Литература

[1] Данг Зуй Руан, *Сложность конечного автомата, соответствующего обобщенному регулярному выражению*, Докл. АН СССР 213, 1.
[2] В. М. Глушков, *Абстрактная теория автоматов*, Успехи матем. наук. 16, 5 (101) (1961), 3–61.

Presented to the Semester
Discrete Mathematics
(February 15–June 16, 1977)

DISCRETE MATHEMATICS
BANACH CENTER PUBLICATIONS, VOLUME 7
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1982

RATIONAL STOCHASTIC AUTOMATA IN FORMAL LANGUAGE THEORY

PAAVO TURAKAINEN

Department of Mathematics, University of Oulu, Oulu, Finland

The purpose of this paper is to study language families which are obtained by applying arbitrary or bounded or λ-free homomorphisms to languages accepted by rational stochastic automata.

1. Definitions and preliminary results

In what follows, every alphabet X will be a finite subset of a fixed infinite set of abstract symbols. For any word P in X^*, $|P|$ means the length of P, $|P|_a$ is the number of occurrences of the letter a in P, and mi$P$ denotes the mirror image (or the reversal) of P. For the empty word we shall use the symbol λ.

The notions of a pre-AFL, an AFL and a principal AFL are defined as in Ginsburg (1975). The families of linear context-free, context-free, quasi-realtime (Book and Greibach, 1970), deterministic lba and recursively enumerable languages are denoted by $\mathscr{L}_{\mathrm{LIN}}$, $\mathscr{L}_{\mathrm{CF}}$, $\mathscr{L}_{\mathrm{QRT}}$, $\mathscr{L}_{\mathrm{DCS}}$ and $\mathscr{L}_{\mathrm{RE}}$, respectively.

A *stochastic automaton* is a quintuple $A = (X, S, M, \pi, f)$ where X is an alphabet, S is a finite set of states, M is a mapping from X into the set of stochastic $|S| \times |S|$ matrices, π is a stochastic $1 \times |S|$ vector, and f is a $|S| \times 1$ vector consisting of 0's and 1's only. If, in addition, all entries in π and f and in the matrices $M(a)$, a in X, are rational numbers, A is called a *rational stochastic automaton.*

Define $M(\lambda) = E(|S| \times |S|$ identity matrix) and $M(Qa) = M(Q)M(a)$ if Q is in X^* and a is in X. Then A generates a stochastic word-function p_A defined by $p_A(P) = \pi M(P) f$ for all P in X^*. Languages of the form

$$(1) \qquad L(A, \eta) = \{P \in X^* |\ \pi M(P) f > \eta\},$$

where the cut-point η is a real number, are called *stochastic languages*. If A is a rational stochastic automation and η is a rational number, $L(A, \eta)$ is called a *rational stochastic language.* The family of all rational stochastic languages will be denoted by $\mathscr{SL}$. If the sign $>$ in (1) is replaced by the sign $=$ or $\neq$, the corresponding language families for rational stochastic automata and rational cut-points are denoted by $\mathscr{EL}$ and $\mathscr{DL}$, respectively. Exactly the same three families are obtained by using rational pseudostochastic automata in which the elements in π and f and in the matrices $M(a)$, a in X, are allowed to be arbitrary rational numbers (Turakainen 1968, 1969a, 1975b). This is a very useful tool when one wants to study whether a given language is in $\mathscr{SL}$, $\mathscr{DL}$ or $\mathscr{GL}$. In questions like this we very often also use the fact that the set of all rational pseudostochastic word-functions is closed under sum, difference, Hadamard product and multiplication by rational numbers.

As an entertaining example, consider the language

$$L = \{a^{k_{2i}}b^{k_{2i+1}} \mid i \geqslant 0\}$$

where $k_0 = 0$, $k_1 = 1$, $k_{n+1} = k_n + k_{n-1}$ $(n > 0)$ is the Fibonacci sequence. For any letter x, there is a rational pseudostochastic automaton which can count the number of x's in any word. This implies that for some A, $p_A(a^m b^n) = n^2 - mn - m^2$ for all m and n. Hence $L_1 = \{P \mid p_A(P) = 1\} \cap a^*b^*$ is in $\mathscr{SL}$. But by the properties of the Fibonacci sequence we know that $L = L_1$. Therefore, L is in $\mathscr{SL}$.

It is known that $\mathscr{SL} \subset \mathscr{GL}$ and $\mathscr{DL} \subset \mathscr{GL} \subset \mathscr{L}_{\mathrm{DCS}}$ where all inclusions are proper. $\mathscr{DL}$ is an intersection-closed AFL, and $\mathscr{SL}$ is a pre-AFL which is closed under mirror image, sum and intersection, but is not closed under complementation, (λ-free) homomorphism, catenation and catenation closure (Turakainen 1968, 1970, 1971). The family $\mathscr{GL}$ has rather weak closure properties. It is closed under mirror image, inverse homomorphism, complementation (Turakainen, 1969b), union with regular sets and intersection with regular sets. But it is not closed under union or intersection (Lapinš, 1974; Soittola, 1976). Lapinš presented the following counterexample. Let u, v and w be any integers such that $0 < u < v < w$. Then the languages

$$L_1 = \{a^m b^k \mid m^u > k^v > 0\}c^*, \quad L_2 = a^*\{b^k c^n \mid k^v > n^w > 0\}$$

are in $\mathscr{GL}$, but $L_1 \cup L_2$ and $L_1 \cap L_2$ are not stochastic.

Finally, $\mathscr{GL}$ is not closed under (λ-free) homomorphism, catenation and catenation closure (Turakainen, 1970, 1971). Namely, the languages $L_0 = \bigcup_k a^k b(a^*b)^* a^k$ and $L_0 b$ are in $\mathscr{SL}$. Since $\mathscr{SL}$ is a pre-AFL, we know that the marked catenation $L_0 c(a \cup b)^*$ is in $\mathscr{SL}$. Hence these languages are in $\mathscr{GL}$. But $(L_0 b)^*$ and $L_0 b(a \cup b)^*$ are not stochastic. Note that this proves also the above-mentioned result that $\mathscr{SL}$ is not closed under λ-free homomorphism, catenation and catenation closure.

