Acta Cybernetica **23** (2017) 91–111.

Commutative Positive Varieties of Languages*

Jorge Almeida[a], Zoltán Ésik[b], and Jean-Éric Pin[c]

To the memory of Zoltán Ésik.

Abstract

We study the commutative positive varieties of languages closed under various operations: shuffle, renaming and product over one-letter alphabets.

Most monoids considered in this paper are finite. In particular, we use the term *variety of monoids* for *variety of finite monoids*. Similarly, all languages considered in this paper are regular languages and hence their syntactic monoid is finite.

1 Introduction

Eilenberg's variety theorem [12] and its ordered version [17] provide a convenient setting for studying classes of regular languages. It states that positive varieties of languages are in one-to-one correspondence with varieties of finite ordered monoids.

There is a large literature on operations on regular languages. For instance, the closure of [positive] varieties of languages under various operations has been extensively studied: Kleene star [16], concatenation product [7, 19, 25], renaming [1, 4, 8, 23, 26] and shuffle [6, 10, 14]. The ultimate goal would be the complete classification of the positive varieties of languages closed under these operations.

*The last two authors acknowledge support from the cooperation programme CNRS/Magyar Tudomanyos Akadémia. The first author was partially supported by CMUP (UID/MAT/00144/2013), which is funded by FCT (Portugal) with national (MCTES) and European structural funds (FEDER), under the partnership agreement PT2020. The third author was partially funded from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 670624) and by the DeLTA project (ANR-16-CE40-0007)

[a]CMUP, Dep. Matemática, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal. E-mail: `jalmeida@fc.up.pt`

[b]Dept. of Foundations of Computer Science, University of Szeged, Árpád tér 2, H-6720 Szeged, P.O.B. 652 Hungary.

[c]IRIF, CNRS and Université Paris-Diderot, Case 7014, 75205 Paris Cedex 13, France.E-mail: `Jean-Eric.Pin@irif.fr`

DOI: 10.14232/actacyb.23.1.2017.7

The first step in this direction is to understand the commutative case, which is the goal of this paper.

We first show in Theorem 5.1 that every commutative positive *ld*-variety of languages is a positive variety of languages. This means that if a class of commutative languages is closed under Boolean operations and under inverses of length-decreasing morphisms then it is also closed under inverses of morphisms. This result has a curious application in weak arithmetic, stated in Proposition 5.4.

Next we study two operations on languages, shuffle and renaming. These two operations are closely related to the so-called *power operator* on monoids, which associates with each monoid the monoid of its subsets. In its ordered version, it associates with each ordered monoid the ordered monoid of its downsets. We give four equivalent conditions characterizing the commutative positive varieties of languages closed under shuffle (Proposition 6.1) or under renaming (Proposition 6.2).

In order to keep the paper self-contained, prerequisites are presented in some detail in Section 2. Inequalities form the topic of Section 3. We start with their formal definitions, describe their various interpretations and establish some of their properties. General results on renaming are given in Section 4 and more specific results on commutative varieties are proposed in Section 5, including our previously mentioned result on *ld*-varieties. Our characterizations of the positive varieties of languages closed under shuffle or renaming form the meat of Section 6 and are illustrated by three examples in Section 7. Finally, a few research directions are suggested in Section 8.

2 Prerequisites

In this section, we briefly recall the following notions: lattices and (positive) varieties of languages, syntactic ordered monoids, varieties of ordered monoids, stamps, downset monoids, free profinite monoids.

2.1 Languages

Let A be a finite alphabet. Let $[u]$ be the *commutative closure* of a word u, that is, the set of words commutatively equivalent to u. For instance, $[aab] = \{aab, aba, baa\}$. A language L is *commutative* if, for every word $u \in L$, $[u]$ is contained in L.

A *lattice of languages* is a set $\mathcal{L}$ of regular languages of A^* containing $\emptyset$ and A^* and closed under finite union and finite intersection. It is *closed under quotients* if, for each $L \in \mathcal{L}$ and $u \in A^*$, the languages $u^{-1}L$ and Lu^{-1} are also in $\mathcal{L}$.

The *shuffle product* (or simply *shuffle*) of two languages L_1 and L_2 over A is the language

$$\begin{aligned} L_1 \sqcup\!\sqcup L_2 = \{w \in A^* \mid w = u_1v_1 \cdots u_nv_n \text{ for some words } u_1, \ldots, u_n \\ v_1, \ldots, v_n \text{ of } A^* \text{ such that } u_1 \cdots u_n \in L_1 \text{ and } v_1 \cdots v_n \in L_2\} \end{aligned}$$

The shuffle product defines a commutative and associative operation on the set of languages over A.

A *renaming* or *length-preserving morphism* is a morphism φ from A^* into B^*, such that, for each word u, the words u and $\varphi(u)$ have the same length. It is equivalent to require that, for each letter a, $\varphi(a)$ is also a letter, that is, $\varphi(A) \subseteq B$. Similarly, a morphism is *length-decreasing* if the image of each letter is either a letter or the empty word.

A *class of languages* is a correspondence $\mathcal{C}$ which associates with each alphabet A a set $\mathcal{C}(A^*)$ of regular languages of A^*.

A *positive variety of languages* is a class of regular languages $\mathcal{V}$ such that:

(1) for every alphabet A, $\mathcal{V}(A^*)$ is a lattice of languages closed under quotients,

(2) if $\varphi : A^* \to B^*$ is a morphism, $L \in \mathcal{V}(B^*)$ implies $\varphi^{-1}(L) \in \mathcal{V}(A^*)$.

A *variety of languages* is a positive variety $\mathcal{V}$ such that each lattice $\mathcal{V}(A^*)$ is closed under complement. We shall also use two slight variations of these notions. A *positive ld-variety* [*lp-variety*] *of languages* [13, 19] is a class of regular languages $\mathcal{V}$ satisfying (1) and

(2′) if $\varphi : A^* \to B^*$ is a length-decreasing [length-preserving] morphism, then $L \in \mathcal{V}(B^*)$ implies $\varphi^{-1}(L) \in \mathcal{V}(A^*)$.

2.2 Syntactic ordered monoids

An *ordered monoid* is a monoid M equipped with a partial order $\leqslant$ compatible with the product on M: for all $x, y, z \in M$, if $x \leqslant y$ then $zx \leqslant zy$ and $xz \leqslant yz$.

The *ordered syntactic monoid* of a language was first introduced by Schützenberger in [24, p. 10]. Let L be a language of A^*. The *syntactic preorder* of L is the relation $\leqslant_L$ defined on A^* by $u \leqslant_L v$ if, for every $x, y \in A^*$, $xuy \in L$ implies $xvy \in L$. When the language L is clear from the context, we may write $\leqslant$ instead of $\leqslant_L$. As is standard in preorder notation, we write $u < v$ to mean that $u \leqslant v$ holds but $v \leqslant u$ does not.

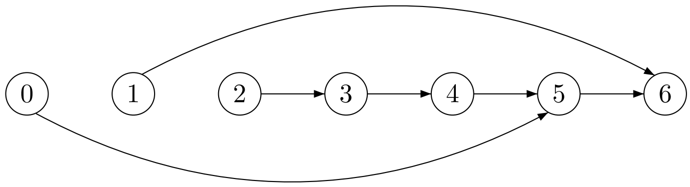
For instance, let $A = \{a\}$. If $L = a + a^3$, then $a^3 \leqslant_L a$, but if $L = a + a^3a^*$, then $a \leqslant_L a^3$.

The associated equivalence relation $\sim_L$, defined by $u \sim_L v$ if $u \leqslant_L v$ and $v \leqslant_L u$, is the *syntactic congruence* of L and the quotient monoid $M(L) = A^*/{\sim_L}$ is the *syntactic monoid* of L. The natural morphism $\eta : A^* \to A^*/{\sim_L}$ is the *syntactic stamp* of L. The *syntactic image* of L is the set $P = \eta(L)$.

The *syntactic order* $\leqslant$ is defined on $M(L)$ as follows: $u \leqslant v$ if and only if for all $x, y \in M$, $xuy \in P$ implies $xvy \in P$. The partial order $\leqslant$ is compatible with multiplication and the resulting ordered monoid $(M, \leqslant)$ is called the *ordered syntactic monoid* of L.

