Rewriting Systems with Low Derivational Complexity *

Yuji Kobayashi

Laboratory of Mathematics and Games
https//: math-game-labo.com

1 Rewriting systems and complexity

Let A be an alphabet, a finite set of letters and let $A^* = \{a_1 a_2 \dots a_n \mid n \geq 0, a_i \in A\}$ be the free monoid generated by A. The empty word in A^* is denoted by 1. We denote by $|x|$ the length n of a word $x = a_1 a_2 \dots a_n \in A^*$.

A rewriting system R on A is a subset of $A^* \times A^*$. An element $r = (u, v)$ of R is called a rule and is written as $u \to v$. R is finite if it is a finite set. For two words x and y in A^*, if $x = x_1 u x_2$, $y = x_1 v x_2$ with $x_1, x_2 \in A^*$, we write as $x \to_r y$. If there are words $x_1, \dots, x_{k-1} \in A^*$ and rules $r_1, \dots, r_k \in R$ such that

$$x = x_0 \to_{r_1} x_1 \to_{r_2} \cdots \to_{r_{k-1}} x_{k-1} \to_{r_k} x_k = y, \tag{1}$$

we write as $x \to_R^k y$ or simply $x \to^k y$. We call (1) a derivation sequence in R of length k and say that y is derived from x for k steps. If there is no sequence of length larger than k starting with x, (1) is called *maximal*.

For $x \in A^*$ the *derivational length* $\delta_R(x)$ of x is the length of a maximal sequence starting with x, that is,

$$\delta_R(x) = \max\{k \mid \exists y \in A^*,\ x \to_R^k y\}.$$

The (*derivational*) *complexity* d_R of R is defined by the function that relates the largest length of derivation sequences in R to the length of starting words;

$$d_R(n) = \max\{\delta_R(x) \mid x \in A^*, |x| = n\}$$

(see [1] and [2]). If $\delta_R(x) < \infty$ for all $x \in A^*$, R is called *terminating*. If R is terminating, d_R is a function from $\mathbb{N}$ to $\mathbb{N}$.

For two functions $f, g : \mathbb{N} \to \mathbb{N}$, we write $f = O(g)$ (resp. $f = \Omega(g)$), if there is a constant $C > 0$ such that $f(n) \leq Cg(n)$ (resp. $f(n) \geq Cg(n)$) for sufficiently large n. We say f and g are *equivalent*, and write as $f \sim g$ or $f = \Theta(g)$ if $f = O(g)$ and $f = \Omega(g)$.

*this is a preliminary version and a full version will appear elsewhere

Example 1.1. (1) The system $R = \{a \to 1\}$ on $\{a\}$ has linear complexity, in fact, $\delta_R(a^n) = n$ and $d_R(n) = n$.

(2) Any nonempty system R has at least linear complexity, that is, $d_R(n) = \Omega(n)$.

(3) The system $R = \{ab \to ba\}$ on $\{a, b\}$ has quadratic complexity. In fact, $\delta_R(a^n b^n) = n^2$ and $d_R(n) = \frac{1}{4}n^2 = \Theta(n^2)$.

(4) The system $R = \{ab \to b^2 a\}$ on $\{a, b\}$ has exponential complexity. In fact, $\delta_R(a^n b^n) = n(2^n - 1)$ and $\Omega(2^n) = d_R(n) = O(3^n)$.

Kobayashi [2] proved that for any real number $\alpha \geq 2$ there is a finite rewriting system with complexity equivalent to n^α, if computational complexity of α is not very high (bounded by C^{2^n} for some $C > 1$), and posed the following problem.

Question 1.2. For a real number α with $1 < \alpha < 2$, is there a finite rewriting system with complexity equivalent to n^α?

Recently, Talambutsa [3] has given a positive answer for any rational α with $1 < \alpha < 2$. That is, for any rational number $\alpha \geq 1$ there is a finite rewriting system with complexity $\Theta(n^\alpha)$.

To his end he constructed a supplementary system which is length-preserving and has complexity $\Theta(n \log n)$. In the next section we give a little different system with this complexity whose mechanism will appear in the system with complexity $\Theta(n \log \log n)$ given in the last section.

