Subhierarchies of the Second Level in the Straubing-Thérien Hierarchy

Ondřej Klíma and Libor Polák *

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

Abstract. In a recent paper we assigned 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 $L_0a_1L_1a_2 \dots a_\ell L_\ell$, where $\ell \leq k$, $a_1, \dots, a_\ell$ are letters and $L_0, \dots, L_\ell$ are from the variety $\mathcal{V}$. For this polynomial operator on a wide class of varieties we gave a certain algebraic counterpart which works with identities satisfied by syntactic (ordered) monoids of considered languages. Here we apply our constructions for particular examples of varieties of languages obtaining four hierarchies of (positive) varieties which have the 3/2 level and the second level of the Straubing-Thérien hierarchy as their limits. We concentrate here on inclusions among such varieties and we also discuss the existence of finite bases of identities for corresponding pseudovarieties of (ordered) monoids.
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

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

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 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 [8]. The main open problem concerning such 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

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

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 $B_i \subseteq A$ ($i = 0, \dots, \ell$). 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 [3], and in a recent paper by the authors [5]. In [6] we considered the restricted case in a general setting and we concentrated on identity problems for corresponding pseudovarieties and on the question whether they are generated by a single (ordered) monoid.

Here we study four hierarchies of languages which result by applying the restricted positive or boolean polynomial operator to the (positive) varieties where the class $\mathcal{V}(A)$ equals 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. Our basic questions are to explore all the inclusion among our varieties and we start to discuss the existence of finite bases for corresponding pseudovarieties of (ordered) monoids. Hopefully our results bring a bit more light into the complexity of the structure of (positive) subvarieties of the second level of the Straubing-Thérien hierarchy. In contrast to other paper dealing with concatenation hierarchies, e.g. Pin [9], which use mainly algebraic tools, our methods belongs rather to combinatorics on words.

Section 2 summarizes the background concerning positive varieties of languages and corresponding classes of (finite) ordered monoids. In Section 3 we overviews the necessary material from [6] dealing with locally finite varieties and polynomial operators. Next section investigates the inclusions among members of our hierarchies. The last section is devoted to the existence of finite bases of identities for pseudovarieties corresponding to our hierarchies in the case $k = 1$.

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. For such a relation, let $\widehat{\rho} = \rho \cap \rho^{\mathsf{d}}$ (sometimes we write also $(\rho)\widehat{\ }$) 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 homomorphisms.

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

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

For a set $Z \subseteq Y$, let $\overline{Z} = \{ u \in Y^* \mid \mathsf{c}(u) = Z \}$.

Now we recall here the basics concerning the Eilenberg type theorems. The boolean case was invented by Eilenberg [4] and the positive case was introduced by Pin [7].

A *positive 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 and finite intersections (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 preimages in 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 *boolean* variety of languages, we use in the first item complements, too.

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 [8]). A *variety* of (ordered) monoids is a class of monoids closed under submonoids, morphic images and arbitrary products. For a variety $\mathbf{V}$ of ordered monoids the class $\mathsf{Fin}\,\mathbf{V}$ of all finite members of $\mathbf{V}$ is a pseudovariety. We call such pseudovarieties *equational*. In fact, in our paper almost all considered pseudovarieties are equational. We fix the set $X = \{x_1, x_2, \dots\}$ of *variables*. An *identity* is an ordered pair $u = v$ $(u \le v)$ of words over X, i.e. $u, v \in X^*$. An identity $u = v$ ($u \le v$, respectively) is *satisfied* in a finite monoid M (ordered monoid $(M, \le)$) if for each morphism $\phi : X^* \to M$ we have $\phi(u) = \phi(v)$ $(\phi(u) \le \phi(v))$. In such a case we write $M \models u = v$, 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 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 $\mathbf{V}$.

A relation γ on X^* is

- *compatible* (with the multiplication), if for every $u, v, w \in X^*$, we have

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

- *fully invariant* if, for every morphism $\varphi : X^* \to X^*$ and $u, v \in X^*$, we have

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

Result 1 (see [2], [1]) *The operators* Id *and* Mod *are pairwise inverse bijections between varieties of ordered monoids and fully invariant compatible quasiorders on* X^*.

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 2 (Eilenberg [4], Pin [7]) *Boolean varieties (positive varieties) of languages correspond to pseudovarieties of finite monoids (ordered monoids). The correspondence, written* $\mathscr{V} \longleftrightarrow \mathbf{V}$ *(*$\mathscr{P} \longleftrightarrow \mathbf{P}$*), is given by the following relationship: for* $L \subseteq A^*$ *we have*

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

Examples (see Pin [8]).

1. 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. This class is a positive variety of languages and the corresponding (equational) pseudovariety of ordered monoids is given by

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

We speak about *semilattices with the smallest element 1.*

2. 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. This class is a boolean variety of languages and the corresponding (equational) pseudovariety of monoids is given by

$$\mathbf{S} = \mathsf{Fin\,Mod}(\, x^2 = x,\, xy = yx\,)$$

We speak about *semilattices.*

3 Locally Finite Varieties of Languages and Polynomial Operators of Bounded Length

Here we overview the necessary material from [6]. In that and in this paper we concentrate on concrete 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 (see below).

A relation γ on X^* is *a finite characteristic* if it is a fully invariant compatible quasiorder on the monoid X^* satisfying “for each finite subset Y of the set X, the set Y^* intersects only finitely many classes of $X^*/\,\widehat{\gamma}$ ”.

We fix notation for finite characteristics of the classes $\mathcal{S}^+$ and $\mathcal{S}$:

$$\sigma = \mathsf{Id}\,\mathbf{S}^+ = \{\,(u,v) \in X^* \times X^* \mid \mathsf{c}(u) \subseteq \mathsf{c}(v)\,\}\ ,$$

$$\widehat{\sigma} = \mathsf{Id}\,\mathbf{S} = \{\,(u,v) \in X^* \times X^* \mid \mathsf{c}(u) = \mathsf{c}(v)\,\}\ .$$

We can define *a natural adaptation* γ_A (or sometimes $\gamma(A)$) of a finite characteristic γ on each finite alphabet A by an identification of A with a subset of X (since γ is fully invariant, the definition does not depend on an identification we choose).

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 result is quite obvious. One can find its proof together with other results from this section in the author's manuscript [6].

Result 3 *Let $\mathcal{V}$ be a class of languages and γ be a finite characteristic of $\mathcal{V}$. Then $\mathcal{V}$ is a positive variety of languages, corresponding to the pseudovariety* $\mathbf{V} = \mathsf{Fin\,Mod}(\gamma)$

Result 4 *Let $\mathcal{V}$ be a positive variety of languages. Then the following conditions are equivalent.*
(i) For each finite alphabet A, the set $\mathcal{V}(A)$ is finite.
(ii) There exists a finite characteristic of $\mathcal{V}$.

A positive variety $\mathcal{V}$ is called *locally finite* if it satisfies the condition (i) from Result 4.

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

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

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

Result 5 *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.*

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

$$f = (u_0, a_1, u_1, a_2, u_2, \dots, u_{\ell-1}, a_\ell, u_\ell)$$

is *a factorization* of u of the 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 the length at most k of the word u is denoted by $\mathsf{Fact}_k(u)$. For a factorization $f = (u_0, a_1, u_1, \ldots, a_\ell, u_\ell)$ of a word $u \in A^*$ and a factorization $g = (v_0, b_1, v_1, \ldots, b_\ell, v_\ell)$ 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

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

Note that the relation $(\mathsf{p}_k(\alpha))_X$ is a finite characteristic and therefore the relation $(\mathsf{p}_k(\alpha))_A$ is equal to $((\mathsf{p}_k(\alpha))_X)_A$. We write $\mathsf{p}_k(\alpha)$ instead of $(\mathsf{p}_k(\alpha))_X$. Further we denote $\mathsf{b}_k(\alpha) = (\mathsf{p}_k(\alpha))\widehat{\;}$.

