Polynomial Operators on Classes of Regular Languages

Ondřej Klíma and Libor Polák*

Department of Mathematics and Statistics, Masaryk University
Kotlářská 2, 611 37 Brno, Czech Republic

Abstract. We assign to each positive variety $\mathcal{V}$ and each natural number k the class of all (positive) Boolean combinations of the restricted polynomials, i.e. the languages of the form $L_0a_1L_1a_2\dots a_\ell L_\ell$, where $\ell \leq k$, $a_1,\dots,a_\ell$ are letters and $L_0,\dots,L_\ell$ are languages from the variety $\mathcal{V}$. For this polynomial operator we give a certain algebraic counterpart which works with identities satisfied by syntactic (ordered) monoids of languages considered. We also characterize the property that a variety of languages is generated by a finite number of languages. We apply our constructions to particular examples of varieties of languages which are crucial for a certain famous open problem concerning concatenation hierarchies.

Keywords: positive varieties of languages, polynomial operators.

2000 Classification: 68Q45 Formal languages and automata.

1 Introduction

The polynomial operator assigns to each positive variety of languages $\mathcal{V}$ the class of all (positive) Boolean combinations of the languages of the form

$$L_0a_1L_1a_2\dots a_\ell L_\ell \ , \qquad (*)$$

where A is an alphabet, $a_1,\dots,a_\ell \in A$, $L_0,\dots,L_\ell \in \mathcal{V}(A)$ (i.e. they are over A). Such operators on classes of languages lead to several concatenation hierarchies. Well-known cases are the Straubing-Thérien and the group hierarchies. Concatenation hierarchies has been intensively studied by many authors – see Section 8 of the Pin's Chapter [8]. The main open problem concerning concatenation hierarchies, which is in fact one of the most interesting open problem in the theory of regular languages, is the membership problem for the level 2 in the Straubing-Thérien hierarchy, i.e. the decision problem whether a given regular language can be written as a Boolean combination of polynomials over languages

* Both authors were supported by the Ministry of Education of the Czech Republic under the project MSM 0021622409 and by the Grants no. 201/06/0936 and 201/09/1313 of the Grant Agency of the Czech Republic.

S. Bozapalidis and G. Rahonis (Eds.): CAI 2009, LNCS 5725, pp. 260–277, 2009.

from level 1 in that hierarchy. It is known that a language is of this type if and only if it is a Boolean combination of polynomials with languages $L_i = B_i^*$ where each $B_i \subseteq A$ $(i = 0, \ldots, \ell)$. So this instance of polynomial operator is the most important case to study.

In the restricted case we fix a natural number k and we allow only $\ell \leq k$ in $(*)$. This operator was considered mainly in the case that $\mathcal{V}$ is the trivial variety by Simon in [10], in a series of papers by Blanchet-Sadri, see for instance [4], and in a recent paper by the authors [6].

The basic question both for general and restricted polynomial operator is to translate the construction on languages to the corresponding pseudovarieties of (ordered) monoids. A crucial tool is the Schützenberger product of (ordered) monoids (see Pin [9]). Other important questions for varieties resulting by the polynomial operator concern the existence of finite basis of (pseudo)identities for the corresponding pseudovarieties of (ordered) monoids and the possibility to generate such pseudovariety by a single monoid (see Volkov [11]).

In the present paper we continue our research from [6]. We concentrate here on identity problems for corresponding pseudovarieties and on the question whether they are generated by a single (ordered) monoid. In our basic examples the class $\mathcal{V}(A)$ equals to $\{\emptyset, A^*\}$ or to finite unions of $B^*,\ B \subseteq A$ or to finite unions of $\overline{B},\ B \subseteq A$ where $\overline{B}$ is the set of all words over A containing exactly the letters from B.

In the next section we recall the necessary background and we introduce there four examples which we will follow thorough the whole paper. We show in Section 3 that the locally finite positive varieties of languages (i.e. such that each $\mathcal{V}(A)$ is finite) correspond to the so-called finite characteristics which are certain relations on $\{x_1, x_2, \ldots\}^*$. Section 4 contains the main result which effectively translates the polynomial operation on languages to an operator on finite characteristics. The last section studies the varieties of languages which are generated by a finite number of languages. In fact, this is equivalent to the property that corresponding pseudovariety of (ordered) monoids is generated by a single monoid. We transfer this property to finite characteristics. We conclude here to by investigating this "finiteness condition" on our basic examples.

2 Preliminaries

For a relation ρ on a set S we define its *dual* relation $\rho^{\mathsf{d}} = \{\, (v,u) \in S \times S \mid (u,v) \in \rho \,\}$. A *quasiorder* ρ on a set S is a reflexive and transitive relation. Let $\widehat{\rho} = \rho \cap \rho^{\mathsf{d}}$ be the corresponding equivalence relation.

An *ordered monoid* is a structure $(M, \cdot, \leq)$ where $(M, \cdot)$ is a monoid and $\leq$ is a *compatible* order on $(M, \cdot)$, i.e. $a \leq b$ implies both $a \cdot c \leq b \cdot c,\ c \cdot a \leq c \cdot b$, for all $a, b, c \in M$. *Morphisms* of ordered monoids are isotone monoid morphisms.

Let $(M, \cdot, \leq)$ be an ordered monoid and let $\preceq$ be a compatible quasiorder on $(M, \cdot)$ satisfying $\leq\ \subseteq\ \preceq$. Then the relation $\leq_{\preceq}$ defined by

$$a\widehat{\preceq}\ \leq_{\preceq}\ b\widehat{\preceq} \quad \text{if and only if} \quad a \preceq b,\ \text{for all } a, b \in M$$

is a compatible order on $(M/\widehat{\preceq}, \cdot)$ and the mapping $a \mapsto a\widehat{\preceq}$ is a morphism of $(M, \cdot, \leq)$ onto $(M/\widehat{\preceq}, \cdot, \leq_{\preceq})$.

Let Y^* be the set of all words over an alphabet Y including the empty one, denoted by λ. For a word $u \in Y^*$, let

$$\mathsf{cont}(u) = \{\, y \in Y \mid u = u'yu'' \text{ for some } u', u'' \in Y^* \,\}\ .$$

For a set $Z \subseteq Y$, let $\overline{Z} = \{\, u \in Y^* \mid \mathsf{cont}(u) = Z \,\}$. Let $|u|_y$ be the number of occurrences of a letter $y \in Y$ in $u \in Y^*$.

An *ideal* I of an ordered set $(M, \leq)$ is a subset of M satisfying $b \leq a \in I$ implies $b \in I$, for all $a, b \in M$. For $a \in M$, we write $(a] = \{\, b \in M \mid b \leq a \,\}$. A language L over an alphabet A is *recognized* by a finite ordered monoid $(M, \cdot, \leq)$ if there exist a morphism $\phi : A^* \to M$ and an ideal I of $(M, \leq)$ such that $L = \phi^{-1}(I)$.

We recall now some basic facts about Eilenberg-type theorems. The Boolean case was invented by Eilenberg [5] and the positive case was introduced by Pin [7].

A *Boolean variety of languages* $\mathcal{V}$ associates to every finite alphabet A a class $\mathcal{V}(A)$ of regular languages over A in such a way that

- $\mathcal{V}(A)$ is closed under finite unions, finite intersections and complements (in particular $\emptyset, A^* \in \mathcal{V}(A)$),
- $\mathcal{V}(A)$ is closed under derivatives, i.e. $L \in \mathcal{V}(A),\ u, v \in A^*$ implies $u^{-1}Lv^{-1} = \{\, w \in A^* \mid uwv \in L \,\} \in \mathcal{V}(A)$,
- $\mathcal{V}$ is closed under inverse morphisms, i.e. $f : B^* \to A^*,\ L \in \mathcal{V}(A)$ implies $f^{-1}(L) = \{\, v \in B^* \mid f(v) \in L \,\} \in \mathcal{V}(B)$.

To get the notion of a *positive* variety of languages, we use in the first item only intersections and unions (not complements). In fact in this paper we consider mainly positive varieties and the Boolean ones are treated as special cases.

The meaning of $\mathcal{V} \subseteq \mathcal{W}$ is that $\mathcal{V}(A) \subseteq \mathcal{W}(A)$, for each finite alphabet A. Similarly, $\bigcup_{i \in I}\ \mathcal{V}_i$ means that $(\bigcup_{i \in I}\ \mathcal{V}_i)(A) = \bigcup_{i \in I}\ \mathcal{V}_i(A)$, for each finite A and arbitrary set I.

A *pseudovariety* of finite monoids is a class of finite monoids closed under taking submonoids, morphic images and products of finite families. Similarly for ordered monoids (see [8]). When defining a *variety* of (ordered) monoids we use arbitrary products.