2 System with complexity $n \log n$

Consider an alphabet

$$A_1 = \{a, \bar{a}, h, p, v, w\}$$

and a system

$$R_0 = \{a^2 h \to h\bar{a},\ wh \to wp,\ p\bar{a} \to ap\}$$

over A_1. Let $x = wa^n hv$ with even number $n \geq 0$, then we have a derivation sequence

$$x = wa^n hv \to\ wa^{n-2}h\bar{a}v \to^{\frac{n}{2}-1} wh\bar{a}^{\frac{n}{2}}v \to wp\bar{a}^{\frac{n}{2}}v \to^{\frac{n}{2}} wa^{\frac{n}{2}}pv$$

in R_0. This is a maximal sequence starting with x, in which h travels for $n/2$ steps from right to left, and at the left end it changes to p and returns to the original position (the pair (h, p) *shuttles* once between v and w). Thus,

$$\delta_{R_0}(x) = n + 1.$$

Adding a new rule $r_0 = (apv, ahv)$ to R_0, set

$$R_1 = R_0 \cup \{apv \to ahv\}.$$

Suppose $n = 2^i$ with $i \geq 1$ and let $x = wa^n hv$, then we have a maximal derivation sequence

$$x \to_{R_0}^{n+1} wa^{\frac{n}{2}}pv \to_{r_0} wa^{\frac{n}{2}}hv \to_{R_0}^{\frac{n}{2}+1} wa^{\frac{n}{4}}pv \to_{r_0} wa^{\frac{n}{4}}hv \to_{R_0} \cdots \to_{r_0} wahv$$

in R_1. In this sequence the pair (h, p) shuttles $i = \log_2 n$ times between v and w, and we have

$$\delta_{R_1}(x) = 2^i + 2^{i-1} + \cdots + 2 + 2i = 2^{i+1} + 2i - 2 = \Theta(n). \quad (2)$$

Next, let $A_2 = \{b, \bar{b}, f, q, v, w\}$, and consider a system

$$R_2 = \{fb \to \bar{b}f,\ fw \to qw,\ \bar{b}q \to qb\}.$$

For a word $x = vfb^n w\,(n \geq 1)$ we have a maximal sequence

$$x \to v\bar{b}fb^{n-1}w \to^{n-1} v\bar{b}^n fw \to v\bar{b}^n qw \to^n vqb^n w$$

in R_2. In the sequence the pair (f, q) shuttles once between w and v, and we have

$$\delta_{R_2}(x) = 2n + 1. \quad (3)$$

Now let

$$A_3 = A_1 \cup A_2 = \{a, \bar{a}, b, \bar{b}, h, p, f, q, v, w\},$$

and define a system R_3 by adding a rule $r_1 = (apvq, ahvf)$ to the union of R_0 and R_1, that is,

$$\begin{aligned} R_3 &= R_0 \cup R_2 \cup \{r_1\} \\ &= \{a^2h \to h\bar{a}, wh \to wp, p\bar{a} \to ap, fb \to \bar{b}f, fw \to qw, \bar{b}q \to qb, apvq \to ahvf\}. \end{aligned}$$

Let $n = 2^i\,(i \geq 1)$ and $x = wa^n hvfb^n w \in A_2^*$. We have a maximal sequence

$$\begin{aligned} x \quad &\to_{R_0}^{n+1} wa^{\frac{n}{2}}pvfb^n w \to_{R_2}^{2n+1} wa^{\frac{n}{2}}pvqb^n w \to_{r_1} wa^{\frac{n}{2}}hvfb^n w \\ &\to_{R_0}^{\frac{n}{2}+1} wa^{\frac{n}{4}}pvfb^n w \to_{R_2}^{2n+1} wa^{\frac{n}{4}}pvqb^n w \to_{r_1} wa^{\frac{n}{4}}hvfb^n w \\ &\to_{R_0} \ldots \to_{r_1} wahvfb^n w \to_{R_2}^{2n+1} wahvqb^n w. \end{aligned} \quad (4)$$

in R_3. In (4) the movements in the left side and in the right of v synchronize, one shuttle of (h, p) in the left corresponds to one shuttle of (f, q) in the right. The number of the shuttlings of (h, p) is $i = \log_2 n$ and the number of derivation steps in them is $O(n)$ by (2) above. The number of applications of the rule r_1 is i, and the number of shuttlings of (f, q) in the right side is also i. Hence, the number of steps in the shuttling s of (f, g) is $(2n + 1)\log_2 n$ by (3). The length of the sequence (4) is the sum of these numbers of steps and is dominated by the last number, and hence we see $\delta_{R_3}(x) = \Theta(n \log n)$. Because (4) gives the maximum length relative to the length of the starting word among all sequences in R_3 (the details are omitted), we see

$$d_{R_3}(n) = \Theta(n \log n).$$

Talambutsa asked about the existence of a finite system with complexity strictly between $\Theta(n)$ and $\Theta(n \log n)$. In the next section we give a system with complexity $n \log \log n$.

