Journal of Universal Computer Science, vol. 3, no. 10 (1997), 1114-1120
submitted: 11/4/97, accepted: 11/7/97, appeared: 28/10/97 © Springer Pub. Co.

On N-algebraic Parikh slender power series

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

Abstract: In a recent paper we introduced Parikh slender languages and series as a generalization of slender languages defined and studied by Andraşiu, Dassow, Păun and Salomaa. Results concerning Parikh slender series can be applied in ambiguity proofs of context-free languages. In this paper an algorithm is presented for deciding whether or not a given **N**-algebraic series is Parikh slender.

Category: F.4.3

1 Introduction

Length considerations are often useful in language theory. For example, Flajolet [6] has shown that the inherent ambiguity of many context-free languages can be deduced from the transcendentality of their generating functions. Other deep results based on length considerations are well known, e.g., in the theory of Lindenmayer systems (see, e.g., Rozenberg and Salomaa [19] and Ruohonen [20]).

The notion of a Parikh slender language (resp. Parikh slender power series) was introduced in Honkala [8]. The idea is to count the words with the same Parikh vector (resp. the sum of the coefficients of the words with the same Parikh vector). The language (resp. series) is Parikh slender if this number (resp. sum) is bounded from above. For basic results concerning Parikh slender languages and series see Honkala [8]. We mention only that Parikh slender series can be used to give a new simple proof of the result of Autebert, Flajolet and Gabarro [2] concerning the inherent ambiguity of coprefix languages of infinite words.

The notion of a Parikh slender language is a generalization of the notion of a slender language due to Andraşiu, Dassow, Păun and Salomaa [1]. For slender languages see also Păun and Salomaa [15-17], Dassow, Păun and Salomaa [4], Ilie [11], Raz [18] and Nishida and Salomaa [14].

The purpose of this note is to prove that it is decidable whether or not a given **N**-algebraic series is Parikh slender. As a byproduct we get information concerning Parikh slender **N**-algebraic series.

2 Definitions and results

Let Σ be an alphabet. The *free monoid* (resp. *the free commutative monoid*) generated by Σ is denoted by Σ^* (resp. $\Sigma^{\oplus}$). The set of **N**-algebraic (resp. **N**-rational) series with noncommuting variables in Σ is denoted by $\mathbf{N}^{\mathrm{alg}} \ll \Sigma^* \gg$ (resp. $\mathbf{N}^{\mathrm{rat}} \ll \Sigma^* \gg$). (Here **N** is the semiring of nonnegative integers.) We consider also **N**-algebraic and **N**-rational series with commuting variables in

Σ. The corresponding sets are denoted by $\mathbf{N}^{\mathrm{alg}} \ll \Sigma^{\oplus} \gg$ and $\mathbf{N}^{\mathrm{rat}} \ll \Sigma^{\oplus} \gg$, respectively. Furthermore, denote by c the canonical morphism $c : \mathbf{N} \ll \Sigma^* \gg \rightarrow \mathbf{N} \ll \Sigma^{\oplus} \gg$. Hence

$$\mathbf{N}^{\mathrm{alg}} \ll \Sigma^{\oplus} \gg = \{c(r) | r \in \mathbf{N}^{\mathrm{alg}} \ll \Sigma^* \gg\}$$

and

$$\mathbf{N}^{\mathrm{rat}} \ll \Sigma^{\oplus} \gg = \{c(r) | r \in \mathbf{N}^{\mathrm{rat}} \ll \Sigma^* \gg\}.$$

A power series $r \in \mathbf{N} \ll \Sigma^* \gg$ is said to be *Parikh slender* if there exists a positive integer k such that

$$(c(r), w) \leq k \text{ for all } w \in \Sigma^{\oplus}.$$

Hence, a series $r \in \mathbf{N} \ll \Sigma^* \gg$ is Parikh slender if and only if the coefficients of $c(r)$ are bounded from above by a constant. The definition of a Parikh slender language is now obtained as a special case. If $L \subseteq \Sigma^*$ is a language, denote by $\mathrm{char}(L) \in \mathbf{N} \ll \Sigma^* \gg$ the characteristic series of L. A language L is said to be *Parikh slender* if the series $\mathrm{char}(L)$ is Parikh slender.

