Journal of Universal Computer Science, vol. 2, no. 4 (1996), 217-223
submitted: 21/2/96, accepted: 9/4/96, appeared: 28/4/96 © Springer Pub. Co.

On Images of Algebraic Series

Juha Honkala
Department of Mathematics
University of Turku
SF-20500 Turku, Finland
jhonkala@sara.cc.utu.fi

Abstract: We show that it is decidable whether or not the set of coefficients of a given **Q**-algebraic sequence is finite. The same question is undecidable for **Q**-algebraic series. We consider also prime factors of algebraic series.
Category: F.4.3

1 Introduction

Formal power series play an important role in many diverse areas of theoretical computer science and mathematics, see [Berstel and Reutenauer 88], [Kuich and Salomaa 86] and [Salomaa and Soittola 78]. The classes of power series studied most often in connection with automata, grammars and languages are the rational and algebraic series.

In language theory formal power series often provide a powerful tool for obtaining deep decidability results, see [Kuich and Salomaa 86] and [Salomaa and Soittola 78]. A brilliant example is the solution of the equivalence problem for finite deterministic multitape automata given in [Harju and Karhumäki 91].

In this paper we consider decision problems concerning algebraic sequences and series. For earlier decidability results see [Kuich and Salomaa 86]. We show first that it is decidable whether or not the set of coefficients of a given **Q**-algebraic sequence is finite. We show that the same question is undecidable for series in $\mathbf{N}^{\mathrm{alg}} \ll X^* \gg$. Next we consider algebraic series with commuting variables. We show that it is decidable, given a positive integer k and a series $r \in \mathbf{Q}^{\mathrm{alg}} \ll X^{\oplus} \gg$, whether or not the set of coefficients of r has cardinality at most k. (Here $X^{\oplus}$ is the free commutative monoid generated by X.) We also apply the methods of our decidability proofs to study the prime factors of **Q**-algebraic series.

The questions studied in this paper are closely related to the study of thin and slender languages and their generalizations, see [Andraşiu, Dassow, Păun and Salomaa 93], [Păun and Salomaa 92], [Păun and Salomaa 93], [Păun and Salomaa 95], [Dassow, Păun and Salomaa 93], [Ilie 94], [Raz 00], [Nishida and Salomaa 00] and [Honkala 00].

Standard terminology and notation concerning formal languages and power series will be used in this paper. Whenever necessary, the reader may consult [Salomaa 73], [Salomaa and Soittola 78], [Kuich and Salomaa 86] and [Berstel and Reutenauer 88].

2 Images of algebraic series

Let X be an alphabet. The *free monoid* (resp. the *free commutative monoid*) generated by X is denoted by X^* (resp. $X^\oplus$). The set of $\mathbf{Q}$-*rational* (resp. $\mathbf{Q}$-*algebraic*) *series* with noncommuting variables in X is denoted by $\mathbf{Q}^{\text{rat}} \ll X^* \gg$ (resp. $\mathbf{Q}^{\text{alg}} \ll X^* \gg$). (Here $\mathbf{Q}$ is the field of rational numbers.) We consider also $\mathbf{Q}$-rational and $\mathbf{Q}$-algebraic series with commuting variables in X. The corresponding sets are denoted by $\mathbf{Q}^{\text{rat}} \ll X^\oplus \gg$ and $\mathbf{Q}^{\text{alg}} \ll X^\oplus \gg$, respectively. Furthermore, denote by c the canonical morphism $c : \mathbf{Q} \ll X^* \gg \to \mathbf{Q} \ll X^\oplus \gg$. Hence,

$$\mathbf{Q}^{\text{rat}} \ll X^\oplus \gg = \{c(r) | r \in \mathbf{Q}^{\text{rat}} \ll X^* \gg\}$$

and

$$\mathbf{Q}^{\text{alg}} \ll X^\oplus \gg = \{c(r) | r \in \mathbf{Q}^{\text{alg}} \ll X^* \gg\}.$$

By definition, the *image* of a series is the set of its coefficients. Hence, if $r = \sum (r, w) w \in \mathbf{Q} \ll X^* \gg$, the image of r equals the set

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

The following basic decidability result concerning images of $\mathbf{Q}$-rational series was established in [Jacob 78].

Theorem 1. (Jacob) *It is decidable whether or not a given rational series $r \in \mathbf{Q}^{rat} \ll X^* \gg$ has a finite image.*

In this paper we discuss the possibilities to generalize this result to $\mathbf{Q}$-algebraic series. We first establish a lemma concerning $\mathbf{Q}$-algebraic series with commuting variables. Its proof relies heavily on earlier deep results in [Kuich and Salomaa 86] and [Semenov 77].

