

RATIONAL SERIES WITH HIGH IMAGE COMPLEXITY

JUHA HONKALA¹

Abstract. By using the universal Diophantine representation of recursively enumerable sets of positive integers due to Matiyasevich we construct a \mathbb{Z} -rational series r over a binary alphabet X which has a maximal image complexity in the sense that all recursively enumerable sets of positive integers are obtained as the sets of positive coefficients of the series $w^{-1}r$ where $w \in X^*$. As a consequence we obtain various undecidability results for \mathbb{Z} -rational series.

Mathematics Subject Classification. 68Q45, 03B25, 11U05.

1. INTRODUCTION

We study \mathbb{Z} -rational power series in noncommutative variables, their images and their positive images. By definition, the image $V(r)$ of a series r is the set of its coefficients and the positive image $V_+(r)$ of r is the set of its positive coefficients. It is well known that the images of \mathbb{Z} -rational series can be very complicated. By using the universal Diophantine representation of recursively enumerable sets of positive integers due to Matiyasevich we show that there is a \mathbb{Z} -rational series r over a binary alphabet X such that all recursively enumerable sets of positive integers are obtained as positive images of the series $w^{-1}r$ where $w \in X^*$. Since the positive images of \mathbb{Z} -rational series are always recursively enumerable, r can be said to have maximal image complexity. We will use this series to give a version of Rice's theorem (stating that no nontrivial property is decidable for recursively enumerable sets) for the positive images of \mathbb{Z} -rational series. Finally, we will use this series to prove improved versions of previously known undecidability results for rational series. More precisely, we show that to reach undecidability it suffices to consider quotients by words of a suitably chosen specific series.

We assume that the reader is familiar with the basics concerning rational series (see [1, 2, 5, 8]) and recursively enumerable sets (see [7]). The approach of this paper was used in [4] to study products of matrices.

2. BASIC DEFINITIONS

As usual, \mathbb{Z} is the set of integers and \mathbb{R} is the set of real numbers. The sets of positive integers and positive real numbers are denoted by \mathbb{Z}_+ and \mathbb{R}_+ , respectively. Hence

$$\mathbb{Z}_+ = \{a \in \mathbb{Z} \mid a > 0\} \quad \text{and} \quad \mathbb{R}_+ = \{a \in \mathbb{R} \mid a > 0\}.$$

Observe that $0 \notin \mathbb{Z}_+$ and $0 \notin \mathbb{R}_+$.

Keywords and phrases. Rational series, recursively enumerable set, Diophantine representation, undecidability.

¹ Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland. juha.honkala@utu.fi

Let X be a finite nonempty set of variables and let A be a semiring. The set of *formal power series* with noncommutative variables in X and coefficients in A is denoted by $A\langle\langle X \rangle\rangle$. If $r \in A\langle\langle X \rangle\rangle$, r is a mapping from the free monoid X^* generated by X into A . The image by r of a word $w \in X^*$ is denoted by (r, w) and r is written as

$$r = \sum_{w \in X^*} (r, w)w.$$

Here (r, w) is called the *coefficient* of w in r . The *support* of r is the set

$$\text{supp}(r) = \{w \in X^* \mid (r, w) \neq 0\}.$$

A power series $r \in A\langle\langle X \rangle\rangle$ is called *proper* if $(r, \varepsilon) = 0$, where ε is the empty word.

Let $r \in A\langle\langle X \rangle\rangle$. Then the *image* $V(r)$ of r is the set

$$V(r) = \{(r, w) \mid w \in X^*\}.$$

If $A \subseteq \mathbb{R}$, then the *positive image* $V_+(r)$ of r is the set

$$V_+(r) = V(r) \cap \mathbb{R}_+.$$

In other words, $V(r)$ is the set of all coefficients of r and $V_+(r)$ is the set of all positive coefficients of r . In the literature the image of r is often denoted by $\text{Im}(r)$.