After these negative results the following questions arise: How large are the families which are obtained by λ-free or arbitrary homomorphisms from rational stochastic languages? What are their closure properties? For these questions we introduce the following notations. For any language families $\mathscr{L}$ and $\mathscr{L}_1$, define

$$\mathscr{L} \wedge \mathscr{L}_1 = \{L \cap L_1 \mid L \text{ in } \mathscr{L} \text{ and } L_1 \text{ in } \mathscr{L}_1\},$$

$$H^\lambda(\mathscr{L}) = \{h(L) \mid L \text{ in } \mathscr{L},\ h \text{ a homomorphism on } L\},$$

$$H(\mathscr{L}) = \{h(L) \mid L \text{ in } \mathscr{L},\ h \text{ a } \lambda\text{-free homomorphism on } L\}.$$

If $\mathscr{L}_s$ denotes the family of all stochastic languages, $H^\lambda(\mathscr{L}_s)$ contains all languages, and $H(\mathscr{L}_s)$ contains all languages over one letter (Soittola, 1976)! For rational stochastic languages the situation is entirely different. We know that $H(\mathscr{SL}) = H(\mathscr{GL}) \subseteq \mathscr{L}_{\mathrm{DCS}}$ and $H^\lambda(\mathscr{SL}) = H^\lambda(\mathscr{GL}) = \mathscr{L}_{\mathrm{RE}}$ (Turakainen, 1976). Moreover, $H(\mathscr{GL})$ is an AFL closed under mirror image, intersection, λ-free substitution and λ-limited homomorphism as well as λ-limited (nondeterministic) gsm mappings (defined below). These closure properties will very often be used in the following considerations.

It is possible to characterize the families $H(\mathscr{GL})$ and $H^\lambda(\mathscr{GL})$ by means of the so-called *rational multistochastic automata* (Turakainen, 1973, 1975a) for which instead of just one mapping M we have a finite set of mappings M_i. In the definition of $L(A, \eta)$, $\pi M(a_1)\ldots M(a_k)f$ is replaced by the least upper bound of all numbers $\pi M_{i_1}(a_1) M_{i_2}(a_2) \ldots M_{i_k}(a_k) f$, a_i's in $X \cup \{\lambda\}$ (see the last section of this paper). The families $H(\mathscr{GL})$ and $H^\lambda(\mathscr{GL})$ are also obtained as the output languages of rational stochastic sequential machines (Turakainen, 1975a).

A homomorphism h is *λ-limited* on a language L if there there exists a constant k such that $PQR \in L$ and $h(Q) = \lambda$ always imply that $|Q| < k$. In an analogous manner, we can define λ-limited gsm mappings.

Let f be a function from N into N where $N = \{0, 1, 2, \ldots\}$. A homomorphism h is *f-bounded* on L is there exists $k > 0$ such that $|P| < kf(|P|)$ for all P in L. If f is the identity function and h is f-bounded on L, we say that h is a *linear erasing* (or *k-linear erasing*) on L.

We conclude this introductory section by listing some open problems:

1. Do the families $H(\mathscr{L}_s)$ and $H^\lambda(\mathscr{L}_s)$ coincide?
2. Is $H(\mathscr{GL})$ a proper subfamily of $\mathscr{L}_{\mathrm{DCS}}$?
3. Are all context-free languages in $H(\mathscr{GL})$?
4. Is $H(\mathscr{GL})$ a principal AFL?
5. Is $H(\mathscr{GL})$ closed under linear erasing?
6. Is $H(\mathscr{GL})$ closed under complementation?

Some results related to these questions will be presented in the following sections.

2. Context-free languages and the family $H(\mathscr{GL})$

A language is called *deterministic linear* (Nasu and Honda, 1969) if it is generated by a linear context-free grammar $G = (V_N, V_T, X_0, F)$ such that all productions in F are of the two forms $X \to aYP$, $X \to a$, a in V_T, Y in V_N, P in V_T^*, and for any X in V_N and a in V_T there is at most one production of type $X \to aQ$, $Q \in (V_N \cup V_T)^*$, in F.

THEOREM 1. *All deterministic linear languages are in $\mathscr{SL}$. Hence they are rational stochastic. There exists a nonstochastic linear language. All linear languages are in $H(\mathscr{GL})$.*

Proof. (The original proof of the first assertion is due to Nasu and Honda, 1969.) Let $G = (\{X_1, \ldots, X_n\}, V_T, X_1, F)$ be any deterministic linear grammar. We may denote the letters in V_T by $1, 2, \ldots, m-1$. We may assume that, for each pair X in V_N, a in V_T, F contains a production of type $X \to aQ$, because otherwise we can take a new nonterminal X_{n+1} and additional productions $X \to aX_{n+1}$ and $X_{n+1} \to bX_{n+1}$ for all b in V_T.

There exists a two-state rational stochastic automaton $(V_T, S_1, M_1, (1,0), (0,1)^T)$ generating the word-function p for which $p(\lambda) = 0$ and

$$p(a_1 a_2 \ldots a_k) = 0.a_k a_{k-1} \ldots a_1 \quad \text{for all } a_1 a_2 \ldots a_k \in V_T^+$$

where the right side is an m-adic expansion. For each a in V_T, define

$$M(a) = \begin{bmatrix} A_{11} & \ldots & A_{1n} & B_1 \\ A_{21} & \ldots & A_{2n} & B_2 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \ldots & 0 & E \end{bmatrix}$$

where E is the 2×2 identity matrix and

$$A_{ij} = \begin{cases} M_1(\mathrm{mi}P) & \text{if } X_i \to aX_jP \text{ is in } F \text{ for some } P, \\ 0 & \text{otherwise,} \end{cases}$$

$$B_i = \begin{cases} E & \text{if } X_i \to a \text{ is in } F, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\pi = (1, 0, \ldots, 0)$ and $f = (0, \ldots, 0, 1)^T$. Clearly each $M(a)$ is a rational stochastic matrix. It is now easy to see that if $X_1 \Rightarrow a_1 Y_1 P_1 \Rightarrow \ldots \Rightarrow a_1 \ldots a_r Y_r P_r \ldots \ldots P_1 \Rightarrow a_1 \ldots a_r a_{r+1} P_r \ldots P_1$, then $\pi M(a_1 a_2 \ldots a_k) f = 0.P_r P_{r-1} \ldots P_1$ for any $k > r$. Otherwise we have $\pi M(P) f = 0$.

