Recent results on the semilinear formal power series

Ion Petre

Abstract

We continue in this paper the investigations on the notion of semilinearity for formal power series (in commuting variables), recently introduced in Petre, 1999. We prove several results connected to the difference operation on semilinear power series, as well as results on possible decompositions of semilinear series into finite sums of linear series with disjoint supports.

Résumé

Nous poursuivons dans cet article les investigations sur la notion de semilinéarité des séries formelles (en variables commutatives), récemment introduit par Petre, 1999. Nous prouvons plusieurs résultats concernant la différence de deux séries semilinéaires, ainsi que d'autres résultats portant sur de possibles décompositions de séries semilinéaires en sommes finies de séries linéaires ayant supports disjoints.

1 Introduction

The semilinearity is a central notion in the theory of formal languages, which has been considered only recently for formal power series. The family of semilinear formal power series has been introduced in [8] as a natural generalization of the notion of semilinear subsets of a commutative monoid (see [3]). As noticed already in [8], the semilinear power series (in commuting variables) have in general similar behavior as the semilinear languages over a commutative monoid: they are closed under rational operations (and thus coincide with the family of rational power series) if the coefficients are taken in an idempotent, commutative semiring, are closed under morphisms, and the well known Parikh's Theorem holds for an idempotent, commutative, ω-continuous semiring of coefficients (see [5]). There are however cases when the semilinear power series behave differently than the semilinear sets: the $\mathbb{N}$-semilinear power series over $\Sigma^{\oplus}$ (see Section 2 for notations) are not closed under rational operations and Hadamard product, and moreover, Parikh's Theorem does not hold in this case (we refer to [8] for more details).

It is proved in [3] that the family of semilinear subsets of a commutative monoid is closed under difference. Moreover, it is proved in [4] that the family of semilinear sets of vectors with nonnegative integers as components, is closed under difference, and the result is effectively computable. The problem seems to be more difficult for formal power series. To our knowledge, no result is known for the difference of two rational power series, in either commuting, or noncommuting variables. In [2], one considers on $\mathbb{N}$-rational series an operation of *quasi-difference* $r \dot{-} s$ defined as $(r \dot{-} s, x) = \max(0, (r, x) - (s, x))$. In the same paper it is proved that if r and s are rational series and s has bounded multiplicities, then the quasi-difference $r \dot{-} s$ is also rational. We prove in this paper a result of the same type: if r and s are semilinear power series (in commuting variables), s has bounded coefficients, and $r \geq s$, then $r - s$ is a rational series (of an "almost semilinear" form). However, the commutativity makes the problem completely different, and we prove our result by combinatorial means, avoiding the use of a deeply algebraic tool as the cross-section theorem in [2].

As it is well known (see, e.g., [4]), the semilinear sets of vectors with nonnegative integers as components can be written as finite unions of linear sets with linearly independent periods, are closed under intersection, and are closed under difference. While these properties are proved or believed to be false in general for semilinear power series, we prove that they hold for semilinear power series with bounded coefficients. These series can be decomposed in finite sums of linear series with disjoint supports, are closed under Hadamard product, and are closed under difference.

The paper is organized as follows. In Section 2, we fix the notations and recall the notions of semilinear sets, semilinear formal power series, and some of their basic properties. In Section 3, we consider the difference operation on semilinear power series. In Section 4, we prove the strong closure properties of the semilinear power series with bounded coefficients.

This paper extends the results in [9] and fills the gap in one of its proofs.

2 Preliminaries

For an alphabet Σ, we denote by $\Sigma^{\oplus}$ the *free commutative monoid* generated by Σ. If $\Sigma = \{a_1, \ldots, a_m\}$, then $\Sigma^{\oplus}$ is the direct product $a_1^* \times \ldots \times a_m^*$. In the sequel, we will call *word* any element of $\Sigma^{\oplus}$. We will denote the empty word by 1.

We recall that given a semiring K and a monoid M, the set of all formal power series over M, with coefficients in K, is denoted by $K\langle\langle M\rangle\rangle$. For a series r and a word u we denote by (r, u) the coefficient of u in r. A series r is called *proper* if $(r, 1) = 0$. The set $\{u \in \Sigma^{\oplus} \mid (r, u) \neq 0\}$ is called the support of r. The subset of $K\langle\langle M\rangle\rangle$ consisting of all series with finite support is denoted by $K\langle M\rangle$ and its elements are referred to as *polynomials*. A polynomial having a singleton as its support is called *monomial*.

For further definitions and results in the theory of formal power series, we refer to [6], [7], and [10].

We denote the set of nonnegative integers by $\mathbb{N}$.

