INFORMATIQUE THÉORIQUE ET APPLICATIONS

J. HONKALA

On lindenmayerian rational subsets of monoids

Informatique théorique et applications, tome 31, nº 1 (1997), p. 81-96
<http://www.numdam.org/item?id=ITA_1997__31_1_81_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 31, n° 1, 1997, pp. 81-96)

ON LINDENMAYERIAN RATIONAL SUBSETS OF MONOIDS (*)

by J. Honkala ([1])

Communicated by J. Berstel

Abstract. – *We define the family of L rational subsets in an arbitrary monoid. We discuss also L rational relations, L rational transductions and L rational star height.*

Résumé. – *Nous définissons la famille des parties L-rationnelles d'un monoïde quelconque. Nous discutons également les relations L-rationnelles, les transductions L-rationnelles, et la hauteur d'étoile L-rationnelle.*

1. INTRODUCTION

Automata and language theory have close connections to the study of semigroups and monoids. In the study of free monoids and their subsets considerations concerning larger classes of monoids are often useful. For example, Eilenberg showed how varieties of monoids can be used to classify various classes of regular languages.

Based on ideas from automata and language theory Eilenberg defined in an arbitrary monoid the classes of recognizable and rational subsets. The purpose of this paper is to establish another link between language and semigroup theory by defining in an arbitrary monoid the class of L rational subsets. This definition is again based on language theory, more precisely, the theory of Lindenmayer systems (*see* Rozenberg and Salomaa [4]). The resulting family appears very natural also from an algebraic point of view. The difference between the definitions of rational and L rational subsets is that the Kleene closure is replaced by morphic closure.

A brief outline of the contents of the paper follows. Section 2 contains the definition of L rational subsets of a monoid. In Section 3 we discuss the

(*) Received September 1996, accepted February 1997.
([1]) Department of Mathematics, University of Turku, SF-20500 Turku, Finland.
E-mail: jhonkala@sara.cc.utu.fi

connections between L rational sets and HDTOL languages. It is seen that in the framework of L rationality, DTOL and HDTOL languages play the same role as regular languages in the case of rational subsets. In Section 4 we define L rational relations and transductions. We establish an analogue of Nivat's theorem for L rational transductions and give many examples. We also show that rational transductions are L rational transductions. Finally, in Section 5 we define the star height of an L rational set and show that star height induces an infinite hierarchy in the general case.

We assume that the reader is familiar with the basics concerning rational sets and transductions (*see* Berstel [1]) and Lindenmayer systems (*see* Rozenberg and Salomaa [4]).

2. DEFINITIONS AND EXAMPLES

Suppose M is a monoid. If A, $B \subseteq M$ and $h_1, \ldots, h_s$ are endomorphisms of M we denote

$$AB = \{ab \mid a \in A,\, b \in B\},$$
$$(h_1 + \ldots + h_s)^+ (A) = \bigcup_{k \geq 1,\, 1 \leq i_1, \ldots, i_k \leq s} h_{i_1} h_{i_2} \ldots h_{i_k} (A)$$

and

$$(h_1 + \ldots + h_s)^* (A) = \bigcup_{k \geq 0,\, 1 \leq i_1, \ldots, i_k \leq s} h_{i_1} h_{i_2} \ldots h_{i_k} (A).$$

DEFINITION 2.1: *Suppose M is a monoid and $\mathcal{H}$ is a set of endomorphisms of M. The family $\mathrm{L}_{\mathcal{H}}\,\mathrm{Rat}\,(M)$ of Lindenmayerian rational subsets of M (shortly, L rational subsets of M) with respect to $\mathcal{H}$ is the least family $\mathcal{R}$ of subsets of M satisfying the following conditions:*

(i) $\emptyset \in \mathcal{R}$, $\{m\} \in \mathcal{R}$ *for all* $m \in M$;

(ii) *if* A, $B \in \mathcal{R}$ *and* $h \in \mathcal{H}$, *then* $A \cup B \in \mathcal{R}$, $AB \in \mathcal{R}$ *and* $h(A) \in \mathcal{R}$;

(iii) *if* $A \in \mathcal{R}$ *and* $h_1, \ldots, h_s \in \mathcal{H}$, *then* $(h_1 + \ldots + h_s)^* (A) \in \mathcal{R}$.

Hence, $\mathrm{L}_{\mathcal{H}}\,\mathrm{Rat}\,(M)$ is the least family containing the finite subsets of M and closed under finite union, product, $\mathcal{H}$-morphic image and $\mathcal{H}$-morphic star. Union, product, $\mathcal{H}$-morphic image and $\mathcal{H}$-morphic star are called the L rational operations with respect to $\mathcal{H}$. If $\mathcal{H} = \mathrm{End}\,(M)$, the set of endomorphisms of M, we denote $\mathrm{LRat}\,(M) = \mathrm{L}_{\mathcal{H}}\,\mathrm{Rat}\,(M)$.