The following result is due to Honkala [10].

Theorem 1. *It is decidable whether or not a given context-free language is Parikh slender.*

For a proof of the following theorem see Honkala [8].

Theorem 2. *It is decidable whether or not a given series* $r \in \mathbf{N}^{rat} \ll \Sigma^* \gg$ *is Parikh slender.*

The purpose of this note is to prove the following result.

Theorem 3. *It is decidable whether or not a given series* $r \in \mathbf{N}^{alg} \ll \Sigma^* \gg$ *is Parikh slender.*

Recall that the *image* of a series is the set of its coefficients. Hence, Theorem 3 can be restated as follows: it is decidable whether or not a given series $r \in \mathbf{N}^{\mathrm{alg}} \ll \Sigma^{\oplus} \gg$ with commuting variables has a finite image. The finiteness of the image is undecidable for series $r \in \mathbf{N}^{\mathrm{alg}} \ll \Sigma^* \gg$ with noncommuting variables (see Honkala [9]). Note that although Theorems 1 and 3 are related, neither implies the other. Both imply the decidability of Parikh slenderness for unambiguous context-free languages.

The reader is referred to Ginsburg [7] and Salomaa [21] for results concerning context-free languages and to Salomaa and Soittola [22], Kuich and Salomaa [13], Berstel and Reutenauer [3] and Kuich [12] for results concerning formal power series.

3 Proofs

In this section we prove Theorem 3. Recall that a language $L \subseteq \Sigma^*$ is *bounded* if there exist words $w_1, w_2, \ldots, w_m \in \Sigma^*$ such that $L \subseteq w_1^* w_2^* \ldots w_m^*$.

Lemma 4. *If $r \in \mathbf{N}^{alg} \ll \Sigma^* \gg$ is Parikh slender, there exist nonempty words $w_1, w_2, \ldots, w_m \in \Sigma^*$ such that*

$$supp(r) \subseteq w_1^* w_2^* \ldots w_m^*. \tag{1}$$

Proof. If r is Parikh slender, so is $\operatorname{supp}(r)$. Also, $\operatorname{supp}(r)$ is context-free (see Salomaa and Soittola [22]). Hence the claim follows by Theorem 4.1 in Honkala [8] stating that a Parikh slender context-free language is bounded. □

Suppose now that (1) holds. Let $\Delta = \{a_1, \ldots, a_m\}$ be a new alphabet with m letters and define the morphism $h : \Delta^* \to \Sigma^*$ by $h(a_i) = w_i$, $1 \le i \le m$. By the Cross-Section Theorem due to Eilenberg [5], there exists a rational language $R \subseteq a_1^* a_2^* \ldots a_m^*$ such that h maps R bijectively onto $w_1^* w_2^* \ldots w_m^*$. Define the series $s \in \mathbf{N} \ll \Delta^* \gg$ by

$$s = h^{-1}(r) \odot \operatorname{char}(R).$$

By the closure properties of $\mathbf{N}$-algebraic series we have $s \in \mathbf{N}^{\mathrm{alg}} \ll \Delta^* \gg$. Note that $h(s) = r$. Consequently, if r is Parikh slender, so is s. Furthermore, $\operatorname{supp}(s) \subseteq a_1^* a_2^* \ldots a_m^*$.