There exists a 3-state rational stochastic automaton $(V_T, S_2, M_2, (1,0,0), (0,0,1)^T)$ generating the word-function q for which $q(\lambda) = 0$ and $q(P) = 0.P$ for all P in V_T^+. For each a in V_T, define

$$M'(a) = \left[\begin{array}{c|cccc} & b_1 & 0 & \ldots & 0 \\ [a_{ij}] & \cdot & \cdot & \cdot & \cdot \\ & b_n & 0 & \ldots & 0 \\ \hline 0 & & M_2(a) & & \end{array}\right]$$

where

$$a_{ij} = \begin{cases} 1 & \text{if } X_i \to aX_jP \text{ is in } F \text{ for some } P, \\ 0 & \text{otherwise,} \end{cases}$$

$$b_i = \begin{cases} 1 & \text{if } X_i \to a \text{ is in } F, \\ 0 & \text{otherwise.} \end{cases}$$

Each $M'(a)$ is a rational stochastic matrix. Let $\pi' = (1, 0, \ldots, 0)$ and $f' = (0, \ldots \ldots, 0, 1)^T$. These definitions imply that if $X_1 \overset{*}{\Rightarrow} a_1 \ldots a_r Y_r P \Rightarrow a_1 \ldots a_r a_{r+1} P$; then we have $\pi' M'(a_1 \ldots a_r a_{r+1} \ldots a_k) f' = 0.a_{r+2} \ldots a_k$ for any $k > r$ (for $k = r+1$ the value is 0). Otherwise this probability is 0. Now it follows that

$$L(G) = \{P \in V_T^* \mid \pi M(P) f = \pi' M'(P) f'\} \cap \{P \in V_T^* \mid \pi' M'(P) f' > 0\} \cup \cup \{P \in V_T^* \mid G \text{ generates } P \text{ by right-linear productions}\}.$$

The first language in the intersection is in $\mathscr{EL}$. Two other languages are regular. Hence $L(G)$ is in $\mathscr{EL}$.

The second assertion follows from the fact that the mirror image of the language $L_0 b(a \cup b)^*$ (see Section 1) is not stochastic, but it is generated by a linear grammar whose productions are $X \to aX$, $X \to bX$, $X \to bY$, $Y \to aYa$, $Y \to b$, $Y \to bZ$, $Z \to aZ$, $Z \to bZ$, $Z \to b$.

Finally, let L be any λ-free linear language $\subseteq V_T^*$. We may assume that $L = L(G)$ where the productions of G are of the forms $X_i \to aX_j$, $X_i \to X_j a$, $X_i \to a$ where a is in V_T. Since $H(\mathscr{GL})$ is closed under λ-free homomorphism, we may assume that the set of the right-linear productions is deterministic linear. For each pair X_i, X_j such that $X_i \overset{*}{\Rightarrow} X_j P$ for some P in V_T^+, let a_{ij} and b_{ij} be new letters. We construct a deterministic linear grammar G' from G by omitting all productions of type $X \to Ya$ and taking new productions $X_i \to a_{ij} X_j b_{ij}$. Then we have

$$L(G) = s\big(h(L(G') \cap (V'_T \cup CV'_T)^*)\big)$$

where V'_T consists of V_T and of the letters b_{ij}, and C consists of the letters a_{ij}; $h(a_{ij}) = \lambda$, $h(x) = x$ for all x in V'_T, and s is the λ-free regular substitution

$$s(b_{ij}) = \{P \in V_T^+ \mid X_i \overset{*}{\underset{G}{\Rightarrow}} X_j P\},$$

$$s(a) = \{a\} \quad \text{for each } a \text{ in } V_T.$$

Since h is λ-limited on the above intersection, we conclude that $L(G)$ is in $H(\mathscr{GL})$. Hence the same holds for $L \cup \{\lambda\}$. Another proof for this is obtained from AFL-theory, for $\mathscr{L}_{\mathrm{LIN}}$ is a principal semi-AFL generated by the mirror image language which is in $\mathscr{EL}$.

COROLLARY 1. REVBD(Lin) $\subseteq H(\mathscr{GL})$ *where* REVBD(Lin) *is the family of all languages accepted by multipushdown automata which operate in such a way that in every computation each pushdown store makes at most a bounded number of reversals and which run in linear time.*

Proof. Every language in REVBD(Lin) is of the form $h(L_1 \cap L_2 \cap L_3)$ where h is λ-free and L_i's are linear (Book, Nivat and Paterson, 1974).

COROLLARY 2. $\mathscr{L}_{\mathrm{DB}} \subset H(\mathscr{GL})$ *where* $\mathscr{L}_{\mathrm{DB}}$ *is the family of all derivation-bounded languages.*

Proof. $\mathscr{L}_{\mathrm{DB}}$ is the smallest AFL containing all linear languages and closed under λ-free substitution (see Ginsburg, 1975).

Let D_1 and D_2 be the Dyck languages over $X_1 = \{a, b\}$ and $X_2 = \{a_1, b_1, a_2, b_2\}$, respectively.

THEOREM 2. *The language $X_1^*-D_1$ is in $H(\mathscr{GL})$. If $D_1 \in H(\mathscr{GL})$, then $X_2^*-D_2$ is in $H(\mathscr{GL})$. Hence if D_1 is not in $H(\mathscr{GL})$ or if D_2 is not in $H(\mathscr{GL})$, then $H(\mathscr{GL})$ is not closed under complementation. Consequently, if $H(\mathscr{GL})$ is closed under complementation, then $\mathscr{L}_{\mathrm{CF}} \subset \mathscr{L}_{\mathrm{QRT}} \subseteq H(\mathscr{GL}) \subseteq \mathscr{L}_{\mathrm{DCS}}$.*

Proof. The last assertion follows from the fact that $\mathscr{L}_{\mathrm{CF}}$ is the principal AFL generated by D_2 and every quasi-realtime language is obtained by a λ-free homomorphism from an intersection of three context-free languages (Book and Greibach, 1970).

Let

$$L_1 = \{P \in X_1^* \mid |P|_a = |P|_b\},$$

$$L_2 = \{P \in X_1^* \mid |Q|_a \geqslant |Q|_b \text{ for every prefix } Q \text{ of } P\}.$$

Clearly, $D_1 = L_1 \cap L_2$ so that $X_1^*-D_1 = (X_1^*-L_1)\cup(X_1^*-L_2)$. Here L_1 is in $\mathscr{EL}$ and hence $X_1^*-L_1$ is in $\mathscr{GL}$. The language $L_3 = \{P \in X_1^* \mid |P|_a < |P|_b\}$ is in $\mathscr{GL}$ and $X_1^*-L_2 = L_3X_1^*$ so that $X_1^*-L_2$ is in $H(\mathscr{GL})$. Consequently, $X_1^*-D_1$ is in $H(\mathscr{GL})$.

The assertion concerning $X_2^*-D_2$ follows from the following identity (Greibach, 1975):

$$X_2^*-D_2 = h^{-1}(X_1^*-D_1)\cup X_2^*(a_1h^{-1}(D_1)b_2\cup a_2h^{-1}(D_1)b_1)X_2^*$$

where $h(a_i) = a$ and $h(b_i) = b$, $i = 1, 2$. Thus, the proof is complete.

