<table>
<thead>
<tr>
<th>Title</th>
<th>A Language Equation and Its Applications (Algorithms in Algebraic Systems and Computation Theory)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Ito, Masami</td>
</tr>
<tr>
<td>Citation</td>
<td>数理解析研究所講究録 (2002), 1268: 7-10</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2002-06</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/2433/42128">http://hdl.handle.net/2433/42128</a></td>
</tr>
<tr>
<td>Type</td>
<td>Departmental Bulletin Paper</td>
</tr>
<tr>
<td>Textversion</td>
<td>publisher</td>
</tr>
<tr>
<td></td>
<td>Kyoto University</td>
</tr>
</tbody>
</table>
A Language Equation and Its Applications

Masami Ito
Department of Mathematics
Kyoto Sangyo University
Kyoto 603-8555, Japan
Email: ito@ksuvx0.kyoto-su.ac.jp

Let $u, v \in X^*$ be words over an alphabet $X$. Then the set $\{u_1v_1u_2v_2\ldots u_nv_n \mid u = u_1u_2\ldots u_n, v = v_1v_2\ldots v_n, u_1, u_2, v_1, v_2, \ldots, u_n, v_n \in X^*, n \geq 1\}$ is called the shuffle product of $u$ and $v$, and denoted by $u \circ v$. For languages $A, B \subseteq X^*$, the set $A \circ B = \bigcup_{u \in A, v \in B} u \circ v$ is called the shuffle product of $A$ and $B$. In this paper, we consider the following problem: Let $A, B \subseteq X^*$ be regular languages. Then can we obtain a solution $C \subseteq X^*$ of the language equation $A = B \circ C$? Obviously, this problem is equivalent to the shuffle decomposition problem for regular languages. Regarding definitions and notations concerning formal languages and automata, not defined in this paper, refer, for instance, to [1].

Now let $A = (S, X, \delta, s_0, F)$ be a finite automaton with $T(A) = A$ and let $B = (T, X, \gamma, t_0, G)$ be a finite automaton with $T(B) = B$. We will look for a regular language $C$ over $X$ such that $A = B \circ C$. By $\overline{X}$, we denote the language $\{\overline{a} \mid a \in X\}$ with $X \cap \overline{X} = \emptyset$. Let $\overline{B} = (T, X \cup \overline{X} \cup \{\#\}, \overline{\gamma}, t_0, G)$ where $\overline{\gamma}$ is defined as follows:

For $t \in T$ and $a \in X$, $\overline{\gamma}(t, a) = t$, $\overline{\gamma}(t, \overline{a}) = \gamma(t, a)$. Moreover, $\overline{\gamma}(t, \#) = t$ if $t \in G$.

Then the following can be easily shown.

**Fact 1** Let $a_1a_2\ldots a_n \in X^*$ where $a_i \in X, i = 1, 2, \ldots, n$. Then $a_1a_2\ldots a_n \in T(B)$ if and only if $u_1\overline{a}_1u_2\overline{a}_2\ldots u_n\overline{a}_n\overline{a}_{n+1}\# \in T(\overline{B})$ where $u_1, u_2, \ldots, u_n \in X^*$.

Let $A_1 = (\overline{S}, X \cup \overline{X} \cup \{\#\}, \overline{\delta}, s_0, \{\alpha, \omega\})$ and let $A_2 = (\overline{S}, X \cup \overline{X} \cup \{\#\}, \overline{\delta}, s_0, \{\alpha\})$ where $\overline{S} = (\cup_{a \in X \cup \{\epsilon\}} S^{(a)}) \cup \{\alpha, \omega\}$. Here $S^{(\epsilon)}$ is regarded as $S$ where $\epsilon$ is the empty word. For $s \in S, t \in S \setminus F, t' \in F, a \in X \cup \{\epsilon\}, b \in X$ and $\{\#\}, \overline{\delta}$ is defined as follows:
\[ \bar{\delta}(s^{(a)}, b) = \delta(s, b)^{(a)}, \quad \bar{\delta}(s^{(a)}, \#) = \delta(s, \#)^{(a)} \]

We consider the following two automata:
\[ C_1 = (S \times T, X \cup \overline{X} \cup \{\#\}, \delta \times \overline{\gamma}, (s_0, t_0), \{\alpha, \omega\} \times G), \quad C_2 = (S \times T, X \cup \overline{X} \cup \{\#\}, \delta \times \overline{\gamma}, (s_0, t_0), \{\alpha\} \times G) \]
where \( \delta \times \overline{\gamma}((s, t), a) = (\delta(s), \overline{\gamma}(t, a)) \)
for \((s, t) \in S \times T\) and \(a \in X\).