Our next goal is to derive a normal form for Parikh slender $\mathbf{N} \ll \Delta^* \gg$-algebraic series s satisfying $\operatorname{supp}(s) \subseteq a_1^* a_2^* \ldots a_m^*$. First, recall the Substitution Lemma for algebraic systems: Assume that

$$y_1 = q_1 + n_1 \alpha_1 y_j \alpha_2, \quad y_i = p_i, \quad 2 \le i \le n$$

where $n_1 \in \mathbf{N}$, $\alpha_1, \alpha_2 \in (\Delta \cup Y)^*$, $1 \le j \le n$ is a proper $\mathbf{N} \ll \Delta^* \gg$-algebraic system having the quasiregular solution $(s_1, \ldots, s_n)$. Then also the modified system

$$y_1 = q_1 + n_1 \alpha p_j \alpha_2, \quad y_i = p_i, \quad 2 \le i \le n,$$

is proper and has the solution $(s_1, \ldots, s_n)$.

Now, consider a proper $\mathbf{N} \ll \Delta^* \gg$-algebraic system $y_i = p_i$, $1 \le i \le n$ and define the relation $\Longrightarrow$ on $Y = \{y_1, \ldots, y_n\}$ by $y_\alpha \Longrightarrow y_\beta$ if and only if y_β has an occurrence in p_α, $1 \le \alpha, \beta \le n$. Let $\Longrightarrow^*$ be the reflexive transitive closure of $\Longrightarrow$. We say that the system $y_i = p_i$ with the strong solution $(s_1, \ldots, s_n)$ is *reduced* if the following conditions hold:

1) $y_1 \Longrightarrow^* y$ for every $y \in Y$,
2) for any $i, i \neq 1$, the variable y_i has an occurrence in p_i,
3) if $i \neq 1$ then s_i is not a polynomial.

If $y_i = p_i$, $1 \le i \le n$, is a given proper $\mathbf{N} \ll \Delta^* \gg$-algebraic system defining the series s, we can construct a reduced proper $\mathbf{N} \ll \Delta^* \gg$-algebraic system defining s by repeated applications of the Substitution Lemma. Next we establish some properties of the relation $\Longrightarrow^*$.

Lemma 5. *Suppose $s \in \mathbf{N}^{alg} \ll \Delta^* \gg$ is a nonzero Parikh slender series with supp$(s) \subseteq a_1^* a_2^* \ldots a_m^*$ and $y_i = p_i$, $1 \leq i \leq n$, is a proper $\mathbf{N} \ll \Delta^* \gg$-algebraic system defining s. Furthermore, assume that no component of the strong solution of $y_i = p_i$ equals zero. Denote $Y_1 = \{y \in Y | y \Longrightarrow^* y_1\}$. Then p_1 contains at most one occurrence of a variable in Y_1.*

Proof. Assume the contrary. By repeated applications of the Substitution Lemma we see that s is defined by a system $y_i = q_i$, $1 \leq i \leq n$, such that q_1 contains at least two occurrences of y_1. Assume first that q_1 has a term which contains at least two occurrences of y_1. Then there exist words $u_1, u_2, u_3 \in \Delta^*$ such that $u_1 s u_2 s u_3 \leq s$. (For $s_1, s_2 \in \mathbf{N} \ll \Delta^* \gg$ we denote $s_1 \leq s_2$ if and only if $(s_1, w) \leq (s_2, w)$ for every $w \in \Delta^*$.) This is not possible if the minimal alphabet of supp(s) has more than one letter. Therefore, suppose that supp$(s) \subseteq a_1^*$ say. Then $u_1 u_2 u_3 s^2 \leq s$ which implies that s has arbitrarily large coefficients. Since this is not possible, the occurrences of y_1 in q_1 have to be in distinct terms of q_1. This implies the existence of words $u_1, u_2, v_1, v_2 \in \Delta^*$ such that

$$u_1 s v_1 + u_2 s v_2 \leq s.$$

Therefore

$$\begin{aligned} s &\geq u_1(u_1 s v_1 + u_2 s v_2) v_1 + u_2(u_1 s v_1 + u_2 s v_2) v_2 \\ &\geq u_1 u_2 s v_2 v_1 + u_2 u_1 s v_1 v_2 = 2 u_1 u_2 s v_1 v_2. \end{aligned} \tag{2}$$