An interesting open problem related to the complementation problem is whether the following languages are in $H(\mathscr{GL})$ or in $\mathscr{L}_{\mathrm{QRT}}$:

$$K_1 = \{a^p \mid p \text{ is a prime number}\},$$

$$K_2 = \{P_1cP_2c \ldots P_rc \mid r > 0,\ P_i \in (a\cup b)^*,\ i \neq j \text{ implies } P_i \neq P_j\}.$$

The complements of these languages are in $H(\mathscr{GL})$, because

$$a^+-K_1 = h(\{a^nb^kc^{k(n-1)} \mid n > 1,\ k > 0\}),$$

$$X^*-K_2 = X^*c^*\{PcQPc \mid P \in X^*,\ Q \in (X^*c)^*\}Y^*\cup Y^*X$$

where $h(a) = h(b) = h(c) = a$, $X = \{a, b\}$ and $Y = \{a, b, c\}$. It is also known that the language

$$K_3 = \{PcQPR \mid P, Q \text{ and } R \text{ in } X^*\}$$

is nonstochastic and belongs to $H(\mathscr{GL})$, but we do not know if its complement is in $H(\mathscr{GL})$.

3. Polynomial-bounded homomorphisms

For any function f and any language family $\mathscr{L}$, let $H_f(\mathscr{L})$ denote the image of $\mathscr{L}$ under f-bounded homomorphisms. We know that $H_f(H(\mathscr{GL})) = H_f(\mathscr{EL}) = H_f(\mathscr{GL})$. This implies that if f is superadditive and semihomogeneous, then $H_f(\mathscr{GL})$ is an AFL which is closed under linear erasing (Book, Greibach and Wegbreit, 1970; Book and Wegbreit, 1971). Hence, by choosing $f(n) = n$ or $f(n) = n^2$,

we conclude that $H_n(\mathscr{GL})$ and $H_{n^2}(\mathscr{GL})$ are AFL's which are closed under linear erasing. We have now the following inclusions:

$$H(\mathscr{GL}) \subseteq H_n(\mathscr{GL}) \subseteq \mathscr{L}_{\mathrm{DCS}},$$

$$H_n(\mathscr{GL}) \subseteq H_{n^2}(\mathscr{GL}) \subseteq \mathrm{NP}.$$

Here NP means the family of languages accepted by nondeterministic Turing machines operating in polynomial time. The last inclusion follows from the facts that $\mathscr{GL} \subseteq \mathrm{NP}$ (Phan Dinh Diêu, 1971) and that NP is closed under polynomial-bounded homomorphism (Book, 1972). Next we shall show that $\mathscr{L}_{\mathrm{CF}}$ is a subfamily of $H_{n^2}(\mathscr{EL})$. We prove first the following theorem:

THEOREM 3. *If L is any context-free language and c is a new letter, then*

$$\{c^{|P|^2}P \mid P \text{ in } L\} \text{ is in } H(\mathscr{GL})$$

and

$$\{(Pc)^{|P|} \mid P \text{ in } L\} \text{ is in } H(\mathscr{GL}).$$

Proof. It suffices to consider the case where λ is not in L. Hence $L = L(G)$ where $G = (V_N, V_T, X_0, F)$ is in Greibach normal form. Define

$$L_1 = \{PcQc \mid P \in V^*V_NV^*,\ P \Rightarrow Q\}$$

where c is not in $V = V_N\cup V_T$. Now L_1 is in $H(\mathscr{GL})$, because it is of the form $\mathrm{GSM}(L_1')$ where GSM is a λ-free nondeterministic gsm and $L_1' = \{PcPc \mid P \text{ in } V^*V_NV^*\}$ which is in $\mathscr{EL}$. Clearly

$$L_1^+(\lambda\cup V^+c)\cap V^+cL_1^*(\lambda\cup V^+c)\cap X_0c(V^+c)^*V_T^+c$$
$$= \{X_0cP_1c \ldots P_kc \mid k > 0,\ P_k \text{ in } V_T^+,\ X_0 \Rightarrow P_1 \Rightarrow \ldots \Rightarrow P_k\}.$$

Denote this language by L_2. Hence L_2 is in $H(\mathscr{GL})$. By using a gsm λ-limited on L_2 we see that the language

$$L_3 = \{d^{|P_1|}c \ldots d^{|P_{k-1}|}cP_k \mid k > 0,\ P_k \text{ in } V_T^+,\ X_0 \Rightarrow P_1 \Rightarrow \ldots \Rightarrow P_k\}$$

is in $H(\mathscr{GL})$ (d is a new letter). Here $|P_i| \leqslant |P_{i+1}|$ and $|P_k| = k$. By the λ-free substitution $s(d) = \bar{V}_T^+$ ($\bar{V}_T$ is a renaming of V_T), $s(x) = x$ for all x in $V_T\cup\{c\}$, we form the language $s(L_3)$. It is in $H(\mathscr{GL})$. It can be verified that the language

$$L_4 = \{(\bar{Q}c)^nQ \mid n \geqslant 0,\ Q \text{ in } V_T^*\}$$

where $\bar{Q} \in \bar{V}_T^*$ is the renaming of Q, is in $H(\mathscr{GL})$. This implies that

(1) $$s(L_3)\cap L_4 = \{(\bar{P}c)^{|P|-1}P \mid P \text{ in } L\}$$

is in $H(\mathscr{GL})$ whence the last assertion of our theorem follows. Moreover, (1) implies that $H(\mathscr{GL})$ contains the language consisting of the words c^kP, P in L, $k = |P|^2-1$. This yields the first assertion.

COROLLARY 1. *$\mathscr{L}_{\mathrm{QRT}}$ is a subfamily of $H_{n^2}(\mathscr{EL})$.*

Proof. Each quasi-realtime language L is of the form $h(L_1\cap L_2\cap L_3)$ where L_i's are context-free and h is length-preserving (Book and Greibach, 1970). Since $H(\mathscr{GL})$ is closed under intersection and λ-free homomorphism, Theorem 3 shows that $\{c^k P \mid P \text{ in } L,\ k = |P|^2\}$ is in $H(\mathscr{GL})$. Thus L is in $H_{n^2}(H(\mathscr{GL}))$ which is equal to $H_{n^2}(\mathscr{EL})$.

COROLLARY 2. $H^\lambda(\mathscr{EL}) = \mathscr{L}_{\mathrm{RE}}$. (See also Theorems 7 and 8.)

Proof. By Theorem 3, $\{c^k P \mid P \text{ in } L_1 \cap L_2,\ k = |P|^2\}$ is in $H(\mathscr{GL})$ for any context-free languages L_1 and L_2. Hence, $h(L_1 \cap L_2)$ is in $H^\lambda(H(\mathscr{GL})) = H^\lambda(\mathscr{GL}) = H^\lambda(\mathscr{EL})$ for any homomorphism h. On the other hand, $\mathscr{L}_{\mathrm{RE}} = H^\lambda(\mathscr{L}_{\mathrm{CF}} \wedge \mathscr{L}_{\mathrm{CF}})$.