Result 6 *Let $\mathcal{V}$ be a locally finite positive variety of languages and α be a 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 $\mathsf{b}_k(\alpha)$.*

In this paper we study the hierarchies $\mathsf{PPol}_k(\mathcal{S}^+)$, $\mathsf{PPol}_k(\mathcal{S})$, $\mathsf{BPol}_k(\mathcal{S}^+)$ and $\mathsf{BPol}_k(\mathcal{S})$. We denote

$$\pi_k^+ = \mathsf{p}_k(\sigma) \;\; \text{and} \;\; \pi_k = \mathsf{p}_k(\widehat{\sigma}) \; .$$

Next we present finite characteristic for our first two hierarchies explicitly. Let $u, v \in X^*$, then

$$u \; \pi_k^+ \; v \;\; \text{iff} \;\; \forall \; (g_0, a_1, \ldots, g_\ell) \in \mathsf{Fact}_k(v) \; \exists \; (f_0, a_1, \ldots, f_\ell) \in \mathsf{Fact}_k(u)$$

$$\text{such that} \;\; \mathsf{c}(f_0) \subseteq \mathsf{c}(g_0), \ldots, c(f_\ell) \subseteq \mathsf{c}(g_\ell) \; ,$$

$$u \; \pi_k \; v \;\; \text{iff} \;\; \forall \; (g_0, a_1, \ldots, g_\ell) \in \mathsf{Fact}_k(v) \; \exists \; (f_0, a_1, \ldots, f_\ell) \in \mathsf{Fact}_k(u)$$

$$\text{such that} \;\; \mathsf{c}(f_0) = \mathsf{c}(g_0), \ldots, c(f_\ell) = \mathsf{c}(g_\ell) \; ,$$

For the remaining two hierarchies we use the equivalence closures of the relations above or we can write

$$u \; (\pi_k^+)\widehat{\;} v \;\; \text{iff the sets of minimal elements of } \mathsf{Fact}_k(u) \text{ and } \mathsf{Fact}_k(v) \text{ are equal} \, ,$$

$$u \; \widehat{\pi_k} \; v \;\; \text{iff} \;\; \{\, (\mathsf{c}(f_0), a_1, \mathsf{c}(f_1), \ldots, c(f_\ell)) \mid (f_0, a_1, f_1, \ldots, f_\ell) \in \mathsf{Fact}_k(u) \,\}$$

$$= \{\, (\mathsf{c}(g_0), a_1, \mathsf{c}(g_1), \ldots, c(g_\ell)) \mid (g_0, a_1, g_1, \ldots, g_\ell) \in \mathsf{Fact}_k(v) \,\} \; .$$

4 Inclusions between our subhierarchies

It is natural to look for all possible inclusions among members of our hierarchies. Clearly, for each k,

$$\mathsf{PPol}_k(\mathcal{S}^+) \subseteq \mathsf{PPol}_k(\mathcal{S}),\ \mathsf{PPol}_k(\mathcal{S}^+) \subseteq \mathsf{BPol}_k(\mathcal{S}^+)\ ,$$

$$\mathsf{PPol}_k(\mathcal{S}) \subseteq \mathsf{BPol}_k(\mathcal{S}),\ \mathsf{BPol}_k(\mathcal{S}^+) \subseteq \mathsf{BPol}_k(\mathcal{S})\ .$$

Proposition 1. *The hiearchies* $\mathsf{PPol}_k(\mathcal{S}^+)$, $\mathsf{PPol}_k(\mathcal{S})$, $\mathsf{BPol}_k(\mathcal{S}^+)$ *and* $\mathsf{BPol}_k(\mathcal{S})$ *are strict.*

Proof. It suffices to consider one-element alphabet $\{a\}$. Let k be a natural number. Then

$$(a^{k+2}, a^{k+1}) \in \pi_k^+ \setminus \pi_{k+1}^+,\ (a^{2k+1}, a^{2k+2}) \in \pi_k \setminus \pi_{k+1},$$

$$(a^{k+2}, a^{k+1}) \in (\pi_k^+)\widehat{\ } \setminus (\pi_{k+1}^+)\widehat{\ },\ (a^{2k+1}, a^{2k+2}) \in \widehat{\pi_k} \setminus \widehat{\pi_{k+1}}\ .$$

□

Proposition 2. *For each k, the varieties* $\mathsf{PPol}_k(\mathcal{S}^+)$, $\mathsf{PPol}_k(\mathcal{S})$, $\mathsf{BPol}_k(\mathcal{S}^+)$ *and* $\mathsf{BPol}_k(\mathcal{S})$ *are pairwise different.*

Proof. Fix k and take again $A = \{a\}$. Clearly,

$$(a^{k+2}, a^{k+1}) \in (\pi_k \cap (\pi_k^+)\widehat{\ }) \setminus \widehat{\pi_k}\ ,$$

$$(a^k, a^{k+1}) \in \pi_k^+ \text{ but not in } \pi_k \text{ nor in } (\pi_k^+)\widehat{\ },$$

$$(a^{k+1}, a^{k+2}) \in (\pi_k^+)\widehat{\ } \setminus \pi_k\ .$$

Finally one can show that for $A = \{a, b\}$,

$$(ab)^k aa(ab)^k, (ab)^k) \in \pi_k \setminus (\pi_k^+)\widehat{\ }.$$

□

One could expect that the members of our four hierarchies form a lattice isomorphic to the product of the square of the two-element lattice with the chain of natural numbers. We show here that the situation is more complicated. At present the work on the exact description is in progress.

Next we will discuss the question whether $\mathsf{PPol}_k(\mathcal{S}) \subseteq \mathsf{PPol}_n(\mathcal{S}^+)$ or $\mathsf{BPol}_k(\mathcal{S}) \subseteq \mathsf{BPol}_n(\mathcal{S}^+)$ are satisfied for some natural numbers k, n.

In paper [6] authors proved that $\mathsf{PPol}_1(\mathcal{S})$ is generated by a finite number of languages. Each of them belongs to some $\mathsf{PPol}_k(\mathcal{S}^+)$ by Proposition 7 and consequently we obtain the inclusion $\mathsf{PPol}_1(\mathcal{S}) \subseteq \mathsf{PPol}_k(\mathcal{S}^+)$ for some k. More precisely, it was shown in [6] that languages which generate the variety $\mathsf{PPol}_1(\mathcal{S})$ can be considered over six-element alphabet. When we use the techniques from the proof of Proposition 7 we see that each of these languages belongs to $\mathsf{PPol}_k(\mathcal{S}^+)$ for $k = 13$. This rough approximation can be improved by the following observation.

Proposition 3. $\mathsf{PPol}_1(\mathcal{S}) \subseteq \mathsf{PPol}_3(\mathcal{S}^+)$ *and* $\mathsf{BPol}_1(\mathcal{S}) \subseteq \mathsf{BPol}_3(\mathcal{S}^+)$.