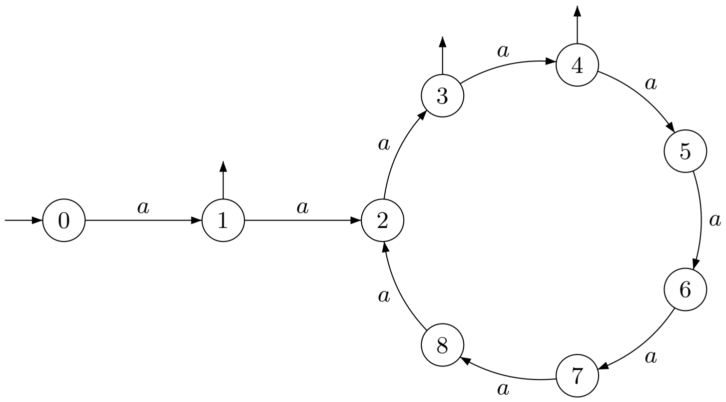
Example 2.1. Let L be the language $1 + a$. The syntactic monoid of L is the commutative monoid $\{1, a, 0\}$ satisfying $a^2 = 0$. The syntactic order is $0 < a < 1$. Indeed, one has $a \leqslant 1$ since, for each $r \geqslant 0$, the condition $a^r a \in L$ implies $a^r \in L$. Similarly, one has $0 \leqslant a$ since, for each $r \geqslant 0$, the condition $a^r a^2 \in L$ implies $a^r a \in L$. However, $1 \not\leqslant a$ and $a \not\leqslant 0$ since $a \in L$ but $a^2 \notin L$.

Example 2.2. Let L be the language $a + a^6a^*$. The syntactic monoid of L may be identified with the commutative monoid $\{0, 1, \ldots, 6\}$ equipped with the operation $xy = \min\{x + y, 6\}$. In particular, 0 and 6 are the unique idempotents. The syntactic order is represented as follows (a path from i to j means that $i < j$):



For instance, one has $1 < 6$ since, for each $r \geqslant 0$, the condition $aa^r \in L$ implies $a^6a^r \in L$. Similarly, one has $0 < 5$ since, for each $r \geqslant 0$, the condition $a^r \in L$ implies $a^5a^r \in L$. But $1 \not< 5$ since $a \in L$ but $a^5 \notin L$.

Example 2.3. Let L be the language $a + (a^3 + a^4)(a^7)^*$. Its minimal automaton is represented below.



The syntactic monoid of L is the monoid presented by $\langle a \mid a^9 = a^2 \rangle$. The syntatic order is the equality relation.

2.3 Stamps

Monoids and ordered monoids are used to recognise languages, but there is a slightly more restricted notion. A *stamp* is a surjective monoid morphism $\varphi : A^* \to M$ from a finitely generated free monoid A^* onto a finite monoid M. If M is an ordered monoid, φ is called an *ordered stamp*.

The *restricted direct product* of two [ordered] stamps $\varphi_1 : A^* \to M_1$ and $\varphi_2 : A^* \to M_2$ is the stamp φ with domain A^* defined by $\varphi(a) = (\varphi_1(a), \varphi_2(a))$ (see Figure 1). The image of φ is an [ordered] submonoid of the [ordered] monoid $M_1 \times M_2$.

$$
\begin{tikzcd}
 & & M_2 \\
A^* \arrow[rru, "\varphi_2"] \arrow[rr, "\varphi"] \arrow[rrd, "\varphi_1"'] & & \mathrm{Im}(\varphi) \subseteq M_1 \times M_2 \arrow[u, "\pi_2"'] \arrow[d, "\pi_1"] \\
 & & M_1
\end{tikzcd}
$$

Figure 1: The restricted direct product of two stamps.

Recall that an *upset* of an ordered set E is a subset U of E such that the conditions $x \in U$ and $x \leqslant y$ imply $y \in U$. A language L of A^* is *recognised by a stamp* $\varphi : A^* \to M$ if there exists a subset P of M such that $L = \varphi^{-1}(P)$. It is *recognised by an ordered stamp* $\varphi : A^* \to M$ if there exists an upset U of M such that $L = \varphi^{-1}(U)$.

It is easy to see that if two languages L_0 and L_1 of A^* are recognised by the [ordered] stamps φ_0 and φ_1, respectively, then $L_0 \cap L_1$ and $L_0 \cup L_1$ are both recognised by the restricted product of φ_0 and φ_1.

2.4 Varieties

Varieties of languages and their avatars all admit an algebraic characterization. We first describe the corresponding algebraic objects and summarize the correspondence results at the end of this section. See [18] for more details.

[Positive] varieties of languages correspond to varieties of [ordered] monoids. A *variety of monoids* is a class of monoids closed under taking submonoids, quotients and finite direct products. *Varieties of ordered monoids* are defined analogously.

The description of the algebraic objects corresponding to positive *lp*- and *ld*-varieties of languages is more complex and relies on the notion of stamp defined in Section 2.3. An *lp-morphism* from a stamp $\varphi : A^* \to M$ to a stamp $\psi : B^* \to N$ is a pair (f, α), where $f : A^* \to B^*$ is length-preserving, $\alpha : M \to N$ is a morphism of [ordered] monoids, and $\psi \circ f = \alpha \circ \varphi$.

$$
\begin{tikzcd}
A^* \arrow[r, "f"] \arrow[d, "\varphi"'] & B^* \arrow[d, "\psi"] \\
M \arrow[r, "\alpha"] & N
\end{tikzcd}
$$

The *lp*-morphism (f, α) is an *lp-projection* if f is surjective. It is an *lp-inclusion* if α is injective.

An [*ordered*] *lp-variety of stamps* is a class of [ordered] stamps closed under *lp*-projections, *lp*-inclusions and finite restricted direct products. [*Ordered*] *ld-*

varieties of stamps are defined in the same way, just by replacing *lp* by *ld* and length-preserving by length-decreasing everywhere in the definition.

Here are the announced correspondence results. Eilenberg's variety theorem [12] and its ordered counterpart [17] give a bijective correspondence between varieties of [ordered] monoids and positive varieties of languages. Let $\mathbf{V}$ be a variety of finite [ordered] monoids and, for each alphabet A, let $\mathcal{V}(A^*)$ be the set of all languages of A^* whose [ordered] syntactic monoid is in $\mathbf{V}$. Then $\mathcal{V}$ is a [positive] variety of languages. Furthermore, the correspondence $\mathbf{V} \to \mathcal{V}$ is a bijection between varieties of [ordered] monoids and [positive] varieties of languages.

There is a similar correspondence for *lp*-varieties of [ordered] stamps [13, 27]. Let $\mathbf{V}$ be an *lp*-variety of [ordered] stamps. For each alphabet A, let $\mathcal{V}(A^*)$ be the set of all languages of A^* whose [ordered] syntactic stamp is in $\mathbf{V}$. Then $\mathcal{V}$ is a [positive] *lp*-variety of languages. Furthermore, the correspondence $\mathbf{V} \to \mathcal{V}$ is a bijection between *lp*-varieties of [ordered] stamps and [positive] *lp*-varieties of languages.

Finally, there is a similar statement for *ld*-varieties of [ordered] stamps.

2.5 Downset monoids

Let $(M, \leqslant)$ be an ordered monoid. A *downset* of M is a subset F of M such that if $x \in F$ and $y \leqslant x$ then $y \in F$. The *product of two downsets* X and Y is the downset

$$XY = \{z \in M \mid \text{there exist } x \in X \text{ and } y \in Y \text{ such that } z \leqslant xy\}$$

This operation makes the set of nonempty downsets of M a monoid, denoted by $\mathcal{P}^{\downarrow}(M)$ and called the *downset monoid* of M. Its identity element is $\downarrow 1$. If one also considers the empty set, one gets a monoid with zero, denoted $\mathcal{P}_0^{\downarrow}(M)$, in which the empty set is the zero. For instance, if M is the trivial monoid, $\mathcal{P}_0^{\downarrow}(M)$ is isomorphic to the ordered monoid $\{0, 1\}$, consisting of an identity 1 and a zero 0, ordered by $0 < 1$. This monoid will be denoted by $U_1^{\downarrow}$.

The monoids $\mathcal{P}_0^{\downarrow}(M)$ and $\mathcal{P}^{\downarrow}(M)$ are closely related. First, $\mathcal{P}^{\downarrow}(M)$ is a submonoid of $\mathcal{P}_0^{\downarrow}(M)$. Secondly, as shown in [10, Proposition 5.1, p. 452], $\mathcal{P}_0^{\downarrow}(M)$ is isomorphic to a quotient monoid of $\mathcal{P}^{\downarrow}(M) \times U_1^{\downarrow}$.

The monoids $\mathcal{P}^{\downarrow}(M)$ and $\mathcal{P}_0^{\downarrow}(M)$ are naturally ordered by inclusion, denoted by $\leqslant$. Note that $X \leqslant Y$ if and only if, for each $x \in X$, there exists $y \in Y$ such that $x \leqslant y$.

Given a variety of ordered monoids $\mathbf{V}$, let $\mathbf{P}^{\downarrow}\mathbf{V}$ $[\mathbf{P}_0^{\downarrow}\mathbf{V}]$ denote the variety of ordered monoids generated by the monoids of the form $\mathcal{P}^{\downarrow}(M)$ $[\mathcal{P}_0^{\downarrow}(M)]$, where $M \in \mathbf{V}$. The operator $\mathbf{P}^{\downarrow}$ was intensively studied in [4]. In particular, it is known that both $\mathbf{P}^{\downarrow}$ and $\mathbf{P}_0^{\downarrow}$ are idempotent operators.

