

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 a fixed natural number k the class of all (positive) boolean combinations of the restricted polynomials, i.e. the languages of the form $L_0 a_1 L_1 a_2 \dots a_\ell L_\ell$, where $\ell \leq k$, a_1, \dots, a_ℓ are letters and L_0, \dots, L_ℓ are 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 considered languages. We also characterize the property that a variety of languages is generated by a finite number of languages. We apply our constructions for 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 the languages of the form

$$L_0 a_1 L_1 a_2 \dots a_\ell L_\ell, \quad (*)$$

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 operator on classes of languages leads to several concatenation hierarchies. Well-known cases are the Straubing-Thérien and the group ones. Concatenation hierarchies has been intensively studied by many authors – see Section 8 of the Pin’s Chapter [9]. 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 a membership problem for the level 2 in the Straubing-Thérien hierarchy, i.e. a decision problem whether a given regular language can be written as a boolean combination of polynomials over languages 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 $B_i \subseteq A$ ($i =$

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

$0, \dots, \ell$). So this instance of polynomial operator is the most important case to study.

The restricted case, i.e. the case when we fix a natural number k and we allow only $\ell \leq k$ in $(*)$, mainly in the case that \mathcal{V} is the trivial variety was considered 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. 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 characteristic which are certain relations on $\{x_1, x_2, \dots\}^*$. 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 $\rho^d = \{(v, u) \in S \times S \mid (u, v) \in \rho\}$. A *quasiorder* ρ on a set S is a reflexive and transitive relation. Let $\hat{\rho} = \rho \cap \rho^d$ be the corresponding equivalence relation. For a mapping $\phi : S \rightarrow T$, let $\text{im } \phi = \{\phi(s) \mid s \in S\}$.

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

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

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

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 $b \in M$. 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 $\phi : A^* \rightarrow M$ and an ideal I of (M, \leq) such that $L = \phi^{-1}(I)$.

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

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

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

Now we recall here the basics concerning the Eilenberg type theorems. The boolean case was invented by Eilenberg [5] and the positive case was introduced by Pin [8].

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 $u^{-1}Lv^{-1} = \{w \in A^* \mid u w v \in L\} \in \mathcal{V}(A)$,
- \mathcal{V} is closed under preimages in morphisms, i.e.
 $f : B^* \rightarrow A^*$, $L \in \mathcal{V}(A)$ 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 .

A *pseudovariety* of finite monoids is a class of finite monoids closed under submonoids, morphic images and products of finite families. Similarly for ordered monoids (see [9]). 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 $\mathbf{M}(L) = A^*/\sim_L$ 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 $\mathbf{M}(L) = A^*/\sim_L$, 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 $\mathbf{O}(L)$.

Result 1 (Eilenberg [5], Pin [8]) *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*

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

The pseudovarieties of ordered monoids can be characterized by pseudoidentities (see e.g. [1] or [7]). 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 finite monoid M (ordered monoid (M, \leq)) if for each morphism $\phi : X^* \rightarrow M$ we have $\phi(u) = \phi(v)$ ($\phi(u) \leq \phi(v)$). In such a case we write $M \models u = v$, and for a set of identities Π , we define $\text{Mod}(\Pi) = \{M \mid (\forall \pi \in \Pi) M \models \pi\}$. For a class \mathcal{M} of monoids, the meaning of $\mathcal{M} \models \Pi$ is that, for each $M \in \mathcal{M}$, we have $M \models \Pi$. Let $\text{Id}(\mathbf{V})$ be the set of all identities which are satisfied in a variety of ordered monoids \mathbf{V} .

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

Result 2 *For a positive variety \mathcal{V} the following holds.*

- (i) \mathcal{V}^c is a positive variety.
- (ii) Let $\mathcal{V} \vee \mathcal{V}^c$ be the smallest positive variety containing both \mathcal{V} and \mathcal{V}^c . Then $(\mathcal{V} \vee \mathcal{V}^c)(A)$ consists of all positive boolean combinations of the languages from $\mathcal{V}(A) \cup \mathcal{V}^c(A)$.
- (iii) The class $\mathcal{V} \vee \mathcal{V}^c$ 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.