Next we recall the definitions of recognizable and rational series. Assume A is a semiring. If m and n are positive integers, then the set of $m \times n$ matrices having entries in A is denoted by $A^{m \times n}$.

A series $r \in A\langle\langle X \rangle\rangle$ is called *A-recognizable* if there exist an integer $n \geq 1$, a monoid morphism

$$\mu : X^* \rightarrow A^{n \times n}$$

and two matrices $\lambda \in A^{1 \times n}$ and $\gamma \in A^{n \times 1}$ such that for all $w \in X^*$,

$$(r, w) = \lambda \mu(w) \gamma.$$

Then the triple (λ, μ, γ) is called a *linear representation* of r .

To define the family of *A-rational* series we first recall what is meant by a rationally closed subset of $A\langle\langle X \rangle\rangle$.

If $r \in A\langle\langle X \rangle\rangle$ is a proper series, the *star* r^* of r is defined by

$$r^* = \sum_{n=0}^{\infty} r^n.$$

A subset S of $A\langle\langle X \rangle\rangle$ is called *rationally closed* if the following conditions hold:

- (i) If $r, s \in S$ and $a \in A$, then $r + s \in S$, $rs \in S$, $ar \in S$ and $ra \in S$.
- (ii) If $r \in S$ is a proper series, then $r^* \in S$.

Now, a power series $r \in A\langle\langle X \rangle\rangle$ is called *A-rational* if r belongs to the smallest subset of $A\langle\langle X \rangle\rangle$ which contains all polynomials and is rationally closed.

By a theorem of Schützenberger, a power series is *A-recognizable* if and only if it is *A-rational* (see [1, 2, 5, 8]).

If $r \in A\langle\langle X \rangle\rangle$ and $w \in X^*$, the series $w^{-1}r$ in $A\langle\langle X \rangle\rangle$ is defined by

$$(w^{-1}r, u) = (r, wu)$$

for all $u \in X^*$.

3. A RATIONAL SERIES WITH MAXIMAL IMAGE COMPLEXITY

In this section we study the positive images of \mathbb{Z} -rational series. If $r \in \mathbb{Z}\langle\langle X \rangle\rangle$ is \mathbb{Z} -rational, then $V_+(r)$ is always recursively enumerable. We will show that nothing more can be said in general. More precisely, we will construct a \mathbb{Z} -rational series r over a two-letter alphabet X such that the set $\{V_+(w^{-1}r) \mid w \in X^*\}$ equals the set of all recursively enumerable subsets of \mathbb{Z}_+ . Our main tool is the universal Diophantine representation of recursively enumerable subsets of \mathbb{Z}_+ proved by Matiyasevich (see [6]).

Fix an enumeration

$$K_0, K_1, K_2, \dots$$

of the recursively enumerable subsets of \mathbb{Z}_+ . Such an enumeration can be obtained by enumerating the Turing machines which define the recursively enumerable sets. Then the result of Matiyasevich can be stated as follows.

Theorem 3.1. *There is a positive integer t and a polynomial $p(z_1, \dots, z_t)$ with integer coefficients such that for all $i \geq 0$ and $m \in \mathbb{Z}_+$ we have*

$$m \in K_i$$

if and only if there exist nonnegative integers m_3, \dots, m_t such that

$$p(i, m, m_3, \dots, m_t) = 0.$$

Furthermore, we can compute such a polynomial $p(z_1, \dots, z_t)$.

For the proof of Theorem 3.1 see the monograph [6].

We will also need the following lemma which has often been used in proving undecidability results for rational series.

Lemma 3.2. *Let $X = \{x, y\}$ be a two-letter alphabet and let $q(z_1, \dots, z_t)$ be a polynomial with integer coefficients. Then there is a \mathbb{Z} -rational series $r \in \mathbb{Z}\langle\langle X \rangle\rangle$ such that*

$$(r, x^{m_1}yx^{m_2} \dots x^{m_{t-1}}yx^{m_t}) = q(m_1, \dots, m_t) \quad (3.1)$$

for all nonnegative integers m_1, \dots, m_t and

$$\text{supp}(r) \subseteq (x^*y)^{t-1}x^*. \quad (3.2)$$

Furthermore, we can compute a linear representation of r .