For a formal power series $r \in \mathbb{N}\langle\langle\Sigma^\oplus\rangle\rangle$, we say that r has *bounded coefficients* if there is $K \in \mathbb{N}$ such that for all $u \in \Sigma^\oplus$, $(r,u) < K$. For two formal power series $r, s \in \mathbb{N}\langle\langle\Sigma^\oplus\rangle\rangle$, we say that r is smaller than s, and write $r \leq s$, if for all $u \in \Sigma^\oplus$, $(r,u) \leq (s,u)$.

In [3], Eilenberg and Schützenberger introduced the notion of semilinearity in a commutative monoid $(M, \cdot)$ as follows. A subset X of M,

$$X = aB^*,$$

with $a \in M$ and B a finite subset of M, is called a *linear* set. One can easily notice that if $B = \{b_1, \ldots, b_r\}$, then

$$X = \{ab_1^{n_1} \ldots b_r^{n_r} \mid n_1, \ldots, n_r \in \mathbb{N}\}.$$

A finite union of linear sets is called a *semilinear* set.

The semilinear formal power series are defined in a similar way in [8]. Let $(M, \cdot)$ be a *commutative monoid* and $(K, +, \cdot)$ a semiring. A formal power series $r \in K\langle\langle M\rangle\rangle$ is called a *linear series* if

$$r = pq^*,$$

with p a monomial in $K\langle M\rangle$, and q a proper polynomial in $K\langle M\rangle$. A finite sum of linear series is called a *semilinear series*.

Note that if we consider series with coefficients in the boolean semiring B, and we identify the subsets of M with their characteristic series, we obtain the old notion of a semilinear set.

Throughout this paper, we will always consider series over $\Sigma^\oplus$, for an arbitrary alphabet Σ, with coefficients in $\mathbb{N}$. We denote the family of such formal power series by $\mathbb{N}\langle\langle\Sigma^\oplus\rangle\rangle$.

The main emphasis in this paper will be on the semilinear power series with bounded coefficients and on their properties. We describe in the next theorem their general form.

Theorem 1. *Let s be a semilinear series. If s has bounded coefficients, then $s = m_1u_1^* + \ldots + m_ku_k^*$, for some monomials $m_1, \ldots, m_k$ and some words $u_1, \ldots, u_k$.*

Proof. If s is a semilinear series, then it is of the form

$$s = m_1p_1^* + \ldots + m_kp_k^*,$$

for some monomials $m_1, \ldots, m_k$, and some polynomials $p_1, \ldots, p_k$. Consider the polynomial $p_1 = \alpha_1u_1 + \ldots + \alpha_iu_i$, with $\alpha_1, \ldots, \alpha_i \in \mathbb{N}\backslash\{0\}$, and $u_1, \ldots, u_i \in \Sigma^\oplus$. All the coefficients $\alpha_1, \ldots, \alpha_i$ must be equal to 1 since otherwise s does not have bounded coefficients. Moreover, if p_1 is not a monomial, i.e. $i \geq 2$, then $(p_1, u_1^{j_1}u_2^{j_2}) \geq \binom{j_1+j_2}{j_1}$, and this is not bounded for $j_1, j_2 \geq 0$. Consequently, $p_1 = u_1$, for some $u_1 \in \Sigma^\oplus$. Similarly, one can prove that $p_i = u_i$, for some $u_i \in \Sigma^\oplus$, for all $2 \leq i \leq k$. ∎

One important tool used often throughout the paper is provided by the following decomposition result.

Lemma 2. *Let* $r = pq^*$ *be a linear series, for a monomial* p *and a proper polynomial* q*. For any* $v \in \Sigma^{\oplus}$*, if* $(r, v) > 0$*, then* $r = vq^* + p'q^* + p''$*, for some polynomials* p' *and* p''*.*

Proof. If $(r, v) > 0$, then $v = v_1 v_2$, with $(p, v_1) > 0$, and $(q^*, v_2) > 0$, i.e., $(q^m, v_2) > 0$, for some $m \in \mathbb{N}$. Then

$$\begin{aligned} r &= p(1 + q + \ldots + q^{m-1} + q^m q^*) = \\ &= p(1 + q + \ldots + q^{m-1}) + vq^* + (pq^m - v)q^*. \end{aligned}$$

■

In computing the difference between a semilinear series and a semilinear series with bounded coefficients, we will be typically interested in knowing how the words of the form vu^n, $n \in \mathbb{N}$, are generated by a linear series, for some $u, v \in \Sigma^{\oplus}$. To this aim, we define the notion of base of a linear series, and we prove the following result.