Examples.

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. Let $\mathcal{A}_m(A)$ be the set of all boolean combinations of the languages of the form $L(a, r) = \{u \in A^* \mid |u|_a \equiv r \pmod{m}\}$, 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, we have

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

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

The names for the (ordered) monoids of the pseudovarieties $\mathbf{T}, \mathbf{S}^+, \mathbf{S}, \mathbf{A}_m$ are *trivial monoids* (semilattices with the smallest element 1, semilattices and abelian groups of index m , respectively) – see Pin [9].

3 Locally Finite Varieties of Languages

In this paper we concentrate on positive varieties of languages which corresponds to locally finite pseudovarieties of ordered monoids. Each such pseudovariety is formed by finite members of locally finite (i.e. finitely generated ordered monoids are finite) variety of ordered monoids 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 and which is formally defined in the following definition.

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 every $u, v, w \in X^*$ we have

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

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

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

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

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

$$u \gamma_A v \quad \text{if and only if} \quad \text{for all } \varphi : A^* \rightarrow X^*, \text{ we have } \varphi(u) \gamma \varphi(v) .$$

It follows from the property (iii) in Definition 1 that in the previous definition of γ_A we can use just one morphism given by a fixed injective mapping $\phi : A \rightarrow X$. 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. $A^*/\widehat{\gamma_A}$ is finite).

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

The pseudovariety \mathbf{FinV}_γ of all finite members from \mathbf{V}_γ corresponds to the positive variety \mathcal{V}_γ of languages by

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

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

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

We present the finite characteristics for our four basic examples.

Examples (a Continuation 1).

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

Lemma 1. *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}_γ ;
- (ii) γ^d is a finite characteristic of the positive variety \mathcal{V}^c ;
- (iii) $\hat{\gamma}$ is a finite characteristic of the positive variety $\mathcal{V} \vee \mathcal{V}^c$;

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 $\mathbf{O}(L)$ is a morphic image of A^*/γ_A . The last is equivalent to $\mathbf{O}(L) \in \text{Fin } V_\gamma$, which means $L \in \mathcal{V}_\gamma(A)$.

“(ii)” It is clear that γ^d is a finite characteristic and the statement follows from the fact that $\preceq_{L^c} = (\preceq_L)^d$.

“(iii)” It is clear that $\hat{\gamma}$ is a finite characteristic and the statement follows from Result 2. \square

Proposition 1. *Let \mathcal{V} be a positive variety of languages and \mathbf{V} be a 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 $\phi : A^* \rightarrow M' = \prod_{i \in I} M_i$ be a morphism, and let $\pi_i : M' \rightarrow M_i$ be the i -th projection ($i \in I$). We want to show that $M = \text{im } \phi$ is finite.

For each $m \in M$, we have $\phi^{-1}([m]) = \bigcap_{i \in I} L_i$ where $L_i = (\pi_i \circ \phi)^{-1}([m])$. We have $L_i \in \mathcal{V}(A)$ since 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 $\mathbf{W} = \text{HSP } \mathbf{V}$ be the variety of ordered monoids generated by the pseudovariety \mathbf{V} . We show that the variety \mathbf{W} is locally finite. Indeed, let M be an ordered submonoid of $\prod_{i \in I} M_i$ where each $M_i \in \mathbf{V}$. Let ϕ be a

surjective morphism of M onto an ordered monoid N with a finite generating set G . Then $\phi^{-1}(G)$ generates in M a finite ordered monoid and N is its image.

It follows that $\gamma = \text{Id } \mathbf{W}$ 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 . \square

A positive variety \mathcal{V} is called *locally finite* if it satisfies the condition (i) from 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 $\text{PPol}_k\mathcal{V}$ of *positive polynomials* of length at most k of languages from the class \mathcal{V} . Namely, for a finite alphabet A , $\text{PPol}_k\mathcal{V}(A)$ consists of finite unions of finite intersections of the languages of the form

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

Similarly we define the classes $\text{BPol}_k\mathcal{V}$ of *boolean polynomials* using all finite boolean combinations of languages of the form $(*)$. Clearly $\text{PPol}_k\mathcal{V} \subseteq \text{PPol}_{k'}\mathcal{V}$ for $k \leq k'$. We denote the union of all $\text{PPol}_k\mathcal{V}$'s by $\text{PPol}(\mathcal{V})$. Similarly for $\text{BPol}_k\mathcal{V}$'s.