In the presence of (ii), condition (iii) holds if and only if for all $A \in \mathcal{R}$ and $h_1, \ldots, h_s \in \mathcal{H}$ we have $(h_1 + \ldots + h_s)^+ (A) \in \mathcal{R}$.

Suppose $A \subseteq M$ and $h_1, \ldots, h_s$ are endomorphisms of M. Then the least solution of the equation

$$L = A \cup h_1 (L) \cup \ldots \cup h_s (L)$$

is given by $(h_1 + \ldots + h_s)^* (A)$. This is the basic reason why we allow more than one morphism in (iii) of Definition 2.1. Note that in Rat (M) we can solve the analogous equation

$$L = A \cup B_1 L \cup \ldots \cup B_s L$$

where $A,\ B_1, \ldots, B_s \subseteq M$.

Condition (ii) of Definition 2.1 guarantees that LRat (M) is a subsemiring of $\mathcal{P}(M)$ closed under endomorphisms of M. In the case of rational sets closure under morphisms follows from the other conditions. For L rational sets, on the contrary, this has to be postulated separately. Indeed, denote $X = \{a, b\}$, define the morphisms $f, g : X^* \to X^*$ by $f(a) = f(b) = a$, $g(a) = b^2$, $g(b) = a^2$ and consider the least subsemiring $\mathcal{R}_1$ of $\mathcal{P}(X^*)$ satisfying (i) and (iii) with $\mathcal{H} = \{f, g\}$. It is easy to see that each infinite set in $\mathcal{R}_1$ has minimal alphabet X. Therefore, because the language $\{a, b^2, a^4, b^8, \ldots\}$ belongs to $\mathcal{R}_1$, the family $\mathcal{R}_1$ is not closed under $\mathcal{H}$-morphic image.

The other conditions of (ii) cannot be deleted either. The necessity of $A \cup B \in \mathcal{R}$ is seen by considering the case $\mathcal{H} = \emptyset$. To see the necessity of $AB \in \mathcal{R}$ consider again $X^* = \{a, b\}^*$ and define the morphism $h : X^* \to X^*$ by $h(a) = h(b) = b^2$. Then the least family $\mathcal{R}_2$ satisfying (i), (iii) and the first and third condition of (ii) with $\mathcal{H} = \{h\}$ has the property that each language in $\mathcal{R}_2$ contains only finitely many words having letter a. On the other hand, the set $\{ab, ab^2, ab^4, \ldots\}$ is clearly $\mathrm{L}_{\mathcal{H}}$ rational.

Example 2.1: Let $M = X^*$ be the free monoid generated by the finite nonempty set X. If $G = (X, g_1, \ldots, g_n, w)$ is a DTOL system, then

$$L(G) = (g_1 + \ldots + g_n)^* (w) \in \text{LRat}(M).$$

Furthermore, if $h : X^* \to X_1^*$ is a morphism, then $h(L(G)) \in$ LRat $((X \cup X_1)^*)$. Hence, if L is an HDTOL language, there exists a free monoid Y^* such that $L \in$ LRat (Y^*).

Example 2.2: Let $M = (\mathbf{N}, +, 0)$. Denote by A the set of those nonnegative integers whose binary expansions have precisely three nonzero digits. Hence,

$$A = \{2^{i_1} + 2^{i_2} + 2^{i_3} | 0 \leq i_1 < i_2 < i_3\}.$$

We claim that A is L rational with respect to $\mathcal{H} = \{h\}$ where h is defined by $h(x) = 2x$ for $x \in \mathbf{N}$. First, $\{1\} \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}(\mathbf{N})$. Hence,

$$h^+(\{1\}) = \{2^{i_3} | i_3 \geq 1\} \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}(\mathbf{N})$$

and

$$1 + h^+(\{1\}) = \{1 + 2^{i_3} | i_3 \geq 1\} \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}(\mathbf{N}).$$

Therefore,

$$h^+(1 + h^+(\{1\})) = \{2^{i_2} + 2^{i_3} | 1 \leq i_2 < i_3\} \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}(\mathbf{N}).$$

Finally,

$$h^*(1 + h^+(1 + h^+(\{1\}))) = A \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}(\mathbf{N}).$$

In what follows we consider the empty set to be a DTOL and an HDTOL language. To conclude this section we define the sets corresponding to HDTOL languages in arbitrary monoids.

DEFINITION 2.2: *Suppose M is a monoid and $\mathcal{H} \subseteq \mathrm{End}(M)$. A subset $A \subseteq M$ is called an $\mathcal{H}$-DTOL set if there exist $m \in M$ and $h_1, \ldots, h_s \in \mathcal{H}$ such that*

$$A = (h_1 + \ldots + h_s)^*(m)$$

or $A = \emptyset$. A subset $A \subseteq M$ is called an $\mathcal{H}$-HDTOL set if there exist $m \in M$ and $h, h_1, \ldots, h_s \in \mathcal{H}$ such that

$$A = h((h_1 + \ldots + h_s)^*(m))$$

or $A = \emptyset$.