Proof. It is clear that $\overline{\emptyset} = \emptyset^* = \{\lambda\}$ and $\overline{B} = \bigcap_{b\in B} B^*bB^*$ for $\emptyset \neq B \subseteq A$. Assume now that B and C are non-empty sets. We show the formula

$$\overline{B}a\overline{C} = \bigcap_{b\in B, c\in C} B^*bB^*aC^*cC^* \ .$$

The inclusion "$\subseteq$" is clear. Denote the language on the right-hand side by L and let $u \in L$. We look at the first occurrences of letters from B in u and choose the right-most, say b_0. So we have $u = u_0b_0u'$, where $B\setminus\{b_0\} \subseteq \mathsf{c}(u_0)$. We know that $u \in B^*b_0B^*aC^*cC^*$ for some $c \in C$. We can see that $\mathsf{c}(u_0) \subseteq B$ and therefore $\mathsf{c}(u_0) = B \setminus \{b_0\}$. Now we look at the last occurrences of letters from C in u and choose the left-most, say c_0. Since $u \in B^*b_0B^*aC^*c_0C^*$ it is clear that this occurrence of c_0 and also the occurrence of a belongs to u' and we can write $u = u_0b_0u_1au_2c_0u_3$, where $u_0, u_1 \in B^*$ and $u_2, u_3 \in C^*$. Now it is easy to see that $\mathsf{c}(u_3) = C \setminus \{c_0\}$ because c_0 is the left-most last occurrence. And we can conclude that $\mathsf{c}(u_0b_0u_1) = B$ and $\mathsf{c}(u_2c_0u_3) = C$. Hence $u \in \overline{B}a\overline{C}$.

When $B = \emptyset$ and $C \neq \emptyset$ one can prove the formula

$$\overline{\emptyset}a\overline{C} = \bigcap_{c\in C} \emptyset^*aC^*cC^*$$

and similarly in the other cases when C is the empty set.

Altogether we prove that every language of the form $\overline{B}a\overline{C}$ is from $\mathsf{PPol}_3(\mathcal{S}^+)$ and the statements follow. □

Note that the previous lemma can be prove also in different way, namely one can show that the inclusion $\pi_3^+ \subseteq \pi_1$ holds.

Since $(x^2, x^3) \in \pi_2^+$ but $(x^2, x^3) \notin \pi_1$ we see that the inclusion $\pi_2^+ \subseteq \pi_1$ does not hold and so $\mathsf{PPol}_1(\mathcal{S}) \not\subseteq \mathsf{PPol}_2(\mathcal{S}^+)$. In the case of boolean varieties we have the following observation which is proved in the next section where we introduce an alternative characterization of the relation π_1.

Proposition 4. $\mathsf{BPol}_1(\mathcal{S}) \subseteq \mathsf{BPol}_2(\mathcal{S}^+)$.

A proof is postponed to the end of this paper.

Now we show that such kind of an inclusion cannot be generalized in the case of $\mathsf{PPol}_k\mathcal{S}$ for $k > 1$.

Proposition 5. *There exists no number k such that* $\mathsf{PPol}_2(\mathcal{S}) \subseteq \mathsf{PPol}_k(\mathcal{S}^+)$.

More generally, there are no $n \geq 2$ and k such that $\mathsf{PPol}_n(\mathcal{S}) \subseteq \mathsf{PPol}_k(\mathcal{S}^+)$.

Proof. Let k be an arbitrary natural number.

We fix an alphabet $A = \{x_0, x_1, \dots, x_k, x_{k+1}\}$ containing $k+2$ letters. We show that $\pi_k^+(A) \not\subseteq \pi_2(A)$, i.e. we find a pair of words $u, v \in A^*$ such that $(u, v) \in \pi_k^+(A)$ but $(u, v) \notin \pi_2(A)$.

We start with the following observation.

Lemma 1. *Let $u, v \in A^*$ be words such that every word of length k over A is a factor of both words u and v. Let $w = (x_0 x_1 \ldots x_k x_{k+1})^{k+1}$. Then*

$$(wuw, wvw) \in \pi_k^+(A) \ .$$

Proof. Let $g = (g_0, a_1, g_1, \ldots, a_\ell, g_\ell)$, $\ell \leq k$ be a factorization of wvw. There is an index i such that $\mathsf{c}(g_i) = A$, otherwise the length of each g_i is at most $k+1$ and $|g_0 a_1 g_1 \ldots a_\ell g_\ell| \leq \ell \cdot (k+1) + \ell \leq k(k+2) < |w|$, so g can not be a factorization of wvw. Denote i the smallest index such that $\mathsf{c}(g_i) = A$. Then the word $g_0 a_1 g_1 \ldots g_{i-1} a_i$ is a prefix of $(x_0 x_1 \ldots x_k x_{k+1})^k$. In the same way we denote the largest index j such that $\mathsf{c}(g_j) = A$ and we have that the word $a_{j+1} g_{j+1} \ldots g_\ell$ is a suffix of $(x_0 x_1 \ldots x_k x_{k+1})^k$. Now we distinguish two cases, namely $i = j$ and $i < j$.

If $i = j$ then we construct the factorization $f = (f_0, a_1, f_1, \ldots, a_\ell, f_\ell)$ of the word wuw in the following way. We put $f_m = g_m$ for every $m \neq i$. Then $f_0 a_1 f_1 \ldots f_{i-1} a_i$ is a prefix of $(x_0 x_1 \ldots x_k x_{k+1})^k$ and $a_{i+1} f_{i+1} \ldots a_\ell f_\ell$ is a suffix of $(x_0 x_1 \ldots x_k x_{k+1})^k$. Hence f_i can be choose in a unique way and it is easy to see that $\mathsf{c}(f_i) = A = \mathsf{c}(g_i)$. So $f \leq_\sigma g$.

If $i < j$ then we start to construct the factorization $f = (f_0, a_1, f_1, \ldots, a_\ell, f_\ell)$ in the similar way, i.e. we put $f_m = g_m$ for every $m < i$ and $m > j$. Once again $f_0 a_1 f_1 \ldots f_{i-1} a_i$ is a prefix of $(x_0 x_1 \ldots x_k x_{k+1})^k$. So, we can denote $p \in A^*$ such that $f_0 a_1 f_1 \ldots f_{i-1} a_i p = w$ and $\mathsf{c}(p) = A$. Similarly, there is a word q such that $q a_{j+1} f_{j+1} \ldots a_\ell f_\ell = w$ and $\mathsf{c}(q) = A$. Further we put $f_m = \lambda$ for every $i < m < j$. We consider the word $a_{i+1} a_{i+2} \ldots a_j$ which has the length at most k. Hence it can be find as a factor in the word u, i.e. $u = u' a_{i+1} a_{i+2} \ldots a_j u''$ for some $u', u'' \in A^*$. Finally, we define $f_i = p u'$ and $f_j = u'' q$ for which we have $\mathsf{c}(f_i) = \mathsf{c}(f_j) = A$. Now it is easy to see that $f \leq_\sigma g$. So, for an arbitrary factorization g of wvw of length at most k we found a factorization f of wuw such that $f \leq_\sigma g$, this means that $(wuw, wvw) \in \pi_k^+(A)$. □

Remark 1. It is clear that for u, v satisfying the assumption in Lemma 1 we have even $(wuw, wvw) \in (\pi_k^+ \curlyvee)(A)$.

Now we are going to construct two special words u and v satisfying the assumption of the previous lemma, i.e. such that $(wuw, wvw) \in \pi_k^+(A)$ and for which we will be able to prove $(wuw, wvw) \notin \pi_2(A)$.

First we multiply (in some order) all words over A of the length k which contain the letter x_0 and we denote the resulting word u_0. Then we multiply (in some order) all words over A of the length k which do not contain the letter x_0 and we denote the resulting word u_1. We put $u = u_0 u_1$. Further, let v be a product of all words over A of length $k+2$ in some order.

By the previous lemma we have $(wuw, wvw) \in \pi_k^+(A)$. Since the word $x_0 x_1 x_2 \ldots x_{k-1} x_k x_0$ is a factor of v there is a factorization $g = (g_0, x_0, g_1, x_0, g_2)$ of the word wvw such that $\mathsf{c}(g_0) = \mathsf{c}(g_2) = A$ and $g_1 = x_1 x_2 \ldots x_{k-1} x_k$. We show that there is no factorization $f = (f_0, x_0, f_1, x_0, f_2)$ of wuw such that $\mathsf{c}(f_0) = \mathsf{c}(f_2) = A$ and $\mathsf{c}(f_1) = \{x_1, \ldots, x_k\}$.

