Theoretical Computer Science 58 (1988) 201-208
North-Holland

ON RECOGNIZABLE SUBSETS OF FREE PARTIALLY COMMUTATIVE MONOIDS

Y. METIVIER

U.E.R. de Mathématiques et d'Informatique, Université de Bordeaux I, 33405 Talence, France

Abstract. We show that, in a free partially abelian monoid generated by a finite alphabet A, the subset $[X]$ of A^* containing all the words equivalent to a word of X is recognizable if X is recognizable and if any iterative factor h has a connected noncommutation graph.

0. Introduction

A free partially commutative monoid is a monoid generated by a finite alphabet A where every relation is of the form $ab \sim ba$ (a and b belong to A). These monoids are considered as a model for studying problems occurring in parallelism, in particular, problems of control of parallelism: several processes compute independently (this authorizes concurrency) and can share resources.

Two operations commute if the result of the computation does not depend on the order of their executions.

Flé and Roucairol [8] have considered a finite set of words $X = \{x_1, \ldots, x_n\}$ constructed from disjoint alphabets such that every two letters occurring in the same word do not commute. They have shown, under this hypothesis, that the subset $[X^*]$ of A^* containing all words equivalent to a product of words of X is recognizable.

Cori and Perrin have established combinatorial properties of commutations. They have shown that if X is a recognizable set such that any word x in X is single in its equivalence class, then $[X^*]$ is recognizable [6, 12].

It is undecidable to determine whether the closure of a recognizable set is still recognizable. In this paper we provide a sufficient condition for the recognizability of the closure; and we show the decidability of this condition. More precisely, we show that if X is a recognizable subset of A such that, for any iterative factor h, the graph whose vertices are letters of h and whose edges are the pairs of letters which do not commute is connected, then the set of words equivalent to at least one word of X is recognizable.

1. Free partially commutative monoids

1.1. Basic definitions

Let A be a finite alphabet, we denote by A^* the set of all words over A. Let w be a word, the number of occurrences of the letter a in the word w will be denoted

0304-3975/88/$3.50

by $|w|_a$ and the subset of A formed by the letters actually occurring in w by alph(w): $\text{alph}(w) = \{a \in A \mid |w|_a \geqslant 1\}$.

A *partial commutation over* A^* is a congruence generated by a symmetric and irreflexive relation ϑ on A. More precisely, let ϑ be a binary relation on the set A. We suppose that ϑ is irreflexive and symmetric. For two letters a, b we denote $ab \sim ba$ whenever (a, b) belongs to ϑ.

Two words u and v are congruent, denoted by $u \sim v$, if there exist words $w_1, \ldots, w_k$ such that $w_1 = u$, $w_k = v$ and, for each i $(1 \leqslant i < k)$, $w_i = f_1abf_2$, $w_{i+1} = f_1baf_2$ with $(a, b) \in \vartheta$.

We denote by $M(A, \vartheta)$ the quotient of A^* by the congruence $\sim$. $M(A, \vartheta)$ is called *the free partially commutative monoid generated by A with respect to the relation* ϑ.

Let $\sim$ be a partial commutation over A^*; for x in A^*, the congruence class of x modulo ϑ is $[x] = \{y \in A^* \mid y \sim x\}$. This notation is extended to subset of A^* by $X \subset A^*$, $[X] = \{y \in A^* \mid \exists x \in X \text{ and } y \sim x\}$. The set $[X]$ is called *the closure of X modulo* ϑ.

The *noncommutation graph* associated with a subset B of A, denoted by $(B, \bar{\vartheta})$, is very useful; it is defined by

- its vertices are the letters of B,
- its edges correspond to noncommuting pairs of letters.

Sometimes, instead of saying $(\text{alph}(x), \bar{\vartheta})$ is connected, for short, we will say x is connected.

Let us first review some basic facts concerning recognizable sets and rational sets. Given a monoid M, a subset X of M is said to be recognizable if the family of sets $\{u^{-1}X \mid u \in M\}$ is finite, where $u^{-1}X$ denotes the subset of M defined by $u^{-1}X = \{v \in M \mid uv \in X\}$. The class of recognizable subsets of M is denoted by Rec(M).

