Annales Academiæ Scientiarum Fennicæ
Series A. I. Mathematica
Volumen 10, 1985, 501—510

Commentationes in honorem
Olli Lehto
LX annos nato

ON META-NORMAL FORMS FOR ALGEBRAIC POWER SERIES IN NONCOMMUTING VARIABLES

ARTO SALOMAA

1. Introduction and preliminaries. The theory of formal power series in noncommuting variables was initiated around 1960 — apart from some scattered work done earlier in connection with free groups. Such power series are applicable in a number of areas but, in particular, they have turned out to be an indispensable tool in automata and language theory. Their usefulness in the latter theories is due to the fact that, in a sense, they lead to the arithmetization of the theory.

The purpose of this paper is to establish classes of normal forms for algebraic power series. The normal forms, as well as the definition of algebraic power series in general, are closely connected with the corresponding questions dealing with context-free grammars. In fact, one of the most important open problems concerns the possibility of extending the "terminally balanced" normal form beyond the Boolean semiring. (It is well known how the theory of power series over the Boolean semiring is isomorphic, in a well-defined sense, to language theory.)

The reader is referred to [11] for motivation and background material, as well as for all unexplained notions. We try to follow the notation of [11] whenever possible. In particular,

$$A^{\mathrm{alg}} \ll X^* \gg$$

denotes the family of all A-algebraic power series with variables in X. Throughout this paper, we assume that the semiring A is commutative.

Every series in $A^{\mathrm{alg}} \ll X^* \gg$ can be obtained as the first component of the solution of a proper algebraic system

$$z_i = p_i, \quad i = 1, \ldots, n. \tag{1}$$

Here $Z = \{z_1, \ldots, z_n\}$ is an alphabet disjoint from X, and p_i are polynomials in $A \langle (X \cup Z)^* \rangle$. Moreover, the coefficients of the empty word λ and those of terms z_j in each of the p_i are equal to 0. (This follows because the system (1) is proper.) In what follows, the alphabet Z will be referred to as the alphabet of *variables*.

In this paper we consider the problem of restricting the form of the polynomials p_i in (1) without affecting the family $A^{\mathrm{alg}} \ll X^* \gg$. For instance, the following result was established in [11, p. 128]:

doi:10.5186/aasfm.1985.1056

Lemma 1. *Every A-algebraic series with variables in X equals the first component in the solution for some algebraic system such that the supports of the right sides of the equations are included in the set*

$$X \cup X\bar{Z} \cup X\bar{Z}^2,$$

where $\bar{Z}$ denotes the set of variables in the system.

As regards normal forms for context-free grammars, the normal form obtained from Lemma 1 represents the starting point: all right-hand sides of productions begin with a terminal letter. The next step is "head-and-tail" normal form, [7]: all right-hand sides begin with and end in a terminal letter. The positioning of terminals in an arbitrary fashion was accomplished in [1] and [4]. Of course, [1] and [4] deal only with languages. Related results are contained in [8] and [10]. In the present paper, we establish analogous results for power series.

The positioning of terminals in an arbitrary fashion can be viewed as a "super" or "meta" normal form. In language theory, a very important strengthening of this normal form deals with the balancing of terminals, [6]. The completeness criterion of context-free grammar forms, [5], is based on this result. The extension of this result to algebraic power series will also be considered below. Essentially, the problem remains open.

While the reference [11] constitutes a sufficient background for understanding the results and proofs contained in this paper, the reader is referred to [2], [9] and especially to [3] for a broader spectrum of related results. The main purpose of this paper is to establish the following result.

Theorem 1. *Let k_1, k_2 and k_3 be nonnegative integers. Then every A-algebraic series over X^* (i.e., with variables in X) equals the first component in the solution for some algebraic system such that the supports of the right sides of the equations are included in the set*

$$X^+ \cup X^{k_1} Z X^{k_2} Z X^{k_3},$$

where Z is the alphabet of variables in the system.