THEOREM 4. *Let f: $N \to N$ be any polynomial. If L is in $H(\mathscr{GL})$ and h is an f-bounded homomorphism on L, then*

$$\{c^{f(|P|)} P \mid P \text{ in } h(L)\} \quad \textit{is in } H(\mathscr{GL}).$$

Proof. Thus we assume that $L \subseteq X^*$ is in $H(\mathscr{GL})$ and h: $X^* \to Y^*$ is f-bounded on L. Define $h_1(x) = h(x)$ if $h(x) \neq \lambda$, and $h_1(x) = \beta$ if $h(x) = \lambda$, where β is a new letter. Then $h_1(L) \subseteq (Y \cup \{\beta\})^*$ is in $H(\mathscr{GL})$, and $h(L) = g(h_1(L))$ where $g(\beta) = \lambda$ and $g(y) = y$ if y is in Y. Let d be a new letter. Now we find that

$$h_1(L) d Y^* \cap \{g^{-1}(P) d P \mid P \text{ in } Y^*\} = \{P d g(P) \mid P \text{ in } h_1(L)\}.$$

Both languages in the intersection are in $H(\mathscr{GL})$. Hence the language L_0 on the right-hand side is in $H(\mathscr{GL})$. Since h is f-bounded on L, it easily follows that there is a constant k such that $|P| \leqslant kf(|g(P)|)$ for all P in $h_1(L)$. Hence one can construct a GSM λ-limited on L_0 such that

$$\mathrm{GSM}(L_0) = \{c^{t(|P|)} dP \mid P \text{ in } h(L)\}$$

where c is a new letter and t is a function of $|P|$ such that $t(|P|) \leqslant f(|P|)$ for all P in $h(L)$. Now we have

$$c^* \mathrm{GSM}(L_0) \cap \{c^{f(|P|)} dP \mid P \text{ in } Y^*\} = \{c^{f(|P|)} dP \mid P \text{ in } h(L)\}.$$

Therefore, the proof is complete if we can show that the latter language in the intersection is in $H(\mathscr{GL})$. It is of the form $s(K)$ where $K = \{c^{f(n)} d a^n \mid n \text{ in } N\}$ and $s(a) = Y$, $s(c) = \{c\}$, $s(d) = \{d\}$. Hence it is in $H(\mathscr{GL})$, for K is in $\mathscr{EL}$.

COROLLARY. *If L is in $H(\mathscr{GL})$ and h is a linear erasing on L, then* $\{PdP \mid P \text{ in } h(L)\}$ *is in $H(\mathscr{GL})$. (Thus $H(\mathscr{GL})$ is closed under linear erasing iff it is closed under 2-linear erasing.)*

Proof. Now $f(n) = n$ so that $L_2 = \{c^{|P|} dP \mid P \text{ in } h(L)\}$ is in $H(\mathscr{GL})$. If s is the λ-free finite substitution $s(c) = Y$, $s(y) = y$, y in $Y \cup \{d\}$, we have

$$s(L_2) \cap \{PdP \mid P \text{ in } Y^*\} = \{PdP \mid P \text{ in } h(L)\}.$$

4. Polynomial languages

For any polynomial p: $N \to N$ with degree > 1, the language $\{a^{p(n)} \mid n \text{ in } N\}$ is nonstochastic (Soittola, 1976), but the language $\{a^n b^{p(n)} \mid n \text{ in } N\}$ is rational stochastic. More generally, we have

THEOREM 5. *Let $r > 0$ and $k > 0$ be any integers and $p_1, \ldots, p_r$ any polynomials in $n_1, \ldots, n_k$ with integer coefficients. Then the language*

$$\{a_1^{n_1} \ldots a_k^{n_k} b_1^{p_1(n_1, \ldots, n_k)} \ldots b_r^{p_r(n_1, \ldots, n_k)} \mid p_i(n_1, \ldots, n_k) \geqslant 0\}$$

is in $\mathscr{EL}$ ($a_1, \ldots, a_k, b_1, \ldots, b_r$ are distinct letters).

Proof. Consider words of the form

$$P = a_1^{n_1} \ldots a_k^{n_k} b_1^{m_1} \ldots b_r^{m_r}.$$

Since the set of word-functions generated by rational pseudostochastic automata is closed under sum, Hadamard product and multiplication by rational numbers, it can be verified that there exists a rational pseudostochastic automaton A such that

$$p_A(P) = \sum_i (p_i(n_1, \ldots, n_k) - m_i)^2.$$

Consequently,

$$L = \{Q \mid p_A(Q) = 0\} \cap a_1^* \ldots a_k^* b_1^* \ldots b_r^*$$

which is in $\mathscr{EL}$.

LEMMA. *Let $k_1, \ldots, k_r$ ($r > 0$) be any nonnegative integers. Then*

$$\{a_1^{n^{k_1}} a_2^{n^{k_2}} \ldots a_r^{n^{k_r}} \mid n \text{ in } N\} \quad \textit{is in } H(\mathscr{GL}).$$

Proof. We consider the nontrivial case that $k_s > 0$ for some s. Let $p_s(n) = n^{k_s} - n$ and $p_i(n) = n^{k_i}$ if $i \neq s$. By the same reasons as in the previous proof we can conclude that

$$\{a_1^{p_1(n)} \ldots a_s^{p_s(n)} b^n a_{s+1}^{p_{s+1}(n)} \ldots a_r^{p_r(n)} \mid n \text{ in } N\} \quad \text{is in } \mathscr{EL},$$

whence the lemma follows by a length-preserving homomorphism.

COROLLARY. *Let $p_1(n), \ldots, p_r(n)$ be any polynomials with integer coefficients. Then the language*

$$L = \{a_1^{p_1(n)} a_2^{p_2(n)} \ldots a_r^{p_r(n)} \mid p_i(n) \geqslant 0\}$$

is in $H(\mathscr{GL})$.

Proof. If the leading coefficient of some p_i is negative, L is finite. Therefore, assume that the leading coefficients are all nonnegative. Let $m = n - n_0$. If n_0 is large enough, then $p_i(n)$ can be written as a polynomial $q_i(m)$ with nonnegative coefficients ($i = 1, \ldots, r$). Hence, for some finite language L_f,

$$L = L_f \cup \{a_1^{q_1(m)} \ldots a_r^{q_r(m)} \mid m \text{ in } N\}.$$

Here the polynomial language is of the form $h(L_1)$ where h is length-preserving and L_1 is of the form considered in the previous Lemma.

