Special Families of Matrix Languages and Decidable Problems

A. Mateescu *

Abstract

We investigate some variants of simple matrix grammars. It is proved that the equivalence problem, the inclusion problem and other problems are decidables for this families of grammars. It would be noted that all these problems are undecidable for the family of simple matrix grammars.

1 Definitions and notations.

For an alphabet Σ we denote by Σ^* the free monoid generated by Σ under the operation of concatenation, and λ is the null element. The length of a string $\alpha \in \Sigma^*$ is denoted by $|\alpha|$. The set of natural numbers is denoted by N. If $n \in \mathrm{N}, n \geq 1$, then $[n]$ denotes the set $\{1, 2, \ldots n\}$. If $n \in N, n \geq 1$, and $\varphi : [n] \longrightarrow N$ is a function, then φ is a *n-function* and $|\varphi| = \Sigma_{i=1}^{n} \varphi(i)$. If $\theta = (\varphi, \psi)$, where φ and ψ are n-functions, then $|\theta| = |\varphi| + |\psi|$ and $\theta_i = \varphi(i) + \psi(i), i = 1, 2, \ldots, n$.

In order to obtain certain subfamilies of matrix languages we consider a special case of simple (linear, regular) matrix grammars, see [4] p. 68, definition 1.5.1.

Definition 1.1 *Let $n, k \in \mathrm{N}$ be such that $1 \leq k \leq n$ and let $\theta = (\varphi, \psi)$ be a pair of n-functions. A (n, k, θ)-linear matrix grammar (lmg) is a matrix grammar $G = (V, \Sigma, S, P)$ of degree n, where V is the nonterminal alphabet, Σ is the terminal alphabet, S is the start symbol ($S \notin V \cup \Sigma$) and P is a finite set of matrices of the following form:*

(i) $(S \longrightarrow A_1 A_2 \ldots A_n)$, $\quad A_i \in V$, $\quad i = 1, \ldots, n$

(ii) $(A_1 \longrightarrow \alpha_1 B_1 \beta_1, \ldots, A_n \longrightarrow \alpha_n B_n \beta_n), A_i, B_i \in V, \alpha_i, \beta_i \in \Sigma^*, |\alpha_i| = \varphi(i), |\beta_i| = \psi(i), \quad i = 1, \ldots, n$

(iii) $(A_1 \longrightarrow \alpha_1, \ldots, A_n \longrightarrow \alpha_n), A_i \in V, i = 1, \ldots, n, \alpha_k \in \Sigma^*, \quad 0 \leq |\alpha_k| < |\theta|$ *and* $\alpha_i = \lambda$ *for* $i \neq k, i = 1, \ldots, n$.

A (n, k, θ)-linear matrix grammar is called *(n, k, φ)-regular matrix grammar (rmg)* iff $\psi(i) = 0, i = 1, \ldots, n$.

We define the direct derivation relation $\underset{G}{\Longrightarrow}$ and the derivation relation $\overset{*}{\underset{G}{\Longrightarrow}}$ as usually, see [3].

*University of Bucharest, Faculty of Mathematics, 14 Academiei, R-70109 Bucharest, ROMANIA

The language generated is:

$$L(G) = \{w | w \in \Sigma^*, S \overset{*}{\underset{G}{\Longrightarrow}} w\}$$

Definition 1.2 *The family of (n,k,θ)-linear matrix languages is:*

$$\mathcal{LM}_{n,k,\theta} = \{L | \exists G, (n,k,\theta) - \text{lmg and } L(G) = L\}$$

and the family of (n,k,φ)*-regular matrix languages is:*

$$\mathcal{RM}_{n,k,\varphi} = \{L | \exists G, (n,k,\varphi) - \text{rmg} \quad \text{and} \quad L(G) = L\}$$

