Geometrically Closed Positive Varieties of Languages*

Ondřej Klíma[a,1], Peter Kostolányi[b,2]

[a] *Department of Mathematics and Statistics, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic*
[b] *Department of Computer Science, Comenius University in Bratislava, Mlynská dolina, 842 48 Bratislava, Slovakia*

Abstract

A recently introduced operation of geometrical closure on formal languages is investigated from the viewpoint of algebraic language theory. Positive varieties $\mathcal{V}$ containing exclusively languages with regular geometrical closure are fully characterised by inclusion of $\mathcal{V}$ in $\mathcal{W}$, a known positive variety arising in the study of the commutative closure. It is proved that the geometrical closure of a language from the intersection of $\mathcal{W}$ with the variety of all star-free languages $\mathcal{SF}$ always falls into $\mathcal{R}_{LT}$, which is introduced as a subvariety of $\mathcal{R}$, the variety of languages recognised by R-trivial monoids. All classes between $\mathcal{R}_{LT}$ and $\mathcal{W} \cap \mathcal{SF}$ are thus geometrically closed: for instance, the level 3/2 of the Straubing-Thérien hierarchy, the **DA**-recognisable languages, or the variety $\mathcal{R}$. It is also shown that $\mathcal{W} \cap \mathcal{SF}$ is the largest geometrically closed positive variety of star-free languages, while there is no largest geometrically closed positive variety of regular languages.

Keywords: Geometrical closure, Commutative closure, Variety of languages, Star-free language

1. Introduction

A *geometrical closure* is an operation on formal languages introduced recently by Dubernard, Guaiana, and Mignot [11]. It is defined as follows: Take a language L over some k-letter alphabet and consider the set called the *figure* of L, which consists of all elements of $\mathbb{N}^k$ corresponding to Parikh vectors of prefixes of words from L. The *geometrical closure* of L is the language $\gamma(L)$ of all words w over the same alphabet such that the Parikh vectors of all the prefixes of w lie in the figure of L. This closure operator was inspired by the previous works of Blanpain, Champarnaud, and Dubernard [4] and Béal et al. [3], in which *geometrical languages* are studied – using the terminology of the later article [11], these can be described as languages whose prefix closure is equal to their geometrical closure. Note that this terminology was motivated by the fact that geometrical languages are completely determined, up to prefix closure, by their (geometrical) figures. In the particular case of binary alphabets, these figures were illustrated by plane diagrams in [11].

The class of all regular languages is easily observed not to be geometrically closed – that is, one can find a regular language such that its geometrical closure is not regular [11] (see also the end of Section 2). It is thus natural to ask which regular languages have regular geometrical closures, or to describe some robust classes of languages with this property. Another problem posed in [11] is to find some geometrically closed subclasses of regular languages. We consider all these problems; the latter constitutes the main theme of this article, while for the former two we obtain a complete characterisation at the level of positive varieties.

*A preliminary version [16] of this article appeared in the proceedings of the conference LATA 2020.
Email addresses: `klima@math.muni.cz` (Ondřej Klíma), `kostolanyi@fmph.uniba.sk` (Peter Kostolányi)
[1] The first author was supported by Grant 19-12790S of the Grant Agency of the Czech Republic.
[2] The second author was supported by the grant VEGA 1/0601/20.

A prominent position among subclasses of regular languages is held by *varieties*, and more generally by *positive varieties of languages*, related via Eilenberg correspondence to *pseudovarieties of finite monoids* and *finite ordered monoids*, respectively.[3] Among (positive) varieties playing a significant role in this article, let us mention the variety $\mathcal{SF}$ of all star-free languages, whose relationship with the pseudovariety of finite aperiodic monoids $\mathbf{A}$ is probably the best-known instance of Eilenberg correspondence.[4] The variety $\mathcal{SF}$ can be classified into the *Straubing-Thérien hierarchy* based on polynomial and Boolean operations. In particular, the variety $\mathcal{V}_1$ (*i.e.*, the variety of languages of level 1) consists of all *piecewise testable languages* and the positive variety $\mathcal{V}_{3/2}$ is formed by polynomials built upon languages of level 1. In addition, let us mention the variety $\mathcal{R}$, corresponding to the pseudovariety $\mathbf{R}$ of all finite R-trivial monoids, and the positive variety $\mathcal{W}$ introduced by Cano Gómez and Pin [8, 9], which is by definition the largest positive variety of regular languages not containing the language $(ab)^*$. As shown by Cano, Guaiana, and Pin [7], $\mathcal{W}$ is closed under commutation, and irregularity of the commutative closure of $(ab)^*$ implies that it is the largest positive variety with this property.

It was proved by Dubernard, Guaiana, and Mignot [11] that the class of all binary languages from the positive variety $\mathcal{V}_{3/2}$ is geometrically closed. They have obtained this result by decomposing the plane diagram of the figure of a given language into specific types of basic subdiagrams, and using this decomposition to construct a regular expression for the language $\gamma(L)$. We prove a generalisation of that result using a different strategy. We do not construct a specific regular expression for $\gamma(L)$, but we determine what kind of expression for $\gamma(L)$ one may obtain. Moreover, we consider L from a larger class, namely from the positive variety $\mathcal{W} \cap \mathcal{SF}$, which has a useful property that it is closed under commutation [7].

In particular, we introduce a variety of languages $\mathcal{R}_{LT}$, which is a subvariety of $\mathcal{R}$. Note that there is a transparent description of languages from $\mathcal{R}$ and also an effective characterisation via the so-called acyclic automata (both are recalled in Subsection 5.1). The variety of languages $\mathcal{R}_{LT}$ is then characterised in the same manner: a precise description by specific regular expressions and also an automata-based characterisation are given. The letters LT in the notation $\mathcal{R}_{LT}$ refer to a characteristic property of acyclic automata in which "loops are transferred" along paths. Moreover, we point out the Eilenberg correspondence between $\mathcal{R}_{LT}$ and the pseudovariety of finite monoids $\mathbf{MK}$.

Utilising closure of $\mathcal{W} \cap \mathcal{SF}$ under commutation, we show that the geometrical closure $\gamma(L)$ of a language L from the positive variety $\mathcal{W} \cap \mathcal{SF}$ always falls into the variety $\mathcal{R}_{LT}$. As a consequence, each class of regular languages between $\mathcal{R}_{LT}$ and $\mathcal{W} \cap \mathcal{SF}$ is geometrically closed. In particular, the positive variety $\mathcal{V}_{3/2}$ is geometrically closed regardless of the alphabet, as well as is the variety $\mathcal{R}$ or the variety corresponding to $\mathbf{DA}$.

We also prove a negative result showing that a positive variety containing the language $(ab)^*$ cannot be geometrically closed, and in fact even contains a language with irregular geometrical closure. This immediately yields a full characterisation of positive varieties $\mathcal{V}$ such that $\gamma(L)$ is regular for all L in $\mathcal{V}$: a positive variety $\mathcal{V}$ has this property if and only if $\mathcal{V} \subseteq \mathcal{W}$. As another consequence of the negative result mentioned above, we observe that $\mathcal{W} \cap \mathcal{SF}$ is the largest geometrically closed positive variety of *star-free* languages and the largest geometrically closed variety of star-free languages corresponds to $\mathbf{DA}$. Our main results on star-free languages can be summarised as follows: A positive variety $\mathcal{V}$ of star-free languages can be geometrically closed only if $\mathcal{V} \subseteq \mathcal{W} \cap \mathcal{SF}$; if moreover $\mathcal{R}_{LT} \subseteq \mathcal{V}$, then $\mathcal{V}$ *is* geometrically closed. We leave open a characterisation of geometrically closed positive varieties of star-free languages not containing $\mathcal{R}_{LT}$, but we prove that the variety $\mathcal{V}_1$ of piecewise testable languages is *not* geometrically closed.

When it comes to geometrically closed classes of regular languages not contained in $\mathcal{SF}$, the situation turns out to be far less satisfying. In particular, we prove that there is no geometrically closed positive variety of languages containing the variety corresponding to $\mathbf{DAb}$ and no largest geometrically closed (positive) variety. On the other hand, several varieties such as the variety of all group languages $\mathcal{G}$, or the variety of all commutative languages, are geometrically closed for trivial reasons.

A preliminary version of this article appeared in the proceedings of LATA 2020 [16]. This is extended here by strengthening the main result for star-free languages ("upper bound" shifted from $\mathcal{V}_{3/2}$ to $\mathcal{W} \cap \mathcal{SF}$)

[3] We refer to the survey by Pin [19] for an introduction to varieties and the algebraic theory of regular languages in general.

[4] Standard notation for pseudovarieties of monoids is used throughout this paper. The reader might consult the Appendix for its concise summary.

and simplifying its proof, proving "optimality" of both $\mathcal{R}_{LT}$ and $\mathcal{W} \cap \mathcal{SF}$ in this result, proving that $\mathcal{V}_1$ is not geometrically closed, and identifying the largest geometrically closed (positive) variety of star-free languages. The findings of Section 4 and most results for other than star-free languages are new as well. New material (except strengthening of the result on star-free languages) comprises mostly (sub)sections 4, 5.3, 5.4, and 6.

2. Preliminaries

All automata considered in this article are understood to be deterministic and finite. An *automaton* is thus a five-tuple $\mathcal{A} = (Q, \Sigma, \cdot, \iota, F)$, where Q is a finite set of states, Σ is a non-empty finite alphabet, $\cdot\colon Q \times \Sigma \to Q$ is a complete transition function, $\iota \in Q$ is the unique initial state, and $F \subseteq Q$ is the set of final states. The transition function $\cdot$ of $\mathcal{A}$ extends naturally to a right action $\cdot\colon Q \times \Sigma^* \to Q$ of the monoid Σ^* on the set Q – hence $q \cdot w$ denotes the state reached from $q \in Q$ upon "reading" $w \in \Sigma^*$. The minimal automaton of a given language L is denoted by $\mathcal{D}_L$.

By a (positive) variety of languages, we always understand what is called a (positive) $*$-variety in [19]. Let us briefly recall this notion for convenience of the reader. A *class of languages* $\mathcal{C}$ is an operator, which determines, for each finite non-empty alphabet Σ, a set $\mathcal{C}(\Sigma^*)$ of languages over Σ. *A positive variety* is a class of regular languages $\mathcal{V}$ such that $\mathcal{V}(\Sigma^*)$ contains the languages $\emptyset$ and Σ^* for every non-empty finite alphabet Σ, while $\mathcal{V}(\Sigma^*)$ is closed under union, intersection, and (left and right) quotients (by words), and the whole class $\mathcal{V}$ is closed under preimages in homomorphisms. A positive variety $\mathcal{V}$ is a *variety* if each $\mathcal{V}(\Sigma^*)$ is closed under complementation.