The family of rational subsets of M, denoted Rat(M), is the least family of subsets of M satisfying the following conditions:

(1) $\emptyset \in R, \{m\} \in R\ \forall m \in M$,

(2) $A, B \in R \Rightarrow A \cup B$ and $AB \in R$,

(3) $A \in R \Rightarrow A^* = \bigcup_{n \geqslant 0} A^n \in R$.

Kleene's theorem asserts that if M is free and finitely generated, then the rational sets are precisely the recognizable sets.

Let X be a recognizable set, a word h is said to be an *iterative factor* if

$$\exists x_1 \in A^*\ \exists x_2 \in A^* \text{ such that } \quad x_1h^*x_2 \subset X.$$

1.2. Free partially commutative monoids and words

In this section, we give basic results about equivalent words. The first one is a characterization of equivalent words. For any subset B of A, we denote by $\pi_B(u)$ the projection of the word u on B^* which is obtained by erasing all letters which are not in B. If $B = \{a, b\}$, then π_B will also be denoted by $\pi_{a,b}$.

Proposition 1.1 (Cori and Perrin [6]). *Let u and v be two words of* A^*; *then* $u \sim v$ *if and only if*

(i) $\forall a \in A$: $|u|_a = |v|_a$,

(ii) $\forall (a, b) \notin \vartheta$: $\pi_{a,b}(u) = \pi_{a,b}(v)$.

The following proposition is of great help in order to build the set $[x]$. First we define a directed graph $\Gamma(x)$ associated to a word x. The word x may be considered as a mapping of $\{1, \ldots, m\}$ in A, where m is the length of x; then $\Gamma(x)$ is the directed graph the vertices of which are the numbers $\{1, \ldots, m\}$ and the arcs are given by the couples (i, j) such that $i < j$ and $(x(i), x(j)) \notin \vartheta$. A total order $<_\sigma$ on the set of vertices is said to be compatible with $\Gamma(x)$ if, for all arcs (i, j) of $\Gamma(x)$, $i <_\sigma j$. For any order $<_\sigma$ such that $i_1 <_\sigma i_2 <_\sigma \cdots <_\sigma i_m$, we define the word $x(\sigma) = x(i_1)x(i_2) \ldots x(i_m)$ and we have the following proposition.

Proposition 1.2 (Cori and Métivier [7]). *The set of words equivalent to x is the set of* $x(\sigma)$ *for the orders* $<_\sigma$ *compatible with* $\Gamma(x)$.

The last proposition characterizes the solutions of the equation $xy \sim zt$, and furthermore, it is useful in the proofs about recognizability. We say that two words u and v *absolutely commute*, denoted by $u \mathrel{C} v$ if $\mathrm{alph}(u) \times \mathrm{alph}(v) \subset \vartheta$.

Proposition 1.3 (Cori and Perrin [6]). *Let x, y, z and t be words of* A^*; *then,* $xy \sim zt$ *if and only if* $x \sim uv$, $y \sim rs$, $z \sim ur$ *and* $t \sim vs$ *for some words u, v, r and s with* $v \mathrel{C} r$.

2. Recognizable sets in commutation monoids

In this section we prove our main result.

Theorem 2.1. *If X is a recognizable subset of* A^* *such that each iterative factor of X has a connected noncommutation graph, then the closure of X is recognizable.*

2.1. Preliminaries

Let φ be the natural morphism of A^* into $M(A, \vartheta)$, let X be a subset of A^*; one has $[X] = \varphi^{-1}(\varphi(X))$. Furthermore, φ is surjective; thus we can state the following remark.

Remark 2.2. $[X] \in \mathrm{Rec}(A^*) \Leftrightarrow \varphi(X) \in \mathrm{Rec}(M(A, \varphi))$.

There is a well-known family of partially commutative monoids, the family obtained by the cartesian product of free monoids. From this family we deduce the following remark.

Remark 2.3 (Berstel and Bertoni, Mauri and Sabadini [3, 4]). It is undecidable to determine whether the closure of a recognizable subset of A^* is recognizable.