2. A class of normal forms. For any choice of k_1, k_2 and k_3, Theorem 1 gives a normal form for algebraic systems defining A-algebraic power series. Thus, Theorem 1 constitutes a "super" normal form or a class of normal forms.

The proof of Theorem 1 will be given in this section. The proof is split into a sequence of lemmas.

We make some conventions, valid throughout this paper. The alphabet of variables in an algebraic system is denoted by Z, possibly provided with some indices. The underlying (commutative) semiring will always be denoted by A. The series will have variables in X, i.e., we consider series in $A^{\mathrm{alg}}\ll X^*\gg$. In connection with systems of equations, matrix notation is sometimes used in the natural

fashion. Although not explicitly stated, it is understood that the series defined by a system of equations is the first component of the solution of the system.

We begin with a modification of Lemma 1.

Lemma 2. *Every A-algebraic series is defined by a system of equations, where the supports of the right sides are included in the set*

$$X \cup X(X \cup Z)^* X.$$

Proof. Consider an arbitrary A-algebraic series, defined by a system of equations satisfying Lemma 1. We write the system in the matrix form

$$C_Z = MC_Z + C_X. \tag{2}$$

Here C_Z is the column vector, consisting of all letters in the alphabet Z of variables. The entries in the square matrix M are polynomials whose supports consist of words beginning with a letter of X. (Moreover, each word in the support belongs to X or XZ.) The entries of C_X are polynomials whose supports consist of (possibly several) letters of X.

The solution of (2), regarding C_Z as an unknown column vector, is

$$C_Z = M^* C_X. \tag{3}$$

(By our assumptions, the existence of M^* is obvious.) On the other hand, $M^+ = MM^*$ can be obtained as the unique solution T of the system

$$T = M + M^2 + MTM, \tag{4}$$

regarding T as an unknown matrix.

By (3), the equation (2) can now be replaced by (4) and (5), where (5) is given below:

$$C_Z = C_X + M^+ C_X = C_X + MC_X + MTC_X. \tag{5}$$

The right sides of the equations resulting from (5) are already of the form required. We still have to transform (4) into this form. For this purpose, the following construction is applied.

Consider an entry μ of M. Whenever a word in the support of μ ends in a letter z of Z, that particular occurrence of z is replaced by the right side of the defining equation for z, resulting from (5). Thereby, the distributive laws are applied and the coefficients positioned, after an eventual multiplication with original coefficients, in front of each term. This construction gives rise to a matrix M_1 such that the entries of M_1 are polynomials but the support of each entry consists of words that both begin with and end in a letter of X.

For instance, assume that A is the semiring of integers, $\mu = 3x_1 + 2x_1z_1 + 3x_2z_1$, and that the defining equation for z_1 resulting from (5) is as follows:

$$z_1 = 4x_2 - 2x_1 + 5z_1x_3 + 5z_1t_{14}x_3.$$

Then the entry corresponding to μ in M_1 is

$$\mu_1 = 3x_1+8x_1x_2-4x_1x_1+10x_1z_1x_3+10x_1z_1t_{14}x_3$$
$$+12x_2x_2-6x_2x_1+15x_2z_1x_3+15x_2z_1t_{14}x_3.$$

Here t_{14} refers to a variable in the matrix T. Of course, (4) gives rise to a system of equations whose cardinality equals the square of the dimension of T. When (4) is explicitly written in this fashion, a new variable t_{ij} has to be introduced for each entry of T.

An obvious induction based on (4) and (5) now shows that our original A-algebraic series is defined by the system of equations

$$C_Z = C_X+MC_X+MTC_X,$$
$$T = M_1+MM_1+MTM_1.$$

When the matrix notation is eliminated, the resulting system is in the form required in Lemma 2. □

Lemma 3. *For each integer* $m \geqq 1$, *every A-algebraic series is defined by a system of equations, where the supports of the right sides are included in the set*

$$X^+ \cup X^m(X \cup Z)^*X^m.$$

Proof. We begin with a system of equations satisfying Lemma 2. All occurrences of letters z of Z are replaced by the right side of the equation for z in the system. In this fashion, we obtain an equivalent system, where the supports of the right sides are included in the set

$$X^+ \cup X^2(X \cup Z)^*X^2.$$

(Observe that the alphabets X and Z are not affected.) By $m-1$ similar substitutions, we obtain the form of Lemma 3. □

The term X^+ in the union appearing in Lemma 3 can be replaced by finitely many powers of X. Their number depends on m. An analogous observation applies also below.