If we look at two consecutive occurrences of x_0 in wuw then:

1. the first one is in the prefix w of wuw, and then the word between these two occurrences of x_0 contains x_{k+1};

2. both are in u, and then they are in u_0 and the word between them has the length at most $k-1$;

3. the second one is in the suffix w of wuw, and then the word between these two occurrences of x_0 contains x_{k+1}.

In all cases we see that such a factorization $f = (f_0, x_0, f_1, x_0, f_2)$ satisfying condition $\mathsf{c}(f_1) = \{x_1, \ldots, x_k\}$ does not exist. We can conclude with $(wuw, wvw) \notin \pi_2(A)$.

The second statement follows from $\mathsf{PPol}_2(\mathcal{S}) \subseteq \mathsf{PPol}_n(\mathcal{S})$. □

The following modification of Proposition 5 follows from the proof of the proposition and from the remark.

Proposition 6. *There exists no k such that $\mathsf{BPol}_2\mathcal{S} \subseteq \mathsf{BPol}_k\mathcal{S}^+$. Moreover, there are no $n \geq 2, k$ such that $\mathsf{BPol}_n(\mathcal{S}) \subseteq \mathsf{BPol}_\mathsf{k}(\mathcal{S}^+)$.*

□

Proposition 7. *It holds $\mathsf{PPol}(\mathcal{S}^+) = \mathsf{PPol}(\mathcal{S})$ and $\mathsf{BPol}(\mathcal{S}^+) = \mathsf{BPol}(\mathcal{S})$.*

Proof. Since $\mathcal{S}^+ \subseteq \mathcal{S}$ we have $\mathsf{PPol}(\mathcal{S}^+) \subseteq \mathsf{PPol}(\mathcal{S})$. To the opposite inclusion we have to show that an arbitrary language

$$L = \overline{B_0}a_1\overline{B_1}a_2 \ldots a_\ell\overline{B_\ell},\ a_1, \ldots, a_\ell \in A,\ B_0, \ldots, B_\ell \subseteq A$$

belongs to $\mathsf{PPol}(\mathcal{S}^+)(A)$. First observe that for an subset C of A consisting of the letters $c_1, c_2, \ldots, c_m$ we can write

$$\overline{C} = \bigcup_{\sigma \in \Sigma} C^* c_{\sigma(1)} C^* c_{\sigma(2)} C^* \ldots C^* c_{\sigma(m)} C^* \ , \tag{1}$$

where Σ is the set of all permutations of the set of indices $\{1, \ldots, m\}$. When we replace each $\overline{B_i}$ in the expression of L by the corresponding sum of languages using formula (1) we obtain that $L \in \mathsf{PPol}(\mathcal{S}^+)(A)$.

So $\mathsf{PPol}(\mathcal{S}^+) = \mathsf{PPol}(\mathcal{S})$ and the equality $\mathsf{BPol}(\mathcal{S}^+) = \mathsf{BPol}(\mathcal{S})$ follows. □

Notice that $\mathsf{PPol}(\mathcal{S}^+)$ is the 3/2 level of the Straubing-Thérien hierarchy and that $\mathsf{BPol}(\mathcal{S}^+)$ is the second level of this hierarchy (see Theorem 8.8 of [8]).

5 Bases of identities for varieties on the first step

Our goal now is to find some finite bases of identities for each variety from our hierarchies of varieties of languages. As the results from [3, 5] indicate we can not expect that such a finite basis exists for every k. We try to find them at least in the case $k = 1$.

Note that in this case we should consider factorizations of length ℓ with $\ell \leq 1$. But for $\ell = 0$ the condition $f \leq g$ for factorizations of a pair of words is exactly saying that the content of the considered words is equal or in inclusion. This information is contained in the condition $f \leq g$ for factorizations of length $\ell = 1$. So we need not pay attention to the factors of length $\ell = 0$.

5.1 Identities for $\mathsf{BPol}_1(\mathcal{S}^+)$ and $\mathsf{PPol}_1(\mathcal{S}^+)$

Let x, y be two different letters from X and $u \in X^*$ be a word which contains both x and y, i.e. $x, y \in \mathsf{c}(u)$. Then it is easy to see that $(uxyx, uyx) \in (\pi_1^+)\widehat{}$. Note that the identity

$$uxyx = uyx \ , \quad \text{where} \quad x, y \in \mathsf{c}(u) \tag{2}$$

is equivalent to a pair of identities: we distinguish two cases $u = u_1xu_2yu_3$ and $u = u_1yu_2xu_3$ for some $u_1, u_2, u_3 \in X^*$, so the identity (2) is equivalent to the identities

$$x_1\, x\, x_2\, y\, x_3 \cdot x\, y\, x = x_1\, x\, x_2\, y\, x_3 \cdot y\, x \ ,$$

$$x_1\, y\, x_2\, x\, x_3 \cdot x\, y\, x = x_1\, y\, x_2\, x\, x_3 \cdot y\, x \ .$$

When we put $x_1 = x_2 = y = \lambda$ and $x_3 = v$ in the previous identities then we obtain the identity $xvx^2 = xvx$.

We have also the dual version of the identity (2)

$$xyxu = xyu \quad \text{where} \quad x, y \in \mathsf{c}(u) \ .$$

When we will refer to the identity (2) we mean also this dual version. Other identity which is satisfied in $\mathsf{BPol}_1(\mathcal{S}^+)$ is

$$uxyv = uyxv \ , \quad \text{where} \quad x, y \in \mathsf{c}(u) \cap \mathsf{c}(v) \ . \tag{3}$$

Note that this identity represents in fact four identities.

Altogether we have the set consisting of eight identities represented by identities (2) and (3).

When we work with π_1^+ we observe that the identities

$$yuyx \leq yuxyx \quad \text{and} \quad xyuy \leq xyxuy \tag{4}$$

are satisfied in $\mathsf{PPol}_1(\mathcal{S}^+)$. Note that the identity $x \leq x^2$ follows from the identity (4). Other identity which is satisfied in $\mathsf{PPol}_1(\mathcal{S}^+)$ is

$$xuxvx \leq xuvx \ . \tag{5}$$

Note that the identity (2) is a consequence of the identities (4) and (5).

Proposition 8. *i) The identities (2) and (3) form a finite basis of identities for the variety of monoids corresponding to* $\mathsf{BPol}_1(\mathcal{S}^+)$.

ii) The identities (3), (4) and (5) form a finite basis of identities for the variety of ordered monoids corresponding to $\mathsf{PPol}_1(\mathcal{S}^+)$.

Proof. A proof follows from the following series of lemmas.

We need a bit more notation. For $u \in X^*$ we denote by:

- $\mathsf{first}(u)$ the sequence of the first occurrences of letters of u from the left,
- $\mathsf{last}(u)$ the sequence of the first occurrences of letters of u from the right, and by

- $\mathsf{Sub}_k(u)$ the set of all subwords of u of the length less or equal to k.

Let $u \in X^*$ be a word and $x, y \in \mathsf{c}(u)$, $x \neq y$ be letters such that $xy \notin \mathsf{Sub}_2(u)$ (i.e. the last occurrence of y is before the first occurrence of x in the word u). Then u can be written in the form $u = u_0 y u_1 x u_2$ where $y \notin \mathsf{c}(u_1 x u_2)$ and $x \notin \mathsf{c}(u_0 y u_1)$. We denote $\mathsf{int}_{y,x}(u) = \mathsf{c}(u_1)$.

Further we define the *skeleton* $\mathsf{skel}(u) \in X^*$ in the following way. We remove from u every occurrence of a given letter which is not the first or the last occurrence of this letter in the word u. After deleting of all "interior" occurrences of all letters from u the resulting word is the skeleton $\mathsf{skel}(u)$ of u.