We say that a language L is in a (positive) variety $\mathcal{V}$ if $L \in \mathcal{V}(\Sigma^*)$ for some non-empty finite alphabet Σ. Furthermore, given two (positive) varieties of languages $\mathcal{U}, \mathcal{V}$, it is easy to prove that the class $\mathcal{U} \cap \mathcal{V}$, defined for all Σ by $(\mathcal{U} \cap \mathcal{V})(\Sigma^*) = \mathcal{U}(\Sigma^*) \cap \mathcal{V}(\Sigma^*)$, is again a (positive) variety of languages.

Note that among our results, those showing that certain classes *are* geometrically closed could also be obtained with an alphabet being fixed. Homomorphisms between different alphabets thus play no role there, and we could equally well consider lattices of languages [13] instead of (positive) varieties of languages for such positive results. However, we prefer to stay in the frame of the theory of (positive) varieties of languages as a primary aim of this article is to describe robust classes of languages closed under geometrical closure.

It is well known that varieties of languages are linked to *pseudovarieties* of finite monoids – *i.e.*, classes of finite monoids closed under taking homomorphic images, submonoids, and finite products – via the so-called *Eilenberg correspondence* [12]. The pseudovariety of finite monoids corresponding to a variety of languages $\mathcal{V}$ is generated by syntactic monoids of languages from $\mathcal{V}$. Conversely, the variety of languages corresponding to a pseudovariety of finite monoids $\mathbf{V}$ consists of precisely all languages with syntactic monoids in $\mathbf{V}$. Pseudovarieties of monoids appear in this article as instances of the correspondence described.[5]

Given two words u, v over an alphabet Σ, we write $u \leq v$ if u is a prefix of v. We also write, for each language $L \subseteq \Sigma^*$,

$$\mathrm{pref}^{\uparrow}(L) := \{u \in \Sigma^* \mid \exists v \in L : u \leq v\} = \bigcup_{w \in \Sigma^*} Lw^{-1},$$
$$\mathrm{pref}^{\downarrow}(L) := \{v \in \Sigma^* \mid \forall u \in \Sigma^* : u \leq v \implies u \in L\}.$$

We call these languages the *prefix closure* and the *prefix reduction* of L, respectively. Both are prefix-closed, while $\mathrm{pref}^{\uparrow}(L) \supseteq L$ and $\mathrm{pref}^{\downarrow}(L) \subseteq L$.

Proposition 2.1. *Each positive variety $\mathcal{V}$ is closed under the operator* $\mathrm{pref}^{\uparrow}$.

Proof. It is well known that each regular language has finitely many right quotients by words. Thus, for each non-empty finite alphabet Σ and each $L \in \mathcal{V}(\Sigma^*)$, the language

$$\mathrm{pref}^{\uparrow}(L) = \bigcup_{w \in \Sigma^*} Lw^{-1}$$

[5] Algebraic descriptions of these pseudovarieties are summarised in the Appendix for convenience of the reader.

is a finite union of right quotients of L, and its membership to $\mathcal{V}(\Sigma^*)$ follows. $\square$

Let $\Sigma = \{a_1, \ldots, a_k\}$ be a linearly ordered alphabet. The *Parikh vector* of a word w in Σ^* is then given by $\Psi(w) = (|w|_{a_1}, \ldots, |w|_{a_k})$, where $|w|_a$ denotes the number of occurrences of the letter a in w. This notation extends naturally to languages: we write $\Psi(L) = \{\Psi(w) \mid w \in L\}$ for $L \subseteq \Sigma^*$. We denote by $[w]$ the equivalence class of the kernel relation of Ψ containing w, *i.e.*,

$$[w] = \{u \in \Sigma^* \mid \Psi(u) = \Psi(w)\}.$$

Then we also write, for each $L \subseteq \Sigma^*$,

$$[L] = \bigcup_{w \in L} [w] = \{u \in \Sigma^* \mid \Psi(u) \in \Psi(L)\}$$

and we call the language $[L]$ the *commutative closure* of L. A language L such that $L = [L]$ is called *commutative*. A class of languages $\mathcal{C}$ is said to be *closed under commutation* if for each non-empty finite alphabet Σ, one has $[L] \in \mathcal{C}(\Sigma^*)$ whenever $L \in \mathcal{C}(\Sigma^*)$.

In the previous paragraph we consider the mapping $\Psi \colon \Sigma^* \to \mathbb{N}^k$, where $\mathbb{N}$ is the set of all non-negative integers. Following the ideas of [11], we introduce some technical notations concerning $\mathbb{N}^k$, whose elements are called *vectors*. We denote by $\mathbf{0}$ the null vector of $\mathbb{N}^k$. Let $\mathbf{x} = (x_1, \ldots, x_k)$ and $\mathbf{y} = (y_1, \ldots, y_k)$ be vectors and $s \in \{1, \ldots, k\}$ be an index. We write $\mathbf{x} \to_s \mathbf{y}$ if $y_s - x_s = 1$ and, at the same time, $y_i = x_i$ for all $i \neq s$. Moreover, $\mathbf{x} \to \mathbf{y}$ means that $\mathbf{x} \to_s \mathbf{y}$ for some index s. A *path* in $\mathbb{N}^k$ is a finite sequence $\pi = [\mathbf{x}_0, \ldots, \mathbf{x}_n]$ of vectors from $\mathbb{N}^k$ such that $\mathbf{x}_0 = \mathbf{0}$ and $\mathbf{x}_{i-1} \to \mathbf{x}_i$ for $i = 1, \ldots, n$; more specifically, we say that π is a *path leading to* $\mathbf{x}_n$. This means that a path always begins in $\mathbf{0}$ and each other vector of the path is obtained from the previous one by incrementing exactly one of its coordinates by one. If in addition $\mathbf{x}_0, \ldots, \mathbf{x}_n$ all belong to a set $F \subseteq \mathbb{N}^k$, we say that π is a path in F and write $\pi \sqsubseteq F$.

Given a word $w = a_{i_1} \ldots a_{i_n}$ in Σ^*, we write $\pi(w)$ for the unique path $[\mathbf{x}_0, \ldots, \mathbf{x}_n]$ in $\mathbb{N}^k$ such that $\mathbf{0} = \mathbf{x}_0 \to_{i_1} \mathbf{x}_1 \to_{i_2} \ldots \to_{i_n} \mathbf{x}_n$. Conversely, for each path $\pi = [\mathbf{x}_0, \ldots, \mathbf{x}_n]$ in $\mathbb{N}^k$, there is a unique word w such that $\pi(w) = \pi$. We denote this unique word w by $\|\pi\|$. For each $F \subseteq \mathbb{N}^k$, we denote by $\|F\|$ the set $\{\|\pi\| \mid \pi \sqsubseteq F\}$. Note that the language $\|F\|$ is prefix-closed.

Moreover, we define $\mathrm{fig}(L) = \Psi(\mathrm{pref}^{\uparrow}(L))$ for each $L \subseteq \Sigma^*$. The set $\mathrm{fig}(L) \subseteq \mathbb{N}^k$ is a *connex figure* in the sense of [11], *i.e.*, for each $\mathbf{x} \in \mathrm{fig}(L)$, there is a path π leading to $\mathbf{x}$ such that $\pi \sqsubseteq \mathrm{fig}(L)$.

Finally, the *geometrical closure* of L is the language $\gamma(L) = \|\mathrm{fig}(L)\|$. A class of languages $\mathcal{C}$ is said to be *geometrically closed* if $\gamma(L)$ belongs to $\mathcal{C}(\Sigma^*)$ whenever L does, for each non-empty finite alphabet Σ.

Note that the class of all regular languages is *not* geometrically closed, as observed in [11]. For instance, the language $L = a^*(ab)^*$ is regular, while its geometrical closure

$$\gamma(L) = \{w \in \{a,b\}^* \mid \forall u \leq w : |u|_a \geq |u|_b\}$$

is the prefix closure of the Dyck language.

3. A Characterisation of the Geometrical Closure

We now characterise the operation of geometrical closure in terms of three simpler operations on languages: the prefix closure, the commutative closure, and the prefix reduction. This characterisation is a key to our later considerations.

Proposition 3.1. *If L is a language over Σ, then $\gamma(L) = \mathrm{pref}^{\downarrow}\left([\mathrm{pref}^{\uparrow}(L)]\right)$.*

Proof. By definition,

$$\gamma(L) = \|\mathrm{fig}(L)\| = \left\|\Psi(\mathrm{pref}^{\uparrow}(L))\right\|.$$

If $w \in \gamma(L)$, then there is a path $\pi = [\mathbf{x}_0, \ldots, \mathbf{x}_n] \sqsubseteq \Psi(\mathrm{pref}^{\uparrow}(L))$ such that $w = \|\pi\|$. For an arbitrary prefix u of w, we have $\pi(u) = [\mathbf{x}_0, \ldots, \mathbf{x}_m]$ for some $m \leq n$. It follows that $\Psi(u) = \mathbf{x}_m$ belongs to $\Psi(\mathrm{pref}^{\uparrow}(L))$. Hence $u \in [\mathrm{pref}^{\uparrow}(L)]$ and w belongs to $\mathrm{pref}^{\downarrow}([\mathrm{pref}^{\uparrow}(L)])$.

On the other hand, if w belongs to $\mathrm{pref}^{\downarrow}([\mathrm{pref}^{\uparrow}(L)])$, then all prefixes u of w belong to $[\mathrm{pref}^{\uparrow}(L)]$. Thus $\Psi(u)$ is in $\Psi(\mathrm{pref}^{\uparrow}(L))$ for each $u \leq w$, and $\pi(w)$ is a path in $\Psi(\mathrm{pref}^{\uparrow}(L))$. As a result, we observe that w is in $\|\Psi(\mathrm{pref}^{\uparrow}(L))\| = \gamma(L)$. □

As a direct consequence of Proposition 2.1 and Proposition 3.1, we obtain the following sufficient condition, under which a positive variety of languages is geometrically closed.

Corollary 3.2. *Each positive variety of regular languages closed under prefix reduction and commutation is geometrically closed.*

4. The Basic Negative Result and Positive Varieties with Regular Geometrical Closures

Let us now prove a simple sufficient condition under which a positive variety contains a language with irregular geometrical closure, and thus is not geometrically closed. The following lemma is used repeatedly in this article, as a tool for obtaining negative results.

Lemma 4.1. *Let $\mathcal{V}$ be a positive variety containing the language $(ab)^*$. Then $\mathcal{V}$ contains a language L such that $\gamma(L)$ is not regular. In particular, $\mathcal{V}$ is not geometrically closed.*

Proof. First, if $(ab)^* \in \mathcal{V}(\{a,b\}^*)$, then also $\{a,b\}^* \in \mathcal{V}(\{a,b,c\}^*)$, as $\{a,b\}^* = h_1^{-1}((ab)^*)$ for a homomorphism $h_1 \colon \{a,b,c\}^* \to \{a,b\}^*$ given by $h_1(a) = h_1(b) = ab$ and $h_1(c) = aa$.