Lemma 4. *For each integer* $m \geqq 1$, *every A-algebraic series is defined by a system of equations, where the supports of the right sides are included in the set*

(6) $$X^+ \cup X^m Z^* X^m.$$

Proof. We now begin with a system of equations satisfying Lemma 3. The alphabet Z is extended by introducing a new letter z_x for each letter x of X. Consider some word w belonging to the support of the right side of some of the equations in our system such that w is not of the form required in Lemma 4. This means that w begins with and ends in m letters of X and contains, furthermore, at least one letter of Z and some additional occurrences of letters of X. The latter occurrences we

now replace, in each such word w, by the corresponding letters z_x. Finally, we add the equations $z_x = x$ to the end of the system.

It is obvious that the new system is equivalent to the original one, as regards the power series obtained as the first component of the solution. Moreover, the supports of the right sides are included in the set (6), where Z is understood as the extended alphabet described above. □

The next lemma is crucial in the construction. We now want to say more about the words in Z^* appearing in (6).

Lemma 5. *For each integer* $m \geqq 1$, *every A-algebraic series is defined by a system of equations, where the supports of the right sides are included in the set*

$$X^+ \cup X^m X^* Z X^m X^* \cup X^m X^* Z X^m X^* Z X^m X^* \cup X^m X^* Z X^m X^* Z X^m X^* Z X^m X^*.$$

Proof. Again, we begin with a system S of equations satisfying the previous lemma, this time Lemma 4. Let n be the length of the longest word over the alphabet Z, appearing in some word belonging to the support of some of the right sides of the equations in S. Such a number n exists because there are only finitely many words appearing in the supports.

We first modify the alphabet Z in such a way that the new alphabet will consist of all letters of the form $[z_1 \dots z_i]$, where $1 \leqq i \leqq n$ and the z's are letters of the Z-alphabet (not necessarily distinct) associated with the original system S. Intuitively, $[z_1 \dots z_i]$ corresponds to the product of the power series defined by the variables $z_1, \dots, z_i$. Thus $[z_j]$ behaves as the original z_j but it is notationally convenient to use brackets also in this case.

We now define a new system S' of equations by constructing the right side of the equation for $[z_1 \dots z_i]$, where $[z_1 \dots z_i]$ is an arbitrary one among the newly introduced letters. In the construction u and w, possibly provided with indices, denote words over X such that the length of each word w is at least m. Letters z (provided with indices) belong to the Z-alphabet associated with the original system S. Greek letters are elements of the semiring A. Finally, a, b, c and d are positive integers, and an "impossible" symbol

$$[z_a \dots z_b], \quad a > b,$$

is understood to be the empty word.

The right side of the equation for $[z_1]$ is the same as the right side of the equation for z_1 in S, except that all words over the Z-alphabet are bracketed. (For instance, if $3x_1 z_2' z_2' z_1' x_2$ is a term on the right side of the equation for z_1, the corresponding term in the equation for $[z_1]$ is $3x_1[z_2' z_2' z_1'] x_2$.)

The right side of the equation for $[z_1 \dots z_i]$, where $i \geqq 2$, consists of all terms obtained by the following three rules (i)—(iii).

(i) If $\alpha_j u_j$ is on the right side for z_j, $j = 1, \dots, i$, then the term $\alpha_1 \dots \alpha_i u_1 \dots u_i$ is obtained.