The hereinabove relation between $\mathcal{P}_0^{\downarrow}(M)$ and $\mathcal{P}^{\downarrow}(M)$ can be extended to varieties as follows. Let $\mathbf{Sl}^{\downarrow}$ be the variety of ordered monoids generated by $U_1^{\downarrow}$. It is a well-known fact that $\mathbf{Sl}^{\downarrow} = [\![xy = yx, x = x^2, x \leqslant 1]\!]$. Moreover, the equality

$$\mathbf{P}_0^{\downarrow}\mathbf{V} = \mathbf{P}^{\downarrow}\mathbf{V} \vee \mathbf{Sl}^{\downarrow} \tag{1}$$

holds for any variety of ordered monoids $\mathbf{V}$.

2.6 Free profinite monoid

We refer the reader to [1, 2, 3, 28] for detailed information on profinite completions and we just recall here a few useful facts. Let d be the profinite metric on the free monoid A^*. We let $\widehat{A^*}$ denote the completion of the metric space (A^*, d). The product on A^* is uniformly continuous and hence has a unique continuous extension to $\widehat{A^*}$. It follows that $\widehat{A^*}$ is a compact monoid, called the *free profinite monoid* on A. Furthermore, every stamp $\varphi : A^* \to M$ admits a unique continuous extension $\widehat{\varphi} : \widehat{A^*} \to M$. Similarly, every morphism $f : A^* \to B^*$ admits a unique continuous extension $\widehat{f} : \widehat{A^*} \to \widehat{B^*}$. In the sequel, $\overline{L}$ denotes the closure in $\widehat{A^*}$ of a subset L of A^*.

The length of a word u is denoted by $|u|$. The length map $u \to |u|$ defines a morphism from A^* to the additive semigroup $\mathbb{N}$. If $A = \{a\}$, this morphism is actually an isomorphism, which maps a^n to n. In other words, $(\mathbb{N}, +, 0)$ is the free monoid with a single generator. We let $\widehat{\mathbb{N}}$ denote the profinite completion of $\mathbb{N}$, which is of course isomorphic to $\widehat{a^*}$.

This allows one to define the length $|u|$ of an element u of $\widehat{A^*}$ simply by extending by continuity the length map defined on A^*. The length map is actually a morphism, that is, $|1| = 0$ and $|uv| = |u| + |v|$ for all $u, v \in \widehat{A^*}$.

3 Inequalities and identities

The inequalities [equalities] occurring in this paper are of the form $u \leqslant v$ $[u = v]$, where u and v are both in $\widehat{A^*}$ for some alphabet A. In an ordered context, $u = v$ is often viewed as a shortcut for $u \leqslant v$ and $v \leqslant u$.

However, these inequalities are interpreted in several different contexts, which may confuse the reader. Let us clarify matters by giving precise definitions for each case.

3.1 Inequalities

Ordered monoids. Let M be an ordered monoid, let X be an alphabet and let $u, v \in \widehat{X^*}$. Then M *satisfies the inequality* $u \leqslant v$ if, for each morphism $\psi : X^* \to M$, $\widehat{\psi}(u) \leqslant \widehat{\psi}(v)$.

This is the formal definition but in practice, it is easier to think of u and v as terms in which one substitutes each symbol $x \in X$ for an element of M. For instance, M satisfies the inequality $xy^{\omega+1} \leqslant x^\omega y$ if, for all $x, y \in M$, $xy^{\omega+1} \leqslant x^\omega y$.

Varieties of ordered monoids. Let $\mathbf{V}$ be a variety of ordered monoids, let X be an alphabet and let $u, v \in \widehat{X^*}$. Then $\mathbf{V}$ *satisfies an inequality* $u \leqslant v$ if each ordered monoid of $\mathbf{V}$ satisfies the inequality. In this context, equalities of the form $u = v$ are often called *identities*.

It is proved in [20] that any variety of ordered monoids may be defined by a (possibly infinite) set of such inequalities. This result extends to the ordered case the classical result of Reiterman [22] and Banaschewski [5]: any variety of monoids may be defined by a (possibly infinite) set of identities.

The case of lp-varieties and ld-varieties of ordered stamps. Let $\mathbf{V}$ be an lp-variety [ld-variety] of ordered stamps, let X be an alphabet and let $u, v \in \widehat{X^*}$. Then $\mathbf{V}$ *satisfies the inequality* $u \leqslant v$ if, for each stamp $\varphi : A^* \to M$ of $\mathbf{V}$ and for every length-preserving [length-decreasing] morphism $f : X^* \to A^*$, $\widehat{\varphi}(\widehat{f}(u)) \leqslant \widehat{\varphi}(\widehat{f}(v))$.

The difficulty is to interpret correctly $\widehat{f}(u)$. If f is length-preserving, $\widehat{f}(u)$ is obtained by replacing each symbol $x \in X$ by a letter of A. For instance, an lp-variety $\mathbf{V}$ satisfies the inequality $xy^{\omega+1} \leqslant x^\omega y$ if, for each stamp $\varphi : A^* \to M$ of $\mathbf{V}$ and for all letters $a, b \in A$, $\widehat{\varphi}(ab^{\omega+1}) \leqslant \widehat{\varphi}(a^\omega b)$.

It is proved in [15, 19] that any ordered lp-variety of stamps may be defined by a (possibly infinite) set of such inequalities.

If f is length-decreasing, this is even more tricky. Then $\widehat{f}(u)$ is obtained by replacing each symbol $x \in X$ by either a letter of A or by the empty word. For instance, an ld-variety $\mathbf{V}$ satisfies the inequality $xy^{\omega+1} \leqslant x^\omega y$ if, for each stamp $\varphi : A^* \to M$ of $\mathbf{V}$ and for all letters $a, b \in A$, $\widehat{\varphi}(ab^{\omega+1}) \leqslant \widehat{\varphi}(a^\omega b)$, $\widehat{\varphi}(b^{\omega+1}) \leqslant \widehat{\varphi}(b)$ and $\widehat{\varphi}(a) \leqslant \widehat{\varphi}(a^\omega)$.

It is proved in [15, 19] that any ordered ld-variety of stamps may be defined by a (possibly infinite) set of such inequalities.

We will also need the following elementary result. Recall that a variety of [ordered] monoids is *aperiodic* if it satisfies the identity $x^\omega = x^{\omega+1}$.

Proposition 3.1. *Let* $\mathbf{V}$ *be an aperiodic variety of ordered monoids. Then, for each* $\alpha \in \widehat{\mathbb{N}}$*,* $\mathbf{V}$ *satisfies the identity* $x^\omega = x^\omega x^\alpha$*.*

Proof. Let $\alpha \in \widehat{\mathbb{N}}$. Then $\alpha = \lim_{n\to\infty} k_n$ for some sequence $(k_n)_{n\geqslant 0}$ of nonegative integers. Since $\mathbf{V}$ is aperiodic, it satisfies the identity $x^{\omega+k_n} = x^\omega$ for all n, and hence it also satisfies the identity $x^\omega x^\alpha = x^\omega$. $\square$

4 Renaming

In this section, we give some general results on renaming.

Since any map may be written as the composition of an injective map with a surjective map, one gets immediately:

Lemma 4.1. *A class of languages is closed under renaming if and only if it is closed under injective and surjective renamings.*

The next two results give a simple description of the positive lp-varieties [ld-varieties] of languages closed under injective renaming:

Proposition 4.1. *The following conditions are equivalent for a positive lp-variety of languages* $\mathcal{V}$*:*

(1) *$\mathcal{V}$ is closed under injective renaming,*

(2) *for each alphabet A and each nonempty set $B \subseteq A$, B^* belongs to $\mathcal{V}(A^*)$,*

(3) *for each alphabet A and each set $B \subseteq A$, B^* belongs to $\mathcal{V}(A^*)$.*

Proof. (1) implies (3). Suppose that $\mathcal{V}$ is closed under injective renaming. Let B be a subset of an alphabet A. Since $B^* \in \mathcal{V}(B^*)$ and since the embedding of B^* into A^* is an injective renaming, one also has $B^* \in \mathcal{V}(A^*)$.

(3) implies (2) is trivial.

(2) implies (3). We have to show that for any alphabet A, $\{1\} \in \mathcal{V}(A^*)$. First assume that A has at least two elements. If $A = B_1 \cup B_2$ is a partition of A into two disjoint nonempty sets B_1 and B_2, then both B_1^* and B_2^* are in $\mathcal{V}(A^*)$, so that $\{1\} = B_1^* \cap B_2^*$ is also in $\mathcal{V}(A^*)$. Now consider a one-letter alphabet a and the two-letter alphabet $\{a, b\}$. The inclusion $h : a^* \to \{a, b\}^*$ is length preserving and thus $\{1\} = h^{-1}(\{1\})$ is in $\mathcal{V}(a^*)$. Finally, the result is trivial if A is empty.