If $\mathcal{H} = \mathrm{End}(M)$ and $A \subseteq M$ is an $\mathcal{H}$-HDTOL set, we call A an HDTOL set of M.

3. CONNECTIONS BETWEEN L RATIONAL SETS AND HDTOL SETS

We first characterize the L rational subsets of free monoids.

PROPOSITION 3.1: *Each L rational subset of X^* is an HDTOL language.*

Proof: The proof is by L rational induction. First, $\emptyset$ and $\{w\}$, $w \in X^*$, are HDTOL languages. If A, B are HDTOL languages, so are $A \cup B$ and AB. Also, if A is an HDTOL language, so is $h(A)$ for any $h : X^* \to X^*$. Finally, consider the set

$$(g_1 + \ldots + g_p)^* \, h \, (h_1 + \ldots + h_s)^* \, (\omega).$$

Without loss of generality, we assume that the h_is are endomorphisms of X_1^*, h maps X_1^* into X^* and the g_js are endomorphisms of X^* where X_1 is an alphabet such that $X \cap X_1 = \emptyset$. Extend h_i, h and g_j to endomorphisms of $(X \cup X_1)^*$ by $h_i(x) = h(x) = x$, $g_j(x_1) = x_1$ for $x \in X$, $x_1 \in X_1$, $1 \leq i \leq s$, $1 \leq j \leq p$. Then

$$\begin{aligned}(g_1 + \ldots + g_p + h + h_1 + \ldots + h_s)^* \, (\omega) \\ = (h_1 + \ldots + h_s)^* \, (\omega) \cup (g_1 + \ldots + g_p)^* \, h \, (h_1 + \ldots + h_s)^* \, (\omega).\end{aligned}$$

Hence

$$\begin{aligned}h \, (g_1 + \ldots + g_p + h + h_1 + \ldots + h_s)^* \, (\omega) \\ = (g_1 + \ldots + g_p)^* \, h \, (h_1 + \ldots + h_s)^* \, (\omega). \quad \square\end{aligned}$$

Let Σ_∞ be an infinite alphabet and denote by $\mathcal{L}$(HDTOL) the set of HDTOL languages included in Σ_∞^*.

COROLLARY 3.2: $\bigcup_{X \subseteq \Sigma_\infty, X \, finite} \mathrm{LRat}\,(X^*) = \mathcal{L}\,(\mathrm{HDTOL})$.

COROLLARY 3.3:

$$\bigcup_{X \subseteq \Sigma_\infty, X \, finite} \mathrm{Rat}\,(X^*) \subset \bigcup_{X \subseteq \Sigma_\infty, X \, finite} \mathrm{LRat}\,(X^*).$$

Proof: Inclusion follows because each regular language is an HDTOL language (*see* Culik II [2]). Proper inclusion follows because there are DTOL languages which are not regular. $\square$

In general, L rational subsets of a monoid M are not HDTOL sets of M. Note that Proposition 3.1 only shows that an L rational subset of the free monoid X^* is an HDTOL set in a free monoid Y^* where Y is an alphabet, usually much larger than X.

Example 3.1: Consider the monoid $M = (\mathbf{N}, +, 0)$. Clearly $A \subseteq \mathbf{N}$ is an HDTOL set of M if and only if there exist $k \geq 0$ and $x, y_1, \ldots, y_k \in \mathbf{N}$

such that

$$A = \{xy_1^{i_1}y_2^{i_2}\dots y_k^{i_k} \mid i_j \geq 0 \text{ for } j = 1, \dots, k\}.$$

In M the DTOL sets and HDTOL sets coincide. We claim that the L rational set

$$B = \{1 + 2^i \mid i \geq 1\}$$

is not an HDTOL set of M. Suppose on the contrary that there exist $k \geq 1$, and $x, y_1, \dots, y_k \in \mathbf{N}$ such that

$$B = \{xy_1^{i_1}y_2^{i_2}\dots y_k^{i_k} \mid i_j \geq 0 \text{ for } j = 1, \dots, k\}$$

and

$$y_1, \dots, y_k > 1.$$

Then necessarily $x, y_1, \dots, y_k$ are odd. Hence, for large i, the binary representation of $(1 + 2^i)\,y_1 \in B$ contains more than two nonzero digits. This contradiction proves the claim.

The next theorem establishes the basic connection between L rational subsets of a monoid and DTOL languages.