Lemma 3.2 follows easily from the basic properties of rational series (see [5, 8]). The computation of a linear representation of r is also explained in detail in [3]. We recall briefly its proof from ([5], p. 125). The polynomial $q(z_1, \dots, z_t)$ is obtained by finitely many applications of addition and multiplication to the polynomials 1, -1 and z_j , $1 \leq j \leq t$. The series corresponding to these polynomials are $(x^*y)^{t-1}x^*$, $-(x^*y)^{t-1}x^*$ and $(x^*y)^{j-1}x^+(yx^*)^{t-j}$, respectively. Then, if r_1 and r_2 correspond to $q_1(z_1, \dots, z_t)$ and $q_2(z_1, \dots, z_t)$, then the sum of r_1 and r_2 corresponds to $q_1 + q_2$ and the Hadamard product of r_1 and r_2 corresponds to q_1q_2 .

Now, let $p(z_1, \dots, z_t)$ be the universal polynomial of Theorem 3.1. Define the polynomial $q(z_1, \dots, z_t)$ by

$$q(z_1, \dots, z_t) = (z_2 + 1)(1 - p(z_1, \dots, z_t)^2) - 1.$$

Let $X = \{x, y\}$. By Lemma 3.2 there is a \mathbb{Z} -rational series $r \in \mathbb{Z}\langle\langle X \rangle\rangle$ such that (3.1) holds for all nonnegative integers m_1, \dots, m_t and (3.2) holds.

Lemma 3.3. *Let i be a nonnegative integer. Then we have*

$$V_+((x^i y)^{-1}r) = K_i.$$

Proof. Let i be a nonnegative integer and let $m \in \mathbb{Z}_+$. If $m \in K_i$, then there exist nonnegative integers m_3, \dots, m_t such that $p(i, m, m_3, \dots, m_t) = 0$. This implies that $q(i, m, m_3, \dots, m_t) = m$. Hence

$$(r, x^i y x^m y x^{m_3} \dots x^{m_{t-1}} y x^{m_t}) = m,$$

which implies that $m \in V_+((x^i y)^{-1} r)$.

Conversely, assume that $m \in V_+((x^i y)^{-1} r)$. Then there exist nonnegative integers m_2, \dots, m_t such that $q(i, m_2, \dots, m_t) = m$. Since $m > 0$, we have $p(i, m_2, \dots, m_t) = 0$. Hence $q(i, m_2, \dots, m_t) = m_2$, which implies that $m_2 = m$. By Theorem 3.1 we have $m \in K_i$. \square

Since K_0, K_1, K_2, \dots is an enumeration of the recursively enumerable subsets of \mathbb{Z}_+ , the following result is an immediate consequence of Lemma 3.3.

Theorem 3.4. *Let $X = \{x, y\}$ be a two-letter alphabet. There exists a \mathbb{Z} -rational series $r \in \mathbb{Z}\langle\langle X \rangle\rangle$ such that $\{V_+(w^{-1}r) \mid w \in x^*y\}$ equals the set of all recursively enumerable subsets of \mathbb{Z}_+ . Furthermore, we can compute a linear representation of r .*

Theorem 3.4 can be used to give examples of the complicated behavior of coefficients of \mathbb{Z} -rational series (see Exercise II.7.7 in [8]).

Example 3.5. Let r be the series of Theorem 3.4 and let z be a new letter. Then there is a word $w_1 \in x^*y$ such that

$$V\left(w_1^{-1}r - \sum_{n=0}^{\infty} n z^n\right) = \{a \in \mathbb{Z} \mid a \leq 0 \text{ or } a \text{ is a positive prime}\}.$$

Example 3.6. Let r be the series of Theorem 3.4 and let z be a new letter. Then there is a word $w_2 \in x^*y$ such that

$$V\left(w_2^{-1}r - \sum_{n=0}^{\infty} n z^n\right) = \left\{ a \in \mathbb{Z} \mid a \leq 0 \text{ or } a = n^{n^{n^{\dots^n}}} \text{ for some } n \in \mathbb{Z}_+ \right\},$$

where the shown tower has height n .