(3) implies (1). Suppose that, for each alphabet A and nonempty set $B \subseteq A$, $B^* \in \mathcal{V}(A^*)$. Let $h : B^* \to A^*$ be an injective renaming. Then there is a renaming $f : A^* \to B^*$ such that $f \circ h$ is the identity function on B^*. Since for any $L \subseteq B^*$, $h(L) = f^{-1}(L) \cap (h(B))^*$, we conclude that $h(L) \in \mathcal{V}(A^*)$ whenever $L \in \mathcal{V}(B^*)$. $\square$

Proposition 4.2. *An ld-variety $\mathcal{V}$ is closed under injective renaming if and only if for each one-letter alphabet a, $\{1\}$ belongs to $\mathcal{V}(a^*)$.*

Proof. Since each *ld*-variety is an *lp*-variety, Proposition 4.1 shows that $\mathcal{V}$ is closed under injective renaming if and only if, for each alphabet A and each subset B of A, B^* belongs to $\mathcal{V}(A^*)$. In particular, if $\mathcal{V}$ is closed under injective renaming, then $\{1\}$ belongs to $\mathcal{V}(a^*)$.

Suppose now that $\mathcal{V}(a^*)$ contains $\{1\}$. Let A be any alphabet and let B be a subset of A. The morphism $h : A^* \to a^*$ that maps each element of B to 1 and all elements of $A \setminus B$ to a is length-decreasing. Since $\mathcal{V}$ is an *ld*-variety and $\{1\}$ belongs to $\mathcal{V}(a^*)$, $h^{-1}(\{1\})$ also belongs to $\mathcal{V}(a^*)$. But $B^* = h^{-1}(\{1\})$, and hence $\mathcal{V}(A^*)$ contains B^* as required. $\square$

Let $\mathbf{V}$ be a variety of ordered monoids and let $\mathcal{V}$ be the corresponding positive variety of languages. A description of the positive variety of languages corresponding to $\mathbf{P}^{\downarrow}\mathbf{V}$ was given by Polák [21, Theorem 4.2] and by Cano and Pin [9] and [10, Proposition 6.3]. The following stronger version[1] was given in [8]. For each alphabet A, let us denote by $\Lambda\mathcal{V}(A^*)$ [$\Lambda'\mathcal{V}(A^*)$] the set of all languages of A^* of the form $\varphi(K)$, where φ is a [surjective] renaming from B^* to A^*, B is an arbitrary finite alphabet, and K is a language of $\mathcal{V}(B^*)$.

Theorem 4.1. *The class $\Lambda\mathcal{V}$ [$\Lambda'\mathcal{V}$] is a positive variety of languages and the corresponding variety of ordered monoids is $\mathbf{P}_0^{\downarrow}\mathbf{V}$ [$\mathbf{P}^{\downarrow}\mathbf{V}$].*

Corollary 4.1. *A positive variety of languages $\mathcal{V}$ is closed under [surjective] renaming if and only if $\mathbf{V} = \mathbf{P}_0^{\downarrow}\mathbf{V}$ [$\mathbf{V} = \mathbf{P}^{\downarrow}\mathbf{V}$].*

[1] We warn the reader that a different notation was used in [8].

5 Commutative varieties

A stamp $\varphi : A^* \to M$ is said to be *commutative* if M is commutative. An ld-variety is *commutative* if all its stamps are commutative. A stamp $\varphi : A^* \to M$ is called *monogenic* if A is a singleton alphabet.

Proposition 5.1. *Every commutative ld-variety of* [*ordered*] *stamps is generated by its monogenic* [*ordered*] *stamps.*

Proof. We first give the proof in the unordered case. Let $\mathbf{V}$ be a commutative ld-variety of stamps and let $\varphi : A^* \to M$ be a stamp of $\mathbf{V}$. For each $a \in A$, denote by M_a the submonoid of M generated by $\varphi(a)$ and let $\gamma_a : A^* \to M_a$ be the stamp defined by $\gamma_a(a) = \varphi(a)$ and $\gamma_a(c) = 1$ for $c \neq a$. Let $\mathbf{W}$ be the ld-variety of stamps generated by the stamps γ_a, for $a \in A$. We claim that $\mathbf{V} = \mathbf{W}$.

Let $\pi_a : A^* \to A^*$ be the length-decreasing morphism defined by $\pi_a(a) = a$ and $\pi_a(c) = 1$ for $c \neq a$. Denoting by ι_a the natural embedding from M_a into M, one gets the following commutative diagram:

$$\begin{tikzcd}
A^* \arrow[r, "\pi_a"] \arrow[d, "\gamma_a"'] & A^* \arrow[d, "\varphi"] \\
M_a \arrow[r, "\iota_a"] & M
\end{tikzcd}$$

Therefore (π_a, ι_a) is an ld-inclusion and each stamp γ_a belongs to $\mathbf{V}$. Thus $\mathbf{W} \subseteq \mathbf{V}$.

The restricted product γ of the stamps γ_a also belongs to $\mathbf{W}$. Note that γ is a surjective morphism from A^* onto $\prod_{a\in A} M_a$. Moreover, the function $\alpha : \prod_{a\in A} M_a \to M$ which maps each family $(m_a)_{a\in A}$ onto the product $\prod_{a\in A} m_a$ is a surjective morphism. Since $\alpha \circ \gamma = \varphi$, the stamp φ belongs to $\mathbf{W}$. Thus $\mathbf{V} \subseteq \mathbf{W}$. This proves the claim and the proposition.

In the ordered case, each M_a is an ordered submonoid of M and thus each γ_a is an ordered stamp. Since ι_a clearly preserves the order, the same argument shows that each γ_a is in $\mathbf{V}$ and thus $\mathbf{W} \subseteq \mathbf{V}$. For the reverse inclusion, one basically needs to observe that $\prod_{a\in A} M_a$ is equipped with the product order, and that the map α preserves the order, since M is an ordered monoid. □

A similar but simpler proof would give the following result:

Proposition 5.2. *Every commutative variety of* [*ordered*] *monoids is generated by its monogenic* [*ordered*] *monoids.*

Proposition 5.1 has an interesting consequence in terms of languages. Equivalently, a language is *commutative* if its syntactic monoid is commutative.

Corollary 5.1. *Let $\mathcal{V}_1$ and $\mathcal{V}_2$ be two positive ld-varieties of commutative languages. Then $\mathcal{V}_1 \subseteq \mathcal{V}_2$ if and only if $\mathcal{V}_1(a^*) \subseteq \mathcal{V}_2(a^*)$.*

Corollary 5.1 shows that a positive commutative ld-variety of languages is entirely determined by its languages on a one-letter alphabet. Here is a more explicit version of this result.

Proposition 5.3. *Let $\mathcal{V}$ be a commutative positive ld-variety of languages. Then for each alphabet $A = \{a_1, \dots, a_k\}$, $\mathcal{V}(A^*)$ consists of all finite unions of languages of the form $L_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup L_k$ where, for $1 \leqslant i \leqslant k$, $L_i \in \mathcal{V}(a_i^*)$.*

Proof. Let $A = \{a_1, \dots, a_k\}$ be an alphabet. Let $\mathcal{W}(A^*)$ consist of all finite unions of languages of the form $L_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup L_k$ where, for $1 \leqslant i \leqslant k$, $L_i \in \mathcal{V}(a_i^*)$. Let us first prove a lemma.

Lemma 5.1. *The class $\mathcal{W}$ is a commutative positive ld-variety of languages.*

Proof. By construction, every language of $\mathcal{W}$ is commutative. Furthermore, $\mathcal{W}(A^*)$ is closed under union. To prove that $\mathcal{W}(A^*)$ is closed under intersection, it suffices to show that the intersection of any two languages $L = L_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup L_k$ and $L' = L'_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup L'_k$ with $L_i, L'_i \in \mathcal{V}(a_i^*)$ is in $\mathcal{W}(A^*)$. We claim that

$$L \cap L' = (L_1 \cap L'_1) \sqcup\!\sqcup \cdots \sqcup\!\sqcup (L_k \cap L'_k) \tag{2}$$

Let R be the right hand side of (2). The inclusion $R \subseteq L \cap L'$ is clear. Moreover, if $u \in L \cap L'$, then $u \in (a_1^{n_1} \sqcup\!\sqcup \cdots \sqcup\!\sqcup a_k^{n_k}) \cap (a_1^{n'_1} \sqcup\!\sqcup \cdots \sqcup\!\sqcup a_k^{n'_k})$, with $a_i^{n_i} \in L_i$ and $a_i^{n'_i} \in L'_i$ for $1 \leqslant i \leqslant k$. This forces $n_i = n'_i$ and hence $u \in R$, which proves the claim.

