JOURNAL OF COMPUTER AND SYSTEM SCIENCES 11, 375–381 (1975)

On Some Questions of Rationality and Decidability

EDUARDO D. SONTAG*

Center for Mathematical System Theory, Department of Mathematics, University of Florida, Gainesville, Florida 32611

Received August 26, 1974

Some results are given in the theory of rational power series over a broad class of semirings. In particular, it is shown that for unambiguous sets the notion of rationality is independent of the semiring over which representations are defined. The undecidability of the rationality of probabilistic word functions is also established.

INTRODUCTION

In this note we begin by resolving (in the negative) a question posed by Paz [6, Open Problem 1, p. 65] concerning the existence of an effective procedure for determining whether a recursively specified "probabilistic input–output relation" is of finite rank. To show that such an algorithm does not exist, it is enough to prove the undecidability of the corresponding problem for "probabilistic word functions" (take an input alphabet of one letter).

The undecidability of rationality for general word functions is used in the proof of the above. This was proved by Paz [7, Corollary E3]. This problem, however, can be posed in much more generality: One might ask about the rationality of power series with coefficients in more general semirings. We give in Part 2 a completely new proof of the undecidability result of Paz. This proof extends readily to the more general situation. For this we note that over a large class of semirings R (namely, those embeddable in commutative rings), unambiguous R-rational power series are recognizable languages. Particular cases of this latter result were already known. For example, the one-letter case with coefficients in a field of characteristic zero follows from results on supports as in [2, Proposition I. 4.1.1]. The case of arbitrary alphabets and positive semirings is treated in [1, Corollary VIII. 4.3]. For the real numbers the result depends on the theory of isolated cutpoints as in [6, Theorem III. B. 2.3].

The proof of the recognizability of R-rational series rests upon some new facts

* This research was supported in part by U.S. Army Research Grant DA-ARO-D-31-124-72-G114 through the Center for Mathematical System Theory, University of Florida, Gainesville Fla. 32611.

about Hankel matrices, extending the results of [3]. A short example shows that in the case R is not commutative the problem is not even well posed.

The author wishes to thank Dr. P. Turakainen, who suggested looking at the first of the problems discussed here.

Preliminaries

X will always denote a finite alphabet, while X^* is the free monoid generated by X. The empty word is denoted by λ; X^+ is XX^*. We write $|w|$ for the length of w in X^*. A semiring will always have an identity $1 \neq 0$, and (except otherwise stated) will be *commutative*. Given a semiring R and a set I, the set of all functions $I \to R$ is R^I, which can be also thought of as "sequences" of elements of R or as "R-subsets" of I. The ith coordinate (i in I) of an f in R^I will be written $f(i)$. By R^n we denote the free R-module on n generators, and will not distinguish between its endomorphisms and n by n matrices (expressed with respect to a canonical basis). In general, an R^I is naturally an R-module under coordinatewise operations (see [1]).

The main objects of study are the R-subsets of X^* for an arbitrary semiring R; these objects are also called *power series* [2] and *word functions* [6]. We say an R-subset f is *R-recognizable*—equivalently, *rational*, for our monoids are free—iff there is an integer n and matrices g in $R^{n\times 1}$, $F(x)$ in $R^{n\times n}$ for each x in X, and h in $R^{1\times n}$, with $f(w) = hF(w)g$ for each $w = x_{i_1} \cdots x_{i_s}$, where $F(w)$ is the product $F(x_{i_1}) \cdots F(x_{i_s})$. If f is an R-subset such that $f(w)$ is always 0 or 1, it is *unambiguous*; we identify it with the subset of X^* (i.e., language) of which it is the characteristic function. When R is the 2-element Boolean semiring, R-recognizable R-subsets are called simply *recognizable*, and they are of course the languages accepted by finite automata. This approach began with Schützenberger [9].