THEOREM 3.4: *Suppose M is a finitely generated monoid and $A \in$ LRat(M). Then there exist a finite set X, a* DTOL *language $L \subseteq X^*$ and a morphism $h : X^* \to M$ such that $A = h(L)$.*

Proof: Let Y be a set with the same cardinality as some generating set of M and denote by g the canonical morphism $g : Y^* \to M$. Then, if $h \in \operatorname{End}(M)$ there exists $h' \in \operatorname{End}(Y^*)$ such that $gh' = hg$. We first claim that if $A \in \operatorname{LRat}(M)$ there exists a set $L \in \operatorname{LRat}(Y^*)$ such that $g(L) = A$. The proof is by L rational induction. The claim is trivial if $A = \emptyset$ or $A = \{m\}$ for $m \in M$. Next, if $A_1, A_2 \in \operatorname{LRat}(M)$ and $g(L_1) = A_1$, $g(L_2) = A_2$ where $L_1, L_2 \in \operatorname{LRat}(Y^*)$, then $g(L_1 \cup L_2) = A_1 \cup A_2$ and $g(L_1L_2) = A_1A_2$. Suppose then that $A \in \operatorname{LRat}(M)$, $h \in \operatorname{End}(M)$ and $A = g(L)$ where $L \in \operatorname{LRat}(Y^*)$. If $h' \in \operatorname{End}(Y^*)$ satisfies $gh' = hg$, we have

$$g(h'(L)) = hg(L) = h(A).$$

Finally, suppose $A \in \mathrm{LRat}(M)$, $h_1, \ldots, h_s \in \mathrm{End}(M)$ and $A = g(L)$ where $L \in \mathrm{LRat}(Y^*)$. Then there exist $h'_1, \ldots, h'_s \in \mathrm{End}(Y^*)$ such that $gh'_i = h_i g$ for $1 \leq i \leq s$. Therefore

$$g((h'_1 + \ldots + h'_s)^* (L)) = (h_1 + \ldots + h_s)^* (g(L)) = (h_1 + \ldots + h_s)^* (A).$$

This concludes the proof of the claim.

Suppose now that $A = g(L)$ where $L \in \mathrm{LRat}(Y^*)$. By Proposition 3.1 there exist an alphabet X, a DTOL language $L_1 \subseteq X^*$ and a morphism $h : X^* \to Y^*$ such that $h(L_1) = L$. Therefore $A = gh(L_1)$ which proves the theorem. □

Note that Theorem 3.4 is analogous to the following result concerning rational subsets of a monoid M (*see* Berstel [1]). If $A \subseteq M$ is rational there exist an alphabet X, a morphism $h : X^* \to M$ and a regular language L such that $h(L) = A$. Hence, the DTOL and HDTOL languages play the role of regular languages in the framework of L rationality.

4. L RATIONAL TRANSDUCTIONS

To generalize the notion of a rational transduction we first define L rational relations.

DEFINITION 4.1: *Let X and Y be finite alphabets. A subset A of $X^* \times Y^*$ is an L rational relation if there exist alphabets X_1 and Y_1 such that $X \subseteq X_1$, $Y \subseteq Y_1$ and $A \in \mathrm{LRat}(X_1^* \times Y_1^*)$.*

We first establish a counterpart of Nivat's theorem (*see* Berstel [1]) for L rational relations.

THEOREM 4.2: *Suppose X and Y are finite alphabets. The following conditions are equivalent:*

(i) *$A \subseteq X^* \times Y^*$ is an L rational relation.*

(ii) *There exist a finite alphabet Z, two morphisms $\phi : Z^* \to X^*$, $\psi : Z^* \to Y^*$ and a* DTOL *language $K \subseteq Z^*$ such that*

$$A = \{(\phi(k), \psi(k)) | k \in K\}.$$

(iii) *There exist a finite alphabet Z, two alphabetic morphisms $\alpha : Z^* \to X^*$, $\beta : Z^* \to Y^*$ and a* DTOL *language $K \subseteq Z^*$ such that*

$$A = \{(\alpha(k), \beta(k)) | k \in K\}.$$

(iv) *There exist a finite alphabet Z, two morphisms $\phi : Z^* \to X^*$, $\psi : Z^* \to Y^*$ and an* HDTOL *language $K \subseteq Z^*$ such that*

$$A = \{(\phi(k), \psi(k)) | k \in K\}.$$

Furthermore, if $X \cap Y = \emptyset$, condition (i) *is equivalent with the condition* (v) *There exist a finite alphabet Z and a* DTOL *language $K \subseteq Z^*$ such that $X \cup Y \subseteq Z$ and*

$$A = \{(\pi_X(k), \pi_Y(k)) | k \in K\}$$

where π_X and π_Y are the projections of Z^* onto X^* and Y^*, respectively.

Proof: Suppose first that (i) holds. Then there exist finite alphabets X_1 and Y_1 such that $X \subseteq X_1$, $Y \subseteq Y_1$, and $A \in \mathrm{LRat}(X_1^* \times Y_1^*)$. Theorem 3.4 implies that there exist a finite alphabet Z, a DTOL language $K \subseteq Z^*$ and a morphism $h : Z^* \to X_1^* \times Y_1^*$ such that $A = h(K)$. Because $A \subseteq X^* \times Y^*$ we may assume that h is a morphism from Z^* into $X^* \times Y^*$. Define the morphisms $\phi : Z^* \to X^*$ and $\psi : Z^* \to Y^*$ by

$$h(z) = (\phi(z), \psi(z))$$

for $z \in Z$. Then

$$A = \{(\phi(k), \psi(k)) | k \in K\}.$$

Hence (ii) holds true.