Lemma 2. *Let $u, v \in X^*$ be arbitrary words. Then $(u, v) \in \pi_1^+$ if and only if the following conditions are satisfied*

i) $\mathsf{first}(u) = \mathsf{first}(v)$ *and* $\mathsf{last}(u) = \mathsf{last}(v)$*;*

ii) $\mathsf{Sub}_2(u) \subseteq \mathsf{Sub}_2(v)$*;*

iii) for every $x, y \in \mathsf{c}(u)$, $xy \notin \mathsf{Sub}_2(v)$ we have $\mathsf{int}_{y,x}(v) \subseteq \mathsf{int}_{y,x}(u)$*.*

Proof. Assume $(u, v) \in \pi_1^+$. It is easy to see that $\mathsf{c}(u) = \mathsf{c}(v)$.

Let x and y be two different letters from $\mathsf{c}(v)$. Assume that the first occurrence of the letter x is before the first occurrence of the letter y in v. Let $g = (g_0, x, g_1)$ be a factorization of v such that the central x is the first occurrence of the letter x in v, i.e. $x \notin \mathsf{c}(g_0)$. Under our assumption also $y \notin \mathsf{c}(g_0)$. There is a factorization $f = (f_0, x, f_1)$ of u such that $\mathsf{c}(f_0) \subseteq \mathsf{c}(g_0)$ and $\mathsf{c}(f_1) \subseteq \mathsf{c}(g_1)$. We obtain $x \notin \mathsf{c}(f_0)$ and $y \notin \mathsf{c}(f_0)$. Hence the central x in f is the first occurrence of x in u and the first occurrence of y in u can not be before them. Hence $\mathsf{first}(u) = \mathsf{first}(v)$.

The dual argument leads to $\mathsf{last}(u) = \mathsf{last}(v)$.

Observe that for each word w and letters $x, y \in \mathsf{c}(w)$, $x \neq y$ we have $xy \in \mathsf{Sub}_2(w)$ if and only if the first occurrence of x in w is before the last occurrence of y in w. Assume, for a moment, that $xy \in \mathsf{Sub}_2(u)$ but $xy \notin \mathsf{Sub}_2(v)$, i.e. the last occurrence of y in v is before the first occurrence of x. Let $g = (g_0, y, g_1)$ be a factorization of v where the central y is the last occurrence of y in v, i.e. $y \notin \mathsf{c}(g_1)$. Under our assumption also $x \notin \mathsf{c}(g_0)$. There is a factorization $f = (f_0, y, f_1)$ of u such that $\mathsf{c}(f_0) \subseteq \mathsf{c}(g_0)$ and $\mathsf{c}(f_1) \subseteq \mathsf{c}(g_1)$. Hence $x \notin \mathsf{c}(f_0)$ and $y \notin \mathsf{c}(f_1)$. The second condition is saying that central y in the factorization f is the last occurrence of y in u and hence x does not occur before this y. So, the last occurrence of y in u is before the first occurrence of x in u, which is a contradiction. Hence we proved the second condition.

Now let $x, y \in \mathsf{c}(u)$, $xy \notin \mathsf{Sub}_2(v)$. So, $xy \notin \mathsf{Sub}_2(u)$ is also true. We can write $v = v_0 y v_1 x v_2$ where $y \notin \mathsf{c}(v_1 x v_2)$ and $x \notin \mathsf{c}(v_0 y v_1)$ and also $u = u_0 y u_1 x u_2$ where $y \notin \mathsf{c}(u_1 x u_2)$ and $x \notin \mathsf{c}(u_0 y u_1)$. Let $z \in \mathsf{c}(v_1) = \mathsf{int}_{y,x}(v)$. Then we can write $v_1 = v_1' z v_1''$ and we have a factorization $g = (v_0 y v_1', z, v_1'' x v_2)$ of the word v. So, there is a factorization $f = (f_0, z, f_1)$ of the word u such that $\mathsf{c}(f_0) \subseteq \mathsf{c}(v_0 y v_1') \subseteq \mathsf{c}(v_0 y v_1)$ and $\mathsf{c}(f_1) \subseteq \mathsf{c}(v_1'' x v_2) \subseteq \mathsf{c}(v_1 x v_2)$. Hence $x \notin \mathsf{c}(f_0)$ and $y \notin \mathsf{c}(f_1)$. This means that the central z in f is before the first occurrence of the letter x in u and after the last occurrence of the letter y in u. So, $z \in \mathsf{c}(u_1)$ and we proved that $\mathsf{int}_{y,x}(v) = \mathsf{c}(v_1) \subseteq \mathsf{c}(u_1) = \mathsf{int}_{y,x}(u)$. So, the proof of the direct implication is complete.

Now assume that u and v satisfy the conditions i) – iii). Note that from the condition i) we have $\mathsf{c}(u) = \mathsf{c}(v)$. Let $g = (g_0, z, g_1)$ be an arbitrary factorization of v. We distinguish several cases.

Assume that the central z in g is the first occurrence and the last occurrence of this letter in v at the same time. Then it is a unique occurrence of z in v and we have $zz \notin \mathsf{Sub}_2(v)$. Hence $zz \notin \mathsf{Sub}_2(u)$ and there is a unique occurrence of z in u, so we can write $u = u_0 z u_1$ and $z \notin \mathsf{c}(u_0) \cup \mathsf{c}(u_1)$. From the condition i) we have $\mathsf{c}(u_0) = \mathsf{c}(g_0)$ and $\mathsf{c}(u_1) = \mathsf{c}(g_1)$.

Assume that the central z in g is the first occurrence of z in v but it is not the last occurrence of z in v. Then $z \notin \mathsf{c}(g_0)$ and $z \in \mathsf{c}(g_1)$. Let consider $u = u_0 z u_1$ where the central z is the first occurrence of z in u, i.e. $z \notin \mathsf{c}(u_0)$. From the condition i) we have $\mathsf{c}(u_0) = \mathsf{c}(g_0)$ and we would like to show that $\mathsf{c}(u_1) \subseteq \mathsf{c}(g_1)$. So, let $y \in \mathsf{c}(u_1)$. Then $zy \in \mathsf{Sub}_2(u) \subseteq \mathsf{Sub}_2(v)$ and $y \in \mathsf{c}(g_1)$ follows. If the central z in g is the last occurrence of z in v we can use a dual construction.

Finally, assume that the central z in g is not the first occurrence either the last occurrence of z in v. If there is no letter y with the last occurrence in v before our z then it means that $\mathsf{c}(g_1) = \mathsf{c}(v) = \mathsf{c}(u)$ and we can easily find a factorization f of u, namely we choose the first occurrence of z in u as a central letter in f and the inequality $f \leq_\sigma g$ is clear. Dually, in the case when no first occurrence of a letter occurs after our z. So, look at the last occurrence of the a letter y before z in v such that there is no last occurrence of some letter between these occurrences of y and our z. In the same way we look at the first occurrence of a letter x in v which is after our z and there is no first occurrence of some letter between them. So $z \in \mathsf{int}_{y,x}(v) \subseteq \mathsf{int}_{y,x}(u)$ by the condition iii) and we can find the occurrence of z in u between the last occurrence of y and the first occurrence of x, i.e $u = u_1 y u_2 z u_3 x u_4$ where x is the first occurrence of x in v and y is the last occurrence of y. We claim that $f = (u_1 y u_2, z, u_3 x u_4) \leq_\sigma g = (g_0, z, g_1)$. Indeed, if some letter a belongs to $u_1 y u_2$ then the first occurrence of this letter a is before the first occurrence of x in u and by the condition i) the first occurrence of a in v is before the first occurrence of x. Hence the first occurrence of a in v is before our z in v, i.e. $a \in \mathsf{c}(g_0)$. So, we proved $\mathsf{c}(u_1 y u_2) \subseteq \mathsf{c}(g_0)$ and one can prove $\mathsf{c}(u_3 x u_4) \subseteq \mathsf{c}(g_1)$ in the same manner, i.e. $f \leq_\sigma g$. In all cases we found such a factorization f of the word u, hence $(u, v) \in \pi_1^+$. □

From the previous lemma we immediately obtain the analogical characterization for the relation $(\pi_1^+)^\frown$.