Let $r = pq^*$ be a linear series, with p a monomial, and q a proper polynomial, $p = \alpha_0 u_0$, $q = \alpha_1 u_1 + \ldots + \alpha_m u_m$, for some $\alpha_i \in \mathbb{N} \setminus \{0\}$, $u_i \in \Sigma^{\oplus}$, for all $0 \le i \le m$. Let us collect in the vector ω all the words appearing as monomials in q:

$$\omega = (u_1, \ldots, u_m).$$

We call the vector ω *the base* of the series r.

To any word $v \in \Sigma^{\oplus}$, we can associate with respect to ω (or r) a finite set of vectors from $\mathbb{N}^m$ in the following way: if

$$v = u_0 u_1^{n_1} \ldots u_m^{n_m},$$

then the vector $t = (n_1, \ldots, n_m)$ is associated to v. We denote $v = u_0 \omega^t$ and say that t is a *representation of* v *with respect to the series* r. We denote by $\mathcal{R}_r(v)$ the set of representations of v with respect to r.

Note that in general, a word v can have multiple representations with respect to a given linear series r, and that $\mathcal{R}_r(v) \neq \emptyset$ if and only if $(r, v) > 0$.

Example 1. *For* $r = (a + a^2)^*$*, its base is* $\omega = (a, a^2)$*. Moreover,*

$$\mathcal{R}_r(a^4) = \{(4, 0), (2, 1), (0, 2)\} \text{ and } \mathcal{R}_r(a^m b^n) = \emptyset, \text{ for all } m \ge 0 \text{ and } n \ge 1$$

Lemma 3. *Let* $r = pq^*$ *be a linear series, for a monomial* p *and a proper polynomial* q*, and let* $u, v \in \Sigma^{\oplus}$ *such that* $(r, vu^n) > 0$ *for some* $n \ge 0$*. Then*

$$r = (vu^{n_1} + \ldots + vu^{n_k})q^* + p'q^* + p'',$$

for some polynomials p'*,* p'' *and some* $n_1, \ldots, n_k \in \mathbb{N}$*, such that for all* $n \in \mathbb{N}$*,* $(r, vu^n) > 0$ *if and only if* $(vu^{n_i} q^*, vu^n) > 0$*, for some* $1 \le i \le k$*.*

Proof. If $q = \alpha_1 u_1 + \ldots + \alpha_m u_m$, with $\alpha_i \in \mathbb{N} \setminus \{0\}$, $u_i \in \Sigma^{\oplus}$, then the base of r is the vector $\omega = (u_1, \ldots, u_m)$.

Let $\mathcal{R} = \bigcup_{n \geq 0} \mathcal{R}_r(vu^n)$. By König's Lemma (see e.g. [4]), there is only a finite set of minimal vectors in $\mathcal{R}$. Let $\{t_1, \ldots, t_k\}$ be this set, $t_i = (t_{i1}, \ldots, t_{im})$, and let vu^{n_i} be the *unique* word such that $t_i \in \mathcal{R}_r(vu^{n_i})$, for all $1 \leq i \leq k$. Let $t_{i0} = t_{i1} + \ldots + t_{ik}$ for all $1 \leq i \leq k$, and let us assume that $t_{i0} \leq t_{j0}$, for $i \leq j$. Then $(pq^{t_{10}}, vu^{n_1}) > 0$ and

$$r = p(1 + q + \ldots + q^{t_{10}-1} + q^{t_{10}} q^*) = p' + vu^{n_1} q^* + p'' q^*,$$

where $p' = p(1 + q + \ldots + q^{t_{10}-1})$ and $p'' = q^{t_{10}} - vu^{n_1}$.

Note now that for all $2 \leq i \leq k$, $(p'', vu^{n_i}) > 0$. Indeed, if this were not true, then for all $\tau_i \in \mathcal{R}_\omega(vu^{n_i})$, we would have $t_1 \leq \tau_i$, which is impossible since t_1 and t_i are incomparable for all $2 \leq i \leq k$. We continue in the same way for all $2 \leq i \leq k$, with $p''q^*$ instead of pq^*, to obtain the claim. ∎

Using Lemma 3, we can prove now that for any semilinear series r and any word u, the set of powers of u included in the support of r is semilinear. We recall first a well known number theoretical result.