Next, let us observe that we also have $\{a,b\}^*c \in \mathcal{V}(\{a,b,c\}^*)$ under the assumption of $\mathcal{V}$ containing $(ab)^*$. Indeed, closure of $\mathcal{V}$ under quotients implies that $(ab)^*b^{-1} = (ab)^*a \in \mathcal{V}(\{a,b\}^*)$. Let $h_2 \colon \{a,b,c\}^* \to \{a,b\}^*$ be a homomorphism such that $h_2(a) = h_2(b) = ab$ and $h_2(c) = a$. Clearly, $h_2^{-1}((ab)^*a) = \{a,b\}^*c$.

These two observations jointly imply that the positive variety $\mathcal{V}$ contains the language $(ab)^*c$, as

$$(ab)^*c = h_3^{-1}((ab)^*) \cap \{a,b\}^*c,$$

where the homomorphism $h_3 \colon \{a,b,c\}^* \to \{a,b\}^*$ is given by $h_3(a) = a$, $h_3(b) = b$, and $h_3(c) = \varepsilon$.

We may finally conclude that the positive variety $\mathcal{V}$ contains the language

$$L = \{a,b\}^* \cup (ab)^*c.$$

The geometrical closure of L is not regular, as

$$\gamma(L)c^{-1} = \{w \in \{a,b\}^* \mid |w|_a = |w|_b\}.$$

This finishes the proof of the lemma. □

Recall that $\mathcal{W}$ is by definition the largest positive variety not containing $(ab)^*$ [8, 9]. Moreover, it is known that $\mathcal{W}$ is closed under commutation [7]. Lemma 4.1 thus can be used to obtain the following full characterisation of positive varieties $\mathcal{V}$ such that $\gamma(L)$ is regular for all L in $\mathcal{V}$.

Theorem 4.2. *Let $\mathcal{V}$ be a positive variety languages. The following conditions are equivalent:*

(*i*) *The language $\gamma(L)$ is regular for every L in $\mathcal{V}$.*

(*ii*) *The positive variety $\mathcal{V}$ is contained in $\mathcal{W}$.*

(*iii*) *The positive variety $\mathcal{V}$ does not contain the language $(ab)^*$.*

Proof. The equivalence of (*ii*) and (*iii*) is merely a restatement of the fact that $\mathcal{W}$ is the largest positive variety that does not contain $(ab)^*$. Now, if $\mathcal{V} \subseteq \mathcal{W}$, then $[L]$ is regular for all L in $\mathcal{V}$, as $\mathcal{W}$ is closed under commutation. As the class of all regular languages is obviously closed both under $\mathrm{pref}^{\uparrow}$ and $\mathrm{pref}^{\downarrow}$, it follows by the characterisation of Proposition 3.1 that $\gamma(L)$ is regular for all L from $\mathcal{V}$. Conversely, if $\mathcal{V} \not\subseteq \mathcal{W}$, then $\mathcal{V}$ contains $(ab)^*$ and Lemma 4.1 implies that (*i*) does not hold. □

5. Geometrically Closed Classes of Star-Free Languages

We turn our attention to the problem of identifying robust geometrically closed classes of languages. Let us first focus on geometrically closed positive varieties of *star-free languages*. In Subsection 5.2, we prove our main result on star-free languages showing that many important (positive) varieties of star-free languages are geometrically closed. To formulate this result, we need a variety of languages $\mathcal{R}_{LT}$, which we introduce in Subsection 5.1, providing its language-theoretic, automata-theoretic, and algebraic description. We touch upon the case of classes not covered by our main result in Subsection 5.4.

5.1. LT-*Acyclic Automata and the Variety* $\mathcal{R}_{LT}$

We now introduce the class of languages $\mathcal{R}_{LT}$, which plays a central role in our main result. Despite being defined simply as a *class* of languages, $\mathcal{R}_{LT}$ is actually a variety, as we observe later below. The letters *LT* refer to a characteristic property of the corresponding class of automata, explained below as well.

Definition 5.1. We denote by $\mathcal{R}_{LT}$ the class of languages such that $\mathcal{R}_{LT}(\Sigma^*)$ consists, for every non-empty finite alphabet Σ, of all finite unions of languages of the form

$$L = \Sigma_0^* a_1 \Sigma_1^* a_2 \ldots a_n \Sigma_n^*, \tag{1}$$

where n is a natural number, $\Sigma_0 \subseteq \Sigma_1 \subseteq \ldots \subseteq \Sigma_n \subseteq \Sigma$, and $a_i \in \Sigma \setminus \Sigma_{i-1}$ for $i = 1, \ldots, n$.

This definition is similar to definitions of some other classes of languages that have already been studied in literature. First of all, if we omit the condition $\Sigma_0 \subseteq \Sigma_1 \subseteq \ldots \subseteq \Sigma_n$, we get a definition of languages from the variety $\mathcal{R}$ corresponding to R-trivial monoids, which we recall in more details later.[6] Let us point out here that $\mathcal{R}_{LT} \subseteq \mathcal{R}$. Secondly, if we also require $a_i \in \Sigma_i$ in Definition 5.1 for $i = 1, \ldots, n$, then we obtain a variety of languages considered by Pin, Straubing, and Thérien [20] and corresponding to a pseudovariety of finite monoids denoted $\mathbf{R}_1$. Finally, if we drop the condition $a_i \notin \Sigma_{i-1}$ in Definition 5.1 and then we generate a variety, we obtain the variety of languages corresponding to the pseudovariety **JMK** considered by Almeida [1, p. 236].

As our next aim is to characterise languages from $\mathcal{R}_{LT}$ in terms of automata, we recall the characterisation of languages from $\mathcal{R}$ first. An automaton $\mathcal{A} = (Q, \Sigma, \cdot, \iota, F)$ is *acyclic* if every cycle in $\mathcal{A}$ is a self-loop at some of its states. In other words, if $p \cdot w = p$ for some $p \in Q$ and $w \in \Sigma^*$, then also $p \cdot a = p$ for every letter a occurring in w. This condition is equivalent to a possibility of numbering the states in Q as $1, \ldots, |Q|$ in such a way that the state $p \cdot a$, with $p \in Q$ and $a \in \Sigma$, always has a number greater than or equal to the number of p. For this reason, these automata are called *extensive* in [18, p. 93]. It is known that they recognise precisely the languages from $\mathcal{R}$ [6].

We say that an acyclic automaton $\mathcal{A} = (Q, \Sigma, \cdot, \iota, F)$ has a *loop transfer* property, if $p \cdot a = p$ implies $(p \cdot b) \cdot a = p \cdot b$ for every $p \in Q$ and $a, b \in \Sigma$. We then call $\mathcal{A}$ an *LT-acyclic automaton* for short. This means that if there is an a-labelled loop at a state p of an LT-acyclic automaton, then there also is an a-labelled loop at each state reachable from p. We may thus equivalently take $b \in \Sigma^*$ in the definition of the loop-transfer property. Our first aim in what follows is to show that languages recognised by LT-acyclic automata are precisely those from $\mathcal{R}_{LT}$. We do so via a series of elementary lemmas.

Lemma 5.2. *The minimal automaton $\mathcal{D}_L$ for a language L of the form (1) is LT-acyclic.*

Proof. Let $L = \Sigma_0^* a_1 \Sigma_1^* a_2 \ldots a_n \Sigma_n^*$ for some natural number n, alphabets $\Sigma_0 \subseteq \Sigma_1 \subseteq \ldots \subseteq \Sigma_n \subseteq \Sigma$, and letters $a_i \in \Sigma \setminus \Sigma_{i-1}$ for $i = 1, \ldots, n$. Let us denote $\Gamma_{i-1} = \Sigma \setminus (\Sigma_{i-1} \cup \{a_i\})$ for $i = 1, \ldots, n$ and let us take $\Gamma_n = \Sigma \setminus \Sigma_n$. Then it is an easy exercise to show that the automaton in Fig. 1 is the minimal automaton for L and that it is an LT-acyclic automaton. □

[6] The notion of R-trivial monoids is based upon *Green's relations*, a fundamental concept of semigroup theory. Let us recall their definitions on a *monoid* M: for $x, y \in M$, one writes xLy if $Mx = My$, xRy if $xM = yM$, and xJy if $MxM = MyM$. Next $H = L \cap R$ and $D = L \circ R = R \circ L$ is the smallest relation containing both L and R. It is a standard fact that D and J coincide for *finite* monoids [15]. A monoid M is said to be X-*trivial* for $X \in \{L, R, J, H, D\}$ if X is an identity relation on M.

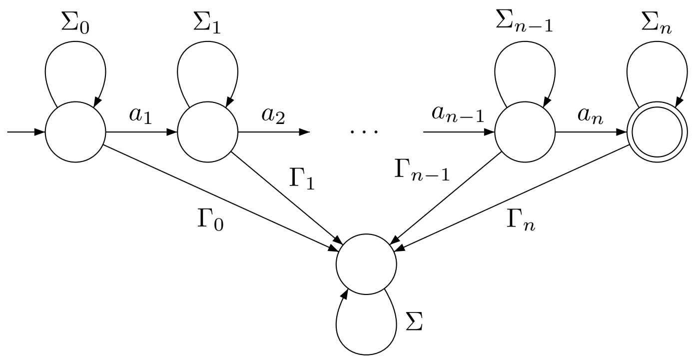


Figure 1: An LT-acyclic automaton for a language of the form (1).

Lemma 5.3. *Let L and K be languages over an alphabet Σ recognised by LT-acyclic automata. Then $L \cup K$ is recognised by an LT-acyclic automaton as well.*

Proof. The language $L \cup K$ can be recognised by the direct product of a pair of automata that recognise the languages L and K, respectively. A routine check shows that a finite direct product of LT-acyclic automata is an LT-acyclic automaton. □

The two lemmas above together show that every language from $\mathcal{R}_{LT}$ is recognised by an LT-acyclic automaton. The following lemma strengthens this observation by showing that the *minimal* automaton of a language from $\mathcal{R}_{LT}$ is LT-acyclic.

Lemma 5.4. *Let L be a language recognised by an LT-acyclic automaton. Then the minimal automaton $\mathcal{D}_L$ for the language L is LT-acyclic as well.*

Proof. Let $\mathcal{A} = (Q, \Sigma, \cdot, \iota, F)$ be an LT-acyclic automaton recognising L. The minimal automaton $\mathcal{D}_L$ is a homomorphic image of the subautomaton of $\mathcal{A}$ [21] consisting of all states reachable from ι.[7] It is clear that a subautomaton of an LT-acyclic automaton is LT-acyclic. Thus we may assume that all states of $\mathcal{A}$ are reachable from the initial state ι.