3 System with complexity $n \log \log n$

Let

$$A_4 = \{b, \bar{b}, \bar{\bar{b}}, c, \bar{c}, f, q, v, w\},$$

and consider a system R_4 over A_4 similar to R_2:

$$R_4 = \{f\bar{b} \to \bar{\bar{b}}f,\ fc \to \bar{c}f,\ fw \to qw,\ \bar{\bar{b}}q \to qb, \bar{c}q \to qc\}.$$

For a word $x = vf\bar{b}^m c^n w\,(m, n \geq 0)$ we have a maximal sequence

$$x \to^{m+n} v\bar{\bar{b}}^m \bar{c}^n fw \to v\bar{\bar{b}}^m \bar{c}^n qw \to^{m+n} vqb^m c^n w. \tag{5}$$

In (5) the pair (f, q) shuttles once between w and v, and we have

$$\delta_{R_4}(x) = 2(m+n) + 1.$$

Next, let

$$A_5 = \{b, \bar{b}, \bar{\bar{b}}, c, g, r, v, w\},$$

and

$$R_5 = \{gb \to \bar{b}g,\ g\bar{b} \to \bar{\bar{b}}g,\ gc \to r\bar{b}^2, \bar{b}r \to rb,\ \bar{\bar{b}}r \to r\bar{b}\}.$$

Let $x = vgb^m c^n w$ with $m \geq 0, n \geq 1$. Then, we have

$$x \to^m v\bar{b}^m gc^n w \to v\bar{b}^m r\bar{b}^2 c^{n-1} w \to^m vrb^m \bar{b}^2 c^{n-1} w.$$

In this sequence the pair (g, r) shuttles once between v and c, and

$$\delta_{R_5}(x) = 2m + 1.$$

Let $A_6 = A_1 \cup A_5$ and let R_6 be the union of R_0 and R_5 adding a rule $r_2 = (apvrb, ahvg)$;

$$R_6 = R_0 \cup R_5 \cup \{apvrb \to ahvg\}.$$

Let $i, j > m \geq 0$ and $n = 2^i$. For a word $x = wa^n hvgb^m c^j w \in A_6^*$ we have

$$\begin{aligned} x \quad &\to_{R_0}^{n+1} wa^{2^{i-1}} pvgb^m c^j w \to_{R_5}^{2m+1} wa^{2^{i-1}} pvrb^m \bar{b}^2 c^{j-1} w \\ &\to_{r_2} wa^{2^{i-1}} hvgb^{m-1} \bar{b}^2 c^{j-1} w \to_{R_5}^{2^{i-1}+2(m+2)} wa^{2^{i-2}} pvrb^{m-1} \bar{b}^4 c^{j-2} w \\ &\to_{r_2} \cdots \to_{R_5} wa^{2^{i-m-1}} pvr\bar{b}^{2m+2} c^{j-m-1} w = y. \end{aligned} \tag{6}$$

In this situation we write $x \Longrightarrow^{(6)} y$. In (6) the pairs (h, p) and (g, r) both shuttle $m+1$ times between v and w, and the number of steps in the shuttlings of (g, r) is

$$\delta_{R_6}(x) = 2(m + (m+1) + \cdots + (m+m)) + m + 1 = \Theta(m^2). \tag{7}$$

Finally, let

$$A_7 = A_1 \cup A_4 \cup A_5 = \{a, \bar{a}, b, \bar{b}, \bar{\bar{b}}, c, \bar{c}, h, p, f, q, g, r, v, w\},$$

and let $r_3 = (apvqb, ahvg)$ and $r_4 = (apvr\bar{b}, ahvf\bar{b})$. Define

$$\begin{aligned} R_7 &= R_0 \cup R_4 \cup R_6 \cup \{r_3, r_4\} \\ &= \{\ a^2h \to h\bar{a},\ wh \to wp,\ p\bar{a} \to ap, \\ &\quad f\bar{b} \to \bar{\bar{b}}f,\ fc \to \bar{c}f,\ fw \to qw,\ \bar{\bar{b}}q \to qb,\ \bar{c}q \to qc, \\ &\quad gb \to \bar{b}g,\ g\bar{b} \to \bar{\bar{b}}g,\ gc \to r\bar{b}^2,\ \bar{b}r \to rb,\ \bar{\bar{b}}r \to r\bar{b}, \\ &\quad apvrb \to ahvg,\ apvqb \to ahvg,\ apvr\bar{b} \to ahvf\bar{b}\ \}. \end{aligned}$$