Lemma 2. *If \mathcal{V} is a positive variety of languages then $\text{PPol}_k\mathcal{V}$ is a positive variety of languages and $\text{BPol}_k\mathcal{V}$ is a boolean variety of languages.*

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

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

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

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

$$f \leq_\alpha g$$

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

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

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

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

Examples. (a Continuation 2)

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

2. and 3. One can show that $\text{PPol } \mathcal{S}^+ = \text{PPol } \mathcal{S}$ is the 3/2-level and $\text{BPol } \mathcal{S}^+ = \text{BPol } \mathcal{S}$ is the second level.

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

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

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

“(i)” The reflexivity of the relation $\mathbf{p}_k(\alpha)$ is trivial. The transitivity follows from the transitivity of the relation \leq_α .

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

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

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

$$h = (u_0, a_1, u_1, \dots, a_i, u'_i v''_i, a_{i+1}, \dots, a_\ell, v_\ell)$$

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

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

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

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

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

$\varphi(u) = \varphi(t_0)\varphi(b_1)\varphi(t_1)\dots\varphi(b_m)\varphi(t_m)$ has a factorization f' such that $f' \leq_\alpha g'$. We can conclude that $(\varphi(u), \varphi(v)) \in \mathbf{p}_k(\alpha)$.

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

“(v)” Denote for simplicity $\mathbf{p}_k(\alpha)_A = \beta$ a compatible quasiorder on A^* with a finite index.

” \implies ” We prove that for every language

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

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 $pvq = v_0 a_1 v_1 \dots a_\ell v_\ell$, where $v_i \in L_i$ for every $i \in \{0, \dots, \ell\}$. 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

$$p = v_0 a_1 \dots v'_i, \quad v = v''_i a_{i+1} \dots a_j v'_j \quad \text{and} \quad q = v''_j a_{j+1} \dots a_\ell v_\ell$$

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

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

In the first case we have $g = (v''_i, a_{i+1}, \dots, a_j, v'_j)$ a factorization of v . We assumed that $u \beta v$, so there is a factorization $f = (u''_i, a_{i+1}, \dots, u'_j)$ of u such that $(u''_i, v''_i) \in \alpha_A, (u_{i+1}, v_{i+1}) \in \alpha_A, \dots, (u'_j, v'_j) \in \alpha_A$. 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 $u_{i+1} \in L_{i+1}, \dots, u_{j-1} \in L_{j-1}$ 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 $u \preceq_L v$.

“ \Leftarrow ”: Let $\beta \subseteq \preceq_L$. This means that L is a finite union of languages of the form

$$\beta v = \{ u \in A^* \mid u \beta v \}.$$

So, it is enough to prove that each βv belongs to $\mathbf{PPol}_k \mathcal{V}(A)$. Consider all possible factorizations of the word v of length at most k , i.e. elements of the set $\mathbf{Fact}_k(v)$. 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 $\{g_1, g_2, \dots, g_m\} = \mathbf{Fact}_k(v)$. 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 $\text{PPol}_k \mathcal{V}$ 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 $(u_{i0}, v_{i0}) \in \alpha_A$. So, there is a factorization f_i of u such that $f_i \leq_\alpha g_i$. Consequently $(u, v) \in \mathbf{p}_k(\alpha)_A$, 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 $i \in \{1, \dots, m\}$, hence $u \in K$. \square

The following lemmas concern the preservation of aperiodicity (i.e. monoids have only trivial subgroups). Their proofs can be found in Appendix.