Lemma 4. *Let $n_1, \ldots, n_k$ be positive integers, and let d be their greatest common divisor. For any large enough integer n, n can be written as a linear combinations of $n_1 \ldots, n_k$ if and only if n is a multiple of d.*

Theorem 5. *For any semilinear series r and $u \in \Sigma^{\oplus}$, the set $P_{r,u} = \{n \in \mathbb{N} \mid (r, u^n) > 0\}$ is a semilinear set of nonnegative integers.*

Proof. Clearly, $P_{r_1+r_2,u} = P_{r_1,u} \cup P_{r_2,u}$ and so, it is enough to prove the claim for a linear series. Assume thus that $r = pq^*$, for a monomial p and a proper polynomial q. By Lemma 3, $r = (u^{n_1} + \ldots + u^{n_k})q^* + p'q^* + p''$, for some $n_1, \ldots, n_k \in \mathbb{N}$ and some polynomials p', p'' such that $(r, u^n) > 0$ iff $(u^{n_i} q^*, u^n) > 0$ for some $1 \leq i \leq k$. Thus, it is enough to prove the claim of the theorem for the series q^*.

Let us assume now that $r = q^*$. By Lemma 3, $r = (u^{n_1} + \ldots + u^{n_k})q^* + p'q^* + p''$, for some $n_1, \ldots, n_k \in \mathbb{N}$ and some polynomials p', p'' such that $(r, u^n) > 0$ iff $(u^{n_i} q^*, u^n) > 0$ for some $1 \leq i \leq k$. In other words, for any $n \in \mathbb{N}$, we have $(r, u^n) > 0$ iff n is a linear combination of $n_1, \ldots, n_k$. The claim follows now by Lemma 4. ∎

3 The difference operation on semilinear series

A basic property of the semilinear subsets of $\mathbb{N}$ is that they are closed under difference:

Lemma 6. *The family of semilinear sets of nonnegative integers is closed under difference.*

There are very few known similar results for formal power series. One of them, for power series in noncommuting variables, is Eilenberg's Theorem ([2]): if r and s are $\mathbb{N}$-rational series in noncommuting variables, and s has bounded coefficients, then their quasi-difference $r \dot{-} s$ is an $\mathbb{N}$-rational series. Nothing seems to be known on the difference of rational formal power series in commuting variables.

For semilinear $\mathbb{N}$-power series, it is an open and apparently difficult problem whether or not they are closed under difference. The smaller family of $\mathbb{N}$-linear series is not closed under this operations, as it can be proved by considering the difference $(2a)^* - a^*$. Note however, that this is a rational series of star height 1, namely $aa^*(2a)^*$ (as one can compute using the techniques developed in this section, see also [1]).

We prove in this section that for two $\mathbb{N}$-semilinear formal power series r and s, such that s has bounded coefficients and $r \geq s$, the difference $r - s$ is an $\mathbb{N}$-rational power series. Moreover, we will describe the precise form of such a difference, proving that in fact, it is very closed to a semilinear form. We will first solve the problem in several simpler instances, to which we will then reduce the general problem.

Lemma 7. *Let $r = pq^*$ be a semilinear series for a polynomial p and a proper polynomial q, and let $s = vu^*$, for some $u, v \in \Sigma^{\oplus}$. If $r \geq s$, then $r - s$ is a rational series. Moreover, $r - s$ is of the form $p_1 + p_2q^* + p_3(u^d)^* + p_4q^*(u^d)^*$, with p_1, p_2, p_3, p_4 polynomials and $d \in \mathbb{N}$, such that $q^* \geq (u^d)^*$.*

Proof. By Lemma 3, we can assume without loss of generality that $v = 1$ and $p = u^{l_1} + \ldots + u^{l_k}$, for some nonnegative integers $k, l_1, \ldots, l_k$. Thus, it is enough to prove the claim for $r = (u^{l_1} + \ldots + u^{l_k})q^*$ and $s = u^*$.

Clearly, there must be an integer $n > 0$ such that $(q^*, u^n) > 0$ and let d be the minimal such integer. If $r \geq s$, then $(r, u^i) > 0$, for all $0 \leq i < d$ and thus, we must have $\{0, 1, \ldots, d-1\} \subseteq \{l_1, \ldots, l_k\}$. Since $\sum_{i=0}^{d-1} u^i q^* \geq u^*$, it is enough to assume that $r = (1 + u + \ldots + u^{d-1})q^*$ and notice that $q^* \geq (u^d)^*$. Since $u^* = (1 + u + \ldots + u^{d-1})(u^d)^*$, $r - s = \sum_{i=0}^{d-1} u^i(q^* - (u^d)^*)$. It is enough to prove now that $q^* - (u^d)^*$ is a rational series.