Next, suppose that (ii) holds. Fix a sufficiently large integer s and choose for each $z \in Z$ new letters $z_1, \ldots, z_s$. Denote $Z_1 = Z \cup \{z_1, \ldots, z_s | z \in Z\}$ and define the morphism $g : Z^* \to Z_1^*$ by $g(z) = zz_1 \ldots z_s$ for $z \in Z$. Furthermore, if $h : Z^* \to Z^*$ is a morphism, define the morphism $\overline{h} : Z_1^* \to Z_1^*$ by $\overline{h}(z) = gh(z)$ if $z \in Z$ and $\overline{h}(z) = \varepsilon$ if $z \in Z_1 - Z$. Suppose $K = (h_1 + \ldots + h_t)^*(w)$ where $h_1, \ldots, h_t : Z^* \to Z^*$ are morphisms and $w \in Z^*$. Denote

$$\overline{K} = (\overline{h}_1 + \ldots + \overline{h}_t)^*(g(w)).$$

Then

$$\overline{K} = g(h_1 + \ldots + h_t)^*(w) = g(K).$$

Because s is sufficiently large there exist alphabetic morphisms $\alpha : Z_1^* \to X^*$ and $\beta : Z_1^* \to Y^*$ such that

$$\alpha(z_1 \ldots z_s) = \phi(z), \quad \beta(z_1 \ldots z_s) = \psi(z), \quad \alpha(z) = \beta(z) = \varepsilon$$

for all $z \in Z$. Then

$$\{(\alpha(k), \beta(k)) | k \in \overline{K}\} = \{(\alpha g(k), \beta g(k)) | k \in K\}$$
$$= \{(\phi(k), \psi(k)) | k \in K\} = A.$$

Hence (iii) holds true.

The equivalence of (ii) and (iv) and is clear.

Next, assume that (ii) holds and $X \cap Y = \emptyset$. We may assume that $Z \cap (X \cup Y) = \emptyset$. Define the morphism $g : Z^* \to (Z \cup X \cup Y)^*$ by $g(z) = z\phi(z)\psi(z)$ for $z \in Z$. As in a previous paragraph it is seen that $g(K)$ is a DTOL language. Hence

$$\{(\pi_X(k), \pi_Y(k)) | k \in g(K)\} = \{(\phi(k), \psi(k)) | k \in K\} = A.$$

Hence (v) holds true.

To conclude the proof it suffices to show that (ii) implies (i). Again, we may assume that $Z \cap (X \cup Y) = \emptyset$. For the proof denote $Z_1 = Z \cup X \cup Y$. If $h : Z^* \to Z^*$ is a morphism, extend h to a morphism $h : Z_1^* \to Z_1^*$ by $h(z) = \varepsilon$ if $z \in Z_1 - Z$ and define the morphism $\overline{h} : Z_1^* \times Y^* \to Z_1^* \times Y^*$ by $\overline{h}(z, y) = (h(z), 1)$ for $z \in Z_1^*, y \in Y^*$. If $K = (h_1 + \ldots + h_s)^*(w)$, denote

$$\overline{K} = (\overline{h}_1 + \ldots + \overline{h}_s)^*(w, 1).$$

Then $\overline{K} = K \times \{1\}$ and $\overline{K}$ is an L rational subset of $Z_1^* \times Y^*$. Now, extend ϕ and ψ to morphisms from Z_1^* into X^* and Y^*, respectively, by $\phi(z) = \psi(z) = \varepsilon$ if $z \in Z_1 - Z$, and define the morphism $g : Z_1^* \times Y^* \to Z_1^* \times Y^*$ by $g(z, y) = (\phi(z), \psi(z))$ for $z \in Z_1^*, y \in Y^*$.

Then

$$g(\overline{K}) = \{(\phi(k), \psi(k)) | k \in K\} = A$$

and hence A is an L rational subset of $Z_1^* \times Y^*$. □

Now we are ready to discuss L rational transductions. In general, a transduction τ from X^* into Y^* is a mapping from X^* into the set of subsets of Y^*. The graph of τ is the relation R defined by

$$R = \{(f, g) \in X^* \times Y^* | g \in \tau(f)\}.$$

DEFINITION 4.3: *A transduction $\tau : X^* \to Y^*$ is L rational if its graph R is an L rational relation.*

The following theorem is a reformulation of Theorem 4.2.

THEOREM 4.4: *Suppose X and Y are finite alphabets. The following conditions are equivalent:*

(i) *$\tau : X^* \to Y^*$ is an L rational transduction.*

(ii) *There exist a finite alphabet Z, two morphisms $\phi : Z^* \to X^*$, $\psi : Z^* \to Y^*$ and a* DTOL *language $K \subseteq Z^*$ such that*

$$\tau(f) = \psi(\phi^{-1}(f) \cap K)$$

for $f \in X^*$.

(iii) *There exist a finite alphabet Z, two alphabetic morphisms $\alpha : Z^* \to X^*$, $\beta : Z^* \to Y^*$ and a* DTOL *language $K \subseteq Z^*$ such that*

$$\tau(f) = \beta(\alpha^{-1}(f) \cap K)$$

for $f \in X^*$.