Remark 1.3 *Let k be such that $1 \leq k \leq n$, let $\theta = (\varphi, \psi)$ be a pair of n-functions and let Σ be an alphabet. We consider the following two alphabets:*

$$\Sigma_1 = \{[\alpha] | \alpha \in \Sigma^* \text{ and } |\alpha| = |\theta|\} \quad \text{and} \quad \Sigma_2 = \{[\beta] | \beta \in \Sigma^* \text{ and } |\beta| < |\theta|\}.$$

For every $w \in \Sigma^*$ there exists and are unique two numbers $p, r \in \mathbf{N}$ such that

$$|w| = p|\theta| + r \quad \text{and} \quad 0 \leq r < |\theta|.$$

It is easy to remark that there exists a unique decomposition of w:

$$w = w_1 w_2 \ldots w_{k-1} u_k \beta v_k w_{k+1} \ldots w_n$$

such that for any $i = 1, \ldots, n, i \neq k, |w_i| = p\theta_i, |u_k| = p\varphi(k), |v_k| = p\psi(k)$, and $|\beta| = r$. Let w_k be the word $u_k v_k$. Then, there are the words $x_i^{(j)}, y_i^{(j)} \in \Sigma^*, j = 1, \ldots, p$ such that $|x_i^{(j)}| = \varphi(i), |y_i^{(j)}| = \psi(i)$ and $w_i = x_i^{(1)} x_i^{(2)} \ldots x_i^{(p)} y_i^{(p)} \ldots y_i^{(2)} y_i^{(1)}$, for all $i = 1, \ldots, n$. Let $z_i^{(j)}$ be the word $x_i^{(j)} y_i^{(j)}, i = 1, \ldots, n, j = 1, \ldots p$.

Using the above notations we shall define the function

$$\tau_\theta^{n,k} : \Sigma^* \longrightarrow \Sigma_1^* \Sigma_2$$

$$\tau_\theta^{n,k}(w) = [z_1^{(1)} z_2^{(1)} \ldots z_n^{(1)}][z_1^{(2)} z_2^{(2)} \ldots z_n^{(2)}] \ldots [z_1^{(p)} z_2^{(p)} \ldots z_n^{(p)}][\beta].$$

Note that for any θ, n and $k, \tau_\theta^{n,k}$ is a bijective function. Let us consider an example.

Example 1.4 *We choose $n = 3, k = 2, \theta = (\varphi, \psi)$ where $\varphi, \psi : [3] \rightarrow N, \varphi(1) = 1, \varphi(2) = 4, \varphi(3) = 1, \psi(1) = 2, \psi(2) = 1, \psi(3) = 3$ and let w be the word $a_1 a_2 \ldots a_{30}$. Note that $|\theta| = 12, |w| = 30$ and therefore $p = 2$ and $r = 6$. It results that:*

$$w = w_1 u_2 \beta v_2 w_2, \text{ where } :$$

$$w_1 = a_1 \ldots a_6, u_2 = a_7 \ldots a_{14}, \beta = a_{15} \ldots a_{20}, v_2 = a_{21} a_{22}, w_2 = a_{23} \ldots a_{30}.$$

Observe that:

$$z_1^{(1)} = a_1 a_5 a_6, z_1^{(2)} = a_2 a_3 a_4, z_2^{(1)} = a_7 \ldots a_{10} a_{22},$$

$$z_2^{(2)} = a_{11} \ldots a_{14} a_{21}, z_3^{(1)} = a_{23} a_{28} a_{29} a_{30}, z_3^{(2)} = a_{24} a_{25} a_{26} a_{27}.$$

From these remarks it follows that:

$$\tau_\theta^{n,k}(w) = [a_1 a_5 a_6 a_7 \ldots a_{10} a_{22} a_{23} a_{28} a_{29} a_{30}][a_2 a_3 a_4 a_{11} \ldots a_{14} a_{21} a_{24} a_{25} a_{26} a_{27}]$$

$$[a_{15} \ldots a_{20}].$$

2 Special properties and closure properties

The next two propositions prove the importance of the functions $\tau_\theta^{n,k}$.