Let us prove that $\mathcal{W}(A^*)$ is closed under quotient by any word u. Setting $n_i = |u|_{a_i}$ for $1 \leqslant i \leqslant k$, it suffices to observe that

$$u^{-1}(L_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup L_k) = (a_1^{n_1})^{-1} L_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup (a_k^{n_k})^{-1} L_k$$

Finally, let $\alpha : B^* \to A^*$ be a length-decreasing morphism. It is proved in [6, Proposition 1.1] that

$$\alpha^{-1}(L_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup L_k) = \alpha^{-1}(L_1) \sqcup\!\sqcup \cdots \sqcup\!\sqcup \alpha^{-1}(L_k) \tag{3}$$

It follows that $\mathcal{W}$ is closed under inverses of ld-morphisms, which concludes the proof. □

Let us now come back to the proof of Proposition 5.3. Since $\mathcal{W}$ is a commutative positive ld-variety by Lemma 5.1, it suffices to prove, by Proposition 5.1, that $\mathcal{V}(a^*) = \mathcal{W}(a^*)$ for each one-letter alphabet a. But this follows from the definition of $\mathcal{W}$. □

Proposition 5.3 has an interesting consequence.

Theorem 5.1. *Every commutative positive ld-variety of languages is a positive variety of languages.*

Proof. Let $\mathcal{V}$ be a commutative positive *ld*-variety of languages and let $\mathcal{W}$ be the positive variety of languages generated by $\mathcal{V}$. We claim that $\mathcal{V} = \mathcal{W}$. Since $\mathcal{V}$ is contained in $\mathcal{W}$, Corollary 5.1 shows that it suffices to prove that $\mathcal{W}(a^*) \subseteq \mathcal{V}(a^*)$ for each one-letter alphabet a. Since inverses of morphisms commute with Boolean operations and quotients, it suffices to prove that if $\varphi : a^* \to A^*$ is a morphism and $L \in \mathcal{V}(A^*)$, then $\varphi^{-1}(L) \in \mathcal{V}(a^*)$.

Let $\varphi(a) = a_1 \cdots a_k$, where $a_1, \ldots, a_k$ are letters of A. Setting $C = \{c_1, \ldots, c_k\}$, where $c_1, \ldots, c_k$ are distinct letters, one may write φ as $\alpha \circ \beta$ where $\beta : a^* \to C^*$ is defined by $\beta(a) = c_1 \cdots c_k$ and $\alpha : C^* \to A^*$ is defined by $\alpha(c_i) = a_i$ for $1 \leqslant i \leqslant k$.

$$\begin{tikzcd} a^* \arrow[r, "\beta"] \arrow[rr, bend right, "\varphi"'] & C^* \arrow[r, "\alpha"] & A^* \end{tikzcd}$$

Since α is length-preserving, the language $K = \alpha^{-1}(L)$ belongs to $\mathcal{V}(C^*)$. It follows by Proposition 5.3 that K is a finite union of languages of the form $L_1 \shuffle \cdots \shuffle L_k$ where, for $1 \leqslant i \leqslant k$, $L_i \in \mathcal{V}(c_i^*)$. Let, for $1 \leqslant i \leqslant k$, β_i be the unique length preserving morphism from a^* to c_i^*, defined by $\beta_i(a^r) = c_i^r$. We claim that

$$\beta^{-1}(L_1 \shuffle \cdots \shuffle L_k) = \beta_1^{-1}(L_1) \cap \cdots \cap \beta_k^{-1}(L_k) \tag{4}$$

Let R be the right hand side of (4). If $a^r \in R$, then $\beta_i(a^r) \in L_i$. Therefore $c_i^r \in L_i$ and since $\beta(a^r) = (c_1 \cdots c_k)^r$, $\beta(a^r) \in L_1 \shuffle \cdots \shuffle L_k$. Thus R is a subset of $\beta^{-1}(L_1 \shuffle \cdots \shuffle L_k)$.

If now $a^r \in \beta^{-1}(L_1 \shuffle \cdots \shuffle L_k)$, then $\beta(a^r) \in c_1^{n_1} \shuffle \cdots \shuffle c_k^{n_k}$ with $c^{n_i} \in L_i$ for $1 \leqslant i \leqslant k$. But since $\beta(a^r) = (c_1 \cdots c_k)^r$, one has $n_1 = \cdots = n_k = r$ and hence $c_i^r \in L_i$. Therefore $a^r \in \beta_i^{-1}(L_i)$ for all i and thus a^r belongs R. This proves (4).

Since $L_i \in \mathcal{V}(c_i^*)$ and β_i is length-preserving, $\beta_i^{-1}(L_i) \in \mathcal{V}(a^*)$. As K is a finite union of languages of the form $L_1 \shuffle \cdots \shuffle L_k$, Formula (4) shows that $\beta^{-1}(K) \in \mathcal{V}(a^*)$. Finally, since $\varphi = \alpha \circ \beta$, one gets $\varphi^{-1}(L) = \beta^{-1}(\alpha^{-1}(L)) = \beta^{-1}(K)$. Therefore $\varphi^{-1}(L) \in \mathcal{V}(a^*)$, which concludes the proof. $\square$

Theorem 5.1 has a curious interpretation on the set of natural numbers, mentioned in [11]. Setting, for each subset L of $\mathbb{N}$ and each positive integer k,

$$\begin{aligned} L - 1 &= \{n \in \mathbb{N} \mid n + 1 \in L\} \\ L \div k &= \{n \in \mathbb{N} \mid kn \in L\} \end{aligned}$$

one gets the following result:

Proposition 5.4. *Let $\mathcal{L}$ be a lattice of finite subsets*[2] *of $\mathbb{N}$ such that if $L \in \mathcal{L}$, then $L - 1 \in \mathcal{L}$. Then for each positive integer k, $L \in \mathcal{L}$ implies $L \div k \in \mathcal{L}$.*

[2] It also works for a lattice of regular subsets of $\mathbb{N}$.

6 Operations on commutative languages

In this section, we compare the expressive power of three operations on commutative languages: product, shuffle and renaming.

6.1 Shuffle

Let us say that a positive variety of languages $\mathcal{V}$ is *closed under product over one-letter alphabets* if, for each one-letter alphabet a, $\mathcal{V}(a^*)$ is closed under product. Commutative positive varieties closed under shuffle may be described in various ways.

Proposition 6.1. *Let $\mathcal{V}$ be a commutative positive variety of languages and let $\mathbf{V}$ be the corresponding variety of ordered monoids. The following conditions are equivalent:*

(1) *$\mathcal{V}$ is closed under surjective renaming,*
(2) *$\mathcal{V}$ is closed under shuffle product,*
(3) *$\mathcal{V}$ is closed under product over one-letter alphabets,*
(4) $\mathbf{V} = \mathbf{P}^{\downarrow}\mathbf{V}$.

Proof. (1) implies (2). Let $B = A \times \{0, 1\}$ and let π_0, π_1 and π be the three morphisms from B^* to A^* defined for all $a \in A$ by

$$\begin{array}{lll} \pi_0(a,0) = a & \pi_1(a,0) = 1 & \pi(a,0) = a \\ \pi_0(a,1) = 1 & \pi_1(a,1) = a & \pi(a,1) = a \end{array}$$

Let L_0 and L_1 be two languages of A^*. Since π is a surjective renaming, the formula $L_0 \sqcup\!\sqcup L_1 = \pi(\pi_0^{-1}(L_0) \cap \pi_1^{-1}(L_1))$ shows that every positive variety closed under surjective renaming is closed under shuffle product.

(2) implies (3) is trivial since on a one-letter alphabet, shuffle product and product are the same.

(3) implies (1). Let $\pi : A^* \to B^*$ be a surjective renaming. For each $b \in B$, let $\gamma_b : b^* \to a^*$ be the renaming which maps b onto a. Let L be a language of $\mathcal{V}(A^*)$. By Proposition 5.3, L is a finite union of languages of the form $\sqcup\!\sqcup_{a\in A} L_a$ where $L_a \in \mathcal{V}(a^*)$ for each $a \in A$. For each $b \in B$, let

$$K_b = \prod_{a\in\pi^{-1}(b)} \gamma_b^{-1}(L_a)$$

If $\mathcal{V}(a^*)$ is closed under product for each one-letter alphabet a, then K_b belongs to $\mathcal{V}(b^*)$. Finally, the formula $\pi(L) = \sqcup\!\sqcup_{b\in B} K_b$ shows that $\pi(L)$ belongs to $\mathcal{V}(B^*)$. Therefore $\mathcal{V}$ is closed under surjective renaming.

Finally, the equivalence of (1) and (4) follows from Corollary 4.1. $\square$

6.2 Renaming

Let us say that a positive variety of languages contains $\{1\}$ if, for every alphabet A, $\mathcal{V}(A^*)$ contains the language $\{1\}$. The following result is a slight variation on Proposition 6.1.