Let $\varphi\colon Q \to P$ be a surjective homomorphism from the automaton $\mathcal{A}$ onto some other automaton $\mathcal{B} = (P, \Sigma, \bullet, \varphi(\iota), \varphi(F))$. We claim that $\mathcal{B}$ is necessarily acyclic. To prove this claim, let $p \in P$ and $w \in \Sigma^*$ be such that $p \bullet w = p$. Then we choose some state q' from $\varphi^{-1}(p)$. For that q', we have $q' \cdot w^m \in \varphi^{-1}(p)$ for every natural number m. As the sequence $q', q' \cdot w, q' \cdot w^2, \ldots$ contains only finitely many states, there are natural numbers n and $m \geq 1$ such that

$$q' \cdot w^{n+m} = q' \cdot w^n = q.$$

As $\mathcal{A}$ is acyclic, we have $q \cdot a = q$ for every letter a occurring in w. Consequently,

$$p \bullet a = \varphi(q) \bullet a = \varphi(q \cdot a) = \varphi(q) = p$$

and $\mathcal{B}$ is acyclic.

Let finally $p \in P$ and $a \in \Sigma$ be such that $p \bullet a = p$. It follows from the previous paragraph that there is a state $q \in \varphi^{-1}(p)$ such that $q \cdot a = q$. As $\mathcal{A}$ is LT-acyclic, we see that $(q \cdot b) \cdot a = q \cdot b$ for every $b \in \Sigma$. Thus

$$p \bullet ba = \varphi(q \cdot ba) = \varphi(q \cdot b) = p \bullet b.$$

As a result, $\mathcal{B}$ is an LT-acyclic automaton. In particular, this property holds for $\mathcal{D}_L$. □

Let us now prove a converse to the statements established above.

[7] Here we understand a *homomorphism* from a (deterministic) automaton $\mathcal{A} = (Q, \Sigma, \cdot, \iota, F)$ to a (deterministic) automaton $\mathcal{B} = (P, \Sigma, \bullet, \nu, G)$ to be a mapping $\varphi\colon Q \to P$ such that $\varphi(\iota) = \nu$, $\varphi(p) \bullet a = \varphi(p \cdot a)$ for all $p \in Q$ and $a \in \Sigma$, and $\varphi(F) = G$. The automata $\mathcal{A}$ and $\mathcal{B}$ are equivalent under these assumptions.

Lemma 5.5. *If $\mathcal{A}$ is an LT-acyclic automaton over Σ, then it recognises a language from $\mathcal{R}_{LT}(\Sigma^*)$.*

Proof. Let $\mathcal{A} = (Q, \Sigma, \cdot, \iota, F)$ and R be the set of all valid runs in the automaton $\mathcal{A}$, which do not use loops:

$$R = \{(q_0, a_1, q_1, a_2, \dots, a_n, q_n) \mid n \in \mathbb{N};\ q_0, \dots, q_n \in Q;\ a_1, \dots, a_n \in \Sigma;\\ q_0 = \iota;\ q_n \in F;\ \forall j \in \{1, \dots, n\} \colon q_{j-1} \neq q_j \ \wedge\ q_{j-1} \cdot a_j = q_j\}.$$

We see that the set R is finite. Moreover, for each q in Q, let Σ_q denote the alphabet $\Sigma_q = \{c \in \Sigma \mid q \cdot c = q\}$, and let us denote by L the language recognised by $\mathcal{A}$. Then

$$L_w := \Sigma_{q_0}^* a_1 \Sigma_{q_1}^* a_2 \dots a_n \Sigma_{q_n}^* \subseteq L$$

is a language of the form (1) for each $w = (q_0, a_1, q_1, a_2, \dots, a_n, q_n)$ in R and

$$L = \bigcup_{w \in R} L_w.$$

Hence the language L recognised by $\mathcal{A}$ belongs to $\mathcal{R}_{LT}(\Sigma^*)$. □

The following theorem, which is a summary of the previous lemmas, thus gives an automata-theoretic characterisation of the class $\mathcal{R}_{LT}$.

Theorem 5.6. *The following statements are equivalent for a language $L \subseteq \Sigma^*$:*

- (i) *L belongs to $\mathcal{R}_{LT}(\Sigma^*)$.*
- (ii) *L is recognised by an LT-acyclic automaton.*
- (iii) *The minimal automaton of L is LT-acyclic.*

Proof. The statement (i) implies (ii) by Lemma 5.2 and Lemma 5.3. The statement (ii) implies (iii) by Lemma 5.4. Finally, (iii) implies (i) by Lemma 5.5. □

One may prove that $\mathcal{R}_{LT}$ is a variety of languages in several different ways. It is possible to prove directly that the class $\mathcal{R}_{LT}$ is closed under basic language operations. It is also possible to prove that the class of LT-acyclic automata forms a variety of actions in the sense of [10]. Here we take an approach, in which we identify the algebraic counterpart of the class $\mathcal{R}_{LT}$ – namely, we characterise the corresponding pseudovariety of finite monoids by pseudoidentities. We do not want to recall the notion of pseudoidentities in general. Let us only recall the implicit operation x^ω here. If we substitute for x some element s in a finite monoid M, then the image of x^ω is s^ω, which is the unique idempotent in the subsemigroup of M generated by s. It could be useful to know that, for a fixed finite monoid M, there always exists a natural number m such that $s^\omega = s^m$ for each $s \in M$.

Theorem 5.7. *Let Σ be an alphabet, $L \subseteq \Sigma^*$, and M_L the syntactic monoid of L. The following statements are equivalent:*

- (i) *L belongs to $\mathcal{R}_{LT}(\Sigma^*)$.*
- (ii) *M_L satisfies the pseudoidentity $x^\omega y x = x^\omega y$.*

Proof. Let $\mathcal{D}_L = (Q, \Sigma, \cdot, \iota, F)$ be the minimal automaton of the language L. Then M_L can be viewed as the transition monoid of $\mathcal{D}_L$ (see [19, p. 692]). The elements of M_L can thus be viewed as changes in states of $\mathcal{D}_L$ determined by words from Σ^*. More formally, for $u \in \Sigma^*$, we denote by f_u the mapping given by the rule $p \mapsto p \cdot u$ for each $p \in Q$. Let m be a natural number such that $s^\omega = s^m$ for each s in M_L.

Let us prove that (i) implies (ii). Suppose that L belongs to $\mathcal{R}_{LT}(\Sigma^*)$. Then $\mathcal{D}_L$ is an LT-acyclic automaton by Theorem 5.6. Let x, y be mapped to elements of M_L corresponding to words $v, w \in \Sigma^*$. To prove (ii), we need to check that $f_{v^m} f_w f_v = f_{v^m} f_w$ holds. As $\mathcal{D}_L$ is acyclic, we have $(p \cdot v^m) \cdot a = p \cdot v^m$

for every $p \in Q$ and $a \in \Sigma$ occurring in v. Moreover, $\mathcal{D}_L$ is an LT-acyclic automaton, which means that a loop labelled by a at a state $p \cdot v^m$ is transferred to every state reachable from $p \cdot v^m$. In particular, for every letter a occurring in v, there is a loop labelled by a at the state $(p \cdot v^m) \cdot w$. The equality $f_{v^m} f_w f_v = f_{v^m} f_w$ follows.

Now, let us show that (ii) implies (i). First of all, we deduce some useful consequences of the pseudoidentity $x^\omega yx = x^\omega y$. Let M be a finite monoid satisfying $x^\omega yx = x^\omega y$, and let m be such that $s^\omega = s^m$ for each $s \in M$. Thus, for each pair of elements $s, t \in M$, we have $s^m ts = s^m t$. If we take $x, y \in M$ and put $s = xy$ and $t = x$ in $s^m ts = s^m t$, we get $(xy)^m xxy = (xy)^m x$, which may be written as

$$(xy)^\omega xxy = (xy)^\omega x. \tag{2}$$

Since x, y in (2) are arbitrary elements of M, we get that (2) holds in M as a pseudoidentity. In this way one may deduce a pseudoidentity from another pseudoidentity, but it is usual to skip the details concerning substitutions used. Moreover, we often interpret variables x, y in pseudoidentities also as elements of M.

Using (2) from right to left repeatedly, we obtain

$$(xy)^\omega x = (xy)^\omega x(xy) = (xy)^\omega x(xy)^2 = \ldots = (xy)^\omega x(xy)^m = (xy)^\omega x(xy)^\omega. \tag{3}$$

By setting $y = 1$ in $x^\omega yx = x^\omega y$, we deduce the defining pseudoidentity $x^{\omega+1} = x^\omega$ of finite aperiodic monoids. Using this pseudoidentity and (3), we get

$$(xy)^\omega = (xy)^\omega xy = (xy)^\omega x(xy)^\omega y. \tag{4}$$

Next, we apply the pseudoidentity (4) on its right-hand side, where the second occurrence of $(xy)^\omega$ is replaced. Iterating this replacement $m-1$ times, we get

$$(xy)^\omega = (xy)^\omega x(xy)^\omega y = ((xy)^\omega x)^2 (xy)^\omega y^2 = \cdots = ((xy)^\omega x)^m (xy)^\omega y^m = ((xy)^\omega x)^\omega (xy)^\omega y^\omega. \tag{5}$$

Multiplying this pseudoidentity by y from the right, the right-hand side does not change, because $y^{\omega+1} = y^\omega$. The original and the new left-hand side thus form the pseudoidentity $(xy)^\omega = (xy)^\omega y$ valid in M. Combining this pseudoidentity with (4) and (3), we finally obtain

$$(xy)^\omega = (xy)^\omega x(xy)^\omega y = (xy)^\omega x(xy)^\omega = (xy)^\omega x.$$

We proved that the defining pseudoidentity $(xy)^\omega x = (xy)^\omega$ of finite R-trivial monoids is a consequence of the pseudoidentity from (ii).