Let $n = 2^i (i \geq 1)$ and $x = wa^n hvf\bar{b}c^n w$. We have a maximal sequence

$$\begin{aligned} x \ &\to_{R_0}^{n+1} wa^{2^{i-1}} pvf\bar{b}c^n w \to_{R_4}^{2n+3} wa^{2^{i-1}} pvqbc^n w \to_{r_3} wa^{2^{i-1}} hvgc^n w \\ &\to_{R_0}^{2^{i-1}+1} wa^{2^{i-2}} pvgc^n w \to_{R_6} wa^{2^{i-2}} pvr\bar{b}^2 c^{n-1} w \to_{r_4} wa^{2^{i-2}} hvf\bar{b}^2 c^{n-1} w \\ &\to_{R_0}^{2^{i-2}+1} wa^{2^{i-3}} pvf\bar{b}^2 c^{n-1} w \to_{R_4}^{2n+3} wa^{2^{i-3}} pvqb^2 c^{n-1} w \\ &\to_{r_3} wa^{2^{i-3}} hvgbc^{n-1} w \Longrightarrow^{(6)} wa^{2^{i-5}} pvr\bar{b}^4 c^{n-3} w \\ &\to_{r_4} wa^{2^{i-5}} hvf\bar{b}^4 c^{n-3} w \to_{R_0} \cdots \to_{R_4} wa^{2^{i-6}} pvqb^4 c^{n-3} w \\ &\to_{r_3} wa^{2^{i-6}} hvgb^3 c^{n-3} w \Longrightarrow^{(6)} wa^{2^{i-10}} pvr\bar{b}^8 c^{n-7} w \\ &\to_{R_7} \cdots \to_{R_7} wahvsb^{2^{j-1}-k} \bar{b}^{2k} c^{n-\ell} w \end{aligned} \tag{8}$$

in R_7. Here, $0 \leq k \leq 2^{j-1}$, j is the number of the shuttlings of the pair (f, q), ℓ is the number of shuttlings of (g, r), and $s = q$ if $k = 0$ and $s = r$ otherwise. Moreover, the pair (g, r) shuttles 2^{t-1} times after the t-th shuttling of (f, q) for $t < j$ and shuttles k times after the last j-th shuttling of (f, q). Thus we see

$$\ell = 1 + 2 + \cdots + 2^{j-2} + k.$$

Now, in the left side of the letter v in (8), the pair (h, p) shuttles $i = \log_2 n$ times, and corresponding to it, in the right side the pairs (f, q) and (g, r) shuttle $i + 1$ times together. Hence,

$$i + 1 = j + \ell = j + 2^{j-1} - 1 + k, \tag{9}$$

and so

$$j = \Theta(\log i) = \Theta(\log \log n).$$

Thus, the number of the steps in the shuttlings of (f, g) in (8) is $(2n + 3)j = \Theta(n \log \log n)$. On the other hand, the number of the steps in the shuttlings of (g, r) is $O(\ell^2)$ by (7) and by (9) it equals $O(2^{2j}) = O(i^2) = O(\log^2 n)$, and the number of the steps in the shuttlings of (h, p) is $O(n)$ by (2). Further, the rules r_2, r_3 and r_4 are applied $i = O(\log n)$ times altogether. To estimate $\delta_{R_7}(x)$, we can ignore these numbers and we may only take the shuttling of (g, r) into account. Thus, we see $\delta_{R_7}(x) = \Theta(n \log \log n)$. Because words of the form of x give the maximum derivation length relative to the length of the words, we finally have

$$d_{R_7}(n) = \Theta(n \log \log n).$$

References

[1] D. Hofbauer, C. Lautermann, Termination proofs and the length of derivations, RTA 1989, LNCS **355**, (1989), 167–177.

[2] Y. Kobayashi, The derivational complexity of string rewriting systems, Theoret. Comp. Sci., **438** (2012), 1–12.

[3] A. Talambutsa, On subquadratic derivational complexity of semi-Thue systems, Proc. Computer Science in Russia-2020, LNCS, to appear.