For a regular language $L \subseteq A^*$, we define the relations $\sim_L$ and $\preceq_L$ on A^* as follows: for $u, v \in A^*$ we have

$$u \sim_L v \text{ if and only if } (\ \forall\, p, q \in A^*\ )\ (\ puq \in L \iff pvq \in L\ )\,,$$

$$u \preceq_L v \text{ if and only if } (\ \forall\, p, q \in A^*\ )\ (\ pvq \in L \implies puq \in L\ )\,.$$

The relation $\sim_L$ is the *syntactic congruence* of L on A^*. It is of finite index (i.e. there are only finitely many classes) and the quotient structure $\mathsf{M}(L) = A^*/\sim_L$ is called the *syntactic monoid* of L.

The relation $\preceq_L$ is the *syntactic quasiorder* of L and we have $\widehat{\preceq_L} = \sim_L$. Hence $\preceq_L$ induces an order on $\mathsf{M}(L) = A^*/\sim_L$, namely: $u\sim_L \leq v\sim_L$ if and only if $u \preceq_L v$. Then we speak about the *syntactic ordered monoid* of L and we denote the structure by $\mathsf{O}(L)$.

Result 1 (Eilenberg[5], Pin[7].) *Boolean varieties (positive varieties) of languages correspond to pseudovarieties of finite monoids (ordered monoids). The correspondence, written* $\mathcal{V} \longleftrightarrow \mathbf{V}$ *(*$\mathcal{P} \longleftrightarrow \mathbf{P}$*), is given by the following relationship: for* $L \subseteq A^*$ *we have*

$$L \in \mathcal{V}(A) \text{ if and only if } \mathsf{M}(L) \in \mathbf{V} \quad (L \in \mathcal{P}(A) \text{ if and only if } \mathsf{O}(L) \in \mathbf{P}) .$$

The pseudovarieties of ordered monoids can be characterized by pseudoidentities (see e.g. [1]). The pseudovarieties we consider here are *equational* – they are given by identities. For the set $X = \{x_1, x_2, \dots\}$, an *identity* is a pair $u = v$ $(u \leq v)$ of words over X, i.e. $u, v \in X^*$. An identity $u = v$ ($u \leq v$, respectively) is *satisfied* in a monoid M (ordered monoid $(M, \leq)$) if for each morphism $\phi : X^* \to M$ we have $\phi(u) = \phi(v)$ $(\phi(u) \leq \phi(v))$. In such a case we write $M \models u = v$ $(M \models u \leq v)$, and for a set of identities Π, we define $\mathsf{Mod}\,\Pi = \{ M \mid (\forall \pi \in \Pi)\ M \models \pi \}$. For a class $\mathbf{M}$ of ordered monoids, the meaning of $\mathbf{M} \models \Pi$ is that, for each $M \in \mathbf{M}$, we have $M \models \Pi$. Let $\mathsf{Id}\,\mathbf{V}$ be the set of all identities which are satisfied in a variety of ordered monoids $\mathbf{V}$. Let $\mathsf{Fin}\,\mathbf{V}$ denote the class of all finite members of a class $\mathbf{V}$.

For a fixed A and $L \subseteq A^*$, let $L^{\mathsf{c}} = A^* \setminus L$ be the complement of L. For a class $\mathcal{V}$ of languages, we define $\mathcal{V}^{\mathsf{c}}$ by $\mathcal{V}^{\mathsf{c}}(A) = \{ L^{\mathsf{c}} \mid L \in \mathcal{V}(A) \}$. The following is obvious.

Lemma 1. *For a positive variety* $\mathcal{V}$ *the following holds.*

(i) $\mathcal{V}^{\mathsf{c}}$ *is a positive variety.*

(ii) Let $\mathcal{V} \vee \mathcal{V}^{\mathsf{c}}$ *be the smallest positive variety containing both* $\mathcal{V}$ *and* $\mathcal{V}^{\mathsf{c}}$*. Then* $(\mathcal{V} \vee \mathcal{V}^{\mathsf{c}})(A)$ *consists of all positive Boolean combinations of the languages from* $\mathcal{V}(A) \cup \mathcal{V}^{\mathsf{c}}(A)$*.*

(iii) The class $\mathcal{V} \vee \mathcal{V}^{\mathsf{c}}$ *is a Boolean variety.*

Next we define the positive varieties of languages $\mathcal{T}$, $\mathcal{S}^+$, $\mathcal{S}$, $\mathcal{A}_m$. We will return to them several times in our paper again.

Example 1. 1. Let $\mathcal{T}(A) = \{\emptyset, A^*\}$ for each finite set A.

2. Let $\mathcal{S}^+(A)$ be the set of all finite unions of the languages of the form B^*, where $B \subseteq A$, for each finite set A.

3. Let $\mathcal{S}(A)$ be the set of all finite unions of the languages of the form $\overline{B}$, where $B \subseteq A$, for each finite set A.

4. Let m be a fixed natural number and let $\mathcal{A}_m(A)$ be the set of all Boolean combinations of the languages of the form $L(a, r) = \{ u \in A^* \mid |u|_a \equiv r \pmod{m} \}$, where $a \in A$ and $0 \leq r < m$, for each finite set A.

Notice that the classes $\mathcal{T}$, $\mathcal{S}$, $\mathcal{A}_m$ are Boolean varieties. Moreover, for the corresponding pseudovarieties of (ordered) monoids consist of finite members of the following varieties:

$$\mathbf{T} = \mathsf{Mod}(\, x = y\,),\ \mathbf{S}^+ = \mathsf{Mod}(\, x^2 = x,\, xy = yx, 1 \leq x\,),$$

$$\mathbf{S} = \mathsf{Mod}(\, x^2 = x,\, xy = yx\,),\ \mathbf{A}_m = \mathsf{Mod}(\, xy = yx,\, x^m = 1\,) .$$

The names for the (ordered) monoids of the varieties $\mathbf{T}$, $\mathbf{S}^+$, $\mathbf{S}$, $\mathbf{A}_m$ are *trivial monoids*, *semilattices with the smallest element 1*, *semilattices* and *Abelian groups of index m*, respectively – see Pin [8].

3 Locally Finite Varieties of Languages

In this paper we concentrate on positive varieties of languages which correspond to locally finite pseudovarieties of ordered monoids. Each such pseudovariety is formed by the finite members of locally finite variety of ordered monoids (i.e. finitely generated ordered monoids are finite), and consequently such a variety of languages can be described by a fully invariant compatible quasiorder on the monoid X^* which has locally finite index; more precisely:

Definition 1. *A relation γ on X^* is* **a finite characteristic** *if it satisfies the following conditions:*
(i) γ is a quasiorder on X^*;
(ii) γ is compatible with the multiplication, i.e. for each $u, v, w \in X^*$ we have

$$u \;\gamma\; v \quad \textit{implies} \quad uw \;\gamma\; vw, \quad wu \;\gamma\; wv \;;$$

(iii) γ is fully invariant, i.e. for each morphism $\varphi : X^* \to X^*$ and each $u, v \in X^*$ we have

$$u \;\gamma\; v \quad \textit{implies} \quad \varphi(u) \;\gamma\; \varphi(v) \;;$$

(iv) for each finite subset Y of the set X, the set Y^* intersects only finitely many classes of $X^*/\widehat{\gamma}$.

For each finite alphabet A, we define the *natural adaptation* γ_A of a finite characteristic γ in the following way. For $u, v \in A^*$, we have

$$u \;\gamma_A\; v \quad \text{if and only if} \quad (\,\forall\, \varphi : A^* \to X^*\,)\; \varphi(u) \;\gamma\; \varphi(v) \;. \tag{†}$$

It follows from the property (iii) in Definition 1 that in (†) we can use just one morphism given by a fixed injective mapping $\phi : A \to X$. In particular, if A is a finite subset of X then γ_A is a restriction of γ on A^*. The condition (iv) from Definition 1 means that γ_A (more precisely $\widehat{\gamma_A}$) has a finite index (i.e. the quotient set $A^*/\,\widehat{\gamma_A}$ is finite).

A relation γ on X^* satisfying the conditions (i) – (iii) is called a *fully invariant compatible quasiorder*. It determines a variety $\mathbf{V}_\gamma$ of ordered monoids; namely γ can be considered as a set of identities and $\mathbf{V}_\gamma = \mathsf{Mod}\,\gamma$. Basics of universal algebra, see [3] and [2], give that Id and Mod are mutually inverse bijections between varieties of ordered monoids and fully invariant compatible quasiorders on X^*. Moreover, for each $Y \subseteq X$, the ordered monoid Y^*/γ_Y is a free ordered monoid in $\mathbf{V}_\gamma$ over Y. The condition (iv) says that the finitely generated free ordered monoids in $\mathbf{V}_\gamma$ are finite. In this case the variety $\mathbf{V}_\gamma$ is *locally finite*, which means that all finitely generated ordered monoids are finite.

The pseudovariety $\mathsf{Fin}\,\mathbf{V}_\gamma$ of all finite members from $\mathbf{V}_\gamma$ corresponds to the positive variety $\mathcal{V}_\gamma$ of languages by

$$L \in \mathcal{V}_\gamma(A) \text{ if and only if } \mathsf{O}(L) \in \mathsf{Fin}\,\mathbf{V}_\gamma, \quad \text{for all finite } A\ .$$

We say that γ is a *finite characteristic of a class of languages* $\mathcal{V}$ if γ is a finite characteristic and for every finite alphabet A we have

$$L \in \mathcal{V}(A) \quad \text{if and only if} \quad \gamma_A \subseteq \preceq_L \ .$$

The following lemma explains the universal algebra point of view.