Finally, in order to prove that (ii) implies (i), suppose that the monoid M_L satisfies the pseudoidentity $x^\omega yx = x^\omega y$. Hence, the monoid M_L is R-trivial and the minimal automaton $\mathcal{D}_L$ of the language L is acyclic. Moreover, let $p \in Q$ and $a \in \Sigma$ be such that $p \cdot a = p$, and take arbitrary $b \in \Sigma$. Then $f_a^\omega f_b$ in M_L maps p to $p \cdot b$. Similarly, $f_a^\omega f_b f_a$ in M_L maps p to $p \cdot ba$. However, taking $x \mapsto f_a$, $y \mapsto f_b$ in $x^\omega yx = x^\omega y$ gives us $f_a^\omega f_b f_a = f_a^\omega f_b$. Therefore, $p \cdot ba = p \cdot b$ and we see that there is a loop labelled by a at the state $p \cdot b$. As a result, $\mathcal{D}_L$ is an LT-acyclic automaton and L belongs to $\mathcal{R}_{LT}(\Sigma^*)$ by Theorem 5.6. □

Corollary 5.8. *The class $\mathcal{R}_{LT}$ is a variety of languages corresponding to the pseudovariety of finite monoids*

$$[\![x^\omega yx = x^\omega y]\!].$$

Let us also note that the pseudoidentity $x^\omega yx = x^\omega y$ is known to describe the pseudovariety of finite monoids **MK**. By definition, this is a pseudovariety generated by monoids of the form S^1, where S is a semigroup from **K**, a pseudovariety of finite semigroups, in which all idempotents act as left zeros. The characterisation via the pseudoidentity $x^\omega yx = x^\omega y$ is due to Pin [17]; cf. also Almeida [1, p. 212]. Therefore, $\mathcal{R}_{LT}$ consists of precisely all **MK**-recognisable languages. This observation strengthens the one from the conference version of this article by taking into account the fact that **MK** is contained in **R**, as observed by Pin [17] (in an article that we have discovered after the conference version was published). The inclusion $\mathbf{MK} \subseteq \mathbf{R}$ is also established in the proof of Theorem 5.7.

5.2. The Main Results for Star-Free Languages

Let us now return to the operation of geometrical closure on languages and prove our main result for star-free languages (Theorem 5.13): *Each class of languages lying between the variety of languages* $\mathcal{R}_{LT}$ *and the positive variety* $\mathcal{W} \cap \mathcal{SF}$ *is geometrically closed.* As $\mathcal{V}_{3/2}$ lies in between these two classes, our result strengthens the one from [11], in which geometrical closure of *binary* languages from $\mathcal{V}_{3/2}$ is proved. Moreover, we show here that $\mathcal{W} \cap \mathcal{SF}$ is the largest geometrically closed *positive variety of star-free languages*, while the largest geometrically closed *variety of star-free languages* corresponds to $\mathbf{DA}$.

The route that we take to arrive at Theorem 5.13 consists of three steps:

1. We recall that the class $\mathcal{W} \cap \mathcal{SF}$ is closed under commutation [7]. In particular, a commutative closure of each language from $\mathcal{W} \cap \mathcal{SF}$ is a commutative star-free language.
2. We observe that each commutative star-free language belongs to $\mathcal{R}_{LT}$.
3. We prove that the variety $\mathcal{R}_{LT}$ is closed under prefix reduction.

These three observations imply that the geometrical closure of a language from $\mathcal{W} \cap \mathcal{SF}$ belongs to $\mathcal{R}_{LT}$, from which our main result follows easily.

Lemma 5.9. *The positive variety* $\mathcal{W} \cap \mathcal{SF}$ *is closed under commutation.*

Proof. Follows directly by a more general result of Cano, Guaiana, and Pin [7, Theorem 6.2], by which $[L]$ belongs to $\mathcal{W}(\Sigma^*)$ for an alphabet Σ whenever $L \in \mathcal{W}(\Sigma^*)$, while the period of $[L]$ divides the period of L. Here, the period of a language K is the smallest positive integer p such that the syntactic monoid M_K satisfies the pseudoidentity $x^{\omega+p} = x^{\omega}$. Hence, star-free languages are precisely languages of period 1, and the result of [7] mentioned above implies that the commutative closure of a star-free language from $\mathcal{W}$ is again a star-free language from $\mathcal{W}$. □

In particular, it follows by Lemma 5.9 that the commutative closure of a language from the positive variety $\mathcal{W} \cap \mathcal{SF}$ is always star-free, while it is trivially commutative. This observation slightly strengthens the one from the conference version of this article, where we have noticed a similar property of $\mathcal{V}_{3/2}$, which is contained in $\mathcal{W} \cap \mathcal{SF}$. We return to this result for $\mathcal{V}_{3/2}$, which can be obtained using tools more elementary in nature than the theorem of [7], in Remark 5.18 below.

We now prove that each commutative star-free language belongs to $\mathcal{R}_{LT}$. The proof given here is algebraic in nature and relies mostly on manipulating pseudoidentities.

Lemma 5.10. *Every commutative star-free language is in* $\mathcal{R}_{LT}$.

Proof. First observe that if L is a commutative language, *i.e.*, if $L = [L]$, then the syntactic monoid M_L of L is commutative as well – that is, $M_L \in \mathbf{Com}$. This is a well-known fact, which can be established, for instance, via the minimal automaton $\mathcal{D}_L = (Q, \Sigma, \cdot, \iota, F)$ of L. Given $q \in Q$ and $u, v \in \Sigma^*$, we see that $q \cdot uv = q \cdot vu$, as commutativity of L implies that for each $w \in \Sigma^*$ one has $wuv \in L$ if and only $wvu \in L$. This implies that M_L, which is the transition monoid of $\mathcal{D}_L$, is commutative.

If L is commutative and star-free, then M_L is commutative and aperiodic. The claim can then be easily proved algebraically, as the pseudovariety of commutative monoids satisfies $\mathbf{Com} = [\![xy = yx]\!]$, the pseudovariety of aperiodic monoids is given by $\mathbf{A} = [\![x^{\omega+1} = x^{\omega}]\!]$, and we have seen in Corollary 5.8 that $\mathcal{R}_{LT}$ corresponds to $\mathbf{MK} = [\![x^{\omega}yx = x^{\omega}y]\!]$. The pseudoidentity $x^{\omega}yx = x^{\omega}y$ follows from $xy = yx$ and $x^{\omega+1} = x^{\omega}$ – if a monoid M satisfies the latter two pseudoidentities, then we see for all $x, y \in M$ that

$$x^{\omega}yx = x^{\omega}xy = x^{\omega+1}y = x^{\omega}y.$$

Thus $M_L \in \mathbf{A} \cap \mathbf{Com}$ implies $M_L \in \mathbf{MK}$, and hence $L \in \mathcal{R}_{LT}(\Sigma^*)$. □

Finally, let us observe that the variety $\mathcal{R}_{LT}$ is closed under prefix reduction. We note that by an *absorbing state* we mean a state p satisfying $p \cdot a = p$ for every $a \in \Sigma$.

Lemma 5.11. *Let $L \in \mathcal{R}_{LT}(\Sigma^*)$ for some alphabet Σ. Then $\mathrm{pref}^{\downarrow}(L) \in \mathcal{R}_{LT}(\Sigma^*)$ as well.*

Proof. Let L be in $\mathcal{R}_{LT}(\Sigma^*)$, which means that this language is recognised by some LT-acyclic automaton $\mathcal{A} = (Q, \Sigma, \cdot, \iota, F)$. If $\iota \notin F$, then L does not contain the empty word, and consequently $\mathrm{pref}^{\downarrow}(L) = \emptyset$, which belongs to $\mathcal{R}_{LT}(\Sigma^*)$. We may thus assume that $\iota \in F$.

Now, roughly stated, we claim that the language $\mathrm{pref}^{\downarrow}(L)$ is recognised by the automaton $\mathcal{A}'$ constructed from $\mathcal{A}$ by replacing all non-final states with a single absorbing non-final state τ. More precisely, we construct an automaton $\mathcal{A}' = (F \cup \{\tau\}, \Sigma, \bullet, \iota, F)$, where τ is a new state, for which we define $\tau \bullet a = \tau$ for each $a \in \Sigma$. Furthermore, for each $p \in F$ and $a \in \Sigma$, we define $p \bullet a = p \cdot a$ if $p \cdot a \in F$, and $p \bullet a = \tau$ otherwise. As $\mathcal{A}$ contains no cycles other than loops, the constructed automaton $\mathcal{A}'$ has the same property. Moreover, any state of $\mathcal{A}'$ reachable in $\mathcal{A}'$ from some p in $F \cup \{\tau\}$ is either reachable from p in $\mathcal{A}$, or it is equal to τ. As $\tau \bullet c = \tau$ for each c in Σ, this implies that $\mathcal{A}'$ is an LT-acyclic automaton and the language $\mathrm{pref}^{\downarrow}(L)$ belongs to $\mathcal{R}_{LT}(\Sigma^*)$ by Theorem 5.6. □

As a consequence of the three lemmas above, we obtain the following theorem, from which our main result follows easily.

Theorem 5.12. *Let Σ be an alphabet and $L \in (\mathcal{W} \cap \mathcal{SF})(\Sigma^*)$. Then $\gamma(L) \in \mathcal{R}_{LT}(\Sigma^*)$.*

Proof. It follows by Proposition 3.1 that the geometrical closure can be written as $\gamma(L) = \mathrm{pref}^{\downarrow}\left(\left[\mathrm{pref}^{\uparrow}(L)\right]\right)$. As $\mathcal{W} \cap \mathcal{SF}$ is a positive variety of languages, Proposition 2.1 tells us that the language $\mathrm{pref}^{\uparrow}(L)$ belongs to $(\mathcal{W} \cap \mathcal{SF})(\Sigma^*)$ whenever L does. The language $\left[\mathrm{pref}^{\uparrow}(L)\right]$ is thus a commutative star-free language by Lemma 5.9 (which is a restatement of a result from [7]). Thus, $\left[\mathrm{pref}^{\uparrow}(L)\right] \in \mathcal{R}_{LT}(\Sigma^*)$ by Lemma 5.10 and $\gamma(L) = \mathrm{pref}^{\downarrow}\left(\left[\mathrm{pref}^{\uparrow}(L)\right]\right) \in \mathcal{R}_{LT}(\Sigma^*)$ by Lemma 5.11. □

We are now prepared to state the main result of this section, which is a direct consequence of the theorem above.

Theorem 5.13. *Let $\mathcal{C}$ be a class of languages containing $\mathcal{R}_{LT}$ and at the same time contained in $\mathcal{W} \cap \mathcal{SF}$. Then $\mathcal{C}$ is geometrically closed.*

There are many important varieties and positive varieties of languages studied in the literature, for which this result can be applied.

