TORSION MATRIX SEMIGROUPS AND RECOGNIZABLE TRANSDUCTIONS.*

Jean-Paul Mascle.
L.I.T.P.
2, Place Jussieu.
75221. PARIS Cedex.

Introduction.

In 1966, J.A. Brzozowski formulated the problem of "limited regular subsets of a free monoid":

"Given a regular language $L$, can we decide whether $L^* = 1 \cup L \cup L^2 \cup \ldots \cup L^n$ for some integer $n"$

Several partial results were published until the year 1978, when K. Hashiguchi and I. Simon independently solved this problem. Hashiguchi's proof is a combinatorial one involving automata whereas Simon's method is based on the decidability of the finiteness of monoids of matrices with entries in a certain semiring $M$. Three years later, and in connection with the starheight problem, Hashiguchi established a deep result [8] that can also be viewed as a decidability result on the same monoids.

It was the idea of Ch. Choffrut to identify this semiring $M$ with a subsemiring of $\text{Rat}(a^*)$. And to notice that generalizations of Simon's and Hashiguchi's results would provide the solution to a well-known problem on rational transductions. In this paper, we are going to extend I. Simon's results over $M$ to $\text{Rat}(a^*)$ and get the following theorem:

**Theorem 3.5.** Every torsion subsemigroup of $((\text{Rat}(a^*))^{n \times n}, \cdot)$ is locally finite.

**Corollary 3.6.** Given a finite subset $X$ of $(\text{Rat}(a^*))^{n \times n}$, it is possible to decide whether the subsemigroup $S$ generated by $X$ is finite.

This corollary contains, as particular cases:

- We can decide whether a square matrix with entries in $\text{Rat}(a^*)$ is torsion.
- Given a finite number of languages in $\text{Rat}(a^*)$, we can decide whether the set of all possible products of these languages is finite.

From an algebraic point of view, some of the results stated hereafter are connected with the "Burnside's problem for semigroups", over which the literature is abundant. Let us just mention a conjecture expressed by H. Straubing [11]), known to be true for rings:

"A finitely generated torsion semigroup of $n \times n$ matrices over a commutative semiring is finite."

Theorem 3.5 states that $\text{Rat}(a^*)$, considered as a commutative semiring with its classical operations of union and product, satisfies the conjecture. $\text{Rat}(a^*)$ gives a new example for which the conjecture is true and which is "very far from being a ring" since its both laws are idempotent.

From the point of view of classical language theory, our main theorem can be formulated as a result concerning recognizable transductions between two free monoids $A^*$ and $a^*$, the second being over a one-letter alphabet. Given a rational transduction $\tau$, we know (see for instance [2] p.85) that $\tau$ can be defined for every word $w$ in $A^*$ by $w \tau = \mu.w\lambda.\nu$, where $\mu$ is a morphism of $A^*$ onto a semigroup of matrices $S$ whose entries lie in $\text{Rat}(a^*)$ and $\lambda$ (resp. $\nu$) is a fixed row- (resp. column-) vector of elements of $\text{Rat}(a^*)$.

When $\tau$ is recognizable, there exist such definitions of $\tau$ for which $S$ is finite, and conversely, whenever $S$ is finite, we are assured that $\tau$ is recognizable ([4], [5]). Our theorem makes possible the effective test of the finiteness of $S$. In that sense, we might see our theorem as an attempt to clarify the following problem, which is open since a long time:

"Given a rational transduction $\tau$ from $A^*$ to $a^*$, can we decide whether $\tau$ is recognizable?"

*This work has been partly supported by the research program "Mathématiques et Informatique" of the Ministère de la Recherche et de la Technologie.
We recall that this problem has been solved by Ginsburg and Spanier in the case when $A$ is also a one-letter alphabet. ([6]; for a recent proof, see [7]). When both alphabets contain at least two letters, the problem is undecidable ([2] p.90). We will conclude by showing this problem solvable in the following particular cases:

**Theorem 3.7.** (a) Let $\tau : A^* \rightarrow a^*$ be a rational substitution. We can decide whether $\tau$ is recognizable.

(b) Let $\tau : a^* \rightarrow A^*$ be a rational substitution. We can decide whether $\tau$ is recognizable.

We shall note that by virtue of results of Ch.Choffrut [4], (b) is equivalent to the decidability of the limitedness of regular sets which was our starting point.