(iv) *There exist a finite alphabet Z, two morphisms $\phi : Z^* \to X^*$, $\psi : Z^* \to Y^*$ and an* HDTOL *language $K \subseteq Z^*$ such that*

$$\tau(f) = \psi(\phi^{-1}(f) \cap K)$$

for $f \in X^*$.

Furthermore, if $X \cap Y = \emptyset$*, condition* (i) *is equivalent with the condition* (v) *There exist a finite alphabet* Z *and a* DTOL *language* $K \subseteq Z^*$ *such that* $X \cup Y \subseteq Z$ *and*

$$\tau(f) = \pi_Y(\pi_X^{-1}(f) \cap K)$$

for $f \in X^*$*, where* π_X *and* π_Y *are the projections of* Z^* *onto* X^* *and* Y^**, respectively.*

COROLLARY 4.5: *Rational transductions are L rational transductions.*

Proof: Suppose $\tau : X^* \to Y^*$ is a rational transduction. By Nivat's theorem there exist an alphabet Z, two morphisms $\phi : Z^* \to X^*$, $\psi : Z^* \to Y^*$ and a regular language $K \subseteq Z^*$ such that

$$\tau(f) = \psi(\phi^{-1}(f) \cap K)$$

for $f \in X^*$. The claim follows by Theorem 4.4 (iv) because K is an HDTOL language. □

COROLLARY 4.6: *If* $\tau : X^* \to Y^*$ *is an L rational transduction and* $A \subseteq X^*$ *is a regular language, then* $\tau(A)$ *is an* EDTOL *language.*

Proof: By Theorem 4.4, we have $\tau(A) = \psi(\phi^{-1}(A) \cap K)$ where K is a DTOL language and ϕ and ψ are morphisms. Because K is an EDTOL language the claim follows by the closure of EDTOL languages with respect to morphic image and intersection with a regular language (*see* Rozenberg and Salomaa [4]). □

Next we give examples of L rational transductions.

Example 4.1: Because $L_1 = \{a^n c b^{2^n} | n \geq 0\}$ is a DOL language, the mapping $\tau_1 : a^* \to a^*$ defined by $\tau_1(a^n) = a^{2^n}$ for $n \geq 0$, is an L rational transduction.

Example 4.2: Suppose $(\omega_n)_{n \geq 0}$ is a DOL sequence over the alphabet X. Choose two new letters a, $c \notin X$. Then $L_2 = \{a^n c \omega_n | n \geq 0\}$ is a DOL language. Therefore the transduction τ_2 defined by $\tau_2(a^n) = a^{|\omega_n|}$ for $n \geq 0$, is an L rational transduction. (Here $|w|$ is the length of the word w). It follows that if $P(x) \in \mathbf{N}[x]$ is a polynomial then the mapping

$$a^n \to a^{P(n)}, \quad n \geq 1,$$

is an L rational transduction.

Example 4.3: Suppose $(u_n)_{n \geq 0}$ and $(v_n)_{n \geq 0}$ are DOL sequences where $u_n, v_n \in X^*$ for $n \geq 0$. Suppose that $\{u_n | n \geq 0\}$ is infinite. Define the transduction τ_3 by

$$\tau_3(u_n) = v_n \text{ for } n \geq 0.$$

Hence, τ_3 translates the sequence (u_n) to the sequence (v_n). We claim that τ_3 is L rational. Let Y be a new alphabet isomorphic to X such that $X \cap Y = \emptyset$ and $(\overline{v}_n)$ be the isomorphic copy of (v_n). Then the set $\{u_n \overline{v}_n | n \geq 0\}$ is a DOL language. Hence the claim follows by Theorem 4.2.

Example 4.4: Suppose X is a finite alphabet. Define the mapping $\tau_4 : X^* \to X^*$ by $\tau_4(w) = \tilde{w}$, $w \in X^*$, where $\tilde{w}$ is the reversal of w. We claim that τ_4 is an L rational transduction. For the proof, suppose Y is an alphabet isomorphic to X such that $X \cap Y = \emptyset$ and denote the isomorphic copy of $w \in X^*$ by w_Y. Choose a letter $c \notin X \cup Y$. Then it is easy to see that $L_4 = \{wc\tilde{w}_Y | w \in X^*\}$ is a DTOL language. This implies the claim.

Example 4.5: Consider the alphabet $X = \{a, b\}$ and define the transduction τ_5 by $\tau_5(a^m b a^n) = a^{mn}$ for $m, n \geq 1$. We show that τ_5 is L rational. For the proof, choose three new letters ω, c, $d \notin X$ and define the morphisms h_1, h_2, h_3 by

$$h_1(\omega) = a\omega cd,$$
$$h_2(\omega) = ba,$$
$$h_3(b) = ba, \quad h_3(c) = cd.$$

(In all unlisted cases h_i acts as the identity.) Then

$$L_5 = h_3^* h_2 h_1^* (a\omega cd) = h_3^* h_2 (\{a^m \omega (cd)^m | m \geq 1\})$$
$$= h_3^* (\{a^m ba (cd)^m | m \geq 1\}) = \{a^m ba^n (cd^n)^m | m, n \geq 1\}.$$

By Proposition 3.1, L_5 is an HDTOL language. Therefore the claim follows by Theorem 4.2.