By Lemma 2, if $q^* \geq (u^d)^*$, then $q^* = 1 + u^d q^* + p_1 q^* + p_2$, for some polynomials p_1 and p_2. Consequently,

$$
\begin{aligned}
q^* - (u^d)^* &= (1 + u^d q^* + p_1 q^* + p_2) - (1 + u^d (u^d)^*) = \\
&= u^d (q^* - (u^d)^*) + p_1 q^* + p_2 = p_1 q^* (u^d)^* + p_2 (u^d)^*,
\end{aligned}
$$

which is a rational series of the form specified in the lemma. ∎

Lemma 8. *Let $r = pq^*u^*$ be a rational series, for a polynomial p, a proper polynomial q, and a word $u \in \Sigma^{\oplus}$, such that $q^* \geq u^*$, and let $v \in \Sigma^{\oplus}$. If $(r, v) > 0$, then also $(pq^*, v) > 0$.*

Proof. If $(r, v) > 0$, then $v = u_1 \cdot u_2 \cdot u^l$, for some $l \in \mathbb{N}$, $u_1, u_2 \in \Sigma^{\oplus}$, such that $(p, u_1) > 0$ and $(q^*, u_2) > 0$. But $q^* \geq u^*$, i.e. $(q^*, u^l) > 0$. Thus, by Lemma 2, $q^* = u^l q^* + p_1 q^* + p_2$, for some polynomials p_1, p_2. Consequently, $(q^*, u^l u_2) > 0$ and so, $(pq^*, v) > 0$. ∎

Lemma 9. *Let r be a semilinear series and $s = vu^*$ for some $u, v \in \Sigma^{\oplus}$. If $r \geq s$, then $r - s$ is a rational series.*

Proof. Similarly as in the proof of Lemma 7, one can reduce the problem to the case when $v = 1$: $r - s = v(r' - u^*) + r''$, where r', r'' are semilinear series and $r' \geq u^*$. Assume thus that $v = 1$, and let $r = p_1 q_1^* + \ldots + p_m q_m^*$, for some polynomials p_i and q_i, for all $1 \leq i \leq m$.

By Theorem 5, the set $P_i = \{n \in \mathbb{N} \mid (p_i q_i^*, u^n) > 0\}$ is semilinear. But then, by Lemma 6, the sets

$$P_i' = P_i - \bigcup_{j=1}^{i-1} P_j$$

are also semilinear sets of nonnegative integers, for all $1 \leq i \leq m$, and $\cup_{i=1}^m P_i' = \cup_{i=1}^m P_i = \mathbb{N}$. The difference $r - s$ can thus be written as

$$r - s = \sum_{i=1}^{m} (p_i q_i^* - \sum_{n \in P_i'} u^n).$$

Note that $\sum_{n \in P_i'} u^n$ is a semilinear series, as P_i' is a semilinear set of nonnegative integers, i.e. $\sum_{n \in P_i'} u^n = u^{l_{i1}}(u^{d_{i1}})^* + \ldots + u^{l_{ik_i}}(u^{d_{ik_i}})^* + p$, for some $l_{ij}, d_{ij} \in \mathbb{N}$, and any polynomial p. It is enough to prove the result for $p = 0$: in general, one can decrease the polynomial p at the end, using the decomposition given by Lemma 2.

Observe now that $p_i q_i^* \geq \sum_{n \in P_i'} u^n$ since $P_i' \subseteq \{n \in \mathbb{N} \mid (p_i q_i^*, u^n) > 0\}$. Thus, we reduced the problem to computing a difference of the form

$$pq^* - (u^{l_1}(u^{d_1})^* + \ldots + u^{l_k}(u^{d_k})^*).$$

Let $r_i = pq^* - (u^{l_1}(u^{d_1})^* + \ldots + u^{l_i}(u^{d_i})^*)$, for all $1 \leq i \leq k$.

By Lemma 7, $pq^* - u^{l_1}(u^{d_1})^* = p_1 + p_2 q^* + p_3 q^*(u^{t_1})^*$, for some polynomials p_1, p_2, p_3 and a nonnegative integer t_1, such that $q^* \geq (u^{t_1})^*$.

Since $r_1 \geq (u^{l_2}(u^{d_2})^* + \ldots + u^{l_k}(u^{d_k})^*)$, we have by Lemma 8 that also $p_1 + p_2 q^* + p_3 q^* \geq (u^{l_2}(u^{d_2})^* + \ldots + u^{l_k}(u^{d_k})^*)$. Moreover,

$$r_1 = p_1 + (p_2 + p_3) q^* + p_3 u^{t_1} q^* (u^{t_1})^*,$$

and

$$r_2 = r_1 - u^{l_2}(u^{d_2})^* = (p_1 + (p_2 + p_3) q^* - u^{l_2}(u^{d_2})^*) + p_3 u^{t_1} q^* (u^{t_1})^*.$$

Continuing in the same way, we obtain that indeed $r - s$ is a rational series of the form

$$r - s = p_0' + \sum_{i=1}^{m} p_i'(q_i')^* + \sum_{i=m+1}^{m+n} p_i'(q_i')^*(u^{d_i})^*,$$

for some $m, n, d_{m+1}, \ldots, d_{m+n} \in \mathbb{N}$ and for some polynomials p_0', p_i', q_i', $1 \leq i \leq m+n$, such that $q_i'^* \geq (u^{d_i})^*$, for all $m+1 \leq i \leq m+n$. ∎

Using the above partial results, we can solve now the general problem.