Lemma 2. *Let $\mathcal{V}$ be a class of languages and γ be a finite characteristic of $\mathcal{V}$. Then*

(i) $\mathcal{V}$ equals to the positive variety of languages $\mathcal{V}_\gamma = \mathsf{Fin}\,\mathsf{Mod}\,\gamma$;

(ii) if $\mathbf{V}$ is the pseudovariety of finite ordered monoids corresponding to $\mathcal{V}$ then $\gamma = \mathsf{Id}\,\mathbf{V}$;

(iii) γ^{d} is a finite characteristic of the positive variety $\mathcal{V}^{\mathsf{c}}$;

(iv) $\widehat{\gamma}$ is a finite characteristic of the Boolean variety $\mathcal{V} \vee \mathcal{V}^{\mathsf{c}}$.

Proof. "(i)" Let A be a finite alphabet. We have to show that $L \in \mathcal{V}(A)$ is equivalent to $L \in \mathcal{V}_\gamma(A)$. The statement on the left hand side is equivalent to $\gamma_A \subseteq \preceq_L$ which is equivalent to the fact that $\mathsf{O}(L)$ is a morphic image of A^*/γ_A. The last is equivalent to $\mathsf{O}(L) \in \mathsf{Fin}\,V_\gamma$, which means $L \in \mathcal{V}_\gamma(A)$.

"(ii)" Notice that $\mathbf{V}_\gamma$ is generated by its finitely generated free ordered monoids which are in $\mathsf{Fin}\,\mathbf{V}_\gamma$.

"(iii)" The statement follows from the fact that $\preceq_{L^{\mathsf{c}}} = (\preceq_L)^{\mathsf{d}}$.

"(iv)" It follows from Lemma 1. □

We present the finite characteristics for our four basic examples.

Example 2. (A continuation of Example 1.)

1. $\mathsf{Id}\,\mathbf{T} = X^* \times X^*$.
2. $\mathsf{Id}\,\mathbf{S}^+ = \{\, (u,v) \in X^* \times X^* \mid \mathsf{cont}(u) \subseteq \mathsf{cont}(v)\,\}$.
3. $\mathsf{Id}\,\mathbf{S} = \{\, (u,v) \in X^* \times X^* \mid \mathsf{cont}(u) = \mathsf{cont}(v)\,\}$.
4. $\mathsf{Id}\,\mathbf{A}_m = \{\, (u,v) \in X^* \times X^* \mid (\forall\, x \in X\,)\ |u|_x \equiv |v|_x \pmod{m}\,\}$.

Proposition 1. *Let $\mathcal{V}$ be a positive variety of languages and $\mathbf{V}$ be the corresponding pseudovariety of ordered monoids. Then the following conditions are equivalent.*

(i) For each finite alphabet A, the set $\mathcal{V}(A)$ is finite.

(ii) The pseudovariety of ordered monoids $\mathbf{V}$ is locally finite, i.e. each finitely generated submonoid of an arbitrary product of ordered monoids from $\mathbf{V}$ is finite.

(iii) There exists a finite characteristic of $\mathcal{V}$.

Proof. "(i) $\implies$ (ii)" Let $(M_i)_{i\in I}$ be an arbitrary family of ordered monoids from the class $\mathbf{V}$. Let A be a finite set, let $\phi : A^* \to M' = \prod_{i\in I} M_i$ be a morphism, and let $\pi_i : M' \to M_i$ be the i-th projection ($i \in I$). We want to show that $M = \phi(A^*)$ is finite.

For each $m \in M$, we have $\phi^{-1}((m]) = \bigcap_{i \in I} L_i$ where $L_i = (\pi_i \phi)^{-1}((\pi_i(m)])$. We have $L_i \in \mathcal{V}(A)$ as L_i is recognized by M_i. Since we have only finitely many languages in $\mathcal{V}(A)$ we intersect only finitely many languages. Consequently $\phi^{-1}((m]) \in \mathcal{V}(A)$. For different $m, n \in M$, the languages $\phi^{-1}((m])$ and $\phi^{-1}((n])$ are different. Now the finiteness of $\mathcal{V}(A)$ gives that M is finite.

"(ii) $\implies$ (iii)" Let $\mathbf{W} = \langle \mathbf{V} \rangle = \mathsf{HSP}\,\mathbf{V}$ be the variety of ordered monoids generated by the pseudovariety $\mathbf{V}$. We claim that the variety $\mathbf{W}$ is locally finite. Indeed, let M be an ordered submonoid of $\prod_{i \in I} M_i$ where each $M_i \in \mathbf{V}$, and let ϕ be a surjective morphism of M onto an ordered monoid N with a finite generating set G. We need to show that N is finite. Let $F \subseteq M$ be a finite set such that $\phi(F) = G$. By assumption (ii), the set F generates in M a finite ordered monoid and N is its image.

It follows that $\gamma = \mathsf{Id}\,\mathbf{W}$ is a finite characteristic for $\mathcal{V}$.

"(iii) $\implies$ (i)" Let γ be a finite characteristic for $\mathcal{V}$. Then $L \in \mathcal{V}(A)$ implies that L is a union of classes of A^*/γ_A. Since the set A^*/γ_A is finite there are only finitely many possibilities for L. □

A positive variety $\mathcal{V}$ is called *locally finite* if it satisfies (i) of Proposition 1.

4 Polynomial Operators of Bounded Length

Let $\mathcal{V}$ be a positive variety of languages and let k be a natural number. We define the class $\mathsf{PPol}_k\mathcal{V}$ of *positive polynomials* of length at most k of languages from the class $\mathcal{V}$. Namely, for a finite alphabet A, $\mathsf{PPol}_k\mathcal{V}(A)$ consists of finite unions of finite intersections of the languages of the form

$$L_0 a_1 L_1 a_2 \dots a_\ell L_\ell, \quad \text{where} \quad \ell \leq k,\, a_1, \dots, a_\ell \in A,\, L_0, \dots, L_\ell \in \mathcal{V}(A)\,. \quad (*)$$

Similarly, we define the classes $\mathsf{BPol}_k\mathcal{V}$ of *Boolean polynomials* using all finite Boolean combinations of languages of the form $(*)$. Clearly, it holds that $\mathsf{PPol}_k\mathcal{V} \subseteq \mathsf{PPol}_{k'}\mathcal{V}$ for $k \leq k'$ and the same for BPol's. We denote the union of all $\mathsf{PPol}_k\mathcal{V}$'s by $\mathsf{PPol}\,\mathcal{V}$. Similarly for $\mathsf{BPol}_k\mathcal{V}$'s.

Example 3. (A continuation of Examples 1 and 2.)

1. The case $\mathcal{V} = \mathcal{T}$ was studied in [6]. Notice only that $\mathsf{PPol}\,\mathcal{T}$ is the 1/2-level of the Straubing-Thérien hierarchy and $\mathsf{BPol}\,\mathcal{T}$ is the first level, i.e. the class of all piecewise testable languages.

2. and 3. One can show that $\mathsf{PPol}\,\mathcal{S}^+ = \mathsf{PPol}\,\mathcal{S}$ is the 3/2-level and $\mathsf{BPol}\,\mathcal{S}^+ = \mathsf{BPol}\,\mathcal{S}$ is the second level – see Theorem 8.8 in [8].

Lemma 3. *If $\mathcal{V}$ is a positive variety of languages then $\mathsf{PPol}_k\mathcal{V}$ is a positive variety of languages and $\mathsf{BPol}_k\mathcal{V}$ is a Boolean variety of languages.*

Proof. One can prove the statements directly. For locally finite varieties it also immediately follows from Theorem 1. □

Let k be a fixed natural number and α be a finite characteristic. Let A be a fixed set; in particular, A can be a finite alphabet or the set X.

For a word $u \in A^*$, we say that

$$f = (u_0, a_1, \ldots, a_\ell, u_\ell)$$

is *a factorization* of u of length ℓ if $u_0, u_1, \ldots, u_\ell \in A^*$, $a_1, a_2, \ldots, a_\ell \in A$ and $u_0 a_1 u_1 \ldots a_\ell u_\ell = u$. The set of all factorizations of length at most k of the word u is denoted by $\mathsf{Fact}_k(u)$. For a factorization $f = (u_0, a_1, \ldots, a_\ell, u_\ell)$ of a word $u \in A^*$ and a factorization $g = (v_0, b_1, v_1, \ldots b_m, v_m)$ of a word $v \in A^*$, we write

$$f \leq_\alpha g$$

if $\ell = m$, $a_i = b_i$ for every $i \in \{1, \ldots, \ell\}$ and $u_i\ \alpha_A\ v_i$ for every $i \in \{0, 1, \ldots, \ell\}$. We define the relation $(\mathsf{p}_k(\alpha))_A$ on the set A^* as follows: for $u, v \in A^*$, we have

$$u\ (\mathsf{p}_k(\alpha))_A\ v \quad \text{if and only if} \quad (\forall\, g \in \mathsf{Fact}_k(v)\,)\ (\exists\, f \in \mathsf{Fact}_k(u)\,)\ f \leq_\alpha g\ .$$

We show in Theorem 1 that the relation $(\mathsf{p}_k(\alpha))_X$ is a finite characteristic and therefore the relation $(\mathsf{p}_k(\alpha))_A$ is equal to $((\mathsf{p}_k(\alpha))_X)_A$ as defined after Definition 1. We write $\mathsf{p}_k(\alpha)$ instead of $(\mathsf{p}_k(\alpha))_X$. Further we denote $\mathsf{b}_k(\alpha) = \widehat{\mathsf{p}_k(\alpha)}$.