Now consider the following homomorphism \( \rho \) of \((X \cup \overline{X} \cup \{\#\})^*\) into \(X^*\):
\[ \rho(a) = a \text{ for } a \in X, \quad \rho(\overline{a}) = \epsilon \text{ for } a \in X \text{ and } \rho(\#) = \epsilon. \]

**Lemma 1** Automata accepting the languages \(\rho(T(C_1))\) and \(\rho(T(C_2))\) can be effectively constructed.

**Proof** Let \(i = 1, 2\). From \(C_i\), we can construct a regular grammar \(G_i\) such that \(L(G_i) = T(C_i)\) with the production rules of the form \(A \rightarrow aB\) (\(A, B\) are variables and \(a \in X \cup \overline{X} \cup \{\#\}\)). Replacing every rule of the form \(A \rightarrow aB\) in \(G_i\) by \(A \rightarrow \rho(a)B\), we can obtain a new grammar \(G'_i\). Then it is clear that \(\rho(L(C_i)) = L(G'_i)\). Using this grammar \(G'_i\), we can construct an automaton \(D_i\) such that \(T(D_i) = T(G'_i)\) i.e. \(\rho(T(C_i)) = T(D_i)\). Notice that all the above procedures are effectively done. This completes the proof of the lemma.

Let \(B, C \subseteq X^*\). By \(B \circ C\) we denote the shuffle product of \(B\) and \(C\), i.e.
\[ \{u_1v_1u_2v_2 \ldots u_nv_n \mid u = u_1u_2 \ldots u_n \in B, v = v_1v_2 \ldots v_n \in A\}. \]

**Proposition 1** Let \(u \in X^*\). Then \(\{u\} \circ B \subseteq A\) if and only if \(u \in \rho(T(C_1)) \setminus \rho(T(C_2))\).

**Proof** (\(\Rightarrow\)) Let \(u = u_1u_2 \ldots u_nu_{n+1} \in X^*\) and let \(a_1a_2 \ldots a_n \in B\) where \(u_1, u_2, \ldots, u_n, u_{n+1} \in X^*\) and \(a_1, a_2, \ldots, a_n \in X\). Then \(\delta \times \overline{\gamma}((s_0, t_0), u_1\overline{a_1}u_2\overline{a_2} \ldots u_n\overline{a_n}u_{n+1}\#) = (\delta(s_0, u_1\overline{a_1}u_2\overline{a_2} \ldots u_n\overline{a_n}u_{n+1}\#), \overline{\gamma}(t_0, u_1\overline{a_1}u_2\overline{a_2} \ldots u_n\overline{a_n}u_{n+1}\#)) \).

\(\overline{\gamma}(t_0, u_1\overline{a_1}u_2\overline{a_2} \ldots u_n\overline{a_n}u_{n+1}\#) = (\delta(s_0, u_1\overline{a_1}u_2\overline{a_2} \ldots u_n\overline{a_n}u_{n+1}\#), \gamma(t_0, a_1a_2 \ldots a_n)) \in \{\omega\} \times G\). Therefore, \(u_1\overline{a_1}u_2\overline{a_2} \ldots u_n\overline{a_n}u_{n+1}\# \in T(C_1) \setminus T(C_2)\). Hence \(u = u_1u_2 \ldots u_nu_{n+1} = \rho(u_1\overline{a_1}u_2\overline{a_2} \ldots u_n\overline{a_n}u_{n+1}\#) \in \rho(T(C_1)) \setminus \rho(T(C_2))\).