Remark 2.3 results from the following classical theorem: Let A and B be two alphabets and let $X \in \mathrm{Rat}(A^* \times B^*)$; it is then undecidable to determine whether X is recognizable.

In order to study the recognizability of the closure of elements of $\mathrm{Rec}(A^*)$, we consider the closure of properties under rational operations: union, product and star operations. First, we have the following proposition.

Proposition 2.4 (Cori and Perrin [6]). *Let X and Y be such that $[X]$ and $[Y]$ belong to* $\mathrm{Rec}(A^*)$; *then $[X \cup Y]$ and $[XY]$ belong to* $\mathrm{Rec}(A^*)$

2.2. Closure under star operation

We show the following proposition.

Proposition 2.5. *Let X be a recognizable subset of A^* such that $[X] = X$ and, for any word x of X,* $(\mathrm{alph}(x), \bar{\vartheta})$ *is connected; then $[X^*]$ is recognizable.*

We need two technical lemmas.

Lemma 2.6. *Let u, v, v' be words satisfying the following properties:*
(i) $(\mathrm{alph}(uv), \bar{\vartheta})$ *is connected,*
(ii) $v' \mathrm{C}\, v$;
then $\mathrm{alph}(v') \neq \mathrm{alph}(u)$.

Proof. If $(\mathrm{alph}(uv), \bar{\vartheta})$ is connected, then there exist $a \in \mathrm{alph}(v)$ and $b \in \mathrm{alph}(u)$ such that $(a, b) \notin \vartheta$. But $v' \mathrm{C}\, v$ thus $b \notin \mathrm{alph}(v')$, i.e., $\mathrm{alph}(v') \neq \mathrm{alph}(u)$. □

Lemma 2.7. *Let X be a subset of A^* such that $[X] = X$ and $\forall x \in X$:* $(\mathrm{alph}(x), \bar{\vartheta})$ *is connected; then the product uv of two words of A^* belongs to $[X^*]$ if and only if $\exists n \geq 0$, $y_i, z_i \in X^*$ $(0 \leq i \leq n)$ and $u_i v_i \in A^*$ $(1 \leq i \leq n)$ with*
(1) $y_0 u_1 y_1 u_2 \ldots u y_n \sim u$,
(2) $z_0 v_1 z_1 v_2 \ldots v_n z_n \sim v$,
(3) $u_i v_i \in X$,
(4) $z_i \mathrm{C}\, (u_{i+1} y_{i+1} \ldots u_n y_n)$ $(0 \leq i \leq n)$,
(5) $v_i \mathrm{C}\, (y_i u_{i+1} \ldots u_n y_n)$ $(1 \leq i \leq n)$,
(6) $\mathrm{alph}(u_i) \neq \mathrm{alph}(u_j)$ $(1 \leq i < j \leq n)$.

Proof. This condition is obviously sufficient.

Conversely, we can use an induction on the minimal integer k such that $uv \in [X^k]$.

If $k = 0$ there is nothing to do.

Otherwise, there exist $x \in X$, $y \in X^{k-1}$ such that $uv \sim xy$. By Proposition 1.3, $x \sim pq$, $y \sim p'q'$, $u \sim pp'$, $v \sim qq'$ and $q \mathrm{C}\, p'$ for some words p, p', q and q'.

As $p'q' \in [X^{k-1}]$, by the induction hypothesis, $\exists n \geq 1$, $\exists y_i, z_i \in X^*$ $(1 \leq i \leq n)$ and $u_i, v_i \in A^*$ $(2 \leq i \leq n)$ with

(1) $y_1 u_2 y_2 u_3 \dots u_n y_n \sim p'$,
(2) $z_1 v_2 z_2 v_3 \dots v_n z_n \sim q'$,
(3) $u_i v_i \in X$ $(2 \leq i \leq n)$,
(4) $z_i \, \mathrm{C} \, (u_{i+1} y_{i+1} \dots u_n y_n)$ $(1 \leq i \leq n)$,
(5) $v_i \, \mathrm{C} \, (y_i u_{i+1} \dots u_n y_n)$ $(2 \leq i \leq n)$,
(6) $\mathrm{alph}(u_i) \neq \mathrm{alph}(u_j)$ $(2 \leq i < j \leq n)$.