Theorem 1. *Let $\mathcal{V}$ be a locally finite positive variety of languages and α be the finite characteristic of $\mathcal{V}$. Then $\mathsf{PPol}_k\mathcal{V}$ is a locally finite positive variety of languages with the finite characteristic $\mathsf{p}_k(\alpha)$ and $\mathsf{BPol}_k\mathcal{V}$ is a locally finite Boolean variety of languages with the finite characteristic $\mathsf{b}_k(\alpha)$.*

Proof. We prove that $\mathsf{p}_k(\alpha)$ is a finite characteristic of $\mathsf{PPol}_k\mathcal{V}$. The rest follows from Lemma 1 and Lemma 2.

We have to check the properties (i) – (iv) from Definition 1 and also the property

(v) $\quad L \in \mathsf{PPol}_k\mathcal{V}(A) \quad \text{if and only if} \quad (\mathsf{p}_k(\alpha))_A \subseteq \preceq_L .$

"(i)" The reflexivity of the relation $\mathsf{p}_k(\alpha)$ is trivial. The transitivity follows from the transitivity of the relation $\leq_\alpha$.

"(ii)" Let $u, v, w \in X^*$ be such that $(u, v) \in \mathsf{p}_k(\alpha)$. We want to show that $(uw, vw) \in \mathsf{p}_k(\alpha)$. Let $g \in \mathsf{Fact}_k(vw)$ be an arbitrary factorization of length at most k of the word vw, i.e. $g = (v_0, a_1, v_1, \ldots, a_\ell, v_\ell)$, where $\ell \leq k$, $a_1, \ldots, a_\ell \in X$, $v_0, \ldots, v_\ell \in X^*$ and there exist $0 \leq i \leq \ell$ and $v_i', v_i'' \in X^*$ such that $v_i' v_i'' = v_i$ and

$$v = v_0 a_1 v_1 \ldots a_i v_i'\ , \quad w = v_i'' a_{i+1} \ldots a_\ell v_\ell\ .$$

From the assumption $(u, v) \in \mathsf{p}_k(\alpha)$ we know that there is a factorization f of the word u such that $f \leq_\alpha (v_0, a_1, v_1, \ldots, a_i, v_i')$, i.e. $f = (u_0, a_1, u_1, \ldots, a_i, u_i')$ such that $u_0\ \alpha\ v_0$, $\ldots$, $u_i'\ \alpha\ v_i'$. Since α is a compatible quasiorder we have $u_i' v_i''\ \alpha\ v_i' v_i''$. Hence

$$h = (u_0, a_1, u_1, \ldots, a_i, u_i' v_i'', a_{i+1}, \ldots, a_\ell, v_\ell)$$

is a factorization of uw such that $h \leq_\alpha g$. This implies $(uw, vw) \in \mathsf{p}_k(\alpha)$.

The proof of the implication "$(u, v) \in \mathsf{p}_k(\alpha) \implies (wu, wv) \in \mathsf{p}_k(\alpha)$" is similar.

"(iii)" Let $u, v \in X^*$ be such that $(u, v) \in \mathsf{p}_k(\alpha)$ and $\varphi : X^* \to X^*$ be an arbitrary morphism. We want to show that $(\varphi(u), \varphi(v)) \in \mathsf{p}_k(\alpha)$. So, let

$$g' = (v_0, a_1, v_1 \ldots, a_\ell, v_\ell) \in \mathsf{Fact}_k(\varphi(v))$$

where $\ell \leq k$, $v_i \in X^*, a_i \in X$ and $v_0 a_1 v_1 \ldots a_\ell v_\ell = \varphi(v)$. We consider a factorization $g = (w_0, b_1, w_1, \ldots, b_m, w_m)$ of v where the occurrences of the letters $b_1, \ldots, b_m$ are such that the corresponding occurrences of $\varphi(b_1), \ldots, \varphi(b_m)$ in $\varphi(v)$ contain all a_i's in the factorization g'. Note that $m \leq \ell$ as $\varphi(b_j)$ can contain more than one a_i. Now $(u, v) \in \mathsf{p}_k(\alpha)$ and there exists a factorization f of u such that $f \leq_\alpha g$, i.e. $f = (t_0, b_1, t_1 \ldots, b_m, t_m)$ where $t_i \;\alpha\; w_i$ for $i \in \{0, \ldots, m\}$. Since α is a finite characteristic we have $\varphi(t_i) \;\alpha\; \varphi(w_i)$. Hence $\varphi(u) = \varphi(t_0)\varphi(b_1)\varphi(t_1) \ldots \varphi(b_m)\varphi(t_m)$ has a factorization f' such that $f' \leq_\alpha g'$. We can conclude that $(\varphi(u), \varphi(v)) \in \mathsf{p}_k(\alpha)$.

"(iv)" Let Y be a finite subset of X. Since $\widehat{\alpha_Y}$ has a finite index, there are only finitely many factorizations of length at most k over Y when identifying the $\widehat{\leq_\alpha}$-related ones. Hence there are only finitely many sets of the form $\mathsf{Fact}_k(u)$ up to $\widehat{\leq_\alpha}$, where $u \in Y^*$. So, $\widehat{\mathsf{p}_k(\alpha)|_Y}$ has a finite index too.

"(v)" For simplicity denote the relation $(\mathsf{p}_k(\alpha))_A$ by β.

" $\implies$ " We prove that for every language

$$L = L_0 a_1 L_1 \ldots a_\ell L_\ell, \quad \text{where} \quad \ell \leq k,\ a_1, \ldots, a_\ell \in A,\ L_0, \ldots, L_\ell \in \mathcal{V}(A)\ ,$$

we have $\beta \subseteq \preceq_L$. This is enough because $\beta \subseteq \preceq_L$ and $\beta \subseteq \preceq_K$ imply $\beta \subseteq \preceq_{L\cap K}$ and $\beta \subseteq \preceq_{L\cup K}$, for each $L, K \subseteq A^*$.

Let L be such a language and let $u, v \in A^*$ satisfy $u \;\beta\; v$. We want to show that $u \preceq_L v$. So, let $p, q \in A^*$ be such that $pvq \in L$. Hence $pvq = v_0 a_1 v_1 \ldots a_\ell v_\ell$, where $v_i \in L_i$ for every $i \in \{0, \ldots, \ell\}$. Then there exist $0 \leq i < j \leq \ell$ and $v_i', v_i'', v_j', v_j'' \in A^*$, such that $v_i' v_i'' = v_i$, $v_j' v_j'' = v_j$ and

$$p = v_0 a_1 \ldots v_i', \quad v = v_i'' a_{i+1} \ldots a_j v_j' \quad \text{and } q = v_j'' a_{j+1} \ldots a_\ell v_\ell$$

or there exist $0 \leq i \leq \ell$ and $v_i', v_i'', v_i''' \in A^*$ such that $v_i' v_i'' v_i''' = v_i$ and

$$p = v_0 a_1 \ldots v_i', \quad v = v_i'' \quad \text{and } q = v_i''' a_{i+1} \ldots a_\ell v_\ell\ .$$

In the first case we have $g = (v_i'', a_{i+1}, \ldots, a_j, v_j')$ a factorization of v. We assumed that $u \;\beta\; v$, so there is a factorization $f = (u_i'', a_{i+1}, \ldots, u_j')$ of u such that $(u_i'', v_i''), (u_{i+1}, v_{i+1}), \ldots, (u_j', v_j') \in \alpha_A$. Since α_A is a compatible quasiorder we have $(v_i' u_i'', v_i' v_i'') \in \alpha_A$ and hence $v_i' u_i'' \preceq_{L_i} v_i' v_i'' = v_i$, so we have $v_i' u_i'' \in L_i$. Similarly $u_{i+1} \in L_{i+1}$, $\ldots$, $u_{j-1} \in L_{j-1}$ and $u_j' v_j'' \in L_j$. Consequently $puq \in L$. The second case is similar and we see that $u \;\beta\; v$ really implies $u \preceq_L v$.