If $w \in X^*$ (or $w \in X^\oplus$), the Parikh vector $\psi(w)$ of w is defined by

$$\psi(w) = (\#_{x_1}(w), \ldots, \#_{x_n}(w)).$$

Here $X = \{x_1, \ldots, x_n\}$ and $\#_x(w)$ stands for the number of occurrences of the letter x in w.

Lemma 2. *If $r \in \mathbf{Q}^{alg} \ll X^\oplus \gg$ has a finite image, then r is a finite $\mathbf{Q}$-linear combination of series in $\mathbf{N}^{rat} \ll X^\oplus \gg$ of the form $uv_1^* \ldots v_m^*$ with pairwise disjoint supports. Here $u, v_1, \ldots, v_m \in X^\oplus$ and the Parikh vectors $\psi(v_1), \ldots, \psi(v_m)$ are linearly independent over $\mathbf{Q}$. In particular, if $r \in \mathbf{Q}^{alg} \ll X^\oplus \gg$ has a finite image then $r \in \mathbf{Q}^{rat} \ll X^\oplus \gg$.*

Proof. Suppose that $r \in \mathbf{Q}^{\text{alg}} \ll X^\oplus \gg$ has a finite image. Without loss of generality we assume that r is quasiregular. Because r has a finite image there exists a positive integer $a \in \mathbf{N}$ such that $ar \in \mathbf{Z} \ll X^\oplus \gg$. By Corollary 16.11 in [Kuich and Salomaa 86] there exists a nonzero polynomial $P(x_1, \ldots, x_n, y) \in \mathbf{Z} < (X \cup y)^\oplus >$ such that

$$P(x_1, \ldots, x_n, ar) = 0. \tag{1}$$

(Here $X = \{x_1, \ldots, x_n\}$.) Next, fix an integer j and denote

$$D_j = \{(i_1, \ldots, i_n) \in \mathbf{N}^n | (ar, x_1^{i_1} \ldots x_n^{i_n}) = j\}.$$

To study the properties of the set D_j choose a large prime p and denote by ν the canonical morphism

$$\nu : \mathbf{Z} \ll X^{\oplus} \gg \rightarrow \mathbf{Z}_p \ll X^{\oplus} \gg .$$

Define the sequence $s : \mathbf{N}^n \rightarrow \mathbf{Z}_p$ by

$$s(i_1, \ldots, i_n) = (\nu(ar), x_1^{i_1} \ldots x_n^{i_n}).$$

It follows from (1) that

$$\nu(P)(x_1, \ldots, x_n, \nu(ar)) = 0$$

or

$$\nu(P)(x_1, \ldots, x_n, \sum_{i_1, \ldots, i_n \geq 0} s(i_1, \ldots, i_n) x_1^{i_1} \ldots x_n^{i_n}) = 0.$$

Hence the sequence s is p-algebraic. By Theorem 5.1 in [Bruyère, Hansel, Michaux and Villemaire 94] the sequence s is p-recognizable. Consequently, the set D'_j defined by

$$D'_j = \{(i_1, \ldots, i_n) \in \mathbf{N}^n | (ar, x_1^{i_1} \ldots x_n^{i_n}) \equiv j \pmod{p}\}$$

is a p-recognizable subset of $\mathbf{N}^n$. Because p is large, $D_j = D'_j$. Hence D_j is a p-recognizable subset of $\mathbf{N}^n$.

Now, by replacing in the argument above the prime p by another large prime q it follows that D_j is also q-recognizable. Therefore, by a deep result of Semenov (see [Semenov 77]), the set D_j is a rational subset of $\mathbf{N}^n$. Denote

$$E_j = \{x_1^{i_1} \ldots x_n^{i_n} | (i_1, \ldots, i_n) \in D_j\}.$$

Clearly, E_j is a rational subset of $X^{\oplus}$. Because $X^{\oplus}$ is a commutative monoid, E_j is an unambiguous rational subset of $X^{\oplus}$ (see [Eilenberg and Schützenberger 69]). It follows that

$$\mathrm{char}(E_j) \in \mathbf{N}^{\mathrm{rat}} \ll X^{\oplus} \gg .$$

Hence $\mathrm{char}(E_j)$ is a finite $\mathbf{N}$-linear combination of series of the form $uv_1^* \ldots v_m^*$ with pairwise disjoint supports, where $u, v_1, \ldots, v_m \in X^{\oplus}$ and the Parikh vectors $\psi(v_1), \ldots, \psi(v_m)$ are linearly independent over $\mathbf{Q}$. Because ar has a finite image, ar is a finite $\mathbf{Z}$-linear combination of series $\mathrm{char}(E_j)$, where j is an integer. This implies the claim. □

In the next theorem, $x \in X$ is a letter.