THEOREM 6. *Let $p_i(n_1, \ldots, n_k)$, $i = 1, \ldots, r$ ($r > 0, k > 0$) be any polynomials with positive integer coefficients. Then the language*

$$\{a_1^{p_1(n_1, \ldots, n_k)} a_2^{p_2(n_1, \ldots, n_k)} \ldots a_r^{p_r(n_1, \ldots, n_k)} \mid n_i\text{'s in } N\}$$

is in $H(\mathscr{GL})$.

Proof. Since $H(\mathscr{GL})$ is closed under union, it suffices to show that

$$L(p_1, \dots, p_r) = \{a_1^{p_1(n_1, \dots, n_k)} \dots a_r^{p_r(n_1, \dots, n_k)} \mid n_i\text{'s} > 0\}$$

is in $H(\mathscr{GL})$.

Denote $A = \{a_1, \dots, a_r\}$ and $B = \{b_1, \dots, b_k\}$. As in the proof of Theorem 5, it can be verified that the language

$$L_1 = \{P \in (A \cup B)^* \mid |P|_{a_i} = p_i(|P|_{b_1}, \dots, |P|_{b_k}), |P|_{b_i} > 0\}$$

is in $\mathscr{EL}$. Since the coefficients of the polynomials p_j are positive and every n_i occurs in some term of the sum $\sum p_j(n_1, \dots, n_k)$, it follows that

$$n_1 + \dots + n_k \leqslant k \sum p_j(n_1, \dots, n_k)$$

whenever all n_i's are greater than 0. Hence every word P in L_1 satisfies the condition

$$\sum_{i=1}^{k} |P|_{b_i} \leqslant k \sum_{i=1}^{r} |P|_{a_i}.$$

Consequently, for any choice of values $n_1 > 0, \dots, n_k > 0$, the language

$$L_2 = L_1 \cap (BA \cup B^2A \cup \dots \cup B^kA)^*A^* \cap (B^*a_1)^*(B^*a_2)^* \dots (B^*a_r)^*$$

contains a word P such that $|P|_{b_i} = n_i$ for all i. Hence, if $h(a_i) = a_i$ and $h(b_j) = \lambda$ for every i and j, then $h(P)$ is the word of $L(p_1, \dots, p_r)$ corresponding to the values $n_1, \dots, n_k$. Consequently, $L(p_1, \dots, p_r) = h(L_2)$. This completes the proof, because h is λ-limited on L_2.

Remark. Theorem 6 does not hold for all polynomials, because every recursively enumerable language $L \subseteq a^*$ is of the form

$$L = \{a^{p(n_1, \dots, n_k)} \mid p(n_1, \dots, n_k) \geqslant 0\}$$

(Matijasevič, 1970).

5. Recursively enumerable languages

We saw in Corollary 2 to Theorem 3 that the families $H^\lambda(\mathscr{EL})$; $H^\lambda(\mathscr{GL})$ and $\mathscr{L}_{\mathrm{RE}}$ are equal. By Theorem 1 we know that $H^\lambda(\mathscr{L}_{\mathrm{LIN}} \wedge \mathscr{L}_{\mathrm{LIN}})$ is a subfamily of $H^\lambda(\mathscr{GL})$. We show now that they are equal.

THEOREM 7. $\mathscr{L}_{\mathrm{RE}} = H^\lambda(\mathscr{L}_{\mathrm{LIN}} \wedge \mathscr{L}_{\mathrm{LIN}})$.

Proof. (The original proof is due to Baker and Book (1974), and it simulates the action of a Turing machine. A sharper version of Theorem 7 with a more complicated proof is presented in Turakainen (1977).) Let L be in $\mathscr{L}_{\mathrm{RE}}$. Hence, $L = L(G)$ where $G = (V_N, V_T, Z_0, F)$ is a formal grammar. For each a in V_T, let $\bar{a}$ be a new letter. For each P in $(V_N \cup V_T)^*$, let $\bar{P}$ be obtained from P by replacing every $a \in V_T$ by $\bar{a}$. Letters in V_N remain unchanged. Consider a linear grammar G_1 with two nonterminals X_0 (the initial letter) and Y_0 and with the following productions:

$$X_0 \to aX_0\bar{a} \quad \text{for all } a \in V_T, \qquad X_0 \to cY_0c,$$
$$Y_0 \to \bar{a}Y_0\bar{a}, \qquad Y_0 \to xY_0x \quad \text{for all } a \in V_T,\ x \in V_N,$$
$$Y_0 \to \bar{Q}Y_0\,\mathrm{mi}\,\bar{P} \quad \text{for all } P \to Q \text{ in } F,$$
$$Y_0 \to cY_0c, \qquad Y_0 \to d \quad (c \text{ and } d \text{ are new letters}).$$

Thus the terminal alphabet of G_1 is the union of V_N, V_T, $\bar{V}_T$ and $\{c, d\}$.

We continue by constructing another linear grammar G_2 with three nonterminals X_0, X_1 and X_2 and with productions

$$X_0 \to X_1, \qquad X_1 \to aX_1 \quad \text{for all } a \in V_T, \qquad X_1 \to cX_2,$$
$$X_2 \to \bar{a}X_2\bar{a}, \qquad X_2 \to xX_2x \quad \text{for all } a \in V_T,\ x \in V_N,$$
$$X_2 \to \bar{P}X_2\,\mathrm{mi}\,\bar{Q} \quad \text{for all } P \to Q \text{ in } F,$$
$$X_2 \to cX_2c, \qquad X_2 \to dZ_0c.$$

Any terminal derivation in the original grammar G is of the form

$$Z_0 \overset{*}{\Rightarrow} P_1 \overset{*}{\Rightarrow} P_2 \overset{*}{\Rightarrow} \dots \overset{*}{\Rightarrow} P_{2k} = P_{2k+1} \in V_T^*$$

where, for every i, P_i generates P_{i+1} in zero or one step. Clearly the corresponding word

(1) $$P_{2k+1}c\bar{P}_{2k-1}c \dots c\bar{P}_1dZ_0c\,\mathrm{mi}\,\bar{P}_2c\,\mathrm{mi}\,\bar{P}_4c \dots c\,\mathrm{mi}\,\bar{P}_{2k}$$

is in $L(G_1) \cap L(G_2)$. Conversely, every word in $L(G_1) \cap L(G_2)$ is of the form (1) where $Z_0 \overset{*}{\Rightarrow} P_1 \overset{*}{\Rightarrow} \dots \overset{*}{\Rightarrow} P_{2k} = P_{2k+1} \in V_T^*$. Hence we can conclude that $L(G) = h(L(G_1) \cap L(G_2))$ where $h(a) = a$ for every a in V_T and $h(x) = \lambda$, otherwise.