(ii) The term $\pi u_1 \dots u_{a-1} w_1 [z_1' \dots z_b'] w_2 u_{a+1} \dots u_i$ is obtained, whenever $\alpha_j u_j$ is on the right side for z_j, for all $j=1, \dots, i$ such that $j \neq a$, and $\beta w_1 z_1' \dots z_b' w_2$ is on the right side for z_a and, finally, π is the product of β and all α_j involved.
(iii) The term

$$\pi u_1 \dots u_{a-1} w_1 [z_1' \dots z_b'] w_2 [z_{a+1} \dots z_{d-1}] w_3 [z_1'' \dots z_c''] w_4 u_{d+1} \dots u_i$$

is obtained, whenever $\alpha_j u_j$ is on the right side for z_j, for all $j=1, \dots, i$ such that $j \neq a$ and $j \neq b$, and $\beta_1 w_1 z_1' \dots z_b' w_2$ (respectively $\beta_2 w_3 z_1'' \dots z_c'' w_4$) is on the right side for z_a (respectively z_b) and, finally, π is the product of the α's and β's involved.

Having defined the system S', we observe first that the right sides are of the form required in Lemma 5. (Observe that two letters from Z can be obtained from (iii) in the case where the brackets in the middle reduce to the empty word.)

Moreover, it is not difficult to establish the equivalence of S' and S. Indeed, (i) corresponds to the case where all the variables $z_1, \dots, z_i$ are terminated. Similarly, (ii) corresponds to the case where all the variables with the exception of one are terminated. Finally, (iii) corresponds to the case where at least two of the variables remain unterminated. □

Lemma 6. *For each integer $m \geqq 1$, every A-algebraic series is defined by a system of equations, where the supports of the right sides are included in the set*

$$X^+ \cup X^m X^* Z X^m X^* Z X^m X^*. \tag{7}$$

Proof. We begin with a system S of equations satisfying Lemma 5 but now we assume that S satisfies Lemma 5 for the constant $2m+1$. To establish Lemma 6, we have to replace S by an equivalent system S' where the supports of the right sides no longer contain words with one or three occurrences of letters of the Z-alphabet.

To eliminate words of the former type, we introduce a new variable z_x for each letter x in X, and add the equation $z_x = x$. (This construction was applied also in the proof of Lemma 4.) Consider now a term $\alpha u_1 z u_2$ appearing on the right side of some equation in S, where z is a letter of Z, and u_1 and u_2 are words of length at least $2m+1$ over X. We write u_1 in the form $u_1 = u_3 x u_4$, where u_3 is of length m and x is a letter of X. The original term is now replaced by the term $\alpha u_3 z_x u_4 z u_2$, and the same procedure is applied to all terms containing one occurrence of a letter of the Z-alphabet. Since in Lemma 6 we are dealing with the constant m rather than $2m+1$, the new terms are in accordance with Lemma 6.

To eliminate words of the latter type (i.e., containing three occurrences of letters of the Z-alphabet), we "pack" two Z-letters into one by the following procedure. Consider an "illegal" term $\alpha u_1 z_1 u_2 z_2 u_3 z_3 u_4$, where the notation is as before. Write u_1 and u_3 in the form

$$u_1 = u_5 u_6 \quad \text{and} \quad u_3 = u_7 u_8,$$

where u_6 and u_7 are words of length m. Introduce the new Z-letter $[u_6 z_1 u_2 z_2 u_7]$ and the equation

$$[u_6 z_1 u_2 z_2 u_7] = u_6 z_1 u_2 z_2 u_7 .$$

It is clear that only finitely many new Z-letters are needed when this procedure is repeated for all "illegal" terms. □

An analysis of the proofs of Lemmas 5 and 6 reveals the reason why the method applied yields only $X^m X^*$ instead of the accurate form X^m appearing in Lemma 4: the new words introduced may make the words altogether longer. Of course, the finiteness of the system and the finiteness of the supports guarantee that only finitely many powers of X can appear.

We are now in a position to establish Theorem 1. Assume that k_1, k_2 and k_3 are given. By Lemma 6, we may assume that the given A-algebraic series is defined by a system S of equations, where the supports of the right sides are included in the set (7). The idea is to choose m sufficiently large so that a suitable part of the words over X can be "packed" together with the Z-letters. For this purpose, the choice

$$m = 9 \max (k_1, k_2, k_3, 1) = 9m'$$

will suffice.