Theorem 3. *It is decidable whether or not a given sequence* $r \in \mathbf{Q}^{alg} \ll x^* \gg$ *has a finite image.*

Proof. First, decide by the method of Theorem 16.13 in [Kuich and Salomaa 86] whether r belongs to $\mathbf{Q}^{\mathrm{rat}} \ll x^* \gg$. If not, Lemma 2 implies that the image of r is infinite. If $r \in \mathbf{Q}^{\mathrm{rat}} \ll x^* \gg$, the finiteness of the image can be decided by Theorem 1. □

Theorem 3 cannot be extended to alphabets with more than one letter.

Theorem 4. *Let X be an alphabet with at least two letters. It is undecidable, given a series $r \in \mathbf{N}^{alg} \ll X^* \gg$, whether or not r has a finite image.*

Proof. Let $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ be two lists of words over an alphabet Σ determining an instance PCP of the Post Correspondence Problem. Choose new letters a, b, c, d and define the series r by

$$r = \sum_{k \geq 1, 1 \leq i_1, \ldots, i_k \leq n} ba^{i_1} ba^{i_2} \ldots ba^{i_k} cu_{i_k} \ldots u_{i_2} u_{i_1} d$$
$$+ \sum_{k \geq 1, 1 \leq i_1, \ldots, i_k \leq n} ba^{i_1} ba^{i_2} \ldots ba^{i_k} cv_{i_k} \ldots v_{i_2} v_{i_1} d.$$

Consider the series r^+. Clearly r^+ is $\mathbf{N}$-algebraic. Now, if PCP has a solution, at least one term of r has coefficient 2. Hence r^+ has an infinite image. On the other hand, if PCP does not possess a solution the set

$$\{ba^{i_1} ba^{i_2} \ldots ba^{i_k} cu_{i_k} \ldots u_{i_2} u_{i_1} d | k \geq 1, 1 \leq i_1, \ldots, i_k \leq n\}$$
$$\cup \{ba^{i_1} ba^{i_2} \ldots ba^{i_k} cv_{i_k} \ldots v_{i_2} v_{i_1} d | k \geq 1, 1 \leq i_1, \ldots, i_k \leq n\},$$

where the union is disjoint, is a prefix code. Therefore, each coefficient of r^+ equals 0 or 1, and the image of r^+ is finite. Consequently, the image of r^+ is finite if and only if PCP does not possess a solution.

Finally, let $h : (\Sigma \cup \{a, b, c, d\})^* \to X^*$ be an injective morphism. Such a morphism exists because X has at least two letters. By the closure properties of algebraic series, $h(r^+)$ belongs to $\mathbf{N}^{\mathrm{alg}} \ll X^* \gg$. Because the injective morphism preserves the image, the claim follows. □

It is an open problem whether or not it is decidable if a given power series $r \in \mathbf{Q}^{\mathrm{alg}} \ll X^{\oplus} \gg$ has a finite image. The following theorem solves a related problem.

Theorem 5. *Given a positive integer k and a series $r \in \mathbf{Q}^{alg} \ll X^{\oplus} \gg$ it is decidable whether or not the image of r has cardinality at most k.*

Proof. First, decide whether or not r belongs to $\mathbf{Q}^{\mathrm{rat}} \ll X^{\oplus} \gg$. If not, r has an infinite image and we are done. If $r \in \mathbf{Q}^{\mathrm{rat}} \ll X^{\oplus} \gg$ we consider two semialgorithms. The first semialgorithm computes successively the coefficients of r and tries to find $k+1$ distinct coefficients. The second semialgorithm tries to express r as a finite $\mathbf{Q}$-linear combination of series of the form $uv_1^* \ldots v_n^*$ with pairwise disjoint supports, where $u, v_1, \ldots, v_n \in X^{\oplus}$ and the Parikh vectors $\psi(v_1), \ldots, \psi(v_n)$ are linearly independent over $\mathbf{Q}$. This semialgorithm terminates, by Lemma 2, if r has a finite image. If it terminates, it can be decided whether or not the image of r has cardinality at most k.

An algorithm for Theorem 5 is now obtained by using concurrently the two semialgorithms. □

3 Prime factors of algebraic series

In this section we use the methods of the previous section to study prime factors of algebraic series.