Corollary 5.14. *The following classes are geometrically closed:*

1. *The positive variety $\mathcal{W} \cap \mathcal{SF}$;*
2. *The positive variety $\mathcal{V}_{3/2}$;*
3. *The variety of all* **DA**-*recognisable languages;*
4. *The variety $\mathcal{R}$;*
5. *The variety of all* **JMK**-*recognisable languages;*
6. *The variety $\mathcal{R}_{LT}$.*

Remark 5.15. Note that the six classes from Corollary 5.14 form a strictly descending chain in the order as listed there. Most of these inclusions are widely known (or can be found in [1]). The only point that requires some attention is the fact that the inclusion between $\mathcal{V}_{3/2}$ and $\mathcal{W} \cap \mathcal{SF}$ is strict: a language from $\mathcal{W} \cap \mathcal{SF}$ that is not in $\mathcal{V}_{3/2}$ is recognised, e.g., by the automaton in Fig. 2. To prove that this language is not in $\mathcal{V}_{3/2}$, one can use an algebraic characterisation via the system of inequalities $x^{\omega} y x^{\omega} \geq x^{\omega}$ for all x, y interpreted as words such that all letters of y do also occur in x [19, p. 725].[8] After using standard techniques to prove that the language recognised by the automaton is star-free, one can prove that it is in $\mathcal{W}$ using the algebraic characterisation [9], by which this positive variety corresponds to a pseudovariety of ordered monoids M satisfying the following property: for any two mutually inverse elements s, t of M and any element z of the minimal ideal of the submonoid of M generated by s, t, one has $(stzst)^{\omega} \geq st$.

[8]Note that these inequalities are actually reversed in [19], as syntactic order is understood there dually to our definition.

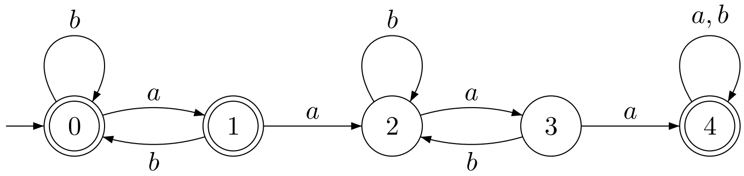


Figure 2: An automaton recognising a language from the difference of $\mathcal{W} \cap \mathcal{SF}$ and $\mathcal{V}_{3/2}$.

The variety of all **DA**-recognisable languages, mentioned in the previous corollary, coincides with the intersection of $\mathcal{V}_{3/2}$ and its dual. This class has a natural interpretation in terms of logical descriptions of levels in the Straubing-Thérien hierarchy (see Section 5 of [22]).

Using Theorem 5.13 and Lemma 4.1, it is easy to identify the largest geometrically closed positive variety and variety of *star-free* languages.

Theorem 5.16. *The largest geometrically closed positive variety of star-free languages is* $\mathcal{W} \cap \mathcal{SF}$.

Proof. The positive variety $\mathcal{W} \cap \mathcal{SF}$ is geometrically closed by Corollary 5.14. Suppose for contradiction that there is a geometrically closed positive variety of star-free languages $\mathcal{V}$ not contained in $\mathcal{W} \cap \mathcal{SF}$. Such positive variety $\mathcal{V}$ is necessarily not contained in $\mathcal{W}$. As $\mathcal{W}$ is the largest positive variety of languages not containing $(ab)^*$, it follows that $\mathcal{V}$ contains $(ab)^*$. Thus $\mathcal{V}$ is not geometrically closed by Lemma 4.1, which contradicts our original assumption. □

Theorem 5.17. *The largest geometrically closed variety of star-free languages consists of precisely all* **DA***-recognisable languages.*

Proof. It is well known that the largest variety of languages not containing the language $(ab)^*$ corresponds to the pseudovariety of finite monoids **DS** [1, 9]. The pseudovariety **DA** is an intersection of **DS** with the pseudovariety **A** of all aperiodic monoids. Hence, it corresponds to the largest variety of *star-free* languages not containing $(ab)^*$. Now, the variety of **DA**-recognisable languages is geometrically closed by Corollary 5.14. On the other hand, any variety of star-free languages not contained in it necessarily contains the language $(ab)^*$, and is not geometrically closed by Lemma 4.1. □

Remark 5.18. We have seen above that the commutative closure $[L]$ of a language L from $\mathcal{W} \cap \mathcal{SF}$ is always a commutative star-free language. This is actually equivalent to saying that $[L]$ is commutative and piecewise testable: commutativity of the syntactic monoid $M_{[L]}$ implies that its H-classes coincide with J-classes and aperiodicity is equivalent to H-triviality [15]. The monoid $M_{[L]}$ is J-trivial as a result, and $[L]$ is a commutative piecewise testable language.

Instead of $\mathcal{W} \cap \mathcal{SF}$, we have worked with its subclass $\mathcal{V}_{3/2}$ in the conference version of this article, for which the above mentioned result can be obtained in a more elementary way. One possibility would be to invoke a known result on partial commutations of Guaiana, Restivo, and Salemi [14], or of Bouajjani, Muscholl, and Touili [5]. However, it can also be seen from a simple direct language-theoretic construction that the commutative closure of a language from $\mathcal{V}_{3/2}$ is piecewise testable (while it is trivially commutative). We now briefly describe this construction, which we have used to obtain our main results in the conference version of this article.

Recall the result of Arfi [2], according to which a language belongs to $\mathcal{V}_{3/2}(\Sigma^*)$ if and only if it is given by a finite union of languages $\Sigma_0^* a_1 \Sigma_1^* a_2 \dots a_n \Sigma_n^*$, where $a_1, \dots, a_n$ are letters from Σ and $\Sigma_0, \dots, \Sigma_n$ are subalphabets of Σ. It suffices to describe the construction for languages of the form $L = \Sigma_0^* a_1 \Sigma_1^* a_2 \dots a_n \Sigma_n^*$, as for finite unions of languages $L_1, \dots, L_m \subseteq \Sigma^*$ we have

$$\left[\bigcup_{i=1}^{m} L_i\right] = \bigcup_{i=1}^{m} [L_i].$$

For L of this form, let $\Sigma' = \Sigma_0 \cup \dots \cup \Sigma_n$ and $x = a_1 \dots a_n$. Then it is not hard to see that

$$[L] = \{w \in \Sigma^* \mid \forall a \in \Sigma' : |w|_a \geq |x|_a;\ \forall b \in \Sigma \setminus \Sigma' : |w|_b = |x|_b\}.$$

This language is piecewise testable, as can be seen from its alternative representation

$$[L] = \bigcap_{a \in \Sigma'} (\Sigma^* a)^{|x|_a} \Sigma^* \cap \bigcap_{b \in \Sigma \setminus \Sigma'} \left((\Sigma^* b)^{|x|_b} \Sigma^* \cap \left((\Sigma^* b)^{|x|_b + 1} \Sigma^* \right)^C \right). \tag{6}$$

This construction gives rise to an alternative proof of Lemma 5.10, which does not make use of pseudoidentities: by what we have noted above, it is sufficient to show that every commutative piecewise testable language L belongs to the variety $\mathcal{R}_{LT}$. As $\mathcal{V}_1$ is contained in $\mathcal{V}_{3/2}$ and as commutativity of L implies $L = [L]$, it follows that L is a finite union of languages of the form (6). For each $a \in \Sigma$ and $k \in \mathbb{N}$, we may write $(\Sigma^* a)^k \Sigma^* = ((\Sigma \setminus \{a\})^* a)^k \Sigma^*$. This shows that the language $(\Sigma^* a)^k \Sigma^*$ belongs to $\mathcal{R}_{LT}$. As $\mathcal{R}_{LT}$ is a variety, it is closed under finite intersection and complementation, which implies that each language of the form (6) is in $\mathcal{R}_{LT}$. Finally, $\mathcal{R}_{LT}$ being a variety also implies its closure under finite union, by which it follows that L is in $\mathcal{R}_{LT}$ as well.

5.3. No Positive Variety Smaller than $\mathcal{R}_{LT}$ Works

Theorem 5.16 tells us that $\mathcal{W} \cap \mathcal{SF}$ is "optimal" in the statement of both Theorem 5.12 and Theorem 5.13, as no larger positive variety of star-free languages can be geometrically closed, which also means that no larger positive variety contained in $\mathcal{SF}$ can exhibit the property of $\mathcal{W} \cap \mathcal{SF}$ from Theorem 5.12. Perhaps it is not surprising that the variety $\mathcal{R}_{LT}$ is *not* the smallest geometrically closed positive variety of star-free languages – in fact, it is easy to see that the *trivial variety* $\mathcal{I}$, given for all non-empty finite Σ by $\mathcal{I}(\Sigma^*) = \{\emptyset, \Sigma^*\}$, *is* geometrically closed. However, also $\mathcal{R}_{LT}$ can be seen as "optimal" for Theorem 5.12: we now prove that the variety $\mathcal{R}_{LT}$ is the *smallest* positive variety of languages $\mathcal{V}$ such that the geometrical closure of every language from $\mathcal{W} \cap \mathcal{SF}$ falls into $\mathcal{V}$. This is equivalent to saying that $\mathcal{R}_{LT}$ is the positive variety generated by the languages $\gamma(L)$ for $L \in \mathcal{W} \cap \mathcal{SF}$. In fact, we prove a stronger statement: the variety $\mathcal{R}_{LT}$ is the positive variety generated by $\gamma(L)$ for $L \in \mathcal{R}_{LT}$. However, we first need the following simple lemma showing that the geometrical closure boils down to the prefix closure for languages of the form (1).

Lemma 5.19. *Let L be of the form*

$$L = \Sigma_0^* a_1 \Sigma_1^* a_2 \ldots a_n \Sigma_n^*,$$

for some natural number n, subsets $\Sigma_0 \subseteq \Sigma_1 \subseteq \ldots \subseteq \Sigma_n$ of some non-empty finite alphabet Σ, and letters $a_i \in \Sigma \setminus \Sigma_{i-1}$ for $i = 1, \ldots, n$. Then $\gamma(L) = \mathrm{pref}^{\uparrow}(L)$.

Proof. The inclusion $\gamma(L) \supseteq \mathrm{pref}^{\uparrow}(L)$ holds trivially for all languages L. It thus remains to prove the opposite inclusion $\gamma(L) \subseteq \mathrm{pref}^{\uparrow}(L)$.

For L from the statement of the lemma, clearly

$$\mathrm{pref}^{\uparrow}(L) = \bigcup_{i=0}^{n} L_i,$$

where L_i is given by $L_i = \Sigma_0^* a_1 \Sigma_1^* a_2 \ldots a_i \Sigma_i^*$ for $i = 0, \ldots, n$. Then, similarly as in Remark 5.18,

$$\left[\mathrm{pref}^{\uparrow}(L)\right] = \left[\bigcup_{i=0}^{n} L_i\right] = \bigcup_{i=0}^{n} [L_i], \tag{7}$$

where

$$[L_i] = \{ w \in \Sigma^* \mid \forall a \in \Sigma_i : |w|_a \geq |a_1 \ldots a_i|_a;\ \forall b \in \Sigma \setminus \Sigma_i : |w|_b = |a_1 \ldots a_i|_b \}.$$

The commutative languages $[L_0], \ldots, [L_n]$ are pairwise disjoint: given $[L_i]$ and $[L_j]$ with $i < j$, we clearly have $a_{i+1} \notin \Sigma_i$, which in turn implies that $|u|_{a_{i+1}} = |a_1 \ldots a_i|_{a_{i+1}}$ for all $u \in [L_i]$. On the other hand, each $v \in [L_j]$ has to satisfy $|v|_{a_{i+1}} \geq |a_1 \ldots a_j|_{a_{i+1}} \geq |a_1 \ldots a_{i+1}|_{a_{i+1}} > |a_1 \ldots a_i|_{a_{i+1}}$.