Given the R-subset f, we denote by $H(f)$ the (generalized) Hankel matrix of f (see [2]), i.e., the infinite matrix with rows and columns indexed by X^* and $f(uv)$ in position (u, v). The uth column is denoted by $H_u = H(f)_u$; it can also be seen as an R-subset of X^* and its vth coordinate $H_{u,v}$ is clearly $f(vu)$. The R-submodule of R^{X^*} generated by the H_u, i.e., the set of all their finite R-linear combinations, is denoted by $\bar{H}(f)$.

If $R = \mathbf{R}$, the real numbers, f is a *probabilistic word function* when (i) $f(X^*)$ is included in $[0, 1]$, (ii) $f(\lambda) = 1$ and (iii) $\sum_{x\in X} f(vx) = f(v)$ for each $v \in X^*$ (see [6, p. 119]). An f as above is also called a (finite) *stochastic process*. When such an f is also $\mathbf{R}$-recognizable it is of *finite rank*.

1. Probabilistic Word Functions

We shall use the following result: *There is no algorithm which decides* $\mathbf{R}$*-recognizability of a recursively specified* $\mathbf{R}$*-subset taking values in* $[0, 1]$. This follows from

[7, Corollary E. 3] or more generally from Part 2 below. Without loss of generality we shall in this section take X to be a two-letter alphabet $\{x_1, x_2\}$. We prove Theorem (1.2) by showing that a decision procedure for probabilistic word functions would imply one for more arbitrary $\mathbf{R}$-subsets; the key is the following fact, also interesting in itself:

PROPOSITION 1.1. *There is an algorithm which, when given a recursive* $\mathbf{R}$*-subset* f *with values in* $[0, 1]$, *constructs a pair of recursive probabilistic word functions* p_i *with the property that* f *is* $\mathbf{R}$*-recognizable iff both* p_i *are.*

Proof. We proceed in several steps:

Step 1. Define f_1, f_2 as follows. $f_i(\lambda) = 0$; for each $w \in X^+$ let $f_i(w)$ be $f(w)$ if w is in X^*x_i and 0 otherwise. Observe that for each w in X^+ either $H(f_i)_w = H(f)_w$ or $H(f_i)_w = 0$ and conversely, at least one of $H(f_1)_w$ or $H(f_2)_w$ is equal to $H(f)_w$. So $\bar{H}(f)$ is finite-dimensional iff both $\bar{H}(f_i)$ are, and by Proposition 2.1 below, or by well-known results for the case of fields, f is $\mathbf{R}$-recognizable iff both f_i are. We shall work with f_1 and construct p_1; the construction of p_2 will be similar. So now assume that $f(w) = 0$ for all w ending in x_2.

Step 2. Observe that f is $\mathbf{R}$-recognizable iff $X^*x_1 + f$ is (where X^*x_1 stands for the corresponding characteristic function), because $\mathbf{R}$-recognizable subsets form a subring. So we will suppose $f(X^*x_1) \subseteq [1, 2]$.

Step 3. Given any $a > 0$, if we define the $\mathbf{R}$-subset f_a by $f_a(w) := a^{|w|}f(w)$, then f_a is $\mathbf{R}$-recognizable iff f is. (Given g, h, $F(x_1), F(x_2)$ representing f, by replacing $aF(x_i)$ for $F(x_i)$ we have a representation for f_a.) Therefore, taking in particular $a := \frac{1}{3}$, we may assume without loss of generality that

$$3^{-|v|} \leqslant f(v) \leqslant 2.3^{-|v|} \qquad \text{for all} \quad v \text{ in } X^*x_1. \tag{$*$}$$