Lemma 3. *Let k be a natural number. If $(x^n, x^{n+1}) \in \alpha$ then $(x^{m-1}, x^m) \in \mathbf{p}_k(\alpha)$ for $m = (k+1)(n+1)$. If $(x^n, x^{n+1}) \in \widehat{\alpha}$ then $(x^{m-1}, x^m) \in \widehat{\mathbf{p}_k(\alpha)}$ for $m = (k+1)(n+1)$.*

Lemma 4. *Let \mathcal{V} be a positive variety with finite characteristic α , such that the corresponding pseudovariety of monoids contains only aperiodic monoids. For a natural number k , the pseudovariety of monoids corresponding to the positive variety of languages $\text{PPol}_k \mathcal{V}$ contains only aperiodic monoids too.*

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 $\text{PPol}_k \mathcal{T}$, k a natural number, are generated by a single ordered monoid. We show such result also for the positive varieties $\text{PPol}_k \mathcal{S}^+$ and we prove that this is not true for the positive varieties $\text{PPol}_k \mathcal{S}$. 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:*

$$((\forall \varphi : B \rightarrow A) \varphi(u) \alpha_A \varphi(v)) \text{ implies } u \alpha_B v .$$

Example 1. The finite characteristic of the positive variety \mathcal{S}^+ was described in Examples. It is finitely determined since it suffices to take $A = \{a, a'\}$, $a \neq a'$. Indeed, for $u, v \in B^*$ with $b \in \text{cont}(u) \setminus \text{cont}(v)$ take ϕ sending b to a and other (possible) elements of B to a' . The same is 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.*

The proof is placed into Appendix.

Example 2. In paper [6] the authors proved that $\text{PPol}_k \mathcal{T}$ is generated by a language $A^* a_1 A^* a_2 \dots a_k A^*$ where a_1, \dots, a_k are different letters and $A = \{a_1, \dots, a_k\}$. Hence the corresponding finite characteristic is finitely determined. One can show that an alphabet with $k + 1$ letters suffices.

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

Proof. Although the direct proof is possible we use Proposition 2 and show that finite characteristic $\beta = \mathbf{p}_k(\alpha)$ of $\text{PPol}_k \mathcal{S}^+$ is finitely determined. Let A be an alphabet containing 2^{2k+1} letters:

$$A = \{a_r \mid r = (r_1, r_2, \dots, r_{2k+1}) \text{ where } r_i \in \{0, 1\} \text{ for each } i \in \{1, 2, \dots, 2k+1\}\}.$$

We prove the property from Definition 2 for this set A . Let B be a finite alphabet and assume that $u, v \in B^*$ are such that

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

We want to prove $u \beta_B v$. So, let $g = (v_0, b_1, v_1, \dots, b_\ell, v_\ell) \in \text{Fact}_k(v)$ be an arbitrary factorization of the 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 $c \in \text{cont}(v_i)$ and $r_{2i} = 1$ iff $c = b_i$. So, we have defined a mapping $\varphi : B \rightarrow A$. Note that if a letter c does not occur in v then $\varphi(c) = a_{(0,0,\dots,0)}$ by this definition. Now $\varphi(u) \beta_A \varphi(v)$ and there exists a factorization f' of $\varphi(u)$ such that $f' \leq_\alpha g' = (\varphi(v_0), \varphi(b_1), \varphi(v_1), \dots, \varphi(b_\ell), \varphi(v_\ell))$. 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 $f = (u_0, b_1, u_1, \dots, b_\ell, u_\ell)$ of u such that $\varphi(u_i) \alpha_A \varphi(v_i)$ for each $i \in \{0, \dots, \ell\}$. We show that this implies $u_i \alpha_B v_i$. First, recall that in general $t \alpha w$ iff $\text{cont}(t) \subseteq \text{cont}(w)$. So, let $d \in \text{cont}(u_i)$ be an arbitrary letter from the alphabet B . Then $\varphi(d) \in \text{cont}(\varphi(u_i)) \subseteq \text{cont}(\varphi(v_i))$. Say that $\varphi(d) = a_r$. Then $a_r \in \text{cont}(\varphi(v_i))$ implies that $r_{2i+1} = 1$. If $d \notin \text{cont}(v_i)$ then $r_{2i+1} = 0$ by the definition of the mapping φ . Hence $d \in \text{cont}(v_i)$. So, we have $u_i \alpha_B v_i$ for each $i = 0, \dots, \ell$. For a given $g \in \text{Fact}_k(v)$ we found $f \in \text{Fact}_k(u)$ such that $f \leq_\alpha g$. This means that we really proved $u \beta_B v$. \square

The proofs of the following two propositions are placed in the Appendix.