"$\Longleftarrow$" Let $\beta \subseteq \preceq_L$. This means that L is a finite union of languages of the form

$$\beta v = \{\, u \in A^* \mid u \;\beta\; v \,\}, \quad \text{where} \;\; v \in A^* \;.$$

It is enough to prove that each βv belongs to $\mathsf{PPol}_k\mathcal{V}(A)$. Consider all possible factorizations of the word v of length at most k, i.e. all elements of the set $\mathsf{Fact}_k(v)$. So, we have

$$g_1 = (v_{10}, a_{11}, \dots, a_{1\ell_1}, v_{1\ell_1}) \;,$$

$$g_2 = (v_{20}, a_{21}, \dots, a_{2\ell_2}, v_{2\ell_2}) \;,$$

$$\vdots$$

$$g_m = (v_{m0}, a_{m1}, \dots, a_{m\ell_m}, v_{m\ell_m}) \;,$$

where for each $i \in \{1, \dots, m\}$ we have $\ell_i \leq k$ and $a_{ij} \in A$ are letters and $v_{ij} \in A^*$ are words and $\{g_1, g_2, \dots, g_m\} = \mathsf{Fact}_k(v)$. For each $i \in \{1, \dots, m\}$ we consider the following language L_i corresponding to the factorization $g_i = (v_{i0}, a_{i1}, \dots, a_{i\ell_i}, v_{i\ell_i})$:

$$L_i = L_{i0}\; a_{i1}\; L_{i1}\; \dots\; a_{i\ell_i}\; L_{i\ell_i} \;,$$

where $L_{ij} = \alpha_A v_{ij} = \{\, u \in A^* \mid u \;\alpha_A\; v_{ij} \,\} \in \mathcal{V}(A)$ for each $j \in \{0, \dots, \ell_i\}$. Then the language

$$K = \bigcap_{i=1}^{m} L_i$$

belongs to $\mathsf{PPol}_k\mathcal{V}(A)$ and we prove that $K = \beta v$.

"$\subseteq$" If $u \in K$ then $u \in L_i$ for each $i \in \{1, \dots, m\}$. This means that for each $i \in \{1, \dots, m\}$ we have

$$u = u_{i0} a_{i1} \dots a_{i\ell_i} u_{i\ell_i} \;,$$

where $(u_{i0}, v_{i0}), \dots, (u_{i\ell_i}, v_{i\ell_i}) \in \alpha_A$. Therefore, there is a factorization f_i of u such that $f_i \leq_\alpha g_i$. Consequently $(u, v) \in (\mathsf{p}_k(\alpha))_A = \beta$, i.e. $u \in \beta v$.

"$\supseteq$" If $u \in \beta v$. Then for each $i \in \{1, \dots, m\}$, we have some factorization f_i of u such that $f_i \leq_\alpha g_i$. This implies that $u \in L_i$ for each $i \in \{1, \dots, m\}$, and hence $u \in K$. □

The following lemmas concern the preservation of aperiodicity (i.e. monoids have only trivial subgroups).

Lemma 4. *Let α be a finite characteristic and let k, n be arbitrary natural numbers. Put $m = (k+1)(n+1)$.*

(i) If $(x^n, x^{n+1}) \in \alpha$ then $(x^{m-1}, x^m) \in \mathsf{p}_k(\alpha)$.

(ii) If $(x^n, x^{n+1}) \in \widehat{\alpha}$ then $(x^{m-1}, x^m) \in \widehat{\mathsf{p}_k(\alpha)}$.

Proof. "(i)" Let g be a factorization of x^m of length $\ell \leq k$, i.e.

$$g = (x^{i_0}, x, x^{i_1}, x, \dots, x, x^{i_\ell})$$

where $i_0+i_1+\cdots+i_\ell+\ell = m$ and $i_0, \dots, i_\ell$ are non-negative integers. Assume that for every $j \in \{0, \dots, \ell\}$ we have $i_j \leq n$, then $i_0+i_1+\cdots+i_\ell+\ell \leq (\ell+1)n+\ell \leq (k+1)n+k < (k+1)(n+1) = m$ a contradiction. Thus, there is $j \in \{0, \dots, \ell\}$ such that $i_j \geq n+1$, hence $x^{i_j-1}\ \alpha\ x^{i_j}$ and consequently there exists a factorization f of x^{m-1} such that $f \leq_\alpha g$. This proves $(x^{m-1}, x^m) \in \mathsf{p}_k(\alpha)$.

"(ii)" With respect to the part (i) it is enough to prove the implication $(x^{n+1}, x^n) \in \alpha \implies (x^m, x^{m-1}) \in \mathsf{p}_k(\alpha)$. This is not a direct consequence of statement (i) since $(\mathsf{p}_k(\alpha))^{\mathsf{d}} \neq \mathsf{p}_k(\alpha^{\mathsf{d}})$ but the implication can be proved in a similar way as part (i). □

Lemma 5. *Let $\mathcal{V}$ be a positive variety with the finite characteristic α, such that the corresponding pseudovariety of ordered monoids contains only aperiodic monoids. Then, for each natural number k, the pseudovariety of ordered monoids corresponding to the positive variety of languages $\mathsf{PPol}_k\mathcal{V}$ contains only aperiodic monoids too.*

Proof. Let $A = \{a\}$ be an alphabet. Then A^*/α_A belongs to the corresponding pseudovariety of monoids, i.e. A^*/α_A is a finite aperiodic monoid. This implies that $(a^n, a^{n+1}) \in \widehat{\alpha_A}$ for some natural number n and $(x^n, x^{n+1}) \in \widehat{\alpha}$ follows. By Lemma 4, we have $(x^{m-1}, x^m) \in \widehat{\mathsf{p}_k(\alpha)}$ for a certain m. Hence for every alphabet B, the monoid B^*/α_B is aperiodic, and consequently the pseudovariety of monoids corresponding to the positive variety of languages $\mathsf{PPol}_k\mathcal{V}$ contains only aperiodic monoids because each of them is a morphic images of the monoid B^*/α_B for some B. □

5 Generating Pseudovarieties by a Single Monoid

It is known (see Volkov [11] or the authors [6]) that the pseudovarieties of ordered monoids corresponding to $\mathsf{PPol}_k\mathcal{T}$, k a natural number, are generated by a single ordered monoid. We show such result also for the positive varieties $\mathsf{PPol}_k\mathcal{S}^+$ and we prove that this is not true for the positive varieties $\mathsf{PPol}_k\mathcal{S}$. At first we define a "finiteness-like" condition concerning finite characteristics.

Definition 2. *Let α be a finite characteristic. We say that α is* **finitely determined** *if there is a finite alphabet A such that for every finite alphabet B and all $u, v \in B^*$ we have:*

$$(\ (\forall\ \varphi : B \to A\ )\ \varphi(u)\ \alpha_A\ \varphi(v)\ ) \quad \textit{implies} \quad u\ \alpha_B\ v\ .$$

The extension of a mapping $\varphi : B \to A$ to a morphism from B^* to A^* is denoted by the same symbol. Clearly, the opposite implication is always true due to Definition 1.

Example 4. The finite characteristic of the positive variety $\mathcal{S}^+$ was described in Example 2.2. It is finitely determined since one can show that the condition from the previous definition is satisfied for $A = \{a, a'\}$, $a \neq a'$. Indeed, for an arbitrary finite alphabet B and $u, v \in B^*$ such that $\mathsf{cont}(u) \nsubseteq \mathsf{cont}(v)$ we can consider a letter $b \in \mathsf{cont}(u) \setminus \mathsf{cont}(v)$. Then we take a mapping $\varphi : B \to A$ sending b to a and (possible) other elements of B to a'. For this φ we have $a \in \mathsf{cont}(\varphi(u)) \setminus \mathsf{cont}(\varphi(v))$.

The same considerations for two element set A are true for $\mathcal{S}$ and for $\mathcal{A}_m$.

Proposition 2. *The following properties for a positive variety $\mathcal{V}$ and the corresponding pseudovariety of ordered monoids $\mathbf{V}$ are equivalent.*

(i) The positive variety $\mathcal{V}$ is generated by a finite number of languages.

(ii) The pseudovariety $\mathbf{V}$ is generated by a single ordered monoid.

(iii) There exists a finite characteristic of $\mathcal{V}$ which is finitely determined.

Proof. "$(i) \implies (ii)$" If $\mathcal{V}$ is generated by a finite number of languages then we can take their syntactic ordered monoids and consider the product of all of them. The resulting ordered monoid generates the pseudovariety of ordered monoids $\mathbf{V}$.

"$(ii) \implies (iii)$" Let the pseudovariety $\mathbf{V}$ be generated by a single finite ordered monoid M. We consider the variety $\mathbf{W} = \langle \mathbf{V} \rangle = \langle M \rangle$ generated by the monoid M. If we take the free ordered monoid F over X in the variety $\mathbf{W}$ and denote α the kernel of the projection from X^* onto F, then this α is a finite characteristic of $\mathcal{V}$. Moreover, for a finite alphabet C, the (finite) structure C^*/α_C is a free ordered monoid over C in $\mathbf{W}$.

Now we put $A = M$ and we prove the property from Definition 2 for this set A. At first, there is a natural morphism $\theta : A^* \to M$ which maps the word $a_1 a_2 \dots a_m \in A^*$ to the product of elements $a_1, a_2, \dots, a_m \in A = M$ in M, i.e. $\theta(a_1 a_2 \dots a_m) = a_1 \cdot a_2 \cdot \dots \cdot a_m$. Note that M is a morphic image of the free ordered monoid A^*/α_A, in other words, α_A is a subset of the kernel of θ.