COROLLARY. *Every L in $\mathscr{L}_{\mathrm{RE}}$ is of the form $(L_1 \cap L_2)/L_R$ where L_1 and L_2 are linear languages and L_R is a regular language.*

Proof. Use the above theorem for the language $L\beta$ where β is a new letter. Then $L = (L(G_1) \cap L(G_2))/\beta A^*$.

Remark. Every L in $\mathscr{L}_{\mathrm{RE}}$ is of the form $L = h(L_1)$ where L_1 is in $\mathscr{EL}$ and, for each letter a, either $h(a) = a$ or $h(a) = \lambda$ (Turakainen, 1976). By the technique used in the proof of Theorem 4, it can be verified that $L_0 = \{h(P)dP \mid P \text{ in } L_1\}$ is in $\mathscr{EL}$. Hence, every $L \in \mathscr{L}_{\mathrm{RE}}$ is of the form $L = L_0/L_R$ where L_0 is in $\mathscr{EL}$ and L_R is regular.

As we mentioned above, every L in $\mathscr{L}_{\mathrm{RE}}$ is obtained from a language in $\mathscr{EL}$ by erasing all occurrences of certain letters. The following theorem shows that the erasing of one or two letters is enough.

THEOREM 8. (i) *The family $\mathscr{EL}$ contains a language $L_0 \subseteq \{a, b\}^*$ such that every recursively enumerable language $L \subseteq a^*$ is of the form $L = h(L_f \cup (L_0 \cap L_R))$ where L_f is a finite language, L_R is regular, and $h(a) = a$, $h(b) = \lambda$.*

(ii) *For any alphabet X, $\mathscr{EL}$ contains a language $L_1 \subseteq (X \cup \{a, b\})^*$ such that every recursively enumerable language L over X is of the form $L = h(L_1 \cap L_R)$ where L_R is regular and $h(a) = h(b) = \lambda$, $h(x) = x$ for all x in X.*

Proof. Matijasevič (1970) has shown that there exists a universal polynomial $U(x, x_1, \ldots, x_r, y)$ with integer coefficients having the following property. For any recursively enumerable $L \subseteq a^*$, there exists a number m_L in N such that

$$L = \{a^n \mid U(m_L, k_1, \ldots, k_r, n) = 0 \text{ for some } k_i\text{'s in } N\}.$$

Define

$$L_0 = \{a^n bb^m ab^{k_1} \ldots ab^{k_r} \mid U(m, k_1, \ldots, k_r, n) = 0\}.$$

It can be verified that L_0 is in $\mathscr{EL}$. Let m' be the value of the indicator m for the language L/a^r. Then we have

$$a^r(L/a^r) = h\big(L_0 \cap a^*bb^{m'}a(a\cup b)^*\big)$$

where $h(a) = a$ and $h(b) = \lambda$, This proves (i).

Let $X = \{a_{r+1}, \ldots, a_{r+s}\}$ be any alphabet. Let a and b be new letters. Define a mapping $g\colon X^* \to N$ as follows:

$$g(\lambda) = r,$$

$$g(a_{i_1} a_{i_2} \ldots a_{i_k}) = i_1 i_2 \ldots i_k \quad ((r+s+1)\text{-ary expansion}).$$

There exists a rational pseudostochastic automaton A such that $p_A(P) = g(P)$ for all P in X^+. This implies that the language

$$K = \{PQ \mid P \text{ in } X^*,\ Q \text{ in } (a\cup b)^*,\ |Q|_a = g(P)\}$$

is in $\mathscr{EL}$.

Let $L \subseteq X^*$ be any language in $\mathscr{L}_{\mathrm{RE}}$. Consider the language $L_g = \{a^{g(P)} \mid P$ in $L\}$. Clearly, L_g is a subset of $a^r a^*$ so that $a^r(L_g/a^r) = L_g$. By the proof of (i) we conclude that $L_g = h(L_0 \cap L_R)$ where $h(a) = a$ and $h(b) = \lambda$ and L_R is regular. This implies that

$$L = h_1(X^*(L_0 \cap L_R) \cap K)$$

where $h_1(a) = h_1(b) = \lambda$, and $h(x) = x$ for all x in X. Since now $X^*(L_0 \cap L_R) \cap K = (X^* L_0 \cap K) \cap X^* L_R$, the proof of (ii) is complete, because $X^* L_0$ is in $\mathscr{EL}$.

Remark. It is not known if every L in $\mathscr{L}_{\mathrm{RE}}$ is obtained from some L_1 in $\mathscr{EL}$ (or in $\mathscr{GL}$) by erasing the occurrences of at most one letter. It is known that any language is obtained with this kind of erasing from a stochastic language (Soittola, 1976).

In the rest part of these notes we deal with rational multistochastic automata $S = (X, S, \{M_1, \ldots, M_k\}, \pi, f)$ in which each M_i is a mapping from $X \cup \{\lambda\}$ into the set of rational stochastic $|S| \times |S|$ matrices. Thus it is possible that, for some values of i, $M_i(\lambda) \neq E$. Let p_A be the word-function generated by A (for formal definition, we refer to Turakainen, 1973, 1975a). It is known that $H^\lambda(\mathscr{GL})$ (that is, $\mathscr{L}_{\mathrm{RE}}$) is the family of all languages of the form

$$L(A, \eta) = \{P \mid p_A(P) > \eta\}$$

where η is a rational number. If for all i, $M\,(\lambda) = E$, we obtain the family $H(\mathscr{GL})$.

THEOREM 9. *There exists a rational multistochastic automaton A and a rational cut-point η such that the following two conditions hold:*

(i) *The language $\{P \mid p_A(P) \leqslant \eta\}$ is not in $H^\lambda(\mathscr{GL})$;*

(ii) *Either $\{P \mid p_A(P) = \eta\}$ is not in $H^\lambda(\mathscr{GL})$ and no infinite subset of it is in $H^\lambda(\mathscr{GL})$, or $\{P \mid p_A(P) < \eta\}$ is not in $H^\lambda(\mathscr{GL})$ and no infinite subset of it is in $H^\lambda(\mathscr{GL})$.*

Proof. Let $L \subseteq a^*$ be a simple set, i.e. L is in $\mathscr{L}_{\mathrm{RE}}$, a^*-L is infinite, and $K \cap L$ is nonempty whenever $K \subseteq a^*$ is an infinite recursively enumerable language. (For the existence of L, see for instance Rogers, 1967.) Since $\mathscr{L}_{\mathrm{RE}} = H^\lambda(\mathscr{GL})$, there exists a rational multistochastic automaton A and a rational cut-point η such that $L = L(A, \eta)$. Since $a^* - L$ is not in $\mathscr{L}_{\mathrm{RE}}$, the first assertion holds.