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

Proposition 5. *For $k = 2$ the positive variety $\text{PPol}_k\mathcal{S}$ is not generated by a finite number of languages.*

Remark 2. 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 $\text{PPol}_k\mathcal{V}$ and $\text{BPol}_k\mathcal{V}$ 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 boolean variety of languages \mathcal{V} the pseudovariety of monoids corresponding to $\text{BPol}_k\mathcal{V}$ is generated by the Schützenberger products of the form $\diamond_{k+1}(M_0, \dots, M_k)$ where $M_0, \dots, M_k \in \mathbf{V}$ (see [7]). In particular cases one can put several restriction on M_i 's.

References

1. J. Almeida, *Finite Semigroups and Universal Algebra*, World Scientific, 1994
2. S. L. Bloom, Varieties of ordered algebras. *J. Comput. Syst. Sci.* **13** (2) (1976), 200–212
3. S. Burris and H. P. Sankappanavar, *A Course in Universal Algebra*, Springer, Berlin, 1981
4. F. Blanchet-Sadri, Equations and monoids varieties of dot-depth one and two, *Theoret. Comput. Sci.* **123** (1994), 239–258
5. S. Eilenberg, *Automata, Languages and Machines*, Vol B, Academic Press, New York, 1976
6. O. Klíma and L. Polák, Hierarchies of piecewise testable languages, in *Developments in Language Theory (DLT 2008)*, Lecture Notes in Computer Science 5257, 479–490, 2008
7. J.-E. Pin, *Varieties of Formal Languages*, North Oxford Academic, Plenum, 1986
8. J.-E. Pin, A variety theorem without complementation, *Russian Mathem. (Iz. VUZ)* **39** (1995), 74–83
9. J.-E. Pin, Syntactic semigroups, Chapter 10 in *Handbook of Formal Languages*, G. Rozenberg and A. Salomaa eds, Springer, 1997
10. I. Simon, Piecewise testable events, in *Proc. ICALP 1975*, LNCS Vol. 33 (1975), 214–222
11. M. V. Volkov, Reflexive relations, extensive transformations and piecewise testable languages of a given height, *Int. J. Algebra and Computation* **14** (2004), 817–827

6 Appendix

Proof. (of Lemma 3) 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 $i_0 + i_1 + \dots + i_\ell + \ell = m$ and i_0, \dots, i_ℓ are non-negative integers. Assume that for every $j \in \{0, \dots, \ell\}$ we have $i_j \leq n$, then $i_0 + i_1 + \dots + i_\ell + \ell \leq (\ell + 1)n + \ell \leq (k + 1)n + k < (k + 1)(n + 1) = m$ a contradiction. So, 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 we have a factorization f of x^{m-1} such that $f \leq_\alpha g$. This proves $(x^{m-1}, x^m) \in \mathbf{p}_k(\alpha)$. The implication $(x^{n+1}, x^n) \in \alpha \implies (x^m, x^{m-1}) \in \mathbf{p}_k(\alpha)$ can be proved in the similar way. \square

Proof. (of Lemma 4) 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 3 we have $(x^{m-1}, x^m) \in \widehat{\mathbf{p}_k(\alpha)}$ for 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 $\mathbf{PPol}_k\mathcal{V}$ contains only aperiodic monoids because each of them is a morphic images of the monoid B^*/α_B for some B . \square

Proof. (of Proposition 2) “(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)” It is well known that if \mathbf{V} is generated by a single ordered monoid (M, \cdot, \leq) then for every finite set B there is a finite free ordered monoid over the set B . If we take the free ordered monoid over X in the variety $\langle M \rangle_{var}$ of finite ordered monoids generated by the monoid M and denote α the kernel of the projection from X^* onto this free ordered monoid over X , then this α is finite characteristic of \mathcal{V} and the free ordered monoid over B can be viewed as B^*/α_B . Now put $A = M$ and we prove the property from Definition 2 for this set A .