Lemma 3. *Let $u, v \in X^*$ be arbitrary words. Then $(u, v) \in (\pi_1^+)^\frown$ if and only if the following conditions are satisfied*

i) $\mathsf{first}(u) = \mathsf{first}(v)$ *and* $\mathsf{last}(u) = \mathsf{last}(v)$*;*

ii) $\mathsf{Sub}_2(u) = \mathsf{Sub}_2(v)$*;*

iii) for every $x, y \in \mathsf{c}(u)$, $xy \notin \mathsf{Sub}_2(v)$ we have $\mathsf{int}_{y,x}(u) = \mathsf{int}_{y,x}(v)$*.* □

It is clear that each letter occurs at most twice in $\mathsf{skel}(u)$ and it occurs exactly once if and only if it occurs exactly once in u.

The following observation follows from the conditions i) and ii) in Lemma 3.

Lemma 4. *Let $u, v \in X^*$ be such that $(u, v) \in (\pi_1^+)^\frown$. Then* $\mathsf{skel}(u) = \mathsf{skel}(v)$.

Now consider the set of variables $X_m = \{x_1, x_2, \dots, x_m\}$. We say that a word $w = b_1 w_1 b_2 w_2 \dots b_{\ell-1} w_{\ell-1} b_\ell \in X_m^*$ is a *canonical word* if the following conditions are satisfied:

- $b_1, b_2, \dots, b_\ell \in X_m$;
- $b_1 b_2 \dots b_\ell = \mathsf{skel}(w)$;
- $w_1, w_2, \dots, w_{\ell-1} \in X_m^*$;
- for every $i = 1, \dots, \ell-1$ there are no indices $j \leq j' \in \{1, \dots, m\}$ such that $x'_j x_j$ is a subword of w_i.

If the following two conditions are also satisfied we speak about a *balanced canonical word.*

- If $i \in \{1, \dots, \ell-1\}$ is such that b_i is not the last occurrence of this letter in w, then $b_i \in \mathsf{c}(w_i)$ and $\mathsf{c}(w_{i-1}) \subseteq \mathsf{c}(w_i)$ if w_{i-1} is defined;
- If $i \in \{2, \dots, \ell\}$ is such that b_i is not the first occurrence of this letter in w, then $b_i \in \mathsf{c}(w_{i-1})$ and $\mathsf{c}(w_i) \subseteq \mathsf{c}(w_{i-1})$ if w_i is defined;

The role of this notion will be clear from the following lemma which completes the proof of the first statement of the proposition.

Lemma 5. *i) Let u, v be balanced canonical words such that $(u, v) \in (\pi_1^+)^\frown$. Then $u = v$.*

ii) Let u be an arbitrary word. Then there exists a canonical word w such that $(u, w) \in (\pi_1^+)^\frown$ and the identity $u = w$ is a consequence of the identities (2) and (3).

Proof. "i)" : If $(u, v) \in (\pi_1^+)^\frown$ then $\mathsf{skel}(u) = \mathsf{skel}(v)$ by Lemma 4. So, let $u = b_1 u_1 b_2 u_2 \dots b_{\ell-1} u_{\ell-1} b_\ell \in X_m^*$ and $v = b_1 v_1 b_2 v_2 \dots b_{\ell-1} v_{\ell-1} b_\ell \in X_m^*$ be balanced canonical words. It is enough to prove that $\mathsf{c}(u_i) = \mathsf{c}(v_i)$ for every $i = 1, \dots, \ell-1$.

Take an arbitrary i. If there is no last occurrence of a letter among the occurrences of the letters $b_1, b_2, \dots, b_i$ then we have $b_1 \in \mathsf{c}(u_1)$, $b_2 \in \mathsf{c}(u_2), \dots, b_i \in \mathsf{c}(u_i)$ and $\mathsf{c}(u_1) \subseteq \mathsf{c}(u_2) \subseteq \dots \subseteq \mathsf{c}(u_i)$. So, we have $\mathsf{c}(u_i) = \{b_1, \dots, b_i\}$. The same is true for v, so we obtain $\mathsf{c}(u_i) = \mathsf{c}(v_i)$.

The dual argument give the equality $\mathsf{c}(u_i) = \mathsf{c}(v_i)$ if there is no first occurrence among $b_{i+1}, \dots, b_\ell$.

Now let p be an index such that b_p is the last occurrence of this letter in u (and v) and there is no last occurrence among $b_{p+1}, \dots, b_i$ and let q be such that b_q is the first occurrence in u (and v) and there is no first occurrence among $b_{i+1}, b_{i+2}, \dots, b_{q-1}$. Then we have $b_{p+1} \in \mathsf{c}(u_{p+1}), \dots b_i \in \mathsf{c}(u_i)$ and $\mathsf{c}(u_p) \subseteq \mathsf{c}(u_{p+1}) \subseteq \dots \subseteq \mathsf{c}(u_i)$. Similarly from right $b_{q-1} \in \mathsf{c}(u_{q-2}), \dots, b_{i+1} \in \mathsf{c}(u_i)$ and $\mathsf{c}(u_{q-1}) \subseteq \dots \subseteq \mathsf{c}(u_{i+1}) \subseteq \mathsf{c}(u_i)$. Altogether we observe that

$$\mathsf{int}_{b_p, b_q}(u) = \mathsf{c}(u_p b_{p+1} u_{p+1} \dots u_{i-1} b_i u_i b_{i+1} u_{i+1} \dots u_{q-2} b_{q-1} u_{q-1}) = \mathsf{c}(u_i) \ .$$

The same equality holds for v and we obtain $\mathsf{c}(u_i) = \mathsf{c}(v_i)$.

"ii)" : Let u be an arbitrary word. It can be written in the form

$$u = b_1 u_1 b_2 u_2 \dots b_{\ell-1} u_{\ell-1} b_\ell, \quad \text{where} \quad b_1 b_2 \dots b_\ell = \mathsf{skel}(u) \ .$$

We can use the identity (3) to commute the letters inside every word u_i. In particular, we can move an arbitrary letter $x \in \mathsf{c}(u_i)$ at the first position or at the last position in such a word u_i. We can add b_i to u_i or u_{i-1} by the identities $xvx^2 = xvx = x^2vx$ (consequences of the identity (2) if needed. Further, we can use the identity (2) to add the letter x to u_i when b_i is not the last occurrence and $x \in \mathsf{c}(u_{i-1}) \setminus \mathsf{c}(u_i)$. Similarly for b_i which is not the first occurrence of a letter. Finally, the identity $xvx^2 = xvx$ can be used to remove from every word u_i all redundant occurrences of letters. So our identities can be used to obtain balanced canonical word which is related to the given u. □

We prove the rest of the statement in two steps which are formulated in the following lemma.

Lemma 6. *i) Let u, v be such that $(u, v) \in \pi_1^+$. Then there exists a word w such that the identity $u \leq w$ is a consequence of the identities (3), (4) and (5), $(w, v) \in \pi_1^+$ and $\mathsf{Sub}_2(w) = \mathsf{Sub}_2(v)$.*

ii) Let w, v be such that $(w, v) \in \pi_1^+$ and $\mathsf{Sub}_2(w) = \mathsf{Sub}_2(v)$. Then the identity $w \leq v$ is a consequence of the identities (3), (4) and (5).