Here the last equality follows because necessarily $u_1 u_2 = u_2 u_1$ and $v_1 v_2 = v_2 v_1$. However, (2) implies that s has arbitrarily large coefficients. This contradiction proves the lemma. □

Lemma 6. *Suppose $s \in \mathbf{N}^{alg} \ll \Delta^* \gg$ is a nonzero Parikh slender series with supp$(s) \subseteq a_1^* a_2^* \ldots a_m^*$. If $y_i = p_i$, $1 \leq i \leq n$, is a reduced proper system defining s, then $\Longrightarrow^*$ is a partial order on $Y = \{y_1, \ldots, y_n\}$.*

Proof. Because $\Longrightarrow^*$ is reflexive and transitive, it suffices to show that if $y_\alpha \Longrightarrow^* y_\beta$ and $y_\beta \Longrightarrow^* y_\alpha$ then $y_\beta = y_\alpha$. Suppose on the contrary that $y_\alpha \Longrightarrow^* y_\beta$ and $y_\beta \Longrightarrow^* y_\alpha$ but $y_\beta \neq y_\alpha$. Denote by $(s_1, \ldots, s_n)$ the strong solution of $y_i = p_i$. Hence $s_1 = s$. For every i, $1 \leq i \leq n$, the series s_i is Parikh slender and satisfies supp$(s_i) \subseteq a_1^* a_2^* \ldots a_m^*$. Without loss of generality, suppose $\alpha \neq 1$. By assumption p_α contains an occurrence of y_α. Consider the other occurrences of the variables in p_α. Because $y_\alpha \Longrightarrow^* y_\beta$, one of the other occurrences of a variable y_γ satisfies $y_\gamma \Longrightarrow^* y_\beta$. But then $y_\alpha \Longrightarrow^* y_\alpha$ and $y_\gamma \Longrightarrow^* y_\alpha$ which contradicts Lemma 5. □

We say that a proper $\mathbf{N} \ll \Delta^* \gg$-algebraic system $y_i = p_i$, $1 \leq i \leq n$, has the triangular form if $\Longrightarrow^*$ is a partial order on $Y = \{y_1, \ldots, y_n\}$ and there exist an integer $a_1 \in \{0, 1\}$, words $u_i, v_i \in \Delta^*$ and polynomials $q_i \in \mathbf{N} < (\Delta \cup Y)^* >$ for $1 \leq i \leq n$ such that

$$\begin{aligned} p_1 &= a_1 u_1 y_1 v_1 + q_1 \\ p_2 &= u_2 y_2 v_2 + q_2 \\ &\ldots \\ p_n &= u_n y_n v_n + q_n, \end{aligned}$$

where for $1 \leq i \leq n$, q_i contains only variables belonging to the set $\{y \in Y | y_i \Longrightarrow^* y, y_i \neq y\}$. (Intuitively, q_i contains only variables which are smaller than y_i.)

Lemma 7. *Suppose $s \in \mathbf{N}^{alg} \ll \Delta^* \gg$ is a nonzero Parikh slender series with $supp(s) \subseteq a_1^* a_2^* \ldots a_m^*$. If $y_i = p_i$, $1 \leq i \leq n$, is a reduced proper $\mathbf{N} \ll \Delta^* \gg$-algebraic system defining s, the system has the triangular form.*