Theorem 10. *Let r and s be semilinear series. If s has bounded coefficients and $r \geq s$, then $r - s$ is a rational series.*

Proof. If s has bounded coefficients, then by Theorem 1, s is of the form

$$s = v_1 u_1^* + \ldots + v_n u_n^*,$$

where $u_i, v_i \in \Sigma^\oplus$, for all $1 \le i \le m$. We prove the claim by induction on n.

If $n = 1$, then $r - s$ is rational by Lemma 9. Assume now that the theorem holds for series s with up to $n-1$ linear components, for some $n > 1$, and let us prove it for n. By Lemma 9, $r_1 = r - v_1 u_1^*$ is of the form

$$r_1 = p_0 + \sum_{i=1}^{m} p_i (q_i)^* + \sum_{i=m+1}^{m+n} p_i (q_i)^* (u^{d_i})^*,$$

for some $m, n, d_{m+1}, \ldots, d_{m+n} \in \mathbb{N}$ and some polynomials p_0, p_i, q_i, $1 \le i \le m+n$, such that $q_i^* \ge (u^{d_i})^*$, for all $m+1 \le i \le m+n$. But then, as $r_1 \ge v_2 u_2^* + \ldots + v_n u_n^*$, we obtain by Lemma 8 that for $r_1' = p_0 + \sum_{i=1}^{m+n} p_i q_i^*$, we also have $r_1' \ge v_2 u_2^* + \ldots + v_n u_n^*$. By the induction hypothesis, $r_1' - (v_2 u_2^* + \ldots + v_n u_n^*)$ is a rational series and thus,

$$\begin{aligned} r - s = r_1 - (v_2 u_2^* + \ldots + v_n u_n^*) = \\ = r_1' - (v_2 u_2^* + \ldots + v_n u_n^*) + \sum_{i=m+1}^{m+n} p_i u_1^{d_i} q_i^* (u_1^{d_i})^* \end{aligned}$$

is itself rational, proving the claim of the theorem. ∎

Example 2. *Let $r = (a+b)^*$ and $s = a^* + bb^*$. Observe that both r and s are semilinear series, s has bounded coefficients, and $r \ge s$. Then, according to Theorem 10, the difference $r - s$ is computed as follows.*

Let $s_1 = a^*$. We first compute $r - s_1$:

$$\begin{aligned} r - s_1 = (a+b)^* - a^* = 1 + a(a+b)^* + b(a+b)^* - 1 - aa^* = \\ = a\left((a+b)^* - a^*\right) + b(a+b)^* = b(a+b)^* a^*. \end{aligned}$$

The difference $r - s$ is then computed as follows:

$$\begin{aligned} r - s = b\left((a+b)^* a^* - b^*\right) = b\left(((a+b)^* - b^*) + a(a+b)^* a^*\right) = \\ = b\left(a(a+b)^* b^* + a(a+b)^* a^*\right) = ab(a+b)^* b^* + ab(a+b)^* a^*. \end{aligned}$$

4 Decomposition of semilinear series

In this section, we take a closer look at the family of semilinear power series with bounded coefficients. We prove that this family has very strong closure properties, similar to the family of semilinear subsets of $\mathbb{N}^n$. We thus prove that bounded multiplicities do not affect essentially the closure properties of the semilinear sets.

As it well known (see, e.g., [4]), any semilinear subset of $\mathbb{N}^n$ can be written as a finite union of linear sets, each of which has linearly independent periods. Furthermore, they are closed under the operation of complementation, intersection, and difference.

We prove in this section that any $\mathbb{N}$-semilinear power series with bounded coefficients can be written as a sum of linear series with disjoint supports. Using similar arguments, one can also prove their closure under Hadamard product and under difference.

Theorem 11. *Any semilinear series with bounded coefficients can be written as a sum of linear series with disjoint supports.*

Proof. Let r and r' be two semilinear series of the form $r = pw^*$, $r' = p'w'^*$, with $w, w' \in \Sigma^\oplus$ and p, p' polynomials, such that for any two monomials μ_1, μ_2 of p, $\mu_1 w^*$ and $\mu_2 w^*$ have disjoint supports, and for any two monomials μ_1' and μ_2' of p', $\mu_1' w'^*$ and $\mu_2' w'^*$ have disjoint supports.