Proof. "i)" : Recall that $(u, v) \in \pi_1^+$ implies $\mathsf{Sub}_2(u) \subseteq \mathsf{Sub}_2(v)$. We prove the statement by an induction with respect to the size of the set $M = \mathsf{Sub}_2(v) \setminus \mathsf{Sub}_2(u)$. If M is the empty set then there is nothing to prove. If $x^2 \in M$ for some $x \in X$, then $u = u_0 x u_1$ where $x \notin \mathsf{c}(u_0) \cup \mathsf{c}(u_1)$. Now we can apply the identity $x \leq x^2$ (consequence of the identity (4)) and obtain $u \leq u_0 x x u_1$. It is easy to see that $(u_0 x x u_1, v) \in \pi_1^+$ when we use the characterization from Lemma 2. So assume that no x^2 belongs to M.

Now from all pairs x, y satisfying $xy \in M$ we choose such that in the corresponding factorization $u = u_0 y u_1 x u_2$ where $y \notin \mathsf{c}(u_1 x u_2)$ and $x \notin \mathsf{c}(u_0 y u_1)$ the word u_1 is a short possible. Since $xy \in M$ we have $v = v_0 x v_1 y v_2$ where $x \notin \mathsf{c}(v_0)$ and $y \notin \mathsf{c}(v_2)$. We know that $\mathsf{first}(v) = \mathsf{first}(u)$, so $y \in \mathsf{c}(v_0)$. In the same manner $x \in \mathsf{c}(v_2)$ follows from $\mathsf{last}(v) = \mathsf{last}(u)$. Namely $x^2, y^2 \in \mathsf{Sub}_2(v)$ and hence $x^2, y^2 \in \mathsf{Sub}_2(u)$. Now if u_1 contain the first occurrence of a letter z then the first occurrence of z is before the first occurrence of x in both u and v and we see that $zy \in M$. This is a contradiction with our choice of the pair x, y. In the same way we can prove that u_1 does not contain the last occurrence of some letter. This means that $\mathsf{c}(u_1) \subseteq \mathsf{c}(u_0)$ and $\mathsf{c}(u_1) \subseteq \mathsf{c}(u_2)$. We use the identity $x \leq x^2$ to introduce one more x into u_1. Then we can commute the letters in u_1 by identity (3) so we obtain $u = u_0 y u_1 x u_2 \leq u_0 y x u_1 x u_2$ and then use the identity (4) to introduce next x immediately before considered occurrence of y, i.e. $u = u_0 y u_1 x u_2 \leq u_0 y x u_1 x u_2 \leq u_0 x y x u_1 x u_2 = w$. It is not hard to see that $(w, v) \in \pi_1^+$ because in the process we almost do not change the invariants of u from Lemma 2.

"ii)": By Lemma 5 we can assume that w and v are balanced canonical words. Recall that the identity (2) is a consequence of the identities (4) and (5). So, let

$w = b_1w_1b_2w_2 \dots b_{\ell-1}w_{\ell-1}b_\ell$ and $v = b_1v_1b_2v_2 \dots b_{\ell-1}v_{\ell-1}b_\ell$ be balanced canonical words with the same skeleton $b_1 \dots b_\ell$. We showed some basic properties of words w_i in balanced canonical words in the proof of the item i) of Lemma 5:
1. If there is no last occurrence among $b_1, \dots, b_i$ then $\mathsf{c}(w_i) = \{b_1, \dots, b_i\}$ and we have $\mathsf{c}(w_i) = \mathsf{c}(v_i)$.
2. Dually if there is no first occurrence among $b_{i+1}, \dots, b_\ell$.
3. If $p < i$ is an index such that b_p is the last occurrence of this letter in w and there is no last occurrence among $b_{p+1}, \dots, b_i$ and if $q > i$ is such that b_q is the first occurrence in w and there is no first occurrence among $b_{i+1}, b_{i+2}, \dots, b_{q-1}$, then $\mathsf{c}(w_i) = \mathsf{int}_{b_p,b_q}(w)$. Hence we can deduce $\mathsf{c}(v_i) \subseteq \mathsf{c}(w_i)$.

So the statement can be prove by an induction with respect to minimal i such that $M_i = \mathsf{c}(w_i) \setminus \mathsf{c}(v_i) \neq \emptyset$ and with respect the size of the set M_i. We can use the identity (5) to add some missing letter to $\mathsf{c}(v_i)$. We obtain a word v' such that $v' \leq v$ is a consequence of the identities from the base. Under our observations concerning $\mathsf{int}_{b_p,b_q}(w)$ we see that the $(w, v') \in \pi_1^+$. The word v' could not be a balanced canonical word but we can use Lemma 5 to rewrite v' to some balanced canonical word v''. On the pair w, v'' we can apply the induction assumption. □

We finished the proof of the proposition. □

5.2 Identities for $\mathsf{BPol}_1(\mathcal{S})$ and $\mathsf{PPol}_1(\mathcal{S})$

The proofs in this part are easier because we can use numerous observations from the previous subsection.

We use the identity (3) again and we introduce new identity

$$xuxxvx = xuxvx \; . \tag{6}$$

It is clear that both these identities are satisfied in $\mathsf{BPol}_1(\mathcal{S})$ and consequently in $\mathsf{PPol}_1(\mathcal{S})$. Also it is easy to see that the identity (5) is satisfied in $\mathsf{PPol}_1(\mathcal{S})$.

Proposition 9. *i) The identities (3) and (6) form the finite base of identities for the variety of monoids corresponding to* $\mathsf{BPol}_1(\mathcal{S})$.

ii) The identities (3), (5), and (6) form the finite base of identities for the variety of ordered monoids corresponding to $\mathsf{PPol}_1(\mathcal{S})$.

Proof. We modify Lemma 5 for relation $\widehat{\pi_1}$.

Lemma 7. *i) Let u, v be canonical words such that $(u, v) \in \widehat{\pi_1}$. Then $u = v$.*

ii) Let u be an arbitrary word. Then there exists a canonical word w such that $(u, w) \in \widehat{\pi_1}$ and the identity $u = w$ is a consequence of the our identities (3) and (6).

Proof. "i)" : We have $\mathsf{skel}(u) = \mathsf{skel}(v)$. So, let $u = b_1u_1b_2u_2 \dots b_{\ell-1}u_{\ell-1}b_\ell$ and $v = b_1v_1b_2v_2 \dots b_{\ell-1}v_{\ell-1}b_\ell$ be canonical words with skeleton $\mathsf{skel}(u) = \mathsf{skel}(v) = b_1 \dots b_\ell$. It is enough to prove that $\mathsf{c}(u_i) = \mathsf{c}(v_i)$ for every $i = 1, \dots, \ell - 1$.

Let $x \in \mathsf{c}(v_i)$. We choose some occurrence of x in v_i and consider a corresponding factorization $g = (g_0, x, g_1)$ of v. Then it is easy to see that $\mathsf{c}(g_0) = \{b_1, \dots, b_i\}$ and $\mathsf{c}(g_1) = \{b_{i+1}, \dots, b_\ell\}$. There is a factorization $f = (f_0, x, f_1)$ such that $\mathsf{c}(f_0) = \mathsf{c}(g_0)$ and $\mathsf{c}(f_1) = \mathsf{c}(g_1)$. We claim that the central x in f is in u_i. Indeed, if b_i is the first occurrence of a letter in u then $\mathsf{c}(f_0) = \{b_1, \dots, b_i\}$ implies that this x is after b_i in u and if b_i is the last occurrence of a letter in u then $b_i \notin \mathsf{c}(f_1) = \{b_{i+1}, \dots, b_\ell\}$ implies the same (note that $x \neq b_i$ because b_i is the last occurrence). The similar argument gives that x is before b_{i+1} in u and we proved that $x \in \mathsf{c}(u_i)$. We have $\mathsf{c}(v_i) \subseteq \mathsf{c}(u_i)$ and the equality $\mathsf{c}(u_i) = \mathsf{c}(v_i)$ follows because $\widehat{\pi_1}$ is a equivalence relation.