Proposition 6.2. *Let $\mathcal{V}$ be a commutative positive variety of languages and let $\mathbf{V}$ be the corresponding variety of ordered monoids. The following conditions are equivalent:*

(1) *$\mathcal{V}$ is closed under renaming,*

(2) *$\mathcal{V}$ is closed under surjective renaming and contains $\{1\}$,*

(3) *$\mathcal{V}$ is closed under shuffle product and contains $\{1\}$,*

(4) *$\mathcal{V}$ is closed under product over one-letter alphabets and contains $\{1\}$,*

(5) $\mathbf{V} = \mathbf{P}_0^{\downarrow}\mathbf{V}$.

Proof. The equivalence of (2)—(4) follows directly from Proposition 6.1. If (2) holds, then $\mathcal{V}$ is closed under injective renaming by Proposition 4.2 and hence is closed under renaming by Lemma 4.1. Thus (2) implies (1).

To show that (1) implies (2), it suffices to show that if $\mathcal{V}$ is closed under renaming then it contains $\{1\}$. Let $A = \{a, b\}$ and let $\pi : A^* \to A^*$ be the renaming defined by $\pi(a) = \pi(b) = a$. Since $A^* \in \mathcal{V}(A^*)$ and $\pi(A^*) = a^*$, one has $a^* \in \mathcal{V}(A^*)$. A similar argument would show that $b^* \in \mathcal{V}(A^*)$ and thus the language $\{1\}$, which is the intersection of a^* and b^* also belongs to $\mathcal{V}(A^*)$. Consider now an alphabet B and the morphism α from B^* to A^* defined by $\alpha(c) = a$ for each $c \in B$. Then $\alpha^{-1}(\{1\}) = \{1\}$ and thus $\mathcal{V}$ contains $\{1\}$.

Finally, the equivalence of (1) and (5) follows from Corollary 4.1. □

7 Three examples

In this section, we study the positive varieties of languages generated by the languages of Examples 2.1, 2.2 and 2.3.

7.1 The language $1 + a$

Let L be the language $1 + a$, let M be its ordered syntactic monoid and let $\mathcal{V}$ be the smallest commutative positive variety such that $\mathcal{V}(a^*)$ contains L. Let $\mathbf{V}$ be the variety of finite ordered monoids corresponding to $\mathcal{V}$.

Since a positive variety of languages is closed under quotients, $\mathcal{V}(a^*)$ contains the language $a^{-1}L = 1$. It follows that $\mathcal{V}(a^*)$ contains 4 languages: $\emptyset$, 1, $1 + a$ and a^*. We claim that

$$\mathbf{V} = [\![\, xy = yx,\ x \leqslant 1 \text{ and } x^2 \leqslant x^3 \,]\!].$$

First, the two inequalities $x \leqslant 1$ and $x^2 \leqslant x^3$ hold in M. Furthermore, the inequality $x \leqslant 1$ implies the inequalities of the form $x^p \leqslant x^q$ with $p > q$ and the inequality $x^2 \leqslant x^3$ implies all the inequalities of the form $x^p \leqslant x^q$ with $2 \leqslant p < q$.

The only other nontrivial inequalities that $\mathbf{V}$ could possibly satisfy are $1 \leqslant x^q$ for $q > 0$ or $x \leqslant x^q$ for $q > 1$. However, M does not satisfy any of these inequalities.

Let $\mathcal{V}'$ be the closure of $\mathcal{V}$ under shuffle, or equivalently, under product over one-letter alphabets. Then $\mathcal{V}'(a^*)$ contains the empty language, the language a^* and all languages of the form $(1+a)^n$ with $n \geqslant 0$. By Theorem 4.1 and Proposition 6.1, $\mathcal{V}'$ corresponds to the variety of ordered monoids $\mathbf{P}^{\downarrow}\mathbf{V}$. We claim that

$$\mathbf{P}^{\downarrow}\mathbf{V} = [\![\, xy = yx \text{ and } x \leqslant 1 \,]\!].$$

Indeed, the ordered syntactic monoids of the languages of $\mathcal{V}'(a^*)$ all satisfy $xy = yx$ and $x \leqslant 1$. Conversely, if the ordered syntactic monoid of a language K of a^* satisfies $x \leqslant 1$, then $x^n \leqslant_K 1$ for every $n \geqslant 0$, and K is closed under taking subwords. If K is infinite, this forces $K = a^*$. If K is finite, it is necessarily of the form $(1 + a)^n$ with $n \geqslant 0$. In both cases, K belongs to $\mathcal{V}'(a^*)$.

Finally, let $\mathbf{W}$ be the variety of ordered monoids corresponding to the closure of $\mathcal{V}$ under renaming. Since $U_1^{\downarrow} \in \mathbf{P}^{\downarrow}\mathbf{V}$, Theorem 4.1 and Formula (1) show that

$$\mathbf{W} = \mathbf{P}_0^{\downarrow}\mathbf{V} = \mathbf{P}^{\downarrow}\mathbf{V} \vee \mathbf{Sl}^{\downarrow} = \mathbf{P}^{\downarrow}\mathbf{V} = [\![\, xy = yx \text{ and } x \leqslant 1 \,]\!].$$

7.2 The language $a + a^6a^*$

Let L be the language $a + a^6a^*$, let M be its ordered syntactic monoid and let $\mathcal{V}$ be the smallest commutative positive variety such that $\mathcal{V}(a^*)$ contains L. Let $\mathbf{V}$ be the variety of finite ordered monoids corresponding to $\mathcal{V}$.

Since a positive variety of languages is closed under quotients, $\mathcal{V}(a^*)$ contains the language $a^{-1}L = 1 + a^5a^*$ and the language $L \cap a^{-1}L = a^6a^*$. It also contains the quotients of this language, which are the languages a^ja^*, for $j \leqslant 6$. Taking the union with L, $a^{-1}L$ or both, one finally concludes that $\mathcal{V}(a^*)$ contains 20 languages: $\emptyset$, a^ia^* for $0 \leqslant i \leqslant 6$, $1 + a^ia^*$ for $1 \leqslant i \leqslant 5$, $a + a^ia^*$ for $3 \leqslant i \leqslant 6$ and $1 + a + a^ia^*$ for $3 \leqslant i \leqslant 5$.

We claim that

$$\mathbf{V} = [\![xy = yx, 1 \leqslant x^5, x^2 \leqslant x^3, x^6 = x^7]\!].$$

Indeed, all defining inequalities hold in M. Since $x^6 = x^7$, the other possible inequalities satisfied by M are equivalent to an inequality of the form $x^p \leqslant x^q$ with $p < q \leqslant 6$. For $p = 0$, the only inequalities of this form satisfied by M are $1 \leqslant x^5$ and $1 \leqslant x^6$, but $1 \leqslant x^6$ is a consequence of $1 \leqslant x^5$ and $x^2 \leqslant x^3$ since $1 \leqslant x^5 = x^3x^2 \leqslant x^3x^3 = x^6$. For $p = 1$, the only inequality of this form satisfied by M is $x \leqslant x^6$, which is a consequence of $1 \leqslant x^5$. Finally, the inequality $x^2 \leqslant x^3$ implies $x^p \leqslant x^q$ for $2 \leqslant p < q \leqslant 6$.

Let $\mathcal{V}'$ be the closure of $\mathcal{V}$ under shuffle, or equivalently, under product over one-letter alphabets. We claim that $\mathcal{V}'(a^*)$ consists of the empty set and the languages of the form

$$a^n(F + a^5a^*) \tag{5}$$

where $n \geqslant 0$ and F is a subset of $(1+a)^4$. First of all, the languages of the form (5) and the empty set form a lattice closed under product, since if $0 \leqslant n \leqslant m$ and F and G are subsets of $(1+a)^4$, then

$$\begin{aligned} a^n(F + a^5a^*) + a^m(G + a^5a^*) &= a^n(F + a^{m-n}G + a^5a^*) \\ a^n(F + a^5a^*) \cap a^m(G + a^5a^*) &= a^m\Big(\big((a^{m-n})^{-1}(F + a^5a^*)\big) \cap G\big) + a^5a^*\Big) \\ a^n(F + a^5a^*)a^m(G + a^5a^*) &= a^{n+m}(FG + a^5a^*) \end{aligned}$$

Since $\mathcal{V}'(a^*)$ is closed under finite unions, it just remains to prove that the languages of the form $a^n(a^k + a^5a^*)$, with $n \geqslant 0$ and $0 \leqslant k \leqslant 4$ all belong to $\mathcal{V}'(a^*)$. But since the languages $a + a^6a^*$ and $1 + a^{5-k}a^*$ are in $\mathcal{V}(a^*)$, this follows from the formula

$$a^n(a^k + a^5a^*) = \big(a + a^6a^*\big)^{n+k}(1 + a^{5-k}a^*)$$

By Theorem 4.1 and Proposition 6.1, $\mathcal{V}'$ corresponds to the variety of ordered monoids $\mathbf{P}^{\downarrow}\mathbf{V}$. We claim that

$$\mathbf{P}^{\downarrow}\mathbf{V} = [\![\, xy = yx \text{ and } 1 \leqslant x^n \text{ for } 5 \leqslant n \leqslant 9 \,]\!].$$

Indeed, the ordered syntactic monoid of any of the languages of the form (5) satisfies all inequalities of the form $1 \leqslant x^n$ for $n \geqslant 5$, but the syntactic ordered monoid of $1 + a^2a^*$ does not satisfy any inequality of the form $x^p \leqslant x^q$ with $p > q$. Moreover, the only inequalities that are not an immediate consequence of an inequality of the form $1 \leqslant x^n$ with $5 \leqslant n \leqslant 9$ are the inequalities $x^i \leqslant x^j$ with $0 \leqslant j - i \leqslant 4$. But none of these inequalities are satisfied by the ordered syntactic monoid of $a^i(1 + a^5a^*)$.