Furthermore, $[X] = X$ thus $pq \in X$. We distinguish three cases.

Case 1: $p = 1$; then $q \in X$. We have $p' \sim u$ and $qq' \sim v$. Let $z = qz_1$, $qz_1 \in X^*$ and $q \, \mathrm{C} \, p'$; then $z \, \mathrm{C} \, (u_2 y_2 \dots u_n y_n)$ and we find the lemma with z instead of z_1.

Case 2: $q = 1$. Let $y = py_1$ and we take y instead of y_1.

Case 3: $p \neq 1$ and $q \neq 1$. The graph $(\mathrm{alph}(pq, \bar{\vartheta})$ is connected and $\forall i$: $q \, \mathrm{C} \, u_i$; by Lemma 2.6, $\forall i$: $\mathrm{alph}(p) \neq \mathrm{alph}(u_i)$. Let $u_1 = p$, $v_1 = q$, $y_0 = z_0 = 1$; this yields the desired decomposition. As $\mathrm{alph}(u_i) \neq \mathrm{alph}(u_j)$, we have $n \leq \mathrm{Card}(A)$. □

Proof of Proposition 2.5. Let $Y = [X^*]$; we now prove that the family $\{u^{-1}Y \mid u \in A^*\}$ is finite. Let σ be defined by $\forall u \in A^*$: $\sigma(u)$ is the family of $3n$-tuples:

$$(\mathrm{alph}(u_1), u_1^{-1}X, \mathrm{alph}(y_1), \dots, \mathrm{alph}(u_n), u_n^{-1}X, \mathrm{alph}(y_n))$$

with

(1) $y_0 u_1 y_1 \dots u_n y_n \sim u$,
(2) $y_i \in X^*$ $(0 \leq i \leq n)$,
(3) $u_i \in A^+$ and $\mathrm{alph}(u_i) \neq \mathrm{alph}(u_j)$ $(1 \leq i < j \leq n)$.

The set X is recognizable, A is finite and $n \leq \mathrm{Card}(A)$; consequently, the family $\{\sigma(u) \mid u \in A^*\}$ is is finite.

Next we prove that $\forall u, t \in A^*$: $\sigma(u) = \sigma(t) \Rightarrow u^{-1}Y = t^{-1}Y$. Assume $\sigma(u) = \sigma(t)$ and $uv \in Y$; we prove that $tv \in Y$. Since $uv \in Y$, by Lemma 2.7,

$$y_0 u_1 y_1 \dots u_n y_n \sim u \quad \text{and} \quad z_0 v_1 z_1 \dots v_n z_n \sim v$$

(with conditions (1)-(6)). Since $\sigma(u) = \sigma(t)$, $t \sim x_0 t_1 x_1 \dots t_n x_n$ with $x_i \in X^*$, $\mathrm{alph}(x_i) = \mathrm{alph}(y_i)$, $t_i^{-1}X = u_i^{-1}X$, $\mathrm{alph}(t_i) = \mathrm{alph}(u_i)$. Since $u_i v_i \in X$ and $t_i^{-1}X = u_i^{-1}X$, it follows that $t_i v_i \in X$. Furthermore, $z_i \, \mathrm{C} \, (u_{i+1} x_{i+1} \dots x_n t_n)$ and $\mathrm{alph}(u_{i+1} \dots u_n y_n) = \mathrm{alph}(t_{i+1} \dots t_n x_n)$, thus $z_i \, \mathrm{C} \, (t_{i+1} x_{i+1} \dots t_n x_n)$ and $v_i \, \mathrm{C} \, (x_i t_{i+1} \dots t_n x_n)$; then $tv \in Y$. Thus the proposition has been proved. □

Note that this proposition has been obtained by successive generalizations of previous results in [8, 6, 7].