Let us now suppose for contradiction that $\gamma(L) \not\subseteq \mathrm{pref}^{\uparrow}(L)$. Then there exists at least one word $w \in \gamma(L) = \mathrm{pref}^{\downarrow}([\mathrm{pref}^{\uparrow}(L)])$ that does not belong to $\mathrm{pref}^{\uparrow}(L)$. Let us assume that this w is of minimal possible length. Existence of this w guarantees nonemptiness of L, from which it follows that $\varepsilon \in \mathrm{pref}^{\uparrow}(L)$. As a result, $w = xc$ for some $x \in \mathrm{pref}^{\uparrow}(L)$ and $c \in \Sigma$.

From $x \in \mathrm{pref}^{\uparrow}(L)$ it follows that $x \in L_k$ for some $k \in \{0, \ldots, n\}$, which is determined uniquely, since pairwise disjointness of $[L_0], \ldots, [L_n]$ implies pairwise disjointness of $L_0, \ldots, L_n$. As $w = xc \notin \mathrm{pref}^{\uparrow}(L)$, we have $c \notin \Sigma_k$ and if $k < n$, then also $c \neq a_{k+1}$. Similarly, $w \in \gamma(L) = \mathrm{pref}^{\downarrow}([\mathrm{pref}^{\uparrow}(L)]) \subseteq [\mathrm{pref}^{\uparrow}(L)]$ implies that there is a unique $m \in \{0, \ldots, n\}$ such that $w \in [L_m]$. We claim that all these assumptions cannot be satisfied at the same moment. We show this by distinguishing following three cases, all of which lead to a contradiction:

1. If $m < k$, then $a_k \notin \Sigma_m$ and $|w|_{a_k} \geq |x|_{a_k} \geq |a_1 \ldots a_k|_{a_k} > |a_1 \ldots a_m|_{a_k} = |w|_{a_k}$ – a contradiction.
2. If $m = k$, then $c \notin \Sigma_k$ and $|w|_c = |xc|_c > |x|_c = |a_1 \ldots a_k|_c$ – a contradiction with $w \in [L_k]$.
3. If $m > k$, then $a_{k+1} \notin \Sigma_k$ and $a_{k+1} \neq c$ give $|w|_{a_{k+1}} \geq |a_1 \ldots a_m|_{a_{k+1}} > |a_1 \ldots a_k|_{a_{k+1}} = |x|_{a_{k+1}} = |xc|_{a_{k+1}} = |w|_{a_{k+1}}$ – a contradiction again. □

Theorem 5.20. *The variety $\mathcal{R}_{LT}$ is generated, as a* positive *variety of languages, by the class $\gamma(\mathcal{R}_{LT})$ defined for every non-empty finite alphabet Σ by $\gamma(\mathcal{R}_{LT})(\Sigma^*) = \{\gamma(L) \mid L \in \mathcal{R}_{LT}(\Sigma^*)\}$.*

Proof. It follows by Corollary 5.14 that $\gamma(\mathcal{R}_{LT})(\Sigma^*) \subseteq \mathcal{R}_{LT}(\Sigma^*)$. It thus remains to prove that every positive variety $\mathcal{V}$ containing the class $\gamma(\mathcal{R}_{LT})$ also contains all languages from $\mathcal{R}_{LT}$.

Let us first show, for each non-empty finite alphabet Σ, that $\mathcal{V}(\Sigma^*)$ contains all languages of the form (1), *i.e.*, all languages

$$L = \Sigma_0^* a_1 \Sigma_1^* a_2 \ldots a_n \Sigma_n^*,$$

where n is a natural number, $\Sigma_0 \subseteq \Sigma_1 \subseteq \ldots \subseteq \Sigma_n \subseteq \Sigma$, and $a_i \in \Sigma \setminus \Sigma_{i-1}$ for $i = 1, \ldots, n$. For any fixed L given like this, let us define the language L' by

$$L' = \Sigma_0^* a_1 \Sigma_1^* a_2 \ldots a_n \Sigma_n^* c \Sigma^*,$$

where $c \notin \Sigma$ is a new symbol. It is obvious that $L' \in \mathcal{R}_{LT}((\Sigma \cup \{c\})^*)$. In fact, L' is of the form (1) in case $\Sigma \cup \{c\}$ is used as an alphabet instead of Σ. It thus follows by Lemma 5.19 that

$$\gamma(L') = \mathrm{pref}^{\uparrow}(L').$$

As a consequence, $\mathrm{pref}^{\uparrow}(L') \in \gamma(\mathcal{R}_{LT})((\Sigma \cup \{c\})^*) \subseteq \mathcal{V}((\Sigma \cup \{c\})^*)$. But $\mathcal{V}$ is a positive variety and thus $\mathcal{V}((\Sigma \cup \{c\})^*)$ also contains the language

$$\mathrm{pref}^{\uparrow}(L')c^{-1} = L.$$

Since L is a language over Σ and $\mathcal{V}$ is a positive variety, it follows that L (which is a preimage of itself under the embedding of Σ^* into $(\Sigma \cup \{c\})^*$) also belongs to $\mathcal{V}(\Sigma^*)$. The set $\mathcal{V}(\Sigma^*)$ thus indeed contains all languages of the form (1). Finally, $\mathcal{V}$ being a positive variety implies that it is closed under finite unions, and it follows that $\mathcal{V}(\Sigma^*)$ contains all languages from $\mathcal{R}_{LT}(\Sigma^*)$. □

5.4. Classes of Star-Free Languages Not Containing $\mathcal{R}_{LT}$

It follows from what has been said above that a positive variety of star-free languages containing $\mathcal{R}_{LT}$ is geometrically closed if and only if it is contained in $\mathcal{W} \cap \mathcal{SF}$. Although we leave a similar characterisation of geometrically closed positive varieties of star-free languages *not* containing $\mathcal{R}_{LT}$ open, we now resolve the question of geometrical closure for several particular classes of this kind. We have already made an easy observation that the trivial variety $\mathcal{I}$, given for each non-empty finite alphabet Σ by $\mathcal{I}(\Sigma^*) = \{\emptyset, \Sigma^*\}$, is geometrically closed. Let us now prove that the variety $\mathcal{V}_1$ of all piecewise testable languages, which is incomparable with $\mathcal{R}_{LT}$,[9] is *not* geometrically closed.

[9] For instance, it is easy to prove that for $\Sigma = \{a, b\}$ one has $a\Sigma^* \in \mathcal{R}_{LT}(\Sigma^*) \setminus \mathcal{V}_1(\Sigma^*)$ and $ab^*a \in \mathcal{V}_1(\Sigma^*) \setminus \mathcal{R}_{LT}(\Sigma^*)$.

Proposition 5.21. *The variety $\mathcal{V}_1$ of all piecewise testable languages is not geometrically closed.*

Proof. Consider the language $L = \{\varepsilon, a, b\} \cup \{aa, bb\}a^*b^*$ over $\Sigma = \{a, b\}$. A routine verification shows that this language is piecewise testable. Moreover, we have $\mathrm{pref}^{\uparrow}(L) = L$, $[\mathrm{pref}^{\uparrow}(L)] = [L] = \Sigma^* \setminus \{ab, ba\}$, and

$$\gamma(L) = \mathrm{pref}^{\downarrow}\left([\mathrm{pref}^{\uparrow}(L)]\right) = \mathrm{pref}^{\downarrow}\left(\Sigma^* \setminus \{ab, ba\}\right) = \{\varepsilon, a, b\} \cup \{aa, bb\}\Sigma^*.$$

This language is easily shown not to be in $\mathcal{V}_1$. □

The following simple lemma goes beyond the scope of star-free languages. Nevertheless, let us state it already at this point, as it can also be used to construct examples of geometrically closed positive varieties of star-free languages.

Lemma 5.22. *Every positive variety $\mathcal{V}$ of commutative languages is geometrically closed.*

Proof. Let $L \in \mathcal{V}(\Sigma^*)$ for some non-empty finite alphabet Σ. By Proposition 2.1, $\mathrm{pref}^{\uparrow}(L) \in \mathcal{V}(\Sigma^*)$ as well. The language $\mathrm{pref}^{\uparrow}(L)$ is clearly commutative if L is, so we have $[\mathrm{pref}^{\uparrow}(L)] = \mathrm{pref}^{\uparrow}(L)$. Finally, this language is prefix-closed and thus not changed by prefix reduction:

$$\gamma(L) = \mathrm{pref}^{\downarrow}([\mathrm{pref}^{\uparrow}(L)]) = \mathrm{pref}^{\uparrow}(L) \in \mathcal{V}(\Sigma^*).$$

The positive variety $\mathcal{V}$ is geometrically closed. □

In particular, it follows by Lemma 5.22 that the variety of languages corresponding to the pseudovariety of semilattices $\mathbf{Sl} = [\![xy = yx, x^2 = x]\!]$, the smallest non-trivial variety of star-free languages [19], is geometricals closed. Another example is the variety of languages corresponding to $\mathbf{A} \cap \mathbf{Com} = \mathbf{J} \cap \mathbf{Com}$.

6. Geometrically Closed Classes Beyond Star-Free Languages

Let us now consider positive varieties of languages that are not necessarily star-free. It turns out that interesting geometrically closed classes are much rarer here than in the star-free case. This is demonstrated mainly by the following theorem, which is our main negative result on other than star-free languages. Recall that **DAb** is the pseudovariety of all finite monoids M such that each regular D-class of M is a commutative group; note that M itself does not have to be commutative.

Theorem 6.1. *There is no geometrically closed positive variety of languages containing the variety of all* **DAb***-recognisable languages.*

Proof. Let $\mathcal{V}$ be a geometrically closed positive variety containing the variety corresponding to **DAb**. We present a **DAb**-recognisable language L such that $\gamma(L)$ is not in $\mathcal{W}$. Since $\mathcal{V}$ is geometrically closed, this means that $\mathcal{V}$ is not contained in $\mathcal{W}$ and hence $\mathcal{V}$ contains some non-regular language by Theorem 4.2, which is a contradiction. It is thus enough to give an example of a **DAb**-recognisable language L with the promised property.