Let B be a finite alphabet and $u, v \in B^*$ be such that for each $\varphi : B \to A$ we have $\varphi(u)\ \alpha_A\ \varphi(v)$. Each mapping $\varphi : B \to A = M$ determines a morphism $\overline{\varphi} = \theta \circ \varphi : B^* \to M$.

Recall that a free monoid over B in $\mathbf{W}$ can be constructed in the following way. There are only finitely many mappings $\varphi : B \to M$; denote Σ the set of all of them. Then we consider the finite product $\prod_{\varphi \in \Sigma} M = M^\Sigma$ and the corresponding morphism $\psi : B^* \to M^\Sigma$ given by $\psi(w) = (\overline{\varphi}(w))_{\varphi \in \Sigma}$. The image of ψ is a free monoid over B in $\mathbf{W}$ and α_B is a kernel of ψ. Now for each $\varphi : B \to A = M$ we have $\varphi(u)\ \alpha_A\ \varphi(v)$. Thus $\varphi(u) \leq \varphi(v)$ in A^*/α_A and $\overline{\varphi}(u) \leq \overline{\varphi}(v)$ in M follows. Hence $\psi(u) \leq \psi(v)$ in the free ordered monoid B^*/α_B and consequently $u\ \alpha_B\ v$.

"$(iii) \implies (i)$" Let α be a finite characteristic of $\mathcal{V}$ which is finitely determined. Let A be the corresponding finite alphabet. Since α_A has a finite index, there are only finitely many languages of the form $\alpha_A v = \{\, u \in A^* \mid u\ \alpha_A\ v \,\}$ where $v \in A^*$. We show that these languages generate $\mathcal{V}$.

Let B be an arbitrary finite alphabet and let $L \in \mathcal{V}(B)$. Since α is a finite characteristic of $\mathcal{V}$ we have $\alpha_B \subseteq \preceq_L$. Hence L is a finite union of languages of the form $\alpha_B w = \{\, t \in B^* \mid t\ \alpha_B\ w \,\}$, where $w \in B^*$.

There are only finitely many mappings from B to A; denote them $\varphi_1, \ldots, \varphi_m$, where $m = |A|^{|B|}$. Now for every $u, v \in B^*$ we have

$$u\ \alpha_B\ v \text{ if and only if } (\ \forall\, i \in \{1, \ldots, m\}\ )\ \ \varphi_i(u)\ \alpha_A\ \varphi_i(v)\ .$$

We show that

$$\alpha_B w = \bigcap_{i=1}^{m} \varphi_i^{-1}(\alpha_A w_i), \quad \text{where} \quad w_i = \varphi_i(w) \quad \text{for} \quad i \in \{1, \ldots, m\}\ . \qquad (\ddagger)$$

Indeed, for $t \in B^*$, it holds $t \in \alpha_B w$ if and only if for each $i \in \{1, \ldots, m\}$ we have $\varphi_i(t)\ \alpha_A\ \varphi_i(w) = w_i$, and this is equivalent to: for each $i \in \{1, \ldots, m\}$ we have $t \in \varphi_i^{-1}(\alpha_A w_i)$.

Equation ($\ddagger$) means that we can obtain each language of the form $\alpha_B w$ from the languages $\alpha_A v$, for $v \in A^*$, when we use inverse morphisms and the operation of intersection. ◻

Example 5. In paper [6] the authors proved that $\mathsf{PPol}_k\mathcal{T}$ is generated by a language $A^*a_1A^*a_2 \ldots a_kA^*$ where $a_1, a_2, \ldots, a_k$ are pairwise different letters and $A = \{a_1, \ldots, a_k\}$. We show that the corresponding finite characteristic $\alpha = \mathsf{p}_k(X^* \times X^*)$ is finitely determined.

Indeed, let $\mathsf{Sub}_k(w)$ denote the set of all subwords of $w \in X^*$ of length at most k. Then $u\ \alpha\ v$ if and only if $\mathsf{Sub}_k(v) \subseteq \mathsf{Sub}_k(u)$. Let $A' = \{a_1, \ldots, a_{k+1}\}$ be of cardinality $k+1$, let B be a finite set, and let $u, v \in B^*$ satisfy $\varphi(u)\ \alpha_{A'}\ \varphi(v)$ for each $\varphi : B \to A'$. Suppose that $u\ \alpha_B\ v$ does not hold. Then there exists $w \in B^*$ of length at most k such that $w \in \mathsf{Sub}_k\,(v) \setminus \mathsf{Sub}_k\,(u)$. Let $C = \mathsf{cont}(w)$. Take an injective mapping $\varphi : C \to \{a_1, \ldots, a_k\}$ and put $\varphi(b) = a_{k+1}$ for $b \notin C$. Thus $\varphi : B \to A'$ and $\varphi(w) \in \mathsf{Sub}_k(\varphi(v)) \setminus \mathsf{Sub}_k(\varphi(u))$ – a contradiction.

Proposition 3. *The positive variety* $\mathsf{PPol}_k\mathcal{S}^+$ *is generated by a finite number of languages.*

Proof. Although a direct proof would be possible we apply Proposition 2. Recall that the finite characteristic α for $\mathcal{S}^+$ is given as follows: for each $u, v \in X^*$, we have $u\ \alpha\ v$ if and only if $\mathsf{cont}(u) \subseteq \mathsf{cont}(v)$. We show that the finite characteristic $\beta = \mathsf{p}_k(\alpha)$ of $\mathsf{PPol}_k\mathcal{S}^+$ is finitely determined.

Let A be an alphabet containing 2^{2k+1} letters:

$$A = \{\, a_r \mid r \in \{0,1\}^{2k+1} \,\}\ .$$

We prove the property from Definition 2. Let B be a finite alphabet and assume that $u, v \in B^*$ satisfy

$$(\ \forall\, \varphi : B \to A\ )\ \ \varphi(u)\ \beta_A\ \varphi(v)\ .$$

We want to prove $u\ \beta_B\ v$. So, let $g = (v_0, b_1, v_1, \ldots, b_\ell, v_\ell) \in \mathsf{Fact}_k(v)$ be an arbitrary factorization of length at most k of the word v. For each letter $c \in B$ we consider the letter $a_r \in A$ where the sequence r has 1 at j-th position if and only if c is at the j-th position in the factorization g. More precisely, $r_{2i+1} = 1$ iff $c \in \mathsf{cont}(v_i)$ and $r_{2i} = 1$ iff $c = b_i$. So, we have defined a mapping $\varphi : B \to A$. Note that if a letter c does not occur in v then $\varphi(c) = a_{(0,0,\ldots,0)}$ by this definition. Now $\varphi(u)\ \beta_A\ \varphi(v)$ and there exists a factorization f' of $\varphi(u)$ such that $f' \leq_\alpha g' = (\varphi(v_0), \varphi(b_1), \varphi(v_1), \ldots, \varphi(b_\ell), \varphi(v_\ell))$. If $\varphi(b_i) = a_r$ then $r_{2i} = 1$ and for this r there is a unique letter $c \in B$, namely b_i, with the property $\varphi(c) = a_r$. Hence we have a factorization $f = (u_0, b_1, u_1, \ldots, b_\ell, u_\ell)$ of u such that $\varphi(u_i)\ \alpha_A\ \varphi(v_i)$ for each $i \in \{0, \ldots, \ell\}$. We show that this implies $u_i\ \alpha_B\ v_i$. Let $d \in \mathsf{cont}(u_i)$ be an arbitrary letter from the alphabet B. Then $\varphi(d) \in \mathsf{cont}(\varphi(u_i)) \subseteq \mathsf{cont}(\varphi(v_i))$. Let $\varphi(d) = a_r$. Then $a_r \in \mathsf{cont}(\varphi(v_i))$ implies that $r_{2i+1} = 1$. If $d \notin \mathsf{cont}(v_i)$ then $r_{2i+1} = 0$ by the definition of the mapping φ. Hence $d \in \mathsf{cont}(v_i)$, and thus, we have $u_i\ \alpha_B\ v_i$ for each $i = 0, \ldots, \ell$. For a given $g \in \mathsf{Fact}_k(v)$, we found $f \in \mathsf{Fact}_k(u)$ such that $f \leq_\alpha g$. This means that we proved $u\ \beta_B\ v$. □

Proposition 4. *The positive variety* $\mathsf{PPol}_1\mathcal{S}$ *is generated by a finite number of languages.*

Proof. Recall that the finite characteristic α for $\mathcal{S}$ is given as follows: for each $u, v \in X^*$, we have $u\ \alpha\ v$ if and only if $\mathsf{cont}(u) = \mathsf{cont}(v)$.