To prove Theorem 2.1, we use the notion of starheight of a rational set; let us recall this notion: Let M be a monoid, and define inductively sets $\mathrm{Rat}_0(M) \subset \mathrm{Rat}_1(M) \subset \cdots$ by

- $X \in \mathrm{Rat}_0(M)$ if X is a finite subset of M,
- $X \in \mathrm{Rat}_{k+1}(M)$ if X is a finite union of sets of the form $Y_1 Y_2 \dots Y_n$ where either Y_i is a singleton or $Y_i = Z_i^*$ for some $Z_i \in \mathrm{Rat}_k(M)$.

We have $\mathrm{Rat}(M)=\bigcup_{k\geq 0}\mathrm{Rat}_k(M)$. The sets in $\mathrm{Rat}_k(M)\backslash\mathrm{Rat}_{k-1}(M)$ are said to have *starheight k*.

Proof of Theorem 2.1. We use an induction on the starheight of X. The result is obvious if $X\in\mathrm{Rat}_0(A^*)$.

Let $X\in\mathrm{Rat}_{k+1}(A^*)$. Thus X is a finite union of sets $X_1\ldots X_n$ where X_i is either a singleton or has the form Y_i^*, Y_i having starheight less than or equal to k.

By Proposition 2.4, it suffices to show that $[X_i]$ is recognizable for any i. If X_i is a singleton, this is true. Otherwise, $X_i=Y_i^*$; any iterative factor of Y_i is an iterative factor of X, thus any iterative factor of Y_i has a connected noncommutation graph, and, by the induction hypothesis, $[Y_i]$ is recognizable. Furthermore, any element y of Y_i is an iterative factor of X, thus y has a connected noncommutation graph; consequently, by Proposition 2.5 $[[Y_i]^*]$ is recognizable. □

Let X be recognizable set, there exists an effective procedure for deciding whether X has connected iterative factors. Indeed, we prove that the alphabets of iterative factors are equal to the alphabets of the loops in the minimal automaton recognizing X. Then it suffices to compute the alphabets of the loops using the McNaughton and Yamada algorithm [10].

Proposition 2.8. *Let $X\in\mathrm{Rec}(A^*)$ and let Δ be the minimal automaton recognizing X; the following conditions are equivalent:*

(i) *every iterative factor of X has a connected noncommutation graph,*

(ii) *the alphabet of any loop of Δ has a connected graph.*

Proof. The implication (i)⇒(ii) is obvious.

We prove (ii)⇒(i). Let $\Delta=\langle A,Q,q_0,F,\delta\rangle$ where Q is the finite set of states, q_0 is the initial state, F is the set of final states and δ is the next-state function ($\delta:Q\times A\to Q$). Let x be an iterative factor of X; then there exist two words x_1 and x_2 such that $x_1x^*x_2\subset X$. Let (f_i) be the sequence of words such that f_i is left factor of x_1x^* and $|f_i|=|X_1|+i$ $(i\geq 0)$. Consider the sequence $(\delta(q_0,f_i))$. Q is finite, thus $\exists j\in Q$ $\exists n$, $p\in\mathbb{N}$, $\exists x_3,x_4,x_5,x_6\in A^*$ such that

- $x_3x_4=x_5x_6=x$,
- $\delta(q_0,x_1x^nx_3)=j$,
- $\sigma(q_0,x_1x^nx_3x_4x^px_5)=j$.

Thus there exists a loop c verifying $\mathrm{alph}(c)=\mathrm{alph}(x)$. If (ii) holds, we deduce that $(\mathrm{alph}(x),\bar{\vartheta})$ is connected. □

3. Commentaries

First we give an example such that X has no a connected iterative factor but $[X]$ is recognizable.

Example 3.1

$$A = \{a, b, c\} \qquad \vartheta = \{(a, b), (b, a), (a, c), (c, a), (b, c), (c, b)\},$$

and $X = (b + c)^* a(bc)^*$. We have bc is not connected and $[X] = (b + c)^* a(b + c)^*$ is recognizable.