Theorem 3.7. *Let $X = \{x, y\}$ be a two-letter alphabet. There is a positive integer k and \mathbb{Z} -rational series $s_1, \dots, s_k \in \mathbb{Z}\langle\langle X \rangle\rangle$ such that the set*

$$\{V_+(a_1 s_1 + \dots + a_k s_k) \mid a_1, \dots, a_k \in \mathbb{Z}\}$$

equals the set of recursively enumerable subsets of \mathbb{Z}_+ .

Proof. Let r be the series of Theorem 3.4. Since r is recognizable, the set $\{w^{-1}r \mid w \in X^*\}$ is included in a \mathbb{Z} -submodule of $\mathbb{Z}\langle\langle X \rangle\rangle$, which is generated by finitely many \mathbb{Z} -rational series (see [1], Prop. 5.1). \square

4. UNDECIDABILITY

The results of Section 3 imply more or less immediately various undecidability results for rational series. In this section $X = \{x, y\}$ and $r \in \mathbb{Z}\langle\langle X \rangle\rangle$ is the series of Theorem 3.4.

Rice's well known theorem states that no nontrivial property is decidable for recursively enumerable sets of positive integers (see [7]). This implies immediately the following result.

Theorem 4.1. *Let \mathcal{A} be any nonempty set of subsets of \mathbb{Z}_+ which is not equal to the set $\mathcal{P}(\mathbb{Z}_+)$ of all subsets of \mathbb{Z}_+ . Then there is no algorithm to decide, given a word $w \in x^*y$ whether*

$$V_+(w^{-1}r) \in \mathcal{A}.$$

Theorem 4.1 implies improved versions of previous undecidability results concerning \mathbb{Z} -rational series (see [5, 8]). Below $\text{char}(X^*)$ is the characteristic series of X^* defined by

$$\text{char}(X^*) = \sum_{w \in X^*} w$$

and y^+ is the series defined by

$$y^+ = \sum_{n=1}^{\infty} y^n.$$

Corollary 4.2. *Let $s = r - \text{char}(X^*)$. It is undecidable, given $w \in x^*y$, whether $w^{-1}s$ has a zero coefficient.*

Proof. The series $w^{-1}s$ has a zero coefficient if and only if $1 \in V_+(w^{-1}r)$. Hence the claim follows by Theorem 4.1. \square

The problem of deciding whether all coefficients of $w^{-1}s$ are nonzero is often called the universality problem for the support of $w^{-1}s$.

Corollary 4.3. *Let $s = ry^+ - \text{char}(X^*)$. It is undecidable, given $w \in x^*y$, whether $w^{-1}s$ has infinitely many zero coefficients.*

Proof. Let $w \in x^*y$. We show that $w^{-1}s$ has infinitely many zero coefficients if and only if $1 \in V_+(w^{-1}r)$.

Assume first that $u \in X^*$ is a word such that $(w^{-1}s, u) = 0$. Then $(s, wu) = 0$ and $(ry^+, wu) = 1$. Let $wu = w_1yw_2y^j$, where w_1 contains $t - 2$ occurrences of y , $w_2 \in x^*$ and $j \geq 1$. Then $1 = (ry^+, wu) = (r, w_1yw_2)(y^+, y^j) = (r, w_1yw_2)$. Since w is a prefix of w_1yw_2 we see that $1 \in V_+(w^{-1}r)$.

Assume conversely that $1 \in V_+(w^{-1}r)$. Let $u \in X^*$ be a word such that $(w^{-1}r, u) = 1$. Then $(r, wu) = 1$. Hence $(ry^+, wuy^j) = 1$ for all $j \geq 1$. This implies that $(s, wuy^j) = 0$ for all $j \geq 1$. Hence $w^{-1}s$ has infinitely many zero coefficients.