At first, there is a natural morphism $\theta : A^* \rightarrow M$ 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. $\theta(a_1 a_2 \dots a_m) = a_1 \cdot a_2 \cdot \dots \cdot a_m$.

Let B be a finite alphabet and $u, v \in B^*$ be such that for each $\varphi : B \rightarrow A$ we have $\varphi(u) \alpha_A \varphi(v)$. We use the symbol φ also for the extension of the mapping $\varphi : B \rightarrow A$ to a morphism $\varphi : B^* \rightarrow A^*$. Moreover, each mapping $\varphi : B \rightarrow A = M$ determines the morphism $\bar{\varphi} : B^* \rightarrow M$, namely $\bar{\varphi} = \theta \circ \varphi$.

Recall that a free monoid over B in \mathbf{V} can be construct in the following way. There are only finitely many mappings $\varphi : B \rightarrow M$; 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 $\psi : B^* \rightarrow M^\Sigma$ given by $\psi(w) = (\psi(w)_\varphi)_{\varphi \in \Sigma}$, where

$\psi(w)_\varphi = \overline{\varphi}(w)$. Homomorphic image of ψ is a free monoid over B in \mathbf{V} and α_B is a kernel of ψ . Now for each $\varphi : B \rightarrow A = M$ we have $\varphi(u) \alpha_A \varphi(v)$, so, $\overline{\varphi}(u) \leq \overline{\varphi}(v)$ in M . 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 $\alpha_B w = \{t \in B^* \mid t \alpha_B w\}$.

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

$$u \alpha_B v \iff (\forall i \in \{1, \dots, m\}) \varphi_i(u) \alpha_A \varphi_i(v).$$

We show that

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

This equation 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 morphic preimages and the operation intersection.

“ \subseteq ” Let $t \in \alpha_B w$. Then $t \alpha_B w$ and we have $\varphi_i(t) \alpha_A \varphi_i(w) = w_i$ for each $i \in \{1, \dots, m\}$. Hence for each i we have $\varphi_i(t) \in \alpha_A w_i$ and consequently $t \in \varphi_i^{-1}(\alpha_A w_i)$.

“ \supseteq ” In fact, the previous implications are equivalences. \square

Proof. (of Prop. 3) Although the direct proof is possible we use Proposition 2 and show that finite characteristic $\beta = \mathbf{p}_k(\alpha)$ of $\mathbf{PPol}_k \mathcal{S}^+$ is finitely determined. Let A be an alphabet containing 2^{2k+1} letters:

$$A = \{a_r \mid r = (r_1, r_2, \dots, r_{2k+1}) \text{ where } r_i \in \{0, 1\} \text{ for each } i \in \{1, 2, \dots, 2k+1\}\}.$$

We prove the property from Definition 2 for this set A . Let B be a finite alphabet and assume that $u, v \in B^*$ are such that

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

We want to prove $u \beta_B v$. So, let $g = (v_0, b_1, v_1, \dots, b_\ell, v_\ell) \in \mathbf{Fact}_k(v)$ be an arbitrary factorization of the 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 $c \in \mathbf{cont}(v_i)$ and $r_{2i} = 1$ iff $c = b_i$. So, we have defined a mapping $\varphi : B \rightarrow A$. Note that if a letter c does not occur in v then $\varphi(c) = a_{(0,0,\dots,0)}$ by this definition. Now $\varphi(u) \beta_A \varphi(v)$ and there exists a factorization f' of $\varphi(u)$