Step 4. Definition of p. Let $p(\lambda) := 1$ and assume by induction on k that $p(v)$ is already defined, satisfying $3^{-|v|} \leqslant p(v) \leqslant 1$ for all v with $|v| \leqslant k$. Given w such that $|w| = k + 1$, it is either in X^*x_1 or X^*x_2. In the first case, let $p(w) := f(w)$; $p(w)$ again satisfies the inductive hypothesis because of $(*)$. If, instead, $w = vx_2$ with $|v| = k$, let $p(w) := p(v) - f(vx_1)$; by hypothesis and because of $(*)$,

$$p(w) \geqslant 3^{-|v|} - 2.3^{-|vx_1|} = 3^{-|v|} - 2.3^{-|v|-1} = 3^{-|v|}(1 - \tfrac{2}{3}) = 3^{-|v|-1} = 3^{-|w|}.$$

Step 5. Proof that p is probabilistic. By construction p satisfies

$$\begin{aligned} &p(vx_2) = p(v) - f(vx_1) = p(v) - p(vx_1) && \text{for all } v \text{ in } X^*, \\ &p(\lambda) = 1, \qquad p(v) \text{ in } [0, 1] && \text{for all } v. \end{aligned} \tag{$**$}$$

Step 6. f is **R**-recognizable iff p is. Observe that (**) says that for all u, v in X^* we have $p(uvx_2) = p(uv) - f(uvx_1)$, i.e.,

$$H(p)_{vx_2} = H(p)_v - H(f)_{vx_1} \qquad \text{for all } v. \qquad (***)$$

Assume now by induction on $|v|$ that $H(p)_v$ is in the submodule generated by $H(p)_\lambda$ and $\bar{H}(f)$. If $w = vx_1$, by definition $p(uvx_1) = f(uvx_1)$ for all u, so $H(p)_{vx_1} = H(f)_{vx_1}$. If $w = vx_2$, apply the induction hypothesis for v and (***). We have then proved that $\bar{H}(p) \subseteq H(p)_\lambda + \bar{H}(f)$.

Conversely, $H(f)_{vx_2} = 0$ and $H(f)_{vx_1} = H(p)_{vx_1}$ for all v, so $\bar{H}(f) \subseteq \bar{H}(p)$. Thus $\bar{H}(p)$ is exactly $\bar{H}(f)$ plus a one-dimensional subspace, and Proposition 2.1 applies again.[1] ▮

THEOREM 1.2. *There is no effective procedure for determining whether recursive probabilistic word functions are of finite rank.*

2. UNAMBIGUOUS R-RECOGNIZABLE SETS

The following proposition generalizes a result well known for fields and certain integral domains (in a much stronger form, see, e.g., [2, Section I. 2.10]). It has been proved before in the simpler one-letter case for arbitrary commutative rings [8, p. 34]. The sufficiency part of the proof below is similar to the latter. As a side remark, note that the result does not remain valid for noncommutative rings (see [10, Part C]).

PROPOSITION 2.1. *Assume R is a commutative ring and f is an R-subset. Then f is R-recognizable if and only if $\bar{H}(f)$ is a finitely generated R-module.*

Proof. ["Only if"]. Assume $f(w) = hF(w)g$ for all w in X^*, and consider the R-subalgebra of $R^{n\times n}$ generated by all matrices $F(x)$ with x in X. As the alphabet is finite, this algebra is also finitely generated as an R-module (see the Appendix). So there is some integer k such that $\{F(u), |u| < k\}$ generates it as a module. In particular,

[1] *Notes added in proof* (October 1975). (a) Using a variant of Lemma 1.1, one may also reduce to Theorem 2.5 (and hence prove unsolvable) the problem of deciding if a given function is the growth function of some Lindenmayer (DOL) system; (b) M. Fliess has pointed out to the author that a recent paper of S. Rao Kosaraju (*Information and Control* **26**, p. 194) approaches Theorem 1.2 independently of Theorem 2.5. Lemma 1.1 shows that both decision problems are in fact equivalent.

for any w in X^*, there are scalars r_u in R with $F(w) = \sum_{|u|<k} r_u F(u)$. So for each v in X^* (using R is commutative):

$$H_{w,v} = f(vw) = hF(vw)g = hF(v)F(w)g = hF(v)\left(\sum_{|u|<k} r_u F(u)\right)g$$
$$= \sum_{|u|<k} r_u H_{u,v} .$$

Hence $H_w = \sum_{|u|<k} r_u H_u$. Therefore the finite set $\{H_u , |u| < k\}$ generates $\bar{H}(f)$.