Now the claim follows by Theorem 4.1. \square

Corollary 4.4. *It is undecidable, given $w \in x^*y$, whether $w^{-1}r$ has a positive coefficient.*

Proof. If $w \in x^*y$, then the series $w^{-1}r$ has a positive coefficient if and only if $V_+(w^{-1}r) \neq \emptyset$, which is undecidable by Theorem 4.1. \square

Corollary 4.5. *Let $s = ry^+$. It is undecidable, given $w \in x^*y$, whether $w^{-1}s$ has infinitely many positive coefficients.*

Proof. If $w \in x^*y$, the series $w^{-1}s$ has infinitely many positive coefficients if and only if $w^{-1}r$ has at least one positive coefficient. Hence the claim follows by Corollary 4.4. \square

Corollary 4.6. *It is undecidable, given $w \in x^*y$, whether $w^{-1}r$ has infinitely many different positive coefficients.*

Proof. If $w \in x^*y$, the series $w^{-1}r$ has infinitely many different positive coefficients if and only if $V_+(w^{-1}r)$ is an infinite set. Hence the claim follows by Theorem 4.1. \square

Corollary 4.7. *It is undecidable, given $w \in x^*y$, whether $V_+(r) = V_+(w^{-1}r)$.*

Proof. Since \mathbb{Z}_+ is a recursively enumerable set, there is a nonnegative integer i such that $V_+((x^iy)^{-1}r) = \mathbb{Z}_+$. Hence $V_+(r) = \mathbb{Z}_+$. Therefore $V_+(r) = V_+(w^{-1}r)$ holds if and only if $V_+(w^{-1}r) = \mathbb{Z}_+$. The claim follows by Theorem 4.1. \square

Corollary 4.8. *Let*

$$s = r - \sum_{m,n=0}^{\infty} nx^m y^{t+n+1}.$$

*It is undecidable, given $w \in x^*y$, whether $V(s) = V(w^{-1}s)$.*

Proof. If $w \in x^*y$, we have $V(s) = V_+(r) \cup \{n \in \mathbb{Z} \mid n \leq 0\}$ and $V(w^{-1}s) = V_+(w^{-1}r) \cup \{n \in \mathbb{Z} \mid n \leq 0\}$. Hence $V(s) = V(w^{-1}s)$ if and only if $V_+(r) = V_+(w^{-1}r)$. The claim follows by Corollary 4.7. \square

Theorem 4.1 shows that no nontrivial property is decidable for the positive images of \mathbb{Z} -rational series. No such result holds if we consider the full images. For example, it is decidable whether a rational series has a finite image (see [1, 2]).

REFERENCES

- [1] J. Berstel and C. Reutenauer, *Rational Series and Their Languages*. Springer, Berlin (1988).
- [2] J. Berstel and C. Reutenauer, *Noncommutative Rational Series With Applications*. Cambridge University Press, Cambridge (2011).
- [3] E. Charlier and J. Honkala, The freeness problem over matrix semigroups and bounded languages. *Inf. Comput.* **237** (2014) 243–256.
- [4] J. Honkala, Products of matrices and recursively enumerable sets. *J. Comput. Syst. Sci.* **81** (2015) 468–472.
- [5] W. Kuich and A. Salomaa, *Semirings, Automata, Languages*. Springer, Berlin (1986).
- [6] Y. Matiyasevich, *Hilbert's Tenth Problem*, MIT Press, Cambridge (1993).
- [7] A. Salomaa, *Computation and Automata*. Springer, Berlin (1985).
- [8] A. Salomaa and M. Soittola, *Automata-Theoretic Aspects of Formal Power Series*. Springer, Berlin (1978).

Communicated by Ch. Choffrut.

Received June 6, 2016. Accepted January 2, 2017.