Finally, (ii) is true, because either $\{P \mid p_A = \eta\}$ is infinite or else $\{P \mid p_A(P) < \eta\}$ is infinite.

Remark. It is not known if Theorem 9 holds for $H(\mathscr{GL})$ and for rational multistochastic automata such that $M_i(\lambda) = E$ for every i.

Note added in proof. Some problems studied in this paper have recently been solved in Turakainen (1981a), (1981b).

References

B. Baker and R. Book (1974), *Reversal-bounded multipushdown machines*, J. Comput. System Sci. 8, 315–332.

R. Book (1972), *On languages accepted in polynomial time*, SIAM J. Comput. 1, 281–287.

R. Book and S. Greibach (1970), *Quasi-realtime languages*, Math. System Theory 4, 97–111.

R. Book, S. Greibach and B. Wegbreit (1970), *Time- and tape-bounded Turing acceptors and* AFL*s*, J. Comput. System Sci. 4, 606–621.

R. Book, M. Nivat and M. Paterson (1974), *Reversal-bounded acceptors and intersections of linear languages*, SIAM J. Comput. 3, 283–295.

R. Book and B. Wegbreit (1971), *A note on* AFL*s and bounded erasing*, Information and Control 19, 18–29.

S. Ginsburg (1975), *Algebraic and automata-theoretic properties of formal languages*, North-Holland Publishing Company.

S. Greibach (1975), *One counter languages and the* IRS *condition*, J. Comput. System Sci. 10, 237–247.

[Ja. K. Lapinš] Я. К. Лапиньш (1974), *О нестохастических языках получаемых как объединение или пересечение стохастических языков*, Автомат. и Вычисл. Техн. 4, 6–13.

Yu. V. Matijasevič (1970), *Enumerable sets are Diophantine*, Soviet Math. Dokl. 11, 354–358.

M. Nasu and N. Honda (1969), *Mappings induced by* PGSM*-mappings and some recursively unsolvable problems of finite probabilistic automata*, Information and Control 15, 250–273.

Phan Dinh Diêu (1971), *On the languages representable by finite probabilistic automata*, Z. Math. Logik Grundlagen Math. 17, 427–442.

H. Rogers, Jr. (1967), *Theory of recursive functions and effective computability*, McGraw-Hill Book Company.

M. Soittola (1976), *On stochastic and Q-stochastic languages*, Ann. Acad. Sci. Fenn. Ser. A I Math. Dissertationes 7.

P. Turakainen (1968), *On probabilistic automata and their generalizations*, ibid. 429.

— (1969a), *Generalized automata and stochastic languages*, Proc. Amer. Math. Soc. 21, 303–309.

— (1969b), *On languages representable in rational probabilistic automata*, Ann. Acad. Sci. Fenn. Ser. A I 439.
— (1970), *The family of stochastic languages is closed neither under catenation nor under homomorphism*, Ann. Univ. Turku. Ser. A I 133.
— (1971), *Some closure properties of the family of stochastic languages*, Information and Control 18, 253–256.
— (1973), *On multistochastic automata*, ibid. 23, 183–203.
— (1975a), *Some remarks on multistochastic automata*, ibid. 27, 75–85.
— (1975b), *Word-functions of stochastic and pseudostochastic automata*, Ann. Acad. Sci. Fenn. Ser. A I 1, 27–37.
— (1976), *On homomorphic images of rational stochastic languages*, Information and Control 30, 96–105.
— (1978), *On characierization of recursively enumerable languages in terms of linear languages and VW-grammars*, Indagationes Math. 40, 145–153.
— (1981a), *On some bounded semi* AFL*s and* AFL*s*, Information Sciences 23, 31–48.
— (1981b), *On nonstochastic languages and homomorphic images of stochastic languages*, to appear in Information Sciences 23.

Presented to the Semester
Discrete Mathematics
(February 15–June 16, 1977)

DISCRETE MATHEMATICS
BANACH CENTER PUBLICATIONS, VOLUME 7
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1982

GENERALIZED IDENTIFICATION EXPERIMENTS FOR FINITE DETERMINISTIC AUTOMATA

HANS-DIETER BURKHARD

Department of Mathematics, Humboldt University, GDR

In the theory of state identification experiments different kinds of experiments can be considered: diagnosing experiments for the determination of initial states, homing experiments for the determination of states at the end of the experiments, and, for instance, control experiments for bringing the automaton in a certain state. In this paper there are introduced generalized identification experiments which can solve all these tasks as special cases, and which can also solve additional tasks that come by connections of diagnose, homing and control. Using the method described in [2] an algorithm is given for the solution of the existence and construction problem.

In [1] were introduced relations of indistinguishability for output words such that the experimentalist can identify only by output words which are distinguishable. Here we consider generalized identification experiments with such indistinguishability relations and solve the existence and construction problem for regular relations.

Let $\mathfrak{A} = [X, Y, Z, \delta, \lambda, Z_0]$ be a finite deterministic synchronous automaton with input set X, output set Y, state set Z, next state function $\delta\colon Z\times X \to Z$, output function $\lambda\colon Z\times X\to Y$ and the set of initial states $Z_0 \subseteq Z$.

(1) Definition. An input word $p\in X^*$ is a *generalized identification experiment* (g.i.e.) for $\mathfrak{A} = [X, Y, Z, \delta, \lambda, Z_0]$ and a given set $\mathcal{M} \subseteq 2^Z\times 2^Z$ iff there is an identification function ζ from $\lambda(Z_0, p)$ in $\mathcal{M}$ such that for all $z\in Z_0$ we have $z \in \varphi_1(\lambda(z,p))$ and $\delta(z,p)\in\varphi_2(\lambda(z,p))$ (φ_1, φ_2 are the components of φ, $\varphi(q) = [\varphi_1(q), \varphi_2(q)]$).

(2) Corollary. *An input word p is a* g.i.e. *for* $\mathfrak{A}$ *and* $\mathcal{M}$ *iff for all* $q\in\lambda(Z_0,p)$ *exists* $[M,N]\in\mathcal{M}$ *with*

$$M_q \underset{\mathrm{Df}}{=} \{z \mid z\in Z_0 \wedge \lambda(z,p) = q\} \subseteq M,$$
$$\delta(M_q, p) \subseteq N.$$

A g.i.e. performs only diagnose in the usual sense if $\mathcal{M} = \{\{z\} \mid z\in Z_0\}\times\{Z\}$, it performs homing if $\mathcal{M} = \{Z_0\}\times\{\{z\} \mid z\in Z\}$ and control respectively if $\mathcal{M} = \{Z_0\}\times\{\{z\}\}$ for the wanted control state z. Additionally a g.i.e. can perform