Proposition 2.1 *If G is a (n,k,θ)-linear matrix grammar, then there is a regular grammar G' such that:*

$$L(G') = \tau_\theta^{n,k}\ (L(G)).$$

Proof. Let $G = (V, \Sigma, S, P)$ be a $(n,k,\theta)-$lmg and we define the regular Chomsky grammar, $G' = (V_N, V_T, S, P')$, see also the notations of remark 1.3

$$V_N = V^n \cup \{S\}, \quad V_T = \Sigma_1 \cup \Sigma_2$$

and the set of rules is:

$$\begin{aligned} P' = \ & \{S \longrightarrow (A_1, \ldots, A_n) | (S \longrightarrow A_1 \ldots A_n) \in P\} \cup \\ \cup \ & \{(A_1, \ldots, A_n) \longrightarrow [\alpha_1\beta_1 \ldots \alpha_n\beta_n](B_1, \ldots, B_n) | \\ & (A_1 \longrightarrow \alpha_1 B_1 \beta_1, \ldots, A_n \longrightarrow \alpha_n B_n \beta_n) \in P\} \cup \\ \cup \ & \{(A_1, \ldots, A_n) \longrightarrow [\alpha_k] | (A_1 \longrightarrow \alpha_1, \ldots, A_n \longrightarrow \alpha_n) \in P \text{ where} \\ & \alpha_i = \lambda, \text{ for any } i = 1, \ldots, n \text{ with } i \neq k\}. \end{aligned}$$

We can prove by induction on the length of the derivations, that:

$$A_1 A_2 \ldots A_n \overset{*}{\underset{G}{\Longrightarrow}} u_1 B_1 v_1 u_2 B_2 v_2 \ldots u_n B_n v_n$$

if and only if:

$$(A_1, A_2, \ldots A_n) \overset{*}{\underset{G'}{\Longrightarrow}} \tau_\theta^{n,k}\ (u_1 v_1 u_2 v_2 \ldots u_n v_n)(B_1, B_2, \ldots, B_n).$$

From the above equivalence and from the definition of $\tau_\theta^{n,k}$ it follows easily our proposition.

Now, we turn to the converse of proposition 2.1.

Proposition 2.2 *Let k be such that $1 \leq k \leq n$ and let $\theta = (\varphi, \psi)$ be a pair of n-functions. If the language $L, L \subseteq \Sigma_1^* \Sigma_2$, is a regular language, then there is a (n,k,θ)-linear matrix grammar, G, such that:*

$$L(G) = \eta_\theta^{n,k}(L),$$

where $\eta_\theta^{n,k}$ is the inverse function of $\tau_\theta^{n,k}$ (see also the remark 1.3).

Proof. Let $G' = (V_N, \Sigma_1 \cup \Sigma_2, S', P')$ be a regular grammar such that $L(G') = L$. Without loossing of generality, we can assume that the nonterminal rules of $P', A \longrightarrow \alpha B$, has the property that $\alpha \in \Sigma_1$ and also we can assume that the terminal rules of $P', A \longrightarrow \alpha$, has the property that $\alpha \in \Sigma_2$.

This assumption follows from the condition $L \subseteq \Sigma_1^*\Sigma_2$. We define the (n,k,θ)-linear matrix grammar $G, G = (V, \Sigma, S, P)$ where: $V = V_N \cup \{S\}$, with S a new symbol and the rules:

$$\begin{aligned} P &= \{(S \longrightarrow \underbrace{S' \dots S'}_{n})\} \cup \{(A \longrightarrow \alpha_1 B \beta_1, \dots, A \longrightarrow \alpha_n B \beta_n) | \\ & \quad |A \longrightarrow [\alpha_1\beta_1 \dots \alpha_n\beta_n] B \in P', |\alpha_i| = \varphi(i), |\beta_i| = \psi(i), i = 1, \dots, n\} \cup \\ & \cup \quad \{(A \longrightarrow \alpha_1, \dots, A \longrightarrow \alpha_n) | A \longrightarrow \beta \in P', \beta \in \Sigma_2, \alpha_i = \lambda, \\ & \quad i = 1, \dots, n, i \neq k \text{ and } \alpha_k = \beta\} \end{aligned}$$

It follows easily that $L(G) = \eta_\theta^{n,k}(L)$.