We conclude this section by showing that L rational transductions are not closed under composition.

First, one can easily construct two DTOL languages L_1 and L_2 such that if $K_1 = L_1 \cap \{a, b\}^*$ and $K_2 = L_2 \cap \{a, b\}^*$ then

$$K_1 = \{(a^m b)^n | m, n \geq 1\},$$
$$K_2 = \{a^m b (a^n b)^{m-1} | m, n \geq 1\}$$

and

$$L_1 \cap L_2 = K_1 \cap K_2 = \{(a^m b)^m | m \geq 1\}.$$

Now, define τ_1 and τ_2 by

$$\tau_1 (f) = \{f\} \cap L_1,$$
$$\tau_2 (f) = \{f\} \cap L_2.$$

By Theorem 4.4, τ_1 and τ_2 are L rational transductions. Furthermore,

$$(\tau_1 \circ \tau_2) (f) = \{f\} \cap L_1 \cap L_2 = \{f\} \cap K_1 \cap K_2.$$

Hence $(\tau_1 \circ \tau_2) (\{a, b\}^*) = K_1 \cap K_2$. Because $K_1 \cap K_2$ is not an ETOL language (*see* Ehrenfeucht and Rozenberg [3]), if follows by Corollary 4.6 that $\tau_1 \circ \tau_2$ is not L rational.

5. L RATIONAL STAR HEIGHT

In this section we define the notion of star height of an L rational set and discuss the infinity of the star height hierarchy in various monoids.

Suppose M is a monoid and $\mathcal{H} \subseteq \text{End}(M)$. Define inductively the sets $\text{L}_{\mathcal{H}}\text{Rat}_i (M)$ for $i \geq 0$ as follows. First, $A \in \text{L}_{\mathcal{H}}\text{Rat}_0 (M)$ if and only if A is a finite subset of M. For $i > 0$, $A \in \text{L}_{\mathcal{H}}\text{Rat}_i (M)$ if and only if A is a finite union of sets of the form $B_1 B_2 \ldots B_n$ where either B_j is a singleton or $B_j = g_1 \ldots g_t (h_1 + \ldots + h_s)^* (C_j)$ for some $g_1, \ldots, g_t$, $h_1, \ldots, h_s \in \mathcal{H}$ and $C_j \in \text{L}_{\mathcal{H}}\text{Rat}_{i-1} (M)$, $1 \leq j \leq n$. It is easy to see that $\text{L}_{\mathcal{H}}\text{Rat}_i (M) \subseteq \text{L}_{\mathcal{H}}\text{Rat}_{i+1} (M)$ for $i \geq 0$. Denote

$$\mathcal{R} = \bigcup_{i \geq 0} \text{L}_{\mathcal{H}}\text{Rat}_i (M).$$

Clearly, $\mathcal{R} \subseteq \mathrm{L}_{\mathcal{H}}\mathrm{Rat}\,(M)$. Furthermore, $\emptyset \in \mathcal{R}$, $\{m\} \in \mathcal{R}$ for each $m \in M$ and $\mathcal{R}$ is closed under union, product, $\mathcal{H}$-morphic image and $\mathcal{H}$-morphic star. Hence

$$\mathrm{L}_{\mathcal{H}}\mathrm{Rat}\,(M) = \bigcup_{i \geq 0} \mathrm{L}_{\mathcal{H}}\mathrm{Rat}_i\,(M).$$

By definition, the *star height* of a set $A \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}\,(M)$ is the smallest i such that $A \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}_i\,(M)$.

If each L rational subset of M is an HDTOL set, then the star height hierarchy collapses and

$$\mathrm{LRat}\,(M) = \bigcup_{i=0}^{1} \mathrm{LRat}_i\,(M).$$

(Here $\mathcal{H} = \mathrm{End}\,(M)$.) Hence, an infinite star height hierarchy implies that there is a large gap between HDTOL sets and L rational sets. Below we give nontrivial examples of an infinite star height hierarchy and a finite star height hierarchy.

To obtain an example of an infinite hierarchy, consider the monoid $M = (\mathbf{N}, +, 0)$ of nonnegative integers and define $\mathcal{H} = \{h\}$ by $h\,(x) = 2x$ for $x \in \mathbf{N}$. We need the following technical notion. A set $A \subseteq \mathbf{N}$ has *width* s, $s \geq 1$, if

$$A \subseteq \{2^{i_1} + 2^{i_2} + \ldots + 2^{i_{s+1}} | i_1 < i_2 < \ldots < i_{s+1}\}$$

and for each $t \geq 1$ the set

$$A \cap \{2^{i_1} + 2^{i_2} + \ldots + 2^{i_{s+1}} \,|\, i_{j+1} - i_j \geq t \text{ for all } 1 \leq j \leq s\}$$

is nonempty.