"ii)" : Let $u = b_1 u_1 b_2 u_2 \dots b_{\ell-1} u_{\ell-1} b_\ell$ be such that $\mathsf{skel}(u) = b_1 \dots b_\ell$. We use the identity (3) to commute letters inside every u_i and we use the identity (6) to remove redundant occurrences of letters. So we can construct w with the required properties. □

Lemma 8. *Let $u, v \in X^*$ be such that $(u, v) \in \pi_1$. Then $\mathsf{skel}(u) = \mathsf{skel}(v)$.*

Proof. If $(u, v) \in \pi_1$ then $(u, v) \in \pi_1^+$ and we have $\mathsf{first}(u) = \mathsf{first}(v)$ and $\mathsf{last}(u) = \mathsf{last}(v)$ by Lemma 2. We claim that $\mathsf{Sub}_2(u) = \mathsf{Sub}_2(v)$. Indeed, the inclusion $\mathsf{Sub}_2(u) \subseteq \mathsf{Sub}_2(v)$ also follows from Lemma 2. Let $xy \in \mathsf{Sub}_2(v)$. If we consider a factorization $g = (v_0, x, v_1)$ of the word v such that the central x is the first occurrence of x in v then $x \notin \mathsf{c}(v_0)$ and $y \in \mathsf{c}(v_1)$. Hence there is a factorization $f = (f_0, x, f_1)$ of u such that $\mathsf{c}(f_0) = \mathsf{c}(v_0)$ and $\mathsf{c}(f_1) = \mathsf{c}(v_1)$. So the central x in f is the first occurrence of x in u and $y \in \mathsf{c}(f_1)$ and we can conclude with $xy \in \mathsf{Sub}_2(u)$. We proved the claim and the statement trivially follows. □

Lemma 9. *Let u, v be words such that $(u, v) \in \pi_1$. Then the identity $u \leq v$ is a consequence of the identities (3), (5) and (6).*

Proof. By Lemma 7 we can assume that both u, v are canonical words. We have $(u, v) \in \pi_1$ and we get $\mathsf{skel}(u) = \mathsf{skel}(v)$. So, let $u = b_1 u_1 b_2 u_2 \dots b_{\ell-1} u_{\ell-1} b_\ell$ and $v = b_1 v_1 b_2 v_2 \dots b_{\ell-1} v_{\ell-1} b_\ell$ be canonical words with the skeleton $\mathsf{skel}(u) = \mathsf{skel}(v) = b_1 \dots b_\ell$. In the proof of the part i) of Lemma 7 we saw that $\mathsf{c}(v_i) \subseteq \mathsf{c}(u_i)$ for every $i = 1, \dots, \ell - 1$. We can add the missing letters by the identity (5). □

We finished the proof of the proposition. □

5.3 Proof of Proposition 4

We want to prove that $\mathsf{BPol}_1(\mathcal{S}) \subseteq \mathsf{BPol}_2(\mathcal{S}^+)$, that is $(\pi_2^+)\widehat{\ } \subseteq \widehat{\pi_1}$. Let $(u, v) \in (\pi_2^+)\widehat{\ }$. Then we have $\mathsf{skel}(u) = \mathsf{skel}(v)$ by Lemma 4. So, let $u = b_1 u_1 b_2 u_2 \dots u_{\ell-1} b_\ell$ and $v = b_1 v_1 b_2 v_2 \dots v_{\ell-1} b_\ell$ with the skeleton $\mathsf{skel}(u) = \mathsf{skel}(v) = b_1 \dots b_\ell$. We would like to prove that For a word $u \in X^*$ we define the skeleton $\mathsf{skel}(u) \in X^*$ in the following way. We remove from u every occurrence of a given letter which is not the first or the last occurrence of this letter in the word u. After deleting of all "interior" occurrences of all letters from u the resulting word is a skeleton $\mathsf{skel}(u)$ of u. $\mathsf{c}(u_i) = \mathsf{c}(v_i)$ for each $i = 1, \dots, \ell - 1$.

We consider a factorization $g = (g_0, b_i, g_1, b_{i+1}, g_2)$ of v, such that $g_0 = b_1 v_1 b_2 v_2 \dots v_{i-1}$, $g_1 = v_i$ and $g_2 = v_{i+1} \dots b_{\ell-1} v_{\ell-1} b_\ell$. Then there is a factorization $f = (f_0, b_i, f_1, b_{i+1}, f_2)$ of u such that $\mathsf{c}(f_0) \subseteq \mathsf{c}(g_0)$, $\mathsf{c}(f_1) \subseteq \mathsf{c}(g_1)$ and $\mathsf{c}(f_2) \subseteq \mathsf{c}(g_2)$. If b_i is the first occurrence of a letter in v then $b_i \notin \mathsf{c}(g_0)$ and $b_i \notin \mathsf{c}(f_0)$ follows. Hence b_i in f is the first occurrence of b_i in u, i.e. it is the occurrence of b_i from the expression $u = b_1 u_1 b_2 u_2 \dots b_{\ell-1} u_{\ell-1} b_\ell$. If b_i is the last occurrence of a letter in v then $b_i \notin \mathsf{c}(g_1 b_{i+1} g_2)$ and $b_i \notin \mathsf{c}(f_1 b_{i+1} f_2)$ follows. Hence b_i in f is the last occurrence of b_i in u, so it is the occurrence of b_i from the expression $u = b_1 u_1 b_2 u_2 \dots b_{\ell-1} u_{\ell-1} b_\ell$. We can prove the same for b_{i+1} and hence $f_1 = u_i$. This means $\mathsf{c}(u_i) = \mathsf{c}(f_1) \subseteq \mathsf{c}(g_1) = \mathsf{c}(v_i)$.

If we exchange the role of v and u we obtain also $\mathsf{c}(v_i) \subseteq \mathsf{c}(u_i)$ and the claim is proved.

Now by Lemma 7 we can see that $(u, v) \in \widehat{\pi_1}$. The proof of Proposition 4 is finished. □

References

1. S. L. Bloom, Varieties of ordered algebras. J. Comput. Syst. Sci. **13** (2) (1976), 200–212
2. S. Burris and H. P. Sankappanavar, *A Course in Universal Algebra*, Springer, Berlin, 1981
3. F. Blanchet-Sadri, Equations and monoids varieties of dot-depth one and two, *Theoret. Comput. Sci.* **123** (1994), 239–258
4. S. Eilenberg, *Automata, Languages and Machines,* Vol B, Academic Press, New York, 1976
5. O. Klíma and L. Polák, Hierarchies of piecewise testable languages, in *Developments in Language Theory (DLT 2008)*, Lecture Notes in Computer Science 5257 (2008), 479–490
6. O. Klíma and L. Polák, Polynomial operators on classes of regular languages, submitted, see `http://www.math.muni.cz/~polak/Grant.html`
7. J.-E. Pin, A variety theorem without complementation, *Russian Mathem. (Iz. VUZ)* **39** (1995), 74–83
8. J.-E. Pin, Syntactic semigroups, Chapter 10 in *Handbook of Formal Languages*, G. Rozenberg and A. Salomaa eds, Springer, 1997
9. J.-E. Pin, Algebraic tools for the concatenation product, Theoretical Computer Science 292, (2003), 317-342
10. I. Simon, Piecewise testable events, in *Proc. ICALP 1975*, LNCS Vol. 33 (1975), 214–222