Proof. By Lemma 6, the relation $\Longrightarrow^*$ is a partial order on $Y = \{y_1, \ldots, y_n\}$. Suppose $1 \leq j \leq n$. By Lemma 5 there exist polynomials $p_{j1}, p_{j2}, q_j \in \mathbf{N} < (\Delta \cup Y)^* >$ such that

$$p_j = p_{j1} y_j p_{j2} + q_j \tag{3}$$

where the only occurrence of y_j in the righthand side is shown. Suppose first that p_{j1} and p_{j2} are nonzero. Denote by $(s_1, \ldots, s_n)$ the strong solution of $y_i = p_i$, $1 \leq i \leq n$. By substituting in (3) s_i for y_i for $1 \leq i \leq n$, and denoting by s_{j1} and s_{j2} the series which are obtained by making the substitution in p_{j1} and p_{j2}, respectively, it is seen that

$$s_{j1} s_j s_{j2} \leq s_j.$$

Hence

$$s_{j1}^k s_j s_{j2}^k \leq s_j$$

for any $k \geq 1$. Because s_j is Parikh slender s_{j1} and s_{j2} have to be words. This is possible only if p_{j1} and p_{j2} are words of Δ^*. Consequently, p_j has the form

$$p_j = a_j u_j y_j v_j + q_j \tag{4}$$

where $u_j, v_j \in \Delta^*$ and $a_j = 1$. If $p_{j1} = 0$ or $p_{j2} = 0$, necessarily $j = 1$ and (4) holds with $a_j = 0$. This implies the claim. □

Algebraic systems in the triangular form are desirable because they are easy to solve.

Lemma 8. *Suppose $y_i = p_i$, $1 \leq i \leq n$, is a proper $\mathbf{N} \ll \Sigma^* \gg$-algebraic system in the triangular form defining the series r. Then a rational series $\overline{r} \in \mathbf{N}^{rat} \ll \Sigma^* \gg$ can be effectively constructed such that $c(\overline{r}) = c(r)$.*

Proof. Denote the strong solution of the system $y_i = p_i$ by $(r_1, \ldots, r_n)$. Suppose first that $y_j \in Y$ is minimal with respect to $\Longrightarrow^*$. Then there exist words $u, v \in \Sigma^*$, $a \in \{0, 1\}$ and a polynomial $q \in \mathbf{N} < \Sigma^* >$ such that

$$p_j = a u y_j v + q.$$

Hence

$$r_j = q + a \sum_{t=1}^{\infty} u^t q v^t.$$

Therefore

$$c(r_j) = c(q) + a c(q) c(uv)^+.$$

Consider then a variable $y_k \in Y$ and suppose that for each $y_l \in Y - \{y_k\}$ with $y_k \Longrightarrow^* y_l$ we have a series $\overline{r}_l \in \mathbf{N}^{\mathrm{rat}} \ll \Sigma^* \gg$ such that $c(\overline{r}_l) = c(r_l)$. Then there exist words $u_1, v_1 \in \Sigma^*$, $a_1 \in \{0, 1\}$ and $q_1 \in \mathbf{N} < (\Sigma \cup Y)^* >$ such that

$$p_k = a_1 u_1 y_k v_1 + q_1(y_{l_1}, \ldots, y_{l_s})$$

and q_1 contains only variables y_l such that $y_k \Longrightarrow^* y_l$ and $y_l \neq y_k$. Hence

$$r_k = q_1(r_{l_1}, \ldots, r_{l_s}) + a_1 \sum_{t=1}^{\infty} u_1^t q_1(r_{l_1}, \ldots, r_{l_s}) v_1^t.$$

Therefore we can choose

$$\overline{r}_k = q_1(\overline{r}_{l_1}, \ldots, \overline{r}_{l_s}) + a_1 q_1(\overline{r}_{l_1}, \ldots, \overline{r}_{l_s}) \cdot (u_1 v_1)^+.$$

Now the existence of $\overline{r}$ follows inductively. □

Now we are ready to conclude the proof of Theorem 3.