(\(\Leftarrow\)) Suppose that \(u \circ B \subseteq A\) does not hold though \(u \in \rho(T(C_1)) \setminus \rho(T(C_2))\). Then there exist \(u = u_1u_2 \ldots u_nu_{n+1} \in X^*\) and \(a_1a_2 \ldots a_n \in B\) such that \(u_1a_1u_2a_2 \ldots u_na_nu_{n+1} \notin A\). Hence \(\overline{\gamma}(t_0, u_1\overline{a_1}u_2\overline{a_2} \ldots u_n\overline{a_n}u_{n+1}\#) = \gamma(t_0, a_1a_2 \ldots a_n) \in G\). On the other hand, since \(u_1a_1u_2a_2 \ldots u_n\overline{a_n}u_{n+1} \notin A\), we have \(\delta(s_0, u_1\overline{a_1}u_2\overline{a_2} \ldots u_n\overline{a_n}u_{n+1}\#) = \overline{\delta}(s_0, u_1\overline{a_1}u_2\overline{a_2} \ldots u_n\overline{a_n}u_{n+1}\#)\)
\(u_{n+1})^{(a_n)}, \#) = \{\alpha\}.\) Hence \(\delta \times \gamma((s_0, t_0), u_1 \overline{a}_1 u_2 \overline{a}_2 \ldots u_n \overline{a}_n u_{n+1} \#) \in \{\alpha\} \times G,\)
i.e. \(u_1 \overline{a}_1 u_2 \overline{a}_2 \ldots u_n \overline{a}_n u_{n+1} \# \in T(C_2).\) Therefore, \(u = \rho(u_1 \overline{a}_1 u_2 \overline{a}_2 \ldots u_n \overline{a}_n u_{n+1} \#) \in \rho(T(C_2)).\) On the other hand, it is obvious that \(u_1 \overline{a}_1 u_2 \overline{a}_2 \ldots u_n \overline{a}_n u_{n+1} \# \in T(C_1).\) Thus \(u \notin \rho(T(C_1)) \setminus \rho(T(C_2)),\) a contradiction. Consequently, the proposition must hold true.

**Corollary** In the above, \(B \circ (\rho(T(C_1)) \setminus \rho(T(C_2))) \subseteq A.\)

Let \(L \subseteq X^*\) be a regular language over \(X.\) By \(#L,\) we denote the number \(\min\{|S| \mid \exists A = (S, X, \delta, s_0, F), L = \mathcal{T}(A)\}\) where \(|S|\) denotes the cardinality of \(S.\) Moreover, \(\mathcal{I}(n, X)\) denotes the class of languages \(\{L \subseteq X^* \mid \#L \leq n\} \).

**Theorem 1** Let \(A \subseteq X^*\) and let \(n\) be a positive integer. Then it is decidable whether there exist nontrivial regular languages \(B \in \mathcal{I}(n, X)\) and \(C \subseteq X^*\) such that \(A = B \circ C.\) Here a language \(D \subseteq X^*\) is said to be nontrivial if \(D \neq \{\epsilon\}.\)

**Proof** Let \(A \subseteq X^*\) be a regular language. Assume that there exist nontrivial regular languages \(B \in \mathcal{I}(n, X)\) and \(C \subseteq X^*\) such that \(A = B \circ C.\) Then, by Proposition 1 and its corollary, \(C \subseteq \rho(T(C_1)) \setminus \rho(T(C_2))\) and \(B \circ (\rho(T(C_1)) \setminus \rho(T(C_2))) \subseteq A.\) Hence \(A = B \circ (\rho(T(C_1)) \setminus \rho(T(C_2))).\) Thus we have the following algorithm: (1) Choose a nontrivial regular language \(B \subseteq X^*\) from \(\mathcal{I}(n, X)\) and construct the language \(\rho(T(C_1)) \setminus \rho(T(C_2))\) (see Lemma 1). (2) Let \(C = \rho(T(C_1)) \setminus \rho(T(C_2)).\) (3) Compute \(B \circ C.\) (4) If \(A = B \circ C,\) then the output is "YES" and "NO", otherwise. (4) If the output is "NO", then choose another element in \(\mathcal{I}(n, X)\) as \(B\) and continue the procedures (1)-(3). (5) Since \(\mathcal{I}(n, X)\) is a finite set, the above process terminates after a finite-step trial. Once one gets the output "YES", then there exist nontrivial regular languages \(B \in \mathcal{I}(n, X)\) and \(C \subseteq X^*\) such that \(A = B \circ C.\) Otherwise, there are no such languages.

Let \(n\) be a positive integer. By \(\mathcal{F}(n, X),\) we denote the class of finite languages \(\{L \subseteq X^* \mid \max\{|u| \mid u \in L\} \leq n\}\) where \(|u|\) is the length of \(u.\) Then the following result by C. Câmpeanu et al. ([2]) can be obtained as a corollary of the above theorem.

**Corollary** For a given positive integer \(n\) and a regular language \(A \subseteq X^*,\)
the problem whether \(A = B \circ C\) for a nontrivial language \(B \in \mathcal{F}(n, X)\) and
a nontrivial regular language \(C \subseteq X^*\) is decidable.
Proof Obvious from the fact that $\mathcal{F}(n, X) \subseteq \mathcal{I}(|X|^{n+1}, X)$.

References