Finally, Theorem 4.1 and Formula (1) show that the variety of ordered monoids corresponding to the closure of $\mathcal{V}$ under renaming is

$$\begin{aligned} \mathbf{P}_0^{\downarrow}\mathbf{V} &= \mathbf{P}^{\downarrow}\mathbf{V} \vee \mathbf{Sl}^{\downarrow} \\ &= [\![\, xy = yx \text{ and } 1 \leqslant x^n \text{ for } 5 \leqslant n \leqslant 9 \,]\!] \vee [\![xy = yx, x^2 = x, x \leqslant 1 \,]\!]. \end{aligned}$$

We claim that $\mathbf{P}_0^{\downarrow}\mathbf{V} = \mathbf{W}$, where

$$\mathbf{W} = [\![\, xy = yx \text{ and } x \leqslant x^n \text{ for } 6 \leqslant n \leqslant 10 \,]\!].$$

First, the inequality $x \leqslant x^n$ is a consequence both of the inequality $1 \leqslant x^{n-1}$ and of the equation $x = x^2$. It follows that $\mathbf{P}_0^{\downarrow}\mathbf{V} \subseteq \mathbf{W}$. To establish the opposite inclusion, it suffices to establish the claim that any inequality of the form $x^p \leqslant x^q$ satisfied by both $\mathbf{P}^{\downarrow}\mathbf{V}$ and $\mathbf{Sl}^{\downarrow}$ is also satisfied by $\mathbf{W}$. If $p = 0$, then the inequality becomes $1 \leqslant x^q$ and it is not satisfied by $\mathbf{Sl}^{\downarrow}$ since $1 \not\leqslant 0$ in $U_1^{\downarrow}$. Moreover, for $p > 0$, the only inequalities of the form $x^p \leqslant x^q$ that are not an immediate consequence of an inequality of the form $x \leqslant x^n$ with $6 \leqslant n \leqslant 10$ are the inequalities $x^p \leqslant x^q$ with $0 \leqslant q - p \leqslant 4$. But we already observed that the ordered syntactic monoid of $a^p(1 + a^5a^*)$ belongs to $\mathbf{P}^{\downarrow}\mathbf{V}$ but does not satisfy any of these inequalities, which proves the claim.

7.3 The language $a + (a^3 + a^4)(a^7)^*$

Let L be the language $a + (a^3 + a^4)(a^7)^*$, let M be its ordered syntactic monoid and let $\mathcal{V}$ be smallest commutative positive variety such that $\mathcal{V}(a^*)$ contains L. Let $\mathbf{V}$ be the variety of finite ordered monoids corresponding to $\mathcal{V}$. One has

$$
\begin{aligned}
(a)^{-1}L &= 1 + (a^2 + a^3)(a^7)^* & (a^2)^{-1}L &= (a + a^2)(a^7)^* \\
(a^3)^{-1}L &= (1 + a)(a^7)^* & (a^4)^{-1}L &= (1 + a^6)(a^7)^* \\
(a^5)^{-1}L &= (a^5 + a^6)(a^7)^* & (a^6)^{-1}L &= (a^4 + a^5)(a^7)^* \\
(a^7)^{-1}L &= (a^3 + a^4)(a^7)^* & (a^8)^{-1}L &= (a^2 + a^3)(a^7)^*
\end{aligned}
$$

The set of final states of the minimal automaton of L is $\{1, 3, 4\}$. The quotients of L are recognised by the same automaton by taking a different set of final states as indicated below

$$
\begin{aligned}
(a)^{-1}L &\to \{0, 2, 3\} & (a^2)^{-1}L &\to \{1, 2, 8\} \\
(a^3)^{-1}L &\to \{0, 1, 7, 8\} & (a^4)^{-1}L &\to \{0, 6, 7\} \\
(a^5)^{-1}L &\to \{5, 6\} & (a^6)^{-1}L &\to \{4, 5\} \\
(a^7)^{-1}L &\to \{3, 4\} & (a^8)^{-1}L &\to \{2, 3\}
\end{aligned}
$$

Observing that

$$
\begin{aligned}
\{0\} &= \{0, 2, 3\} \cap \{0, 6, 7\} & \{1\} &= \{1, 3, 4\} \cap \{1, 2, 8\} \\
\{2\} &= \{0, 2, 3\} \cap \{1, 2, 8\} & \{3\} &= \{1, 3, 4\} \cap \{0, 2, 3\} \\
\{4\} &= \{3, 4\} \cap \{4, 5\} & \{5\} &= \{4, 5\} \cap \{5, 6\} \\
\{6\} &= \{5, 6\} \cap \{0, 6, 7\} & \{0, 7\} &= \{0, 6, 7\} \cap \{0, 1, 7, 8\} \\
\{1, 8\} &= \{1, 2, 8\} \cap \{0, 1, 7, 8\}
\end{aligned}
$$

it follows that a language belongs to the lattice of languages generated by the quotients of L if and only if it is accepted by the minimal automaton of L equipped with a set F of final states satisfying the two conditions

$$7 \in F \implies 0 \in F \quad \text{and} \quad 8 \in F \implies 1 \in F \tag{6}$$

Now, the complement of a set F satisfying (6) also satisfies (6). It follows that the lattice of languages generated by the quotients of L is actually a Boolean algebra and consequently, $\mathcal{V}$ is a variety of languages. It also follows that

$$\mathbf{V} = [\![xy = yx, x^2 = x^9]\!].$$

Moreover, since $U_1 = \{0, 1\}$ belongs to $\mathbf{V}$, it follows that $\mathbf{PV} = \mathbf{P}_0\mathbf{V}$. By [16, Théorème 2.14], $\mathbf{PV}$ is the variety of all commutative monoids whose groups satisfy the identity $x^7 = 1$. Therefore

$$\mathbf{PV} = [\![xy = yx, x^\omega = x^{\omega+7}]\!].$$

The closure of $\mathcal{V}$ under shuffle, or equivalently, under product over one-letter alphabets, and the closure of $\mathcal{V}$ under renaming both correspond to the variety of monoids $\mathbf{PV}$.

8 Conclusion

We gave an algebraic characterization of the commutative positive varieties of languages closed under shuffle product, renaming or product over one-letter alphabets, but several questions might be worth a further study.

First, each commutative variety of ordered monoids can be described by the equality $xy = yx$ and by a set of inequalities in one variable, like $x^p \leqslant x^q$ or more generally $x^\alpha \leqslant x^\beta$ with $\alpha, \beta \in \widehat{\mathbb{N}}$. It would then be interesting to compare these varieties. We just mention a few results of this flavour, which may help in finding bases of inequalities for commutative positive varieties of languages.

Proposition 8.1. *The variety* $[\![xy = yx, x \leqslant x^{n+1}]\!]$ *is contained in the variety* $[\![xy = yx, x \leqslant x^{m+1}]\!]$ *if and only if n divides m.*

Proof. Suppose that n divides m, that is, $m = kn$ for some $k \geqslant 0$. If $x \leqslant x^{n+1}$, then $x \leqslant xx^n$ and by induction, $x \leqslant xx^{kn} = xx^m = x^{m+1}$. Thus $[\![xy = yx, x \leqslant x^{n+1}]\!]$ is contained in the variety $[\![xy = yx, x \leqslant x^{m+1}]\!]$.