Suppose $r \in \mathbf{N}^{\mathrm{alg}} \ll \Sigma^* \gg$ is a quasiregular series. Without restriction we assume that r is not a polynomial. First check whether or not $\mathrm{supp}(r)$ is a bounded language (see Ginsburg [7]). If not, Lemma 4 implies that r is not Parikh slender. We continue with the assumption that there exist nonempty words $w_1, w_2, \ldots, w_m \in \Sigma^*$ such that (1) holds and define the alphabet Δ, the morphism $h : \Delta^* \to \Sigma^*$ and the series s as above. Recall that $h(s) = r$. The construction of s is effective. Let $y_i = p_i$, $1 \leq i \leq n$, be a proper $\mathbf{N} \ll \Delta^* \gg$-algebraic system defining s. As pointed out above, we may assume that the system is reduced. If the system does not have the triangular form, Lemma 7 implies that s, and hence also r is not Parikh slender. We continue with the assumption that $y_i = p_i$ has the triangular form. Because h is a nonerasing morphism, we can construct a proper $\mathbf{N} \ll \Sigma^* \gg$-algebraic system in the triangular form defining r. Then Lemma 8 implies the effective existence of a rational series $\overline{r} \in \mathbf{N}^{\mathrm{rat}} \ll \Sigma^* \gg$ such that

$$c(\overline{r}) = c(r).$$

Now Theorem 3 follows by Theorem 2.

References

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

[2] J.-M. Autebert, P. Flajolet and J. Gabarro, Prefixes of infinite words and ambiguous context-free languages, *Inform. Process. Letters* **25** (1987), 211-216.

[3] J. Berstel and C. Reutenauer, "Rational Series and Their Languages," Springer, Berlin, 1988.

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

[5] S. Eilenberg, "Automata, Languages and Machines," Vol. A, Academic Press, New York, 1974.

[6] P. Flajolet, Analytic models and ambiguity of context-free languages, *Theoret. Comput. Sci.* **49** (1987), 283-309.

[7] S. Ginsburg, "The Mathematical Theory of Context-Free Languages," McGraw-Hill, New York, 1966.

[8] J. Honkala, On Parikh slender languages and power series, *J. Comput. System Sci.* **52** (1996), 185-190.

[9] J. Honkala, On images of algebraic series, *J. Univ. Comput. Sci.* **2** (1996) 217-223.

[10] J. Honkala, A decision method for Parikh slenderness of context-free languages, *Discrete Appl. Math.* **73** (1997), 1-4.

[11] L. Ilie, On a conjecture about slender context-free languages, *Theoret. Comput. Sci.* **132** (1994), 427-434.
[12] W. Kuich, Semirings and formal power series: their relevance to formal languages and automata, in G. Rozenberg and A. Salomaa (eds.): "Handbook of Formal Languages," Springer (1997), pp. 609-677.
[13] W. Kuich and A. Salomaa, "Semirings, Automata, Languages," Springer, Berlin, 1986.
[14] T. Nishida and A. Salomaa, Slender 0L languages, *Theoret. Comput. Sci.* **158** (1996) 161-176.
[15] G. Păun and A. Salomaa, Decision problems concerning the thinness of D0L languages, *EATCS Bulletin* **46** (1992), 171-181.
[16] G. Păun and A. Salomaa, Closure properties of slender languages, *Theoret. Comput. Sci.* **120** (1993), 293-301.
[17] G. Păun and A. Salomaa, Thin and slender languages, *Discrete Appl. Math.* **61** (1995), 257-270.
[18] D. Raz, Length considerations in context-free languages, *Theoret. Comput. Sci.*, to appear.
[19] G. Rozenberg and A. Salomaa, "The Mathematical Theory of L Systems," Academic Press, New York, 1980.
[20] K. Ruohonen, The decidability of the DOL-DTOL equivalence problem, *J. Comput. System Sci.* **22** (1981), 42-52.
[21] A. Salomaa, "Formal Languages," Academic Press, New York, 1973.
[22] A. Salomaa and M. Soittola, "Automata-Theoretic Aspects of Formal Power Series," Springer, Berlin, 1978.