I am indebted to Ch.Choffrut for having built up the framework of this paper, and I must thank J.Sakarovitch for having directed my work and given many useful hints.

**Basic facts and notations.**

Let $A$ be a finite, non-empty alphabet, $K$ a semiring and $Q$ a finite non-empty set. The set $K^{Q\times Q}$ of all $Q\times Q$-matrices with entries in $K$ is a semiring, and in particular, a monoid for the classical matrix multiplication. By definition, a $K$-automaton with dimension $Q$ over the free monoid $A^*$ is a triplex $T = (Q, \mu, \nu)$, where $\lambda$ is a row vector in $K^{1\times Q}$, $\nu$ a column vector in $K^{Q\times 1}$ and $\mu$ a morphism from $A^*$ to $K^{Q\times Q}$.

The behavior of an automaton $T$ is the formal series $S = \sum (\lambda, w\mu, v)w$, $w$ ranging over $A^*$. One can view such a series as a mapping from $A^*$ into $K$, write: $\lambda, w\mu, v = (S, w) = wS$ and if $E$ is any subset $K$, designate $\{w \in A^* \mid wS \in E\}$ by $ES^{-1}$. We can also say that $(\lambda, \mu, \nu)$ is a linear representation of $S$. The set of all the series thus obtained is named the set of the $K$-recognizable series over $A^*$. According to a theorem due to Schützenberger, this set is also the set of $K$-rational series over $A^*$: it can then be indifferently be referred to as $K$-Rec($A^*$) or as $K$-Rat($A^*$).

The language recognized by $T$ is by definition the support of $S$:

$$supp(S) = \{ w \in A^* \mid (\lambda, w\mu, v) \neq 0_K \}.$$ 

If $L$ designates any language over $A^*$, the characteristic series of $L$ is defined by: $L = \sum w$, $w$ ranging over $L$. Note that $supp(L) = L$.

The monoid of an automaton $A$ is the monoid $A^*\mu$.

We shall say that a series $S$ is cofinite if the set of all coefficients of $S$ is finite. For more information about formal power series, two reference books are [1] and [9].

In the following, $B$ designates the boolean ring with two elements; $N$ is the set of non-negative integers. The regular (rational, recognizable...) languages over $A^*$ can be defined as the languages recognized by the $B$-automata over $A^*$. We shall call Rat($A^*$) or Rec($A^*$) the set constituted by these languages. Rat($A^*$) equipped with ordinary union and product is a semiring, which is commutative when $A$ is a one-letter alphabet.

If $R \in$ Rat($a^*$), the syntactic monoid of $R$ is cyclic and must be some $\langle N, + \rangle$, this notation standing for the quotient of $(N, +)$ by the smallest congruence ~ such that $r - r + p$. ($r \geq 0, p \geq 1$ are fixed)

$R$ can be written in a "normalized" way under the form:

$$R = F \cup G(a^p)*$$

where $F \subseteq \{ 1, a, \ldots, a^{r-1} \}$ and $G \subseteq \{ a^r, \ldots, a^{r+p-1} \}$.

We can refer to $r$ as the index and $p$ as the period of $R$. More generally, we shall say that $p$ is one period of the rational set $R$ over $a^*$, or that $R$ is a rational set with period $p$, if $p$ is a non-zero multiple of the period of $R$. The set of all rational sets with period $p$ is a subsemiring of Rat($a^*$).

If $(R_1)_{1 \leq i \leq n}$ is a finite family of rational sets $R_i$ with index $r_i$ and period $p_i$, we can define $r = \max\{ r_i \mid 1 \leq i \leq n \}$ and $p = \lcm\{ p_i \mid 1 \leq i \leq n \}$ and write every $R_i$ under the form:

$$R_i = F_i \cup G_i(a^p)*$$

where $F_i \subseteq \{ 1, a, \ldots, a^{r_i-1} \}$ and $G_i \subseteq \{ a^r, \ldots, a^{r+p-1} \}$.

An element $s$ of a semigroup $S$ is torsion if the semigroup generated by $s$ is finite. In that case, there exists a couple $(r, r+p)$ of two distinct integers of minimum sum such that $s^r = s^{r+p}$; $r$ is called the index of $s$ and $p$ is its period.