\le C(S(1,\ m)) + C(S(m+1,\ m+n))\\ &\le \max(C(\mathcal{A}^{m})) + \max(C(\mathcal{A}^{n})). \qquad \square\end{aligned}$$

Now we obtain the approximation of maximum and average of complexities of sequences in $\mathcal{A}^n$.

Remark 3.8. Let $a_n = M(C(\mathcal{A}^n))$ and $|\mathcal{A}| = k$. Then we can see that $\{a_n\}$ is a lower bounded sequence with $a_{n+m} \le a_n + a_m$ for $n,\ m \in \mathbb{N}$ from Theorem 3.7. Hence by Fekete's Lemma 3.6,

$$\lim_{n->\infty} \frac{M(C(\mathcal{A}^n))}{n} = \inf_{n\in\mathbb{N}} \frac{M(C(\mathcal{A}^n))}{n} = \alpha,$$

where α is dependent on $|\mathcal{A}| = k$. Now, from Theorem 3.7, since $1 - \frac{k}{(k-1)^2} \le K \le \inf_{n\in\mathbb{N}} \frac{M(C(\mathcal{A}^n))}{n} = \alpha \le 1$, we can see that α converges to 1 as $|\mathcal{A}| = k$ increases.

Remark 3.9. Let $b_n = \max(C(\mathcal{A}^n))$. Then we can see that $\{b_n\}$ is a lower bounded sequence with $b_{n+m} \le b_n + b_m$ for $n,\ m \in \mathbb{N}$ from Theorem 3.7. Hence by Fekete's Lemma 3.6,

$$\lim_{n->\infty} \frac{\max(C(\mathcal{A}^n))}{n} = \inf_{n\in\mathbb{N}} \frac{\max(C(\mathcal{A}^n))}{n} = \beta.$$

Now, since the maximum is bigger than the average, $\beta > \alpha \ge K > 0$ when $|\mathcal{A}| \ge 3$.

Remark 3.10. When $|\mathcal{A}| = 2$, we can get $K < 0$ but $\alpha \approx 0.35$ by random sampling.

References

[1] A. Lempel and J. Ziv, On the complexity of finite sequences, IEEE Trans. Inform. Theory IT 22(1) (1976), 75-81.

[2] J.-P. Allouche, Surveying some notions of complexity for finite and infinite sequences, available online at http://www.math.jussieu.fr/allouche/japon.pdf

[3] J.-P. Allouche and J. Shallit, Automatic Sequence, Theory, Applications, Generalization, Cambridge University Press, Cambridge, 2003.