If p is a prime, the *p-adic valuation* ν_p over $\mathbf{Q}$ is defined as follows. If $a, b \in \mathbf{Z}$, $b \neq 0$ and p divides neither a nor b, then $\nu_p(p^n a/b) = n$ for $n \in \mathbf{Z}$. Furthermore, $\nu_p(0) = \infty$. Now, if $r \in \mathbf{Q} \ll X^* \gg$ (or $r \in \mathbf{Q} \ll X^\oplus \gg$), the set $\mathrm{Prime}(r)$ of *prime factors* of r is defined by

$$\mathrm{Prime}(r) = \{p \in \mathbf{N} | p \text{ is a prime number and for some } w \in X^* \text{ we have } \nu_p((r, w)) \neq 0, \infty\}.$$

For the theory of prime factors of $\mathbf{Q}$-rational series, see [Berstel and Reutenauer 88]. By a well known theorem of [Pólya 21], the set of prime factors of a rational series $r \in \mathbf{Q}^{\mathrm{rat}} \ll x^* \gg$ is finite if and only if r is the sum of a polynomial and of a merge of geometric series.

For the next theorem we need two definitions. First, a language $L \subseteq X^*$ is called *commutatively nonrational* if the commutative variant $c(L)$ of L is not a rational subset of $X^\oplus$. Secondly, a language $L \subseteq X^*$ is called *Parikh thin* if $c(w_1) \neq c(w_2)$ whenever w_1 and w_2 are distinct elements of L.

Theorem 6. *Suppose $r \in \mathbf{Q}^{alg} \ll X^* \gg$ is a $\mathbf{Q}$-algebraic series. If supp(r) is commutatively nonrational and Parikh thin, there is at most one prime p such that p is not a prime factor of r.*

Proof. We assume without loss of generality that r is quasiregular. Because r is Parikh thin, the series r and $c(r)$ have the same prime factors. Therefore it suffices to show that there is at most one prime p which is not a prime factor of $c(r)$. Suppose p is such a prime. Denote

$$A = \{a \in \mathbf{Q} | \nu_p(a) \geq 0\},$$

$$I = \{a \in \mathbf{Q} | \nu_p(a) > 0\}.$$

It is well known that A is a ring and I is a maximal ideal of A. Hence $F = A/I$ is a field with p elements. Denote by ν the canonical morphism

$$\nu : A \rightarrow F$$

and its extension

$$\nu : A \ll X^\oplus \gg \rightarrow F \ll X^\oplus \gg .$$

Because p is not a prime factor of $c(r)$, we have $c(r) \in A \ll X^\oplus \gg$. Hence, $\nu(c(r)) \in F \ll X^\oplus \gg$. Furthermore, the supports of $c(r)$ and $\nu(c(r))$ are equal.

Now, by Corollary 16.12 in [Kuich and Salomaa 86], there exists a primitive polynomial $P(x_1, \ldots, x_n, y) \in \mathbf{Z} < (X \cup y)^\oplus >$ such that

$$P(x_1, \ldots, x_n, c(r)) = 0. \tag{2}$$

(Here $X = \{x_1, \ldots, x_n\}$.) Next, regard (2) as an equation in $A \ll X^\oplus \gg$ and apply the morphism ν. It follows that

$$\nu(P)(x_1, \ldots, x_n, \nu(c(r))) = 0.$$

Denote

$$D = \{(i_1, \ldots, i_n) | x_1^{i_1} \ldots x_n^{i_n} \in \text{supp}(c(r))\}.$$

Now, it follows as in the proof of Lemma 2 that D is a p-recognizable subset of $\mathbf{N}^n$. Consequently, we have seen that if p is a prime which is not a prime factor of r, then the set D is p-recognizable.

To conclude the proof, suppose that p and q are distinct primes which are not prime factors of r. Then the set D is both a p-recognizable and a q-recognizable subset of $\mathbf{N}^n$. Hence, by the result of [Semenov 77], D is a rational subset of $\mathbf{N}^n$. Consequently, supp$(c(r))$ is a rational subset of $X^{\oplus}$. This is not possible because supp$(c(r)) = c(\text{supp}(r))$. Hence there cannot be more than one prime which is not a prime factor of r. □

Denote by α the isomorphism $\alpha : X^{\oplus} \to \mathbf{N}^n$ defined by

$$\alpha(x_1^{i_1} \ldots x_n^{i_n}) = (i_1, \ldots, i_n).$$

By definition, a language $L \subseteq X^*$ is *commutatively p-recognizable* if $\alpha(c(L))$ is a p-recognizable subset of $\mathbf{N}^n$.