If the intersection of the supports of r and r' is finite (or empty), then it is not difficult by Lemma 2 to decompose r and r' as $r = r_1 + p$, $r' = r_1' + p'$, for some polynomials p, p' and some semilinear series r_1, r_1' with disjoint supports, of the same form as r and r'.

Assume now that the intersection of the supports of r and r' is an infinite set of words. For the sake of simplicity, we will assume that we have a two letter alphabet $\Sigma = \{a, b\}$. The general case can be treated in a similar way.

If r and r' have bounded coefficients, then by Theorem 1,

$$r = (u_1^{l_1} + \ldots + u_k^{l_k})w^* \quad \text{and} \quad r' = (v_1^{l_1'} + \ldots + v_{k'}^{l_{k'}'})w'^*,$$

for some words $u_1, \ldots, u_k, w, v_1, \ldots, v_{k'}, w' \in \Sigma^\oplus$, and some nonnegative integers $k, k', l_1, \ldots, l_k, l_1', \ldots, l_{k'}'$. For any $1 \le i \le k$, there are $t_{i1}, t_{i2} \in \mathbb{N}$, such that $u_i = a^{t_{i1}} b^{t_{i2}}$, and for any $1 \le j \le k'$, there are $t_{j1}', t_{j2}' \in \mathbb{N}$, such that $v_j = a^{t_{j1}'} b^{t_{j2}'}$. Also, $w = a^{s_1} b^{s_2}$, $w' = a^{s_1'} b^{s_2'}$, for some $s_1, s_2, s_1', s_2' \in \mathbb{N}$. To compute the intersection of the supports of r and r', one has to compute the solution of the following linear system of equations in the nonnegative integers m and n:

$$\begin{cases} t_{i1} + m \cdot s_1 = t_{j1}' + n \cdot s_1' \\ t_{i2} + m \cdot s_2 = t_{j2}' + n \cdot s_2', \end{cases}$$

for $1 \le i \le k$ and $1 \le j \le k'$. As it is well known, the solution of this system is either the empty set, or it is a linear set of pairs of nonnegative integers $\{(l_{ij}^{(0)} + l_{ij}^{(1)} p, l_{ij}^{(3)} + l_{ij}^{(4)} p) \mid p \in \mathbb{N}\}$. It is straightforward to see that in fact $l_{ij}^{(1)}$ and $l_{ij}^{(3)}$ depend only on s_1, s_2, s_1' and s_2', and not on i and j: $l_{ij}^{(1)} = l_1$, $l_{ij}^{(3)} = l_3$. Using the fact that for any $t \in \Sigma^\oplus$ and any $k \in \mathbb{N}$,

$$t^* = (1 + t + \ldots + t^{k-1})(t^k)^*,$$

one can thus derive that

$$r = (u_1^{l_1} + \ldots + u_k^{l_k})(1 + w + \ldots + w^{l_1 - 1})(w^{l_1})^* = p_1(w^{l_1})^* + p_2(w^{l_1})^*,$$

and

$$r' = (v_1^{l_1'} + \ldots + v_{k'}^{l_{k'}'})(1 + w' + \ldots + w'^{l_1' - 1})(w'^{l_1'})^* = p_3(w'^{l_1'})^* + p_4(w'^{l_1'})^*,$$

for some polynomials p_1, p_2, p_3, p_4, where $p_2(w^{l_1})^* = p_4(w'^{l_1'})^*$, and $p_1(w^{l_1})^*$, $p_2(w^{l_1})^*$, $p_3(w'^{l_1'})^*$ have disjoint supports. Thus,

$$r + r' = (p_1 + 2p_2)(w^{l_1})^* + p_3(w'^{l_1'})^*.$$

Let r be now an arbitrary semilinear series with bounded coefficients, $r = p_1w_1^* + p_2w_2^* + \ldots + p_nw_n^*$, where p_i is polynomial and $w_i \in \Sigma^\oplus$, for all $1 \leq i \leq n$. Assuming that we have decomposed $r_i = p_1w_1^* + \ldots + p_iw_i^*$, $i \geq 2$ in a sum of linear series with disjoint supports,

$$r_i = q_1u_1^* + \ldots + q_{k_i}u_{k_i}^* + q_0,$$

for some polynomials $q_0, \ldots, q_{k_i}$ and some words $u_1, \ldots, u_{k_i}$, we continue as follows. We first compute the intersection of the supports of $p_{i+1}w_{i+1}^*$ and obtain as above two series with disjoint supports, of which, one is added to $q_1u_1^*$ and the other, say $s_{i+1,1}$ has disjoint support with $q_1u_1^*$. Then we compute the intersection of the supports of $s_{i+1,1}$ and $q_2u_2^*$, etc. ■