The Z-alphabet will be augmented by some letters defined below. We also modify the right sides of the equations for the original Z-letters to satisfy Theorem 1. The terms on the right sides remain unaltered whenever their support is in X^+.

Consider now a term $\alpha u_1 z_1 u_2 z_2 u_3$, resulting from the second part of the union (7). We write u_1, u_2 and u_3 as follows:

$$u_1 = w_1 w_2, \quad u_2 = w_3 w_4 w_5, \quad u_3 = w_6 w_7,$$

where the lengths of w_1, w_3, w_4 and w_7 are k_1, $4m'$, k_2 and k_3, respectively. (Clearly, this condition uniquely determines the words w_2, w_5 and w_6.) The original term is now replaced by the term

$$\alpha w_1 [w_2 z_1 w_3] w_4 [w_5 z_2 w_6] w_7, \tag{8}$$

where the bracketed letters are new elements of the Z-alphabet. Clearly, (8) is of the proper shape.

We still have to introduce equations for the bracketed letters. To satisfy Theorem 1, the supports of the right sides of these equations have to be of the proper shape. For this purpose, we still augment the Z-alphabet by two new types of letters. It will be obvious that the number of the new Z-letters is finite. In fact, an explicit upper bound, based on the system, can easily be given.

Observe that each of the words w_2, w_3, w_5, w_6 appearing in (8) is of length at least $4m'$. Therefore, we may write w_2 and w_3 in the form

$$w_2 = v_1 v_2 v_3 v_4, \quad w_3 = v_5 v_6 v_7 v_8,$$

where the lengths of v_1, v_3, v_4, v_5, v_7 and v_8 are k_1, k_2, k_1, k_2, k_3 and k_3, respectively.

We now introduce new Z-letters, denoted by brackets and braces as indicated below, as well as the following equations:

$$[w_2 z_1 w_3] = v_1[v_2]v_3\{v_4 z_1 v_5 v_6 v_7\}v_8,$$
$$\{v_4 z_1 v_5 v_6 v_7\} = v_4 z_1 v_5[v_6]v_7,$$
$$[v_2] = v_2, \quad [v_6] = v_6.$$

The second Z-letter $[w_5 z_2 w_6]$ is handled similarly. This concludes the proof of Theorem 1.

The following more general result is now easily obtained.

Theorem 2. *Assume that* $t \geqq 3$ *and that* $m_1, m_2, \ldots, m_t$ *are nonnegative integers. Then every A-algebraic series is defined by a system of equations, where the supports of the right sides are included in the set*

$$X^+ \cup X^{m_1} Z X^{m_2} \ldots Z X^{m_t}. \tag{9}$$

Proof. We define

$$k_1 = m_1 + \ldots + m_{t-2} + t - 3, \quad k_2 = m_{t-1}, \quad k_3 = m_t,$$

and apply Theorem 1. The resulting system of equations can be immediately transformed into the shape required in Theorem 2, by introducing new Z-letters z_x and equations $z_x = x$, similarly to the proof of Lemma 4. The letters z_x can be positioned in such a way that the supports corresponding to the transformed system of equations are contained in the set (9).

3. The terminally balanced case. In Theorems 1 and 2, any word over the "terminal" alphabet X may appear in the supports, i.e., there are no explicit restrictions as regards the first part of the union (9). It is clear that an explicit upper bound in terms of the k's or the m's can be given for the powers of X required. However, no good estimates for such an upper bound are known. Obviously, no upper bound independent of the k's or the m's exists.

Apart from an upper bound, also other types of restrictions may be imposed on the powers of X. A natural restriction, especially from the point of view of language theory, is customarily referred to as the "balancing of terminals". We say that a proper algebraic system defining a series r in $A^{\mathrm{alg}}\ll X^*\gg$ is *terminally balanced* if the supports of the right sides contain only those words over X whose length belongs to the length set of the support of r.