We show that finite characteristic $\beta = \mathsf{p}_1(\alpha)$ of the variety $\mathsf{PPol}_1\mathcal{S}$ is finitely determined on a six-element alphabet $A = \{a_0, a_1, a_2, a_3, a_4, a_5\}$. First we formulate some basic consequences of the assumption $s\ \beta_A\ t$ for a pair of words $s, t \in A^*$. We have $\mathsf{cont}(s) = \mathsf{cont}(t)$ since we can consider (unique) factorizations of s and t of length 0. Further, if we assume that the first occurrence of a letter $a \in A$ in s is before the first occurrence of a letter $a' \in A$ in s then there is a factorization (s_0, a', s_1) of the word s such that $a \in \mathsf{cont}(s_0)$, $s_0, s_1 \in A^*$ but there is no factorization (s_0, a, s_1) of s such that $a' \in \mathsf{cont}(s_0)$, $s_0, s_1 \in A^*$. Thus from $s\ \beta_A\ t$ we can conclude that the sequences of the first occurrences of all letters in s and in t coincide. Equivalently this can be expressed by the equality $\{\, \mathsf{cont}(s') \mid s' \text{ prefix of } s \,\} = \{\, \mathsf{cont}(t') \mid t' \text{ prefix of } t \,\}$. The similar observations can be done for the last occurrences of letters in s and t.

Let B be a finite alphabet containing at least seven letters[1] and assume that for a given pair of words $u, v \in B^*$ we have

$$(\ \forall\ \varphi : B \to A\ )\ \ \varphi(u)\ \beta_A\ \varphi(v)\ .$$

Let $g = (g_0, b, g_1) \in \mathsf{Fact}_1(v)$ be a factorization of v. We need to show that there exists a factorization $f = (f_0, b, f_1) \in \mathsf{Fact}_1(u)$ such that $\mathsf{cont}(f_0) = \mathsf{cont}(g_0)$ and $\mathsf{cont}(f_1) = \mathsf{cont}(g_1)$.

First of all, we take an arbitrary pair of different letters $b_1, b_2 \in B$ and consider the mapping $\varphi_{b_1,b_2} : B \to A$ given by the rules $\varphi_{b_1,b_2}(b_1) = a_1, \varphi_{b_1,b_2}(b_2) = a_2$

[1] For alphabets with at most six letters the statement is trivial.

and $\varphi_{b_1,b_2}(c) = a_0$ for all $c \in B \setminus \{b_1, b_2\}$. Since $(\varphi_{b_1,b_2}(u), \varphi_{b_1,b_2}(v)) \in \beta_A$ we can apply our basic observations concerning β_A and we see that $\mathsf{cont}(\varphi_{b_1,b_2}(u)) = \mathsf{cont}(\varphi_{b_1,b_2}(v))$ from which we observe $b_1 \in \mathsf{cont}(u) \iff b_1 \in \mathsf{cont}(v)$. This is true for each b_1 and thus $\mathsf{cont}(u) = \mathsf{cont}(v)$ follows. Further, the sequence of the first occurrences of letters in $\varphi_{b_1,b_2}(u)$ and $\varphi_{b_1,b_2}(v)$ coincide. Hence the first occurrence of b_1 in the word u is before the first occurrence of b_2 in u if and only if the first occurrence of b_1 in v is before the first occurrence of b_2 in v. This is true for every pair of letters b_1 and b_2 and we can summarize that $\{\, \mathsf{cont}(u') \mid u' \text{ prefix of } u \,\} = \{\, \mathsf{cont}(v') \mid v' \text{ prefix of } v \,\}$. When we consider the same idea from the right we obtain the same observations concerning the last occurrences of letters and finally we obtain the equality $\{\, \mathsf{cont}(u') \mid u' \text{ suffix of } u \,\} = \{\, \mathsf{cont}(v') \mid v' \text{ suffix of } v \,\}$.

There is a prefix u' of the word u such that $\mathsf{cont}(u') = \mathsf{cont}(g_0)$. Let u_1 be the shortest prefix of u with this property and u_2 be the longest prefix of u with this property. Note that u_1 can be the empty word (when $\mathsf{cont}(g_0) = \emptyset$, i.e. in the case $g_0 = \lambda$) and u_2 can be equal to u (when $\mathsf{cont}(g_0) = \mathsf{cont}(v)$). If u_1 is not the empty word then $u_1 = u_1' b_1$ where $b_1 \in B$ and $b_1 \in \mathsf{cont}(u_1) = \mathsf{cont}(g_0)$, $b_1 \notin \mathsf{cont}(u_1')$. A useful consequence is that this b_1 is the first occurrence of b_1 in u. Similarly, if $u_2 \neq u$ then $u = u_2 b_2 u_2'$ where $b_2 \in B$, $u_2' \in B^*$ and $b_2 \notin \mathsf{cont}(u_2) = \mathsf{cont}(g_0)$. Once again this b_2 is the first occurrence of b_2 in u. Note that if b_1 and b_2 are defined then they are different because $b_2 \notin \mathsf{cont}(g_0)$, but one of them can be equal to the letter b. These definitions can be also consider dually from the right. I.e. we can consider the shortest suffix u_3 of u and the longest suffix u_4 of u with the properties $\mathsf{cont}(u_3) = \mathsf{cont}(u_4) = \mathsf{cont}(g_1)$. If $u_3 \neq \lambda$ then we denote its first letter b_3, i.e $u_3 = b_3 u_3'$ and we have $b_3 \in \mathsf{cont}(u_3) = \mathsf{cont}(g_1)$, $b_3 \notin \mathsf{cont}(u_3')$. If $u_4 \neq u$ then we denote $u = u_4' b_4 u_4$ where $b_4 \in B$, $u_4' \in B^*$, $b_4 \notin \mathsf{cont}(u_4) = \mathsf{cont}(g_1)$.

Now we have the subset $B' = \{b, b_1, b_2, b_3, b_4\}$ of the alphabet B which has at most five elements. Note that some of the letters can be equal, some of them can not be defined. We consider some mapping $\varphi : B \to A$ such that $\varphi(c) = a_5$ for every $c \notin B'$, $\varphi(B') \subseteq A \setminus \{a_5\}$, $\varphi(b) = a_0$ and which is injective on B'. Then $(\varphi(g_0), a_0, \varphi(g_1))$ is a factorization of $\varphi(v)$ and there is a factorization $f = (f_0, d, f_1)$ of u such that $(\varphi(f_0), \varphi(d), \varphi(f_1)) \leq_\alpha (\varphi(g_0), \varphi(b), \varphi(g_1))$ where $\varphi(d) = \varphi(b)$, i.e. $d = b$, $\varphi(f_0)\ \alpha_A\ \varphi(g_0)$ and $\varphi(f_0)\ \alpha_A\ \varphi(g_0)$. We show that $\mathsf{cont}(f_0) = \mathsf{cont}(g_0)$ and $\mathsf{cont}(f_1) = \mathsf{cont}(g_1)$.

"$\mathsf{cont}(g_0) \subseteq \mathsf{cont}(f_0)$" If $\mathsf{cont}(g_0) = \emptyset$ then it is clear. If $\mathsf{cont}(g_0) \neq \emptyset$ then $b_1 \in \mathsf{cont}(g_0)$ is defined. Hence $\varphi(b_1) \in \mathsf{cont}(\varphi(g_0)) = \mathsf{cont}(\varphi(f_0))$ and since φ is injective on B' we have $b_1 \in \mathsf{cont}(f_0)$. By the definition of b_1 we can conclude that u_1 is a prefix of f_0, so, $\mathsf{cont}(g_0) = \mathsf{cont}(u_1) \subseteq \mathsf{cont}(f_0)$.

"$\mathsf{cont}(f_0) \subseteq \mathsf{cont}(g_0)$" If $\mathsf{cont}(g_0) = \mathsf{cont}(v) = \mathsf{cont}(u)$ then it is clear. If $\mathsf{cont}(g_0) \neq \mathsf{cont}(v)$ then b_2 is defined. We have $b_2 \notin \mathsf{cont}(g_0)$. Hence $\varphi(b_2) \notin \mathsf{cont}(\varphi(g_0)) = \mathsf{cont}(\varphi(f_0))$ and this implies $b_2 \notin \mathsf{cont}(f_0)$. By the definition of b_2 we can conclude that f_0 is a prefix of u_2, so, $\mathsf{cont}(f_0) \subseteq \mathsf{cont}(u_2) = \mathsf{cont}(g_0)$.