["If"]. In general, given any matrix of the type $H(f)$, we can define for each x in X an R-endomorphism F_x of $\bar{H}(f)$ which extends linearly the map that sends each H_w into H_{xw}. We only need to see that it is well defined ($\bar{H}(f)$ is not *freely* generated by the columns), all the other properties being obvious. So assume there is a relation $\sum r_u H_u = 0$. We claim $\sum r_u H_{xu} = 0$ for all x in X. Indeed, for any v in X^*, the vth coordinate $(\sum r_u H_{xu})_v = \sum r_u f(v(xu)) = \sum r_u f((vx)u) = (\sum r_u H_u)_{vx} = 0$. Denote $\bar{g} := H_\lambda$. Let $\bar{h}\colon \bar{H}(f) \to R$ be the projection on the first component $\sum r_u H_u \mapsto \sum r_u f(u)$. Then for any $w = x_1 \cdots x_n$,

$$\bar{h} \circ F_{x_1} \circ \cdots \circ F_{x_n}(\bar{g}) = \bar{h}(H_{x_1 \cdots x_n}) = f(w).$$

Now suppose that $\bar{H}(f)$ is finitely generated. Let $p\colon R^n \to \bar{H}(f)$ (surjective) be a free presentation. There exist matrices h, g, and $F(x)$ for each x such that $h = \bar{h} \circ p$, $p(g) = \bar{g}$, and $p \circ F(x) = F_x \circ p$. Then for each $w = x_1 \cdots x_n$, $f(w) = hF(x_1) \cdots F(x_n)g$. So f is R-recognizable. ∎

The following, although quite trivial, is crucial.

LEMMA 2.2. *Let R be an arbitrary semiring and S a finite subset of R. Let I be any set. Assume that $w_1 ,..., w_n$ are elements of $S^I \subseteq R^I$ and call M the R-submodule they generate. Then $M \cap S^I$ is also a finite set.*

Proof. If the coordinates can only assume finitely many values, the possible number of vectors $(w_{1,i} ,..., w_{n,i})$ is also finite. So there is a finite subset J of I representing them, i.e., such that for each i in I, there is a $j = j(i)$ in J with $(w_{1,i} ,..., w_{n,i}) = (w_{1,j} ,..., w_{n,j})$. Now, given arbitrary u and v in M, $u = \sum r_k w_k$, $v = \sum s_k w_k$, assume $u_j = v_j$ for all j in J. Take any i in I. Choosing $j = j(i)$ as before, $u_i = \sum r_k w_{k,i} = \sum r_k w_{k,j} = u_j = v_j = v_i$. So coinciding on the indexes in J is enough for equality. The lemma then follows immediately from the observation that S^J is finite. ∎

We then have

THEOREM 2.3. *Let R be a semiring which can be embedded in a commutative ring and f an unambiguous R-subset. Then f is R-recognizable iff it is recognizable.*

Proof. Sufficiency is well known (see, e.g., [1 Proposition VI. 7.14]). Assume now that f is R-recognizable. Without loss of generality, suppose that R is a ring. By Proposition 2.1, $\bar{H}(f)$ is finitely generated, and the generators can be chosen out of the columns H_u, which have all coordinates either 0 or 1. Applying Lemma 2.2, the set $\{H_u, u \in X^*\}$ is finite. But as observed by Fliess, this means that f is recognizable, because the classes of the left congruence in X^* given by "$u \sim v$ iff $f(wu) = f(wv)$ for all w in X^*" are represented by the different columns H_u. ∎

Remark 2.4. It is worth observing that to drop the commutativity assumption in Theorem 2.3 changes the situation completely. In fact, assume that we are given ***any*** subset f of X^*, recognizable or not. Then there exists a ring R (constructed using f) such that f is R-recognizable as an (unambiguous) R-subset.