such that $f' \leq_\alpha g' = (\varphi(v_0), \varphi(b_1), \varphi(v_1), \dots, \varphi(b_\ell), \varphi(v_\ell))$. 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 $f = (u_0, b_1, u_1, \dots, b_\ell, u_\ell)$ of u such that $\varphi(u_i) \alpha_A \varphi(v_i)$ for each $i \in \{0, \dots, \ell\}$. We show that this implies $u_i \alpha_B v_i$. First, recall that in general $t \alpha w$ iff $\text{cont}(t) \subseteq \text{cont}(w)$. So, let $d \in \text{cont}(u_i)$ be an arbitrary letter from the alphabet B . Then $\varphi(d) \in \text{cont}(\varphi(u_i)) \subseteq \text{cont}(\varphi(v_i))$. Say that $\varphi(d) = a_r$. Then $a_r \in \text{cont}(\varphi(v_i))$ implies that $r_{2i+1} = 1$. If $d \notin \text{cont}(v_i)$ then $r_{2i+1} = 0$ by the definition of the mapping φ . Hence $d \in \text{cont}(v_i)$. So, we have $u_i \alpha_B v_i$ for each $i = 0, \dots, \ell$. For a given $g \in \text{Fact}_k(v)$ we found $f \in \text{Fact}_k(u)$ such that $f \leq_\alpha g$. This means that we really proved $u \beta_B v$ \square

Proof. (of Prop. 4) We show that finite characteristic $\beta = \mathbf{p}_k(\alpha)$ of the variety $\text{PPol}_k\mathcal{S}$ where $k = 1$ is finitely determined on a six-element alphabet $A = \{a_0, a_1, a_2, a_3, a_4, a_5\}$.

Let B be a finite alphabet containing at least seven letters¹ and assume that for a given pair of words $u, v \in B^*$ we have

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

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

First of all, we take an arbitrary pair of different letters $b_1, b_2 \in B$ and consider the mapping $\varphi_{b_1, b_2} : B \rightarrow A$ given by the rule $\varphi_{b_1, b_2}(b_1) = a_1, \varphi_{b_1, b_2}(b_2) = a_2$ and $\varphi_{b_1, b_2}(c) = a_0$ for $c \in B \setminus \{b_1, b_2\}$. Since $(\varphi_{b_1, b_2}(u), \varphi_{b_1, b_2}(v)) \in \beta_A$ we see that the first letters in $\varphi_{b_1, b_2}(u)$ and $\varphi_{b_1, b_2}(v)$ coincide. In other words if the first occurrence of b_1 in the word v is before the first occurrence of b_2 in v then the first occurrence of b_1 in u is before the first occurrence of b_2 in u as well. It is also clear that $\text{cont}(\varphi_{b_1, b_2}(v)) = \text{cont}(\varphi_{b_1, b_2}(u))$ from which we observe $b_1 \in \text{cont}(v) \iff b_1 \in \text{cont}(u)$ and $\text{cont}(v) = \text{cont}(u)$ follows. We can summarize that $\{\text{cont}(v') \mid v' \text{ prefix of } v\} = \{\text{cont}(u') \mid u' \text{ prefix of } u\}$. 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 $\{\text{cont}(v') \mid v' \text{ suffix of } v\} = \{\text{cont}(u') \mid u' \text{ suffix of } u\}$.

So, there is a prefix u' of the word u such that $\text{cont}(u') = \text{cont}(g_0)$. 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 $\text{cont}(g_0) = \emptyset$, i.e. in the case $g_0 = \lambda$) and u_2 can be equal to u (when $\text{cont}(g_0) = \text{cont}(v)$). If u_1 is not the empty word then $u_1 = u'_1 b_1$ where $b_1 \in A$ and $b_1 \in \text{cont}(u_1) = \text{cont}(g_0)$, $b_1 \notin \text{cont}(u'_1)$. 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 $b_2 \in A$, $u'_2 \in A^*$ and $b_2 \notin \text{cont}(u_2) = \text{cont}(g_0)$. 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 $b_2 \notin \text{cont}(g_0)$, 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

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

suffix u_4 of u with the properties $\text{cont}(u_3) = \text{cont}(u_4) = \text{cont}(g_1)$. If $u_3 \neq \lambda$ then we denote its first letter b_3 , i.e. $u_3 = b_3 u'_3$ and we have $b_3 \in \text{cont}(u_3) = \text{cont}(g_1)$, $b_3 \notin \text{cont}(u'_3)$. If $u_4 \neq u$ then we denote $u = u'_4 b_4 u_4$ where $b_4 \in A$, $u'_4 \in A^*$, $b_4 \notin \text{cont}(u_4) = \text{cont}(g_1)$.