One can prove the equality $\mathsf{cont}(f_1) = \mathsf{cont}(g_1)$ in the same way using the letters b_3 and b_4. $\square$

Proposition 5. *The positive variety* $\mathsf{PPol}_2\mathcal{S}$ *is not generated by a finite number of languages.*

Proof. For the finite characteristic α for $\mathcal{S}$ we have, for each $u, v \in X^*$, it holds $u\ \alpha\ v$ if and only if $\mathsf{cont}(u) = \mathsf{cont}(v)$

Assume that the finite characteristic $\beta = \mathsf{p}_2(\alpha)$ of the positive variety $\mathsf{PPol}_2\mathcal{S}$ is finitely determined. Let $A = \{c_1, \ldots, c_m\}$ be an alphabet for which the property from Definition 2 is satisfied. Let $B = A \cup \{d\}$, $d \notin A$. Assume that $s_1, \ldots, s_n$ are all words of length at most $m+1$ over the alphabet A such that $\mathsf{cont}(s_j) \neq A$ for $j \in \{1, \ldots, n\}$. Further $t_{j_0 j_1 j_2} = dc_{j_0} ds_{j_1} dc_{j_2} d$ for all $j_1 \in \{1, \ldots, n\}$, $j_0, j_2 \in \{1, \ldots, m\}$ and t be a product of all words $t_{j_0 j_1 j_2}$ in a fixed order. Finally, we denote $s = c_1 \ldots c_m$ and we define a pair of words over the alphabet B:

$$u = sstt\ ttss \quad \text{and} \quad v = sstt\ dsd\ ttss\ .$$

We show that this pair of words contradicts the assumption, namely we show
(i) $(u, v) \notin \beta_B$ and
(ii) for each $\varphi : B \to A$ we have $\varphi(u)\ \beta_A\ \varphi(v)$.

To prove the first claim we can consider the factorization

$$g = (sstt,\ d,\ s,\ d,\ ttss)$$

of the word v. For this g there is no factorization f of the word u such that $f \leq_\alpha g$ because there are no two consecutive occurrences of d in u such that the word between them has a content equal to the set A.

The second claim is more complicated. Let $\varphi : B \to A$ be a mapping. We consider two cases.

I) First assume that there is a letter $c_i \in A$ such that $\varphi(c_i) = \varphi(d)$. Then we consider the mapping $\varphi' : B \to A$ such that $\varphi'|_A$ is the identity mapping and $\varphi'(d) = c_i$ and the mapping $\varphi'' : A \to A$ such that $\varphi''(c) = \varphi(c)$ for each $c \in A$. Then $\varphi = \varphi'' \circ \varphi'$ and it is enough to show that $\varphi'(u)\ \beta_A\ \varphi'(v)$, since the rest is a consequence of the fact that β is fully invariant. Let g be an arbitrary factorization of

$$\varphi'(v) = ss\ \varphi'(t)\varphi'(t)\ c_i s c_i\ \varphi'(t)\varphi'(t)\ ss$$

where $g = (g_0, a, g_1, b, g_2)$ with $a, b \in A$, $g_0, g_1, g_2 \in A^*$. We want to show the existence of a factorization $f = (f_0, a, f_1, b, f_2)$ of $\varphi'(u)$ such that $\mathsf{cont}(f_0) = \mathsf{cont}(g_0)$, $\mathsf{cont}(f_1) = \mathsf{cont}(g_1)$, $\mathsf{cont}(f_2) = \mathsf{cont}(g_2)$ and $f_0 a f_1 b f_2 = \varphi'(u)$. We distinguish several cases:

1a) "$\mathsf{cont}(g_0) \neq A$, $\mathsf{cont}(g_1) \neq A$"
Then $g_0 a g_1 b$ is a prefix of the prefix ss of the word $\varphi'(v)$, i.e. $ss = g_0 a g_1 b h$ for some $h \in A^*$. Hence $\mathsf{cont}(g_2) = A$, and we can put $f_0 = g_0$, $f_1 = g_1$, $f_2 = h\varphi'(t)\varphi'(t)c_i s c_i \varphi'(t)\varphi'(t)ss$.

1b) "$\mathsf{cont}(g_0) \neq A$, $\mathsf{cont}(g_1) = A$, $\mathsf{cont}(g_2) \neq A$"
Then g_0 is a prefix of the first s in $\varphi'(v)$ and g_2 is a suffix in the last s in $\varphi'(v)$. We can put $f_0 = g_0$, $f_2 = g_2$ and f_1 is an appropriate word.

1c) “$\mathsf{cont}(g_0) \neq A$, $\mathsf{cont}(g_1) = A$, $\mathsf{cont}(g_2) = A$”
Then g_0 is a prefix of the first s in $\varphi'(v)$, i.e. we put $f_0 = g_0$ and we can choose b from the last but one s from $\varphi'(u)$ and define f_1 and f_2 adequately.
Altogether we finished the case of $\mathsf{cont}(g_0) \neq A$.

2) Dually we can solve the cases of $\mathsf{cont}(g_2) \neq A$.

3) Assume $\mathsf{cont}(g_0) = \mathsf{cont}(g_2) = A$. And in addition we assume:

3a) “$\mathsf{cont}(g_1) = A$”
Then we can choose a from the second s in $\varphi'(u)$ and b from the last but one s in $\varphi'(u)$ and define f_0, f_1, f_2 in the expected way.

3b) “$\mathsf{cont}(g_1) \neq A$ and $\mathsf{cont}(ag_1b) \neq A$”
Then there is a word f_1 of length at most $m-1$ such that $\mathsf{cont}(f_1) = \mathsf{cont}(g_1)$ and the word af_1b is equal to some s_j. Hence we can find the word af_1b as a factor of the first occurrence $\varphi'(t)$ in $\varphi'(u)$ and then define f_0 and f_2.

3c) “$\mathsf{cont}(g_1) \neq A$ and $\mathsf{cont}(ag_1b) = A$, $c_i \in \mathsf{cont}(g_1)$”
Then we can find some s_j such that $w = c_is_jc_i$ has the property $\mathsf{cont}(w) = \mathsf{cont}(g_1)$. Further ads_jdb is a factor of t, hence we can put $f_1 = w$ and af_1b is a factor of the first occurrence of $\varphi'(t)$ in $\varphi'(u)$. As usually, we denote f_0 and f_2 as needed.

3d) “$\mathsf{cont}(g_1) \neq A$ and $\mathsf{cont}(ag_1b) = A$ and $c_i \notin \mathsf{cont}(g_1)$”
Then $a = c_i$ or $b = c_i$.
If $a = b = c_i$ then we can find s_j such that $\mathsf{cont}(s_j) = \mathsf{cont}(g_1)$ and $c_is_jc_i$ is a factor of the first occurrence of $\varphi'(t)$ in $\varphi'(u)$. Thus we consider the factorization f of u where f_1 is equal to this occurrence of s_j.
If $a = c_i$, $b \neq c_i$ then we can find f_1 such that f_1b is one of s_j with $\mathsf{cont}(f_1b) = \mathsf{cont}(g_1b)$ because $c_i \notin \mathsf{cont}(g_1b)$, i.e. $\mathsf{cont}(s_j) \neq A$. The case $a \neq c_i$, $b = c_i$ is dual.

II) Now assume that there is no such a letter. This means that $\varphi(c_i) = \varphi(c_{i'})$ for some different $i, i' \in \{1, \ldots, m\}$. Considerations are analogous to that of Case I). □

Remark. 1. If a positive variety of languages is locally finite we can generate the corresponding pseudovariety of ordered monoids by finitely generated free monoids. We are able to present effectively the free ordered monoids in pseudovarieties corresponding to $\mathsf{PPol}_k\mathcal{V}$ and $\mathsf{BPol}_k\mathcal{V}$ for $\mathcal{V}$ being any of $\mathcal{T}$, $\mathcal{S}^+$, $\mathcal{S}$, $\mathcal{A}_m$. It would be desirable to put a closer look into their structures.

2. For each positive variety of languages $\mathcal{V}$ the pseudovariety of ordered monoids corresponding to $\mathsf{PPol}_k\mathcal{V}$ is generated by the Schützenberger products of the form $\diamondsuit_{k+1}(M_0, \ldots, M_k)$ where $M_0, \ldots, M_k \in \mathbf{V}$ (see [9]). Notice that our Proposition 3 follows from results from [9].

References

1. Almeida, J.: Finite Semigroups and Universal Algebra. World Scientific, Singapore (1994)
2. Bloom, S.L.: Varieties of ordered algebras. J. Comput. Syst. Sci. 13(2), 200–212 (1976)

3. Burris, S., Sankappanavar, H.P.: A Course in Universal Algebra. Springer, Berlin (1981)
4. Blanchet-Sadri, F.: Equations and monoids varieties of dot-depth one and two. Theoret. Comput. Sci. 123, 239–258 (1994)
5. Eilenberg, S.: Automata, Languages and Machines, vol. B. Academic Press, New York (1976)
6. Klíma, O., Polák, L.: Hierarchies of piecewise testable languages. In: Ito, M., Toyama, M. (eds.) DLT 2008. LNCS, vol. 5257, pp. 479–490. Springer, Heidelberg (2008)
7. Pin, J.-E.: A variety theorem without complementation. Russian Mathem (Iz. VUZ) 39, 74–83 (1995)
8. Pin, J.-E.: Syntactic semigroups. In: Rozenberg, G., Salomaa, A. (eds.) Handbook of Formal Languages, ch. 10. Springer, Heidelberg (1997)
9. Pin, J.-E.: Algebraic tools for the concatenation product. Theoretical Computer Science 292, 317–342 (2003)
10. Simon, I.: Piecewise testable events. In: ICALP 1975. LNCS, vol. 33, pp. 214–222. Springer, Heidelberg (1975)
11. Volkov, M.V.: Reflexive relations, extensive transformations and piecewise testable languages of a given height. Int. J. Algebra and Computation 14, 817–827 (2004)