Remark 2.3 *The propositions 2.1 and 2.2 are also true if G is a (n,k,φ) regular matrix grammar.*

Theorem 2.4 *For every $n, k \in \mathbf{N}$, with $1 \leq k \leq n$ and for every pair of n-functions, $\theta = (\varphi, \psi)$, the family $\mathcal{LM}_{n,k,\theta}$ is closed under union, intersection and complement.*

Proof. The closure under union is obvious. Therefore, it is enough to prove the closure under complement. If $L \in \mathcal{LM}_{n,k,\theta}, L \subseteq \Sigma^*$, then the language $\tau_\theta^{n,k}(L)$ is regular (Proposition 2.1). It follows that the language $L_1 = \Sigma_1^*\Sigma_2 - \tau_\theta^{n,k}(L)$ is also a regular language and $L_1 \subseteq \Sigma_1^*\Sigma_2$. From the proposition 2.2 we deduce that the language $\eta_\theta^{n,k}(L_1)$ is in $\mathcal{LM}_{n,k,\theta}$. But, $\tau_\theta^{n,k}$ is a bijective function and $\eta_\theta^{n,k}$ is the inverse function of $\tau_\theta^{n,k}$. It is easily to observe that $\eta_\theta^{n,k}(L_1) = \Sigma^* - L = CL$ and therefore $CL \in \mathcal{LM}_{n,k,\theta}$.

Corollary 2.5 *For every $n, k \in N$, with $1 \leq k \leq n$ and for every n-function φ, the family $\mathcal{RM}_{n,k,\theta}$ is closed under union, intersection and complement.*

3 Decidable problems

For a general discussion on decidable and undecidable problems in theory of matrix languages see the monography [3].

In the sequel we establish some decidable properties of the families $\mathcal{LM}_{n,k,\theta}$ and $\mathcal{RM}_{n,k,\varphi}$.

Theorem 3.1 *For every family $\mathcal{LM}_{n,k,\theta}$, the following problems are decidable:*

(1) Equivalence $(L_1 = L_2?)$

(2) Inclusion $(L_1 \subseteq L_2?)$

(3) Empty intersection $(L_1 \cap L_2 = \emptyset?)$

(4) Finite intersection (is $L_1 \cap L_2$ a finite set?)

(5) Empty complement $(CL = \emptyset?)$

(6) Finite complement (is CL a finite set?)

Proof.

(1) If $L_1, L_2 \in \mathcal{LM}_{n,k,\theta}$, then the languages $L_i' = \tau_\theta^{n,k}(L_i), i = 1,2$ are regular languages (see proposition 2.1.). But $\tau_\theta^{n,k}$ is a bijective function and therefore $L_1 = L_2$ if and only if $L_1' = L_2'$. The last equality is decidable.

(2) analogously.

(3)-(4). If $L_1, L_2 \in \mathcal{LM}_{n,k,\theta}$, then $L_1 \cap L_2 \in \mathcal{LM}_{n,k,\theta}$ (see theorem 2.4). But, for the family of simple matrix languages the emptiness problem and the finiteness problem are decidable problems (see [3]).

(5)-(6) If $L \in \mathcal{LM}_{n,k,\theta}$, then $CL \in \mathcal{LM}_{n,k,\theta}$ (see theorem 2.4) and the proof follows like in the (3)-(4) cases.

Corollary 3.2 *All problems from theorem 3.1 are decidable for every family* $\mathcal{RM}_{n,k,\varphi}$.