Theorem 7. *Suppose $r \in \mathbf{Q}^{alg} \ll X^* \gg$ is a $\mathbf{Q}$-algebraic series such that supp(r) is Parikh thin. If supp(r) is commutatively p-recognizable for no prime p, then every prime is a prime factor of r.*

Proof. The claim follows by the proof of Theorem 6. □

We conclude with an example of a series satisfying the assumptions of Theorem 7.

Example 1. Denote

$$r = \sum_{n,m \geq 0} (n^2 - m)^2 a^n b^m.$$

The series r belongs to $\mathbf{Q}^{\text{rat}} \ll \{a, b\}^* \gg$. Clearly,

$$\text{supp}(r) = \{a^n b^m | n^2 \neq m \text{ and } n, m \geq 0\}.$$

Hence, supp(r) is Parikh thin. Also, the set $\alpha(c(\text{supp}(r))) = \{(n, m) | n^2 \neq m \text{ and } n, m \geq 0\}$ is p-recognizable for no prime p. Indeed, if $\alpha(c(\text{supp}(r)))$ were p-recognizable so would be the sets $\{(n, m) | n^2 = m \text{ and } n, m \geq 0\}$ and $\{n^2 | n \geq 0\}$. However, the last set is a well known example of a set which is not p-recognizable for any p. Hence r satisfies the assumptions of Theorem 7. Obviously each prime is a prime factor of r.

References

[Andraşiu, Dassow, Păun and Salomaa 93] Andraşiu, M., Dassow, J., Păun, G. and Salomaa, A.: "Language-theoretic problems arising from Richelieu cryptosystems"; Theoret. Comput. Sci. 116 (1993) 339-357.

[Berstel and Reutenauer 88] Berstel, J. and Reutenauer, C.: "Rational Series and Their Languages"; Springer, Berlin (1988).

[Bruyère, Hansel, Michaux and Villemaire 94] Bruyère, V., Hansel, G., Michaux, C. and Villemaire, R.: "Logic and p-recognizable sets of integers"; Bull. Belgian Math. Soc. 1 (1994) 191-237.

[Dassow, Păun and Salomaa 93] Dassow, J., Păun, G. and Salomaa, A.: "On thinness and slenderness of L languages"; EATCS Bulletin 49 (1993) 152-158.

[Eilenberg and Schützenberger 69] Eilenberg, S. and Schützenberger, M. P.: "Rational sets in commutative monoids"; J. Algebra 13 (1969) 173-191.

[Harju and Karhumäki 91] Harju, T. and Karhumäki, J.: "The equivalence problem of multitape finite automata"; Theoret. Comput. Sci. 78 (2) (1991) 347-355.

[Honkala 00] Honkala, J.: "On Parikh slender languages and power series"; J. Comput. System Sci., to appear.

[Ilie 94] Ilie, L.: "On a conjecture about slender context-free languages"; Theoret. Comput. Sci. 132 (1994) 427-434.

[Jacob 78] Jacob, G.: "La finitude des representations lineaires des semi-groupes est decidable"; J. Algebra 52 (1978) 437-459.

[Kuich and Salomaa 86] Kuich, W. and Salomaa, A.: "Semirings, Automata, Languages"; Springer, Berlin (1986).

[Nishida and Salomaa 00] Nishida, T. and Salomaa, A.: "Slender 0L languages"; Theoret. Comput. Sci., to appear.

[Păun and Salomaa 92] Păun, G. and Salomaa, A.: "Decision problems concerning the thinness of D0L languages"; EATCS Bulletin 46 (1992) 171-181.

[Păun and Salomaa 93] Păun, G. and Salomaa, A.: "Closure properties of slender languages"; Theoret. Comput. Sci. 120 (1993) 293-301.

[Păun and Salomaa 95] Păun, G. and Salomaa, A.: "Thin and slender languages"; Discrete Appl. Math. 61 (1995) 257-270.

[Pólya 21] Pólya, G.: "Arithmetische Eigenschaften der Reihenentwicklungen rationaler Funktionen"; J. Reine Angew. Math. 151 (1921) 1-31.

[Raz 00] Raz, D.: "Length considerations in context-free languages"; Theoret. Comput. Sci., to appear.

[Salomaa 73] Salomaa, A.: "Formal Languages"; Academic Press, New York (1973).

[Salomaa and Soittola 78] Salomaa, A. and Soittola, M.: "Automata-Theoretic Aspects of Formal Power Series"; Springer, Berlin (1978).

[Semenov 77] Semenov, A. L.: "Presburgerness of predicates regular in two number systems" (in Russian); Sibirsk. Mat. Zh. 18 (1977) 403-418. English translation: Siberian Math. J. 18 (1977) 289-299.