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 $\varphi : B \rightarrow A$ 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 $\varphi(f_1) \alpha_A \varphi(g_1)$. We show that $\text{cont}(f_0) = \text{cont}(g_0)$ and $\text{cont}(f_1) = \text{cont}(g_1)$.

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

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

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

Proof. (of Prop. 5) Assume that for $k = 2$ the finite characteristic $\beta = \mathbf{p}_k(\alpha)$ of the positive variety $\text{PPol}_k \mathcal{S}$ is finitely determined. Let $A = \{c_1, \dots, c_m\}$ be an alphabet for which the property from Definition 2 is satisfied. Let $B = A \cup \{d\}$. Assume that s_1, \dots, s_n are all words of length at most $m + 1$ over the alphabet A such that $\text{cont}(s_i) \neq A$ for $i \in \{1, \dots, n\}$. Further $t_{i_0 i_1 i_2} = d c_{i_0} d s_{i_1} d c_{i_2} d$ for all $i_1 \in \{1, \dots, n\}$, $i_0, i_2 \in \{1, \dots, m\}$ and t be a product of all words $t_{i_0 i_1 i_2}$ in one fixed order. Finally, we denote $s = c_1 \dots c_m$ 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 $\varphi : B \rightarrow A$ 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 such 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 $\varphi : B \rightarrow A$ 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 $\varphi' : B \rightarrow A$ such that $\varphi'|_A$ is the identity mapping and

the mapping $\varphi'(d) = c_i$ and $\varphi'' : A \rightarrow A$ 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. So, let g be a 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 $\text{cont}(f_0) = \text{cont}(g_0)$, $\text{cont}(f_1) = \text{cont}(g_1)$, $\text{cont}(f_2) = \text{cont}(g_2)$ and $f_0 a f_1 b f_2 = \varphi'(u)$. We distinguish some cases:

1a) “ $\text{cont}(g_0) \neq A$, $\text{cont}(g_1) \neq A$ ”

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 $\text{cont}(g_2) = A$ 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) “ $\text{cont}(g_0) \neq A$, $\text{cont}(g_1) = A$, $\text{cont}(g_2) \neq A$ ”

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) “ $\text{cont}(g_0) \neq A$, $\text{cont}(g_1) = A$, $\text{cont}(g_2) = A$ ”

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 finish the case when $\text{cont}(g_0) \neq A$.

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

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

3a) “ $\text{cont}(g_1) = A$ ”

Then we can choose a from the second s in $\varphi'u$ and b from the last but one s and define f_0, f_1, f_2 in expected way.

3b) “ $\text{cont}(g_1) \neq A$ and $\text{cont}(a g_1 b) \neq A$ ”

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

3c) “ $\text{cont}(g_1) \neq A$ and $\text{cont}(a g_1 b) = A$, $c_i \in \text{cont}(g_1)$ ” Then we can find some s_i such that $w = c_i s_i c_i$ has the property $\text{cont}(w) = \text{cont}(g_1)$. Further $a d s_i d b$ is a factor of t , hence we can put $f_1 = w$ and $a f_1 b$ 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) “ $\text{cont}(g_1) \neq A$ and $\text{cont}(a g_1 b) = A$ and $c_i \notin \text{cont}(g_1)$ ” Then $a = c_i$ or $b = c_i$. If $a = b = c_i$ then we can find s_i such that $\text{cont}(s_i) = \text{cont}(g_1)$ and $c_i s_i c_i$ is a factor of the first occurrence of $\varphi'(t)$ in $\varphi'(u)$. So we consider the factorization f of u where f_1 is equal to this occurrence of s_i .

If $a = c_i$, $b \neq c_i$ then we can find f_1 such that $f_1 b$ is one of s_i with $\text{cont}(f_1 b) = \text{cont}(g_1 b)$ because $c_i \notin \text{cont}(g_1 b)$, i.e. $\text{cont}(s_i) \neq A$. The case $a \neq c_i$, $b = c_i$ is dual.

II) Now assume that there is no such a letter. The proof is analogous to that of Case I). \square