Remark 3.3 *All problems from the theorem 3.1 are undecidable for whole family of simple linear (regular) matrix languages (see [3]).*

In what it follows we establish the relation between the families $\mathcal{LM}_{n,k,\theta}$, $\mathcal{RM}_{n,k,\varphi}$ and the Chomsky families of languages.

Obviously every family $\mathcal{LM}_{n,k,\theta}$ is a proper subfamily of $\mathcal{LM}$, the family of all simple matrix languages. It is well-known that $\mathcal{LM}$ is a proper subfamily of $\mathcal{L}_1$, the family of dependent context languages (see [3], [4]). Therefore, for every $n, k, \in N, 1 \le k \le n$ and for every pair θ of n-functions is true that $\mathcal{LM}_{n,k,\theta} \underset{\neq}{\subset} \mathcal{L}_1$.

Consequently, it follows that $\mathcal{RM}_{n,k,\varphi} \underset{\neq}{\subset} \mathcal{L}_1$ for every $n, k, \in N, 1 \le k \le n$ and every n-function, φ.

Theorem 3.4 *(i) the regular family of languages, $\mathcal{L}_3$, is a proper subfamily of every family $\mathcal{RM}_{n,k,\varphi}(\mathcal{L}_3 \subsetneq \mathcal{RM}_{n,k,\varphi})$.*

(ii) $\mathcal{L}_3$ is a proper subfamily of every family $\mathcal{LM}_{n,k,\theta}.(\mathcal{L}_3 \underset{\neq}{\subset} \mathcal{LM}_{n,k,\theta})$.

Proof. Let L be a regular language, $L \in \mathcal{L}_3$. There is a finite deterministic automaton, $A = (Q, \Sigma, \delta, q_1, F)$ such that $L(A) = L$.

We shall describe only the main constructions.

(i) We define a (n, k, φ)-regular matrix grammar, $G = (V, \Sigma, S, P)$, such that $L(G) = L$.

Let S be a new symbol and consider $V = Q \times Q \cup \{S\}$. The set of rules, P, is:

(1) $(S \longrightarrow (q_1, q_1)(q_2, q_2) \dots (q_n, q_n)), q_i \in Q,\ i = 1, \dots, n$, where q_1 is the initial state of A.

(2) $((p_1, r_1) \longrightarrow \alpha_1(p_1, t_1), (p_2, r_2) \longrightarrow \alpha_2(p_2, t_2), \dots, (p_n, r_n) \longrightarrow \alpha_n(p_n, t_n))$, for every $\alpha_i \in \Sigma^*$ such that $|\alpha_i| = \varphi(i), \delta(r_i, \alpha_i) = t_i$, and $p_i, r_i, t_i \in Q$ for $i = 1, \dots, n$.

(3) $((q_1, q_2) \longrightarrow \lambda, (q_2, q_3) \longrightarrow \lambda, \dots, (q_k, q_k') \longrightarrow \beta, (q_{k+1}, q_{k+2}) \longrightarrow \lambda, \dots, (q_n, p) \longrightarrow \lambda)$, for every $\beta \in \Sigma^*$ such that $|\beta| < |\varphi|, \delta(q_k', \beta) = q_{k+1}, p \in F$ and $q_i \in Q, i = 1, \dots, n, q_k' \in Q$.

One can prove that $L(G) = L$.

(ii) Analogously, we define a (n, k, θ)-linear matrix grammar, $G(V, \Sigma, S, P)$, such that $L(G) = L$.