LEMMA 5.1: *If* $A \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}\,(\mathbf{N})$ *has width greater than or equal to* s, $s \geq 1$, *then the star height of* A *is at least* s.

Proof: If $A \subseteq \mathbf{N}$ has width at least 1, the set A is infinite and hence its star height is at least one. Suppose inductively that the lemma is true for s, $s \geq 1$, and consider a set $A \subseteq \mathbf{N}$ of width $s+t$, $t \geq 1$. We have to prove that $A \notin \mathrm{L}_{\mathcal{H}}\mathrm{Rat}_s\,(\mathbf{N})$. Suppose on the contrary that $A \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}_s\,(\mathbf{N})$. Then A is a finite union of sets of the form $B_1 + B_2^* + \ldots + B_n^*$ where B_1 is a singleton and $B_2, \ldots, B_n \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}_{s-1}\,(\mathbf{N})$ are nonempty sets none of which equals

$\{0\}$. Here we denote $B^* = h^*(B)$. (Notice that $(h + \ldots + h)^*(B) = h^*(B)$ for any $B \subseteq \mathbf{N}$.)

First, suppose that in at least one of the terms of the union $n \geq 3$. Choose $b_i \in B_i$ for $1 \leq i \leq n$ and consider the binary expansions of the numbers b_i. The total number of nonzero digits in the expansions equals $s + t + 1$. Next, choose u and v such that the smallest nonzero term in the binary expansion of b_2 is 2^u and of b_3 is 2^v, respectively. Then

$$b_1 + b_2 \cdot 2^v + b_3 \cdot 2^u + b_4 + \ldots + b_n \in B_1 + B_2^* + \ldots + B_n^* \subseteq A.$$

However, the number of nonzero digits in the binary expansion of $b_1 + b_2 \cdot 2^v + b_3 \cdot 2^u + b_4 + \ldots + b_n$ is less than $s + t + 1$. This contradiction shows that A is a finite union of singletons and sets of the form $B_1 + B_2^*$ where B_1 is a singleton and $B_2 \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}_{s-1}(\mathbf{N})$ is a nonempty set different from $\{0\}$.

Next, consider a set $B_1 + B_2^*$. Suppose first that $B_1 + B_2^*$ has width $s + t$. Then B_2 has width $s + t - 1$ or $s + t$. By the inductive hypothesis, the star height of B_2 is at least s. This is not possible because $B_2 \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}_{s-1}(\mathbf{N})$.

Hence A is a finite union of singletons and sets of the form $B_1 + B_2^*$ none of which has width $s + t$. Therefore the width of A cannot equal $s + t$. This contradiction proves the lemma. □

THEOREM 5.2: *Consider the monoid* $M = (\mathbf{N}, +, 0)$ *and define* $\mathcal{H}$ *as above. Denote*

$$A_s = \{2^{i_1} + 2^{i_2} + \ldots + 2^{i_s} | 0 \leq i_1 < i_2 < \ldots < i_s\}$$

for $s \geq 1$. *Then the star height of* A_s *equals* s.

Proof: Clearly, the set $1 + h(A_s)$ has width s. Hence, by Lemma 5.1, the star height of A_s is at least s. The fact that the star height of A_s is at most s follows inductively by the equations

$$A_1 = h^*(1)$$
$$A_{s+1} = h^*(1 + h(A_s)).$$

Indeed, A_1 has star height one. Furthermore, if A_s has star height at most s, so has $1 + h(A_s)$. Hence, A_{s+1} has star height at most $s + 1$. □

To conclude this section we give a nontrivial example of a finite star height hierarchy.

Let X_∞ be an infinite alphabet and consider the free monoid $M = X_\infty^*$. Let $\mathcal{H}$ be the set of endomorphisms of M such that $h(x) = x$ for almost all $x \in X_\infty$. If $A \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}\,(M)$ then it is easy to see by L rational induction that there exists a finite alphabet X such that $A \in \mathrm{LRat}\,(X^*)$. Hence, by Proposition 3.1, A is an HDTOL language. It follows that the star height of an infinite set $A \in \mathrm{L}_{\mathcal{H}}\mathrm{Rat}\,(M)$ with respect to $\mathcal{H}$ is one. Hence, in this case the star height hierarchy collapses.

REFERENCES

1. J. Berstel, *Transductions and Context-Free Languages*, Teubner, Stuttgart, 1979.
2. K. Culik II, New techniques for proving the decidability of equivalence problems, in: T. Lepistö and A. Salomaa (eds.), *Automata, Languages and Programming*, Springer, Berlin, 1988.
3. A. Ehrenfeucht and G. Rozenberg, On proving that certain languages are not ETOL, *Acta Inform.*, 1976, *6*, pp. 407-415.
4. G. Rozenberg and A. Salomaa, *The Mathematical Theory of L Systems*, Academic Press, New York, 1980.