One way of obtaining R is as follows. Consider the free $\mathbf{Z}$-algebra (i.e., the set of noncommutative polynomials) $\mathbf{Z}\langle X'\rangle$, where $X' := X \cup \{y\}$ for some $y \notin X$. Let I be the ideal generated by both all the $(ywy - 1)$ for which $f(w) = 1$ and all the ywy for which $f(w) = 0$. Let $R := (\mathbf{Z}\langle X'\rangle)/I$; this is a ring with $1 \neq 0$. Denote by $\bar{r}$ the image in R of $r \in \mathbf{Z}\langle X'\rangle$. Define for each i, $F(x_i) := \bar{x}_i \in R^{1\times 1} = R$, $g := h := \bar{y} \in R$. Then for every w in X^*, $hF(w)g = \overline{ywy}$, which is 0 or 1 according to f. ∎

THEOREM 2.5. *Fix a semiring R as in Theorem 2.3. The problem of deciding whether a recursive R-subset is R-recognizable is unsolvable. Moreover, the same conclusion holds with respect to unambiguous ones.*

Proof. If solvable, one would have a decision procedure for recognizability of recursive sets, which gives a contradiction (take for example the language generated by a context-free grammar and apply [4, p. 230]). ∎

APPENDIX

The following result from commutative algebra is used in the proof of Proposition 2.1. It is clearly valid for more general algebras than matrix rings.

PROPOSITION. *If R is a commutative ring and B is the R-subalgebra of $R^{n\times n}$ generated by the matrices $A_1, \ldots, A_m$, then B is finitely generated as an R-module.*

Proof. Let S be the smallest subring of R containing the identity and all the entries of the matrices A_i. Being a finitely generated $\mathbf{Z}$-algebra, it is a Noetherian [5, p. 145] ring. In particular, $S^{n\times n}$ is a Noetherian S-module. Now observe that, by definition of S, all A_i are in $S^{n\times n}$. So let B_S be the S-algebra generated by all A_i.

B_S is also an S-submodule of $S^{n\times n}$, generated as such by all products $A_{i_1} \cdots A_{i_s}$ with $1 \leqslant i_j \leqslant m$ for all j, and with s arbitrary. But by the Noetherian property for

$S^{n \times n}$, there exists some r such that, as before, all products with $s > r$ are an S-linear combination of the products with $s \leqslant r$. In particular, as $S \subseteq R$, they are an R-linear combination, and so $\{A_{i_1} \cdots A_{i_s}, s \leqslant r\}$ generate B. ▮

References

1. S. Eilenberg, "Automata, Languages, and Machines," Vol. A, Academic Press, New York, 1974.
2. M. Fliess, Sur certaines familles de séries formelles, Thèse de Doctorat d'État, University of Paris VII, 1972.
3. M. Fliess, Matrices de Hankel, *J. Math. Pures Appl.* **53** (1974).
4. J. E. Hopcroft and J. D. Ullman, "Formal Languages and Their relation to Automata," Addison–Wesley, Reading, Mass., 1969.
5. S. Lang, "Algebra," Addison-Wesley, Reading, Mass., 1965.
6. A. Paz, "Introduction to Probabilistic Automata," Academic Press, New York, 1971.
7. A. Paz, Formal series, finiteness properties and decision problems, *Ann. Acad. Sci. Fenn., Ser. A* 1 (1971), 493.
8. Y. Rouchaleau, Linear, discrete-time, finite-dimensional dynamical systems over some classes of commutative rings, Ph.D. Dissertation, Stanford University, 1972.
9. M. P. Schützenberger, On the definition of a family of automata, *Information Control* **4** (1961), 245–270.
10. E. Sontag, On linear systems and noncommutative rings, *Math. System Theory* **9** (1975).