Let V be the set $Q^4 \cup \{S\}$, where S is a new symbol. The rules in P are:

(1') $(S \longrightarrow (q_1, q_1, s_1, s_1)(q_2, q_2, s_2, s_2) \dots (q_n, q_n, s_n, s_n))$, for every $q_i, s_i \in Q, i = 1, \dots, n$ (q_1 is the initial state of $\mathcal{A}$)

(2') $((q_i, p_i, s_i, r_i) \longrightarrow \alpha_i(q_i, t_i, s_i, u_i)\beta_i), 1 \leq i \leq n$ for every $\alpha_i, \beta_i \in \Sigma^*$ such that $|\alpha_i| = \varphi(i), |\beta_i| = \psi(i), \delta(p_i, \alpha_i) = t_i, \delta(r_i, \beta_i) = u_i$, for every $q_i, p_i, s_i, r_i, t_i, u_i \in Q, i = 1, \dots, n$.

(3') $((q_1, p_1, p_1, r_1) \longrightarrow \lambda, (r_1, p_2, p_2, r_2) \longrightarrow \lambda, \dots, (r_{k-1}, p_k, t_k, r_k) \longrightarrow \beta, \dots, (r_{n-1}, p_n, p_n, r_n) \longrightarrow \lambda))$, for every $\beta \in \Sigma^*$ such that $|\beta| < |\theta|, \delta(p_k, \beta) = t_k, r_n \in F$ and $q_i, p_i, r_i \in Q, i = 1, \dots, n, t_k \in Q$.

It is not difficult to verify that $L(G) = L$.

Corollary 3.5 *For every family $\mathcal{LM}_{n,k,\theta}(\mathcal{RM}_{n,k,\varphi})$ the equivalence problem between an arbitrary language from the family and an arbitrary regular language is decidable.*

Proof. The proof uses theorem 3.4 theorem 3.1 (1) and corollary 3.2 (1).

Remark 3.6 *The above problem is also undecidable for the family of all linear (regular) simple matrix languages (see [3], [4]).*

For every $n \geq 2$, the family of context free languages, $\mathcal{L}_2$, is incomparable with any family $\mathcal{LM}_{n,k,\theta}$ or $\mathcal{RM}_{n,k,\varphi}$. This follows from the fact that the language $L = \{a^n b^n | n \geq 1\}^*$ is a context free language but L is neither a simple regular matrix language nor a simple linear matrix language (see [3]).

4 Further questions

For every families $\mathcal{LM}_{n,k,\theta}$ or $\mathcal{RM}_{n,k,\varphi}$ one can prove specifically pumping lemmas or other properties.

An interesting open problem arises from the following fact:

In the case $n = 1$ the family $\mathcal{LM}_{1,1,0}$ is the same with the family $\mathcal{L}_{i,j}$, see [1] and [2]. It is known, [5] and [6], that if $\mathcal{L}_{i,j}$ and $\mathcal{L}_{i',j'}$ are different families, then $\mathcal{L}_{i,j} \cap \mathcal{L}_{i',j'} = \mathcal{L}_3$.

From this remark in [5] and [6] it was found an important decidable problem.

For $n > 1$ this problem: "if the families $\mathcal{LM}_{n,k,\theta}$ and $\mathcal{LM}_{n,k,\theta}$, are different families, then $\mathcal{LM}_{n',k',\theta'} \cap \mathcal{LM}_{n,k,\theta}, = \mathcal{L}_3$" is an open problem. Analogously, this problem is open for the families $\mathcal{RM}_{n,k,\varphi}$.

References

[1] Amar, V., Putzolu, G. (1964), On a family of linear grammars, Information and Control 7, 283-291.

[2] Amar, V., Putzolu, G. (1965) Generalization of regular events, Information and Control 8, 56-63.

[3] Dassow, J., Paun, G. - Regulated Rewriting in Formal Language Theory, Akademie-Verlag, Berlin, 1989.

[4] Paun G., Gramatici matriciale, Ed. St. si Enc., Bucuresti, 1981.

[5] Semenov, A. L. (1974) Regularity of languages k-linear for various k, Soviet. Math. Dokl . 15, 492-436.

[6] Yehudai, A. (1980). The decidability of equivalence for a family of linear grammars, Information and Control, 47, 122-136.

(Received June 16, 1989)