The basic motivation behind the definition above is language-theoretic: terminating productions with the length of the right side lying outside the length set of the language are unnatural because they are really not needed. On the other hand, the resulting terminally balanced super normal form, [6], leads to the completeness criterion of context-free grammar forms, [5]. Thus, if we want to exhaust all normal forms for context-free grammars, we must consider the balancing of terminals.

In the special case where the basic semiring A is chosen to be the Boolean semiring B, the next theorem can be established by rather obvious modifications of the methods applied in [6]. Therefore, its proof is omitted.

Theorem 3. *Assume that k_1, k_2 and k_3 are nonnegative integers. Then every series in $B^{alg}\ll X^*\gg$ is defined by a terminally balanced system of equations, where the supports of the right sides are included in the set*

$$X^+ \cup X^{k_1} Z X^{k_2} Z X^{k_3}.$$

A result analogous to Theorem 2 can be obtained also in the terminally balanced case for Boolean semirings. However, as pointed out in [6], this result is not a direct consequence of Theorem 3 in the same fashion as Theorem 2 is a direct consequence of Theorem 1.

4. Discussion and open problems. Many of the constructions presented above are modifications of the corresponding language-theoretic arguments. In general, every construction dealing with defining systems of equations for algebraic power series yields as a special case the corresponding construction for context-free grammars. On the other hand, the latter constructions cannot always be translated into the former ones.

A typical example is the construction, [6], yielding the terminally balanced (k_1, k_2, k_3) normal form for context-free grammars. This construction involves several arguments that seem to be inherently of language-theoretic nature. Therefore, the construction applies only to power series over the Boolean semiring.

There are several open problems as regards the extension of the construction given in [6] to more general semirings. We mention here only the most important one. Is the length set of the support of an A-algebraic series always almost periodic? This result holds true if A is the Boolean semiring but does it hold, for instance, if A is the semiring of integers? It is likely to hold for all positive semirings.

Also, as regards dealing with Theorems 1 and 2, some problems remain open. We mention the problem of deriving an upper bound, as good as possible, for the powers of X required when the triple (k_1, k_2, k_3) is given.

References

[1] BLATTNER, M., and S. GINSBURG: Position-restricted grammar forms and grammars. - Theoret. Comput. Sci. 17, 1982, 1—27.

[2] KUICH, W.: Formal power series, cycle-free automata and algebraic systems. - Institute für Informationsverarbeitung, TU Graz, Bericht F 103, 1982.

[3] KUICH, W., and A. SALOMAA: Semirings, automata, languages. - Springer-Verlag (in preparation).

[4] MAURER, H., A. SALOMAA, and D. WOOD: On generators and generative capacity of EOL forms. - Acta Inform. 13, 1980, 87—107.

[5] MAURER, H., A. SALOMAA, and D. WOOD: Completeness of context-free grammar forms. - J. Comput. System Sci. 23, 1981, 1—10.

[6] MAURER, H., A. SALOMAA, and D. WOOD: A supernormal-form theorem for context-free grammars. - J. Assoc. Comput. Mach. 30, 1983, 95—102.

[7] ROSENKRANTZ, D.: Matrix equations and normal forms for context-free grammars. - Ibid. 14, 1967, 501—507.

[8] ROZENBERG, G., and A. SALOMAA: The mathematical theory of *L* systems. - Academic Press, New York—London—Toronto—Sydney—San Fransisco, 1980.

[9] SALOMAA, A.: Formal power series in noncommuting variables. - 18th Scandinavian Congress of Mathematicians (Aarhus, 1980), Progress in Math. 11. Birkhäuser, Boston, Mass., 1981, 104—124.

[10] SALOMAA, A.: Jewels of formal language theory. - Computer Science Press, Rockville, Md., 1981.

[11] SALOMAA, A., and M. SOITTOLA: Automata-theoretic aspects of formal power series. - Texts and Monographs in Computer Science. Springer-Verlag, New York—Heidelberg. 1978.

University of Turku
Mathematics Department
SF-20500 Turku
Finland

Received 12 March 1984