If we consider a lexicographic order on A^*, we can definne the smallest lexicographic element in each congruence class, denoted by $\mathrm{Inf}(x)$ $(x \in A^*)$. The set of words in lexicographic normal form of a set X is denoted by $\mathrm{Inf}(X)$. There exists a characterization of the smallest lexicographic element in each congruence class [2]: $w = \mathrm{Inf}(w)$ if and only if, for each factorization $w = w_1 b w_2 a w_3$ and $a < b$, there exists a letter $c \in \mathrm{alph}(bw_2)$ such that $(a, c) \notin \vartheta$.

If C_a denotes the set of letters which commute with a, we have

$$\mathrm{Inf}(A^*) = A^* \setminus \bigcup_{a \in A} \bigcup_{\substack{b > a, \\ b \in C_a}} A^* b C_a^* a A^*.$$

Thus, $\mathrm{Inf}(A^*)$ is recognizable.

Ochmanski [13] has shown the following result.

Theorem 3.2 (Ochmanski [13]). *The closure of a recognizable set X is recognizable if and only if* $\mathrm{Inf}(X)$ *is recognizable.*

Proof. Note that we can prove this result of Ochmanski using Theorem 2.1. The proof of this theorem also needs the following lemma ([13, Lemma 4.2]).

Lemma 3.3. *If $w = \mathrm{Inf}(w)$ and w is not connected, then $ww \neq \mathrm{Inf}(ww)$.*

From this lemma we deduce that if $\mathrm{Inf}(X)$ is recognizable, then any iterative factor of $\mathrm{Inf}(X)$ is connected; by Theorem 2.1, $[\mathrm{Inf}(X)] = [X]$ is recognizable. Conversely, if $[X]$ is recognizable $\mathrm{Inf}(X) = \mathrm{Inf}(A^*) \cap [X]$ is also rcognizable. $\square$

References

[1] I.J.J. AAlbersberg and G. Rozenberg, Trace Theory—a survey, *Tech. Rept.*, Institute of Applied Mathematics and Computer Science, University of Leiden (1985).

[2] A.V. Anisimov and D. Knuth, Inhomogeneous sorting, Unpublished manuscript.

[3] J. Berstel, *Transductions and Context-Free Languages* (Teubner, Stuttgart, 1979).

[4] A. Bertoni, G. Mauri and N. Sabadini, Equivalence and membership problems for regular trace languages, in *Proc. 9th ICALP* 1982, Lecture Notes in Computer Science **140** (Springer, Berlin, 1982) 61–71.

[5] P. Cartier and D. Foata, Problèmes combinatoires de commutation et de réarrangements, in: Lecture Notes in Mathematics **85** (Springer, Berlin, 1969).

[6] R. Cori and D. Perrin, Sur la reconnaissabilité dans les monoïdes partiellement commutatifs libres, *RAIRO Inform. Théor.* **19** (1985) 21–32.

[7] R. Cori and Y. Métivier, Recognizable subsets of partially abelian monoids, *Theoret. Comput. Sci.* **35** (1985), 179–189.

[8] M.P. Flé and G. Roucairol, Maximal serializability of iterated transactions, *Theoret. Comput. Sci.* **38** (1985) 1–16.

[9] M. Lothhaire, *Combinatorics on Words* (Addison-Wesley, Reading, MA, 1983).

[10] R. McNaughton and H. Yamada, Regular expressions and state graphs for automata, *IRE Trans. On Electric Comput.* **EC-9** (1960) 38–47.

[11] Y. Métivier, Une condition suffisante de reconnaissabilité dans un monoïde partiellement commutatif, *RAIRO Inform. Théor.* **20** (1986) 121–127.

[12] D. Perrin, Words over a partially commutative alphabet, in: A. Apostolico and Z. Galil, eds., *Combinatorial Algorithms on Words* NATO ASI-Series. **F12** (Springer, Berlin, 1985).

[13] E. Ochmanski, Regular behaviour of concurrent systems, *EATCS Bull.* October 1985.

[14] G. Viennot, Empilement et commutation: théorie combinatoire et applications, in: *Proc. Coll. Combinatoire énumérative*, Montréal 1985, Lecture Notes in Mathematics **1234** (Springer, Berlin, 1987) 321–350.