Suppose now that $[\![xy = yx, x \leqslant x^{n+1}]\!]$ is contained in the variety $[\![xy = yx, x \leqslant x^{m+1}]\!]$. Then the ordered syntactic monoid of $a(a^n)^*$ satisfies the inequality $x \leqslant x^{n+1}$ and thus it also satisfies the inequality $x \leqslant x^{m+1}$. Since $a \in a(a^n)^*$, this means in particular that $a^m \in a(a^n)^*$ and thus that n divides m. □

In fact, a more general result holds. For each set of natural numbers S, let

$$\mathbf{V}_S = [\![\, xy = yx, x \leqslant x^{n+1} \text{ for all } n \in S \,]\!].$$

Let $\langle S \rangle$ denote the additive submonoid of $\mathbb{N}$ generated by S. It is a well-known fact that any additive subsemigroup of $\mathbb{N}$ is finitely generated and consequently, there exists a finite set of natural numbers F_S such that $\langle S \rangle = \langle F_S \rangle$.

Proposition 8.2. *The variety $\mathbf{V}_S$ satisfies the inequality $x \leqslant x^{m+1}$ if and only if m belongs to $\langle S \rangle$.*

Proof. Let T be the set of all natural numbers n such that $\mathbf{V}_S$ satisfies the inequality $x \leqslant x^{n+1}$. First observe that T is an additive submonoid of $\mathbb{N}$. Indeed, if $\mathbf{V}_S$ satisfies the inequalities $x \leqslant xx^m$ and $x \leqslant xx^n$, then it satisfies $x \leqslant xx^m \leqslant (xx^n)x^m = x^{n+m+1}$. Now T contains S by definition and thus also $\langle S \rangle$. It follows that if m belongs to $\langle S \rangle$, then $\mathbf{V}_S$ satisfies the inequality $x \leqslant x^{m+1}$.

Suppose now that $\mathbf{V}_S$ satisfies the inequality $x \leqslant x^{m+1}$ and let

$$L_S = \{a^{n+1} \mid n \in \langle S \rangle\}.$$

Since $\langle S \rangle = \langle F_S \rangle$, one has

$$L_S = a\{a^s \mid s \in F_S\}^*$$

and thus L_S is a regular language.

We claim that the ordered syntactic monoid M of L_S satisfies an inequality of the form $x \leqslant x^{n+1}$ if and only if $n \in \langle S \rangle$. Suppose first that M satisfies $x \leqslant x^{n+1}$. Then the property $a \in L_S$ implies $a^{n+1} \in L_S$ and hence $n \in \langle S \rangle$.

Conversely, let $n \in \langle S \rangle$. We need to prove that M satisfies the inequality $x \leqslant x^{n+1}$, or equivalently, that $a^k \leqslant_{L_S} (a^k)^{n+1}$ for all $k \geqslant 0$. But for each $r \geqslant 0$, the condition $a^r a^k \in L_S$ implies $r+k-1 \in \langle S \rangle$. Since $r+k(n+1)-1 = r+k-1+kn$, one gets $r+k(n+1)-1 \in \langle S \rangle$ and hence $a^r(a^k)^{n+1} \in L_S$ as required. This concludes the proof of the claim.

In particular, since M satisfies all the inequalities $x \leqslant x^{n+1}$ for $n \in S$, M belongs to $\mathbf{V}_S$ and thus also satisfies the inequality $x \leqslant x^{m+1}$, which finally implies that m belongs to $\langle S \rangle$. $\square$

Corollary 8.1. *Let S and T be two sets of natural numbers. Then $V_S = V_T$ if and only if $\langle S \rangle = \langle T \rangle$.*

It would also be interesting to have a systematic approach to treat examples similar to those given in Section 7. That is, find an algorithm which takes as input a monogenic ordered monoid M and outputs a set of inequalities defining respectively $\mathbf{V}$, $\mathbf{P}^{\downarrow}\mathbf{V}$ and $\mathbf{P}_0^{\downarrow}\mathbf{V}$, where $\mathbf{V}$ is the variety of ordered monoids generated by M.

Acknowledgements

We would like to thank the anonymous referees for their helpful comments.

References

[1] Almeida, Jorge. *Finite semigroups and universal algebra. Series in Algebra*, volume 3. World Scientific, Singapore, 1995.

[2] Almeida, Jorge. Finite semigroups: an introduction to a unified theory of pseudovarieties. In *Semigroups, algorithms, automata and languages (Coimbra, 2001)*, pages 3–64. World Sci. Publ., River Edge, NJ, 2002.

[3] Almeida, Jorge. Profinite semigroups and applications. In Kudryavtsev, V. B. and Rosenberg, I. G., editors, *Structural theory of automata, semigroups and universal algebra*, pages 1–45, New York, 2005. Springer.

[4] Almeida, Jorge, Cano, Antonio, Klíma, Ondrej, and Pin, Jean-Éric. Fixed points of the lower set operator. *Internat. J. Algebra Comput.*, 25(1-2):259–292, 2015.

[5] Banaschewski, Bernhard. The Birkhoff theorem for varieties of finite algebras. *Algebra Universalis*, 17(3):360–368, 1983.

[6] Berstel, Jean, Boasson, Luc, Carton, Olivier, Pin, Jean-Éric, and Restivo, Antonio. The expressive power of the shuffle product. *Information and Computation*, 208:1258–1272, 2010.

[7] Branco, Mário J.J. and Pin, Jean-Éric. Equations for the polynomial closure. In Albers, S., Marchetti-Spaccamela, A., Matias, Y., Nikoletseas, S., and Thomas, W., editors, *ICALP 2009, Part II*, volume 5556 of *Lect. Notes Comp. Sci.*, pages 115–126, Berlin, 2009. Springer.

[8] Cano, Antonio and Pin, Jean-Éric. Upper set monoids and length preserving morphisms. *J. of Pure and Applied Algebra*, 216:1178–1183, 2012.

[9] Cano Gómez, Antonio. *Semigroupes ordonnés et opérations sur les langages rationnels*. PhD thesis, Université Paris 7 and Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia, 2003.

[10] Cano Gómez, Antonio and Pin, Jean-Éric. Shuffle on positive varieties of languages. *Theoret. Comput. Sci.*, 312:433–461, 2004.

[11] Cégielski, Patrick, Grigorieff, Serge, and Guessarian, Irène. On lattices of regular sets of natural integers closed under decrementation. *Inform. Process. Lett.*, 114(4):197–202, 2014.

[12] Eilenberg, S. *Automata, Languages and Machines*, volume B. Academic Press, New York, 1976.

[13] Ésik, Zoltán and Ito, Masami. Temporal logic with cyclic counting and the degree of aperiodicity of finite automata. *Acta Cybernetica*, 16:1–28, 2003.

[14] Ésik, Zoltan and Simon, Imre. Modeling literal morphisms by shuffle. *Semigroup Forum*, 56:225–227, 1998.

[15] Kunc, Michal. Equational description of pseudovarieties of homomorphisms. *Theoret. Informatics Appl.*, 37:243–254, 2003.

[16] Perrot, Jean-François. Variétés de langages et operations. *Theoret. Comput. Sci.*, 7:197–210, 1978.

[17] Pin, Jean-Éric. A variety theorem without complementation. *Russian Mathematics (Iz. VUZ)*, 39:80–90, 1995.

[18] Pin, Jean-Éric. Equational descriptions of languages. *Int. J. Found. Comput. S.*, 23:1227–1240, 2012.

[19] Pin, Jean-Éric and Straubing, Howard. Some results on $\mathcal{C}$-varieties. *Theoret. Informatics Appl.*, 39:239–262, 2005.

[20] Pin, Jean-Éric and Weil, Pascal. A Reiterman theorem for pseudovarieties of finite first-order structures. *Algebra Universalis*, 35:577–595, 1996.

[21] Polák, Libor. Operators on classes of regular languages. In Gomes, Gracinda, Pin, Jean-Éric, and Silva, P.V., editors, *Semigroups, Algorithms, Automata and Languages*, pages 407–422. World Scientific, 2002.

[22] Reiterman, Jan. The Birkhoff theorem for finite algebras. *Algebra Universalis*, 14(1):1–10, 1982.

[23] Reutenauer, Christophe. Sur les variétés de langages et de monoïdes. In *Theoretical computer science (Fourth GI Conf., Aachen)*, volume 67 of *Lect. Notes Comp. Sci.*, pages 260–265. Springer, Berlin, 1979.

[24] Schützenberger, Marcel Paul. Une théorie algébrique du codage. *Séminaire Dubreil. Algèbre et théorie des nombres*, 9:1–24, 1955-1956. `http://eudml.org/doc/111094`.

[25] Straubing, Howard. Aperiodic homomorphisms and the concatenation product of recognizable sets. *J. Pure Appl. Algebra*, 15(3):319–327, 1979.

[26] Straubing, Howard. Recognizable sets and power sets of finite semigroups. *Semigroup Forum*, 18:331–340, 1979.

[27] Straubing, Howard. On logical descriptions of regular languages. In *LATIN 2002*, number 2286 in Lect. Notes Comp. Sci., pages 528–538, Berlin, 2002. Springer.

[28] Weil, Pascal. Profinite methods in semigroup theory. *Int. J. Alg. Comput.*, 12:137–178, 2002.