Example 3. *Let $r_1 = a(a^2)^*$, $r_2 = a^3(a^4)^*$, and $r_3 = a^2(a^5)^*$, The semilinear series $r = r_1 + r_2 + r_3$ can be written as a sum of linear series with disjoint supports as follows. We first solve the following equation in the nonnegative integers m and n:*

$$2m + 1 = 4n + 3.$$

Its solution is $\{(2p+1, p) \mid p \in \mathbb{N}\}$. Thus,

$$r_1 + r_2 = a(1 + a^2)(a^4)^* + a^3(a^4)^* = a(a^4)^* + 2a^3(a^4)^*.$$

Then, we solve the following two equations in the nonnegative integers m and n:

$$4m + 1 = 5n + 2 \quad \text{and} \quad 4m + 3 = 5n + 2.$$

Their solutions are the sets $\{(5p+4, 4p+3) \mid p \in \mathbb{N}\}$ and $\{(5p+1, 4p+1) \mid p \in \mathbb{N}\}$, respectively. Thus,

$$\begin{aligned} r = \; & (a + 2a^3)(a^4)^* + a^2(a^5)^* = \\ = \; & (a + 2a^3)(1 + a^4 + a^8 + a^{12} + a^{16})(a^{20})^* + \\ & + a^2(1 + a^5 + a^{10} + a^{15})(a^{20})^* = \\ = \; & (a + a^2 + 2a^3 + a^5 + 3a^7 + a^9 + 2a^{11} + a^{12} + a^{13} + 2a^{15} + 2a^{17})(a^{20})^*. \end{aligned}$$

In the proof of the theorem 11, we have proved essentially that for two linear power series with bounded coefficients, the intersection of their supports is semilinear. Using this argument, one can derive now the closure under Hadamard product.

Theorem 12. *The family of $\mathbb{N}$-semilinear power series with bounded coefficients is closed under Hadamard product.*

Furthermore, it is easy to prove the closure under difference using the techniques in the proof of Theorem 10. The main argument here is that if $r = u_1^*$, $s = u_2^*$, for some $u_1, u_2 \in \Sigma^\oplus$, and $r \geq s$, then $r - s$ is semilinear. Indeed, if $r \geq s$, then there must be $k \in \mathbb{N}$ such that $u_2 = u_1^k$, and thus, $r - s = (1 + u_1 + \ldots + u_1^{k-1})(u_1^k)^* - u_2^* = (u_1 + \ldots + u_1^{k-1})(u_1^k)^*$, a semilinear series (with bounded coefficients). We omit the details here, as they are very similar to the details of Theorem 10.

References

[1] F. Bassino, Nonnegative companion matrices and star-height of $\mathbb{N}$-rational series, Theoret. Comput. Sci. 180 : 61–80, 1997.

[2] S. Eilenberg, *Automata, Languages and Machines*, Academic Press, 1974.

[3] S. Eilenberg, M.P.Schützenberger, Rational sets in commutative monoids, J. Algebra 13, 173-191, 1969.

[4] S. Ginsburg, *The Mathematical Theory of Context-free Languages* , McGraw-Hill Book Company, 1966.

[5] W. Kuich, The Kleene and the Parikh theorem in complete semirings, *Automata, Languages and Programming* (Karlsruhe 1987), 211–215, Lecture Notes in Computer Science, Springer-Verlag 1987.

[6] W. Kuich, *Semirings and formal power series: their relevance to formal languages and automata* in *Handbook of Formal Languages,* (G. Rozenberg, A.Salomaa, eds.), Springer-Verlag, Berlin, Heidelberg,609-678, 1997.

[7] W. Kuich, A. Salomaa, *Semirings, Automata, Languages*, Springer-Verlag, 1986.

[8] I. Petre, On semilinearity in formal power series, in *Proceedings of DLT 1999*, Developments in Language Theory: foundations, applications, and perspectives (G.Rozenberg, W.Thomas, eds.), World Scientific, Singapore: 220–231, 2000.

[9] I. Petre, On the difference problem for semilinear power series, in Grammars and Automata for String Processing: from *Mathematics and Computer Science to Biology*, and Back (V. Mitrana, C. Martin-Vide, eds.), Gordon and Breach, London, 2001.

[10] A. Salomaa, M. Soittola, *Automata-theoretic Aspects of Formal Power Series*, Springer-Verlag, 1978.

Turku Centre for Computer Science (TUCS) and
Department of Mathematics, University of Turku
Turku 20520, Finland
E-mail: `ipetre@cs.utu.fi`