Let us take

$$L = \{a, b\}^* \cup \{wc \mid w \in \{a, b\}^*;\ |w|_a - |w|_b \not\equiv 2 \pmod 3\}.$$

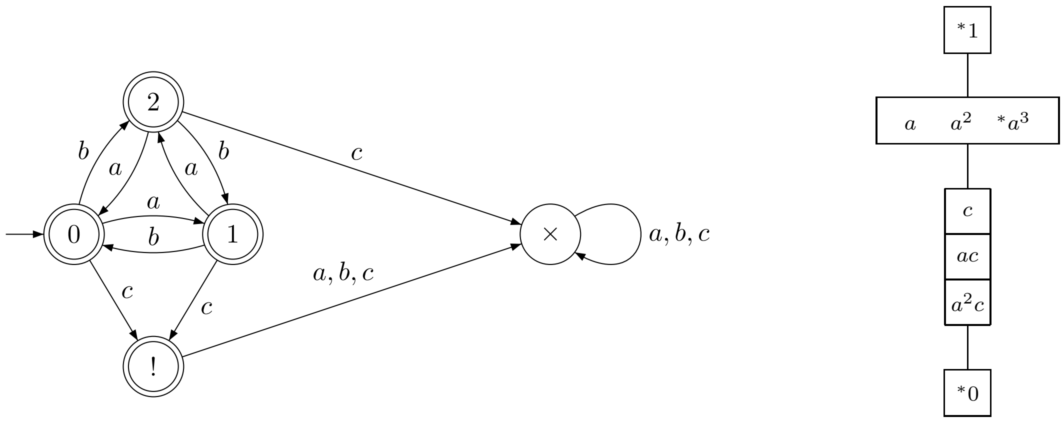
The minimal automaton $\mathcal{D}_L$ for the language L is shown in Fig. 3a.

The syntactic monoid M_L of L is given by the transition monoid of this automaton – we thus have $M_L = \langle a, b, c \mid a = a^4, b = a^2, c = a^3c, ca = 0, cb = 0, cc = 0\rangle$. The egg-box diagram of the monoid M_L is shown in Fig. 3b. We see that all regular D-classes of M_L are commutative groups: the monoid M_L is in **DAb** and L is **DAb**-recognisable.

It remains to show that $\gamma(L) \notin \mathcal{W}$. As L is obviously prefix-closed, we have $\gamma(L) = \mathrm{pref}^{\downarrow}([L])$. The language $[L]$ is given by

$$[L] = \{a, b\}^* \cup \{ucv \mid u, v \in \{a, b\}^*;\ |uv|_a - |uv|_b \not\equiv 2 \pmod 3\}$$

with its minimal automaton shown in Fig. 4.



(a) The minimal automaton $\mathcal{D}_L$ for the language L.

(b) The egg-box diagram for M_L

Figure 3: The automaton $\mathcal{D}_L$ and the egg-box diagram of M_L.

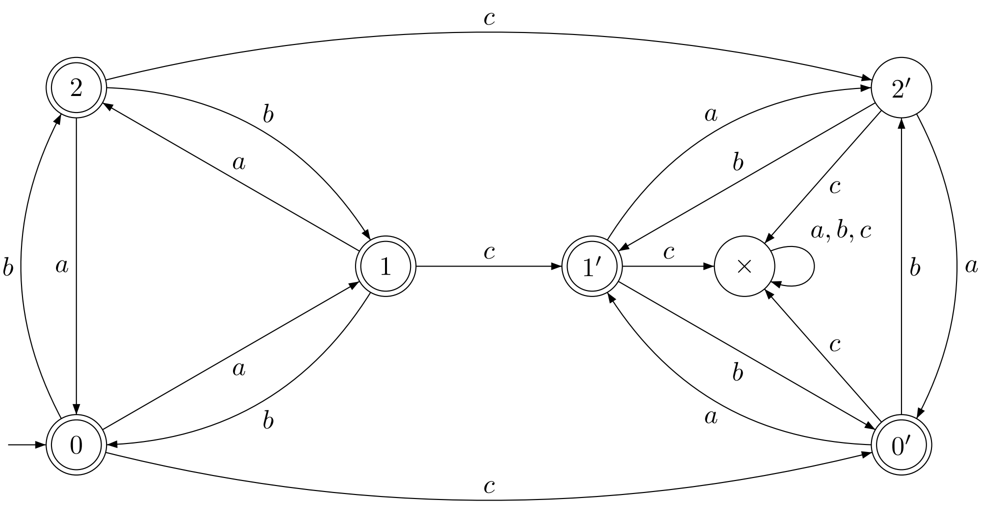


Figure 4: The minimal automaton for the language $[L]$.

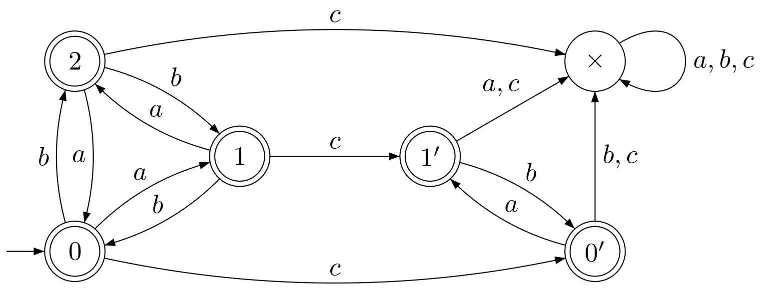


Figure 5: The minimal automaton for the language $\gamma(L)$.

It follows that the minimal automaton for $\gamma(L) = \mathrm{pref}^{\downarrow}([L])$ is given as in Fig. 5. In particular, under the assumption that $\gamma(L)$ is in $\mathcal{W}$, we see that $c^{-1}\gamma(L)a^{-1} = (ab)^* \in \mathcal{W}(\{a,b,c\}^*)$, as $\mathcal{W}$ is a positive variety. This implies that $(ab)^* \in \mathcal{W}(\{a,b\}^*)$, which contradicts the definition of $\mathcal{W}$. Thus $\gamma(L) \notin \mathcal{W}$ as claimed. □

The construction used in the proof of the previous theorem in fact implies a slightly stronger statement: there is no geometrically closed positive variety of languages containing the variety corresponding to **DV** for any pseudovariety of semigroups **V** containing the three-element cyclic group.[10] The same property holds for pseudovarieties **V** containing the cyclic group of arbitrary order $n > 1$ – for $n > 2$, this can be proved using a simple modification of the construction above; for $n = 2$, the reasoning is slightly more technical. This happens whenever **V** contains a nontrivial group.

Corollary 6.2. *None of the following (positive) varieties is geometrically closed:*

1. *The positive variety* $\mathcal{W}$*;*
2. *The variety of all* **DS***-recognisable languages;*
3. *The variety of all* **DG***-recognisable languages;*
4. *The variety of all* **DAb***-recognisable languages.*

On the other hand, some positive varieties of languages $\mathcal{V}$ are geometrically closed for trivial reasons – for instance all $\mathcal{V}$ such that $\gamma(L) = \Sigma^*$ for all non-empty $L \in \mathcal{V}(\Sigma^*)$. In particular, this is the case for L whenever $\mathrm{pref}^{\uparrow}(L) = \Sigma^*$. The proof of the following lemma is easy to see.

Lemma 6.3. *Let L be a regular language over an alphabet Σ and $\mathcal{D}_L$ be the minimal automaton of L. Then the following conditions are equivalent:*

(*i*) $\mathrm{pref}^{\uparrow}(L) = \Sigma^*$*;*

(*ii*) *for each state p in $\mathcal{D}_L$, there exists a final state reachable from p;*

(*iii*) *every absorbing state p in $\mathcal{D}_L$ is final.*

The conditions of Lemma 6.3 are satisfied in particular for all non-empty group languages. The variety $\mathcal{G}$, consisting of all languages L such that the syntactic monoid M_L is a group, is thus geometrically closed. This result can be extended to languages of the form $L = L_0 a_1 L_1 \dots a_\ell L_\ell$, where each a_i is a letter, and each L_i is a non-empty group language. Indeed, for every $u \in \Sigma^*$, there is some $w \in L_0$ such that $u \leq w$, and one can find at least one $w_i \in L_i$ for every $i = 1, \dots, \ell$. Then u is a prefix of the word $w a_1 w_1 \dots a_\ell w_\ell \in L$. This implies that $\mathrm{pref}^{\uparrow}(L) = \Sigma^*$. We may thus conclude that the positive variety $\mathcal{G}_{1/2}$, consisting of languages of level 1/2 in the group hierarchy, is geometrically closed. (The reader not familiar with the group hierarchy is referred to [19].) Moreover, every positive subvariety of $\mathcal{G}_{1/2}$ is geometrically closed as well. It follows as a particular case of this observation that also the positive variety $\mathcal{V}_{1/2}$, forming the level 1/2 of the Straubing-Thérien hierarchy, is geometrically closed. We have thus proved the following statement.

Corollary 6.4. *The following classes are geometrically closed:*

1. *The variety $\mathcal{G}$ of all group languages;*
2. *The positive variety* $\mathcal{G}_{1/2}$*;*
3. *The positive variety* $\mathcal{V}_{1/2}$*.*

In addition to the negative result of Theorem 6.1, let us also prove that contrary to the case of star-free languages, there is no largest geometrically closed positive variety or variety of languages.

Proposition 6.5. *There is no largest geometrically closed positive variety of languages, as well as no largest geometrically closed variety of languages.*

Proof. Suppose that $\mathcal{V}$ is either the largest geometrically closed positive variety, or the largest geometrically closed variety. In both cases, $\mathcal{V}$ has to contain both $\mathcal{G}$ and the variety corresponding to **DA**. It follows that $\mathcal{V}$ contains the language

[10] The definition of the operator **D** is reviewed in the Appendix.

$$L = \{a,b\}^* \cup \{wc \mid w \in \{a,b\}^*;\ |w|_a - |w|_b \not\equiv 2 \pmod 3\},$$

since this can be written as

$$L = \{a,b\}^* \cup (\{a,b\}^*c \ \cap \ \{w \in \{a,b,c\}^* \mid |w|_a - |w|_b \not\equiv 2 \pmod 3\}),$$

where $\{a,b\}^*$ and $\{a,b\}^*c$ are **DA**-recognisable, and $\{w \in \{a,b,c\}^* \mid |w|_a - |w|_b \not\equiv 2 \pmod 3\}$ is in $\mathcal{G}$. However, we have seen in the proof of Theorem 6.1 that there is no geometrically closed positive variety containing L, as such positive variety would also contain $(ab)^*$, which is impossible by Lemma 4.1. □

Finally, let us note that there also is no *smallest* non-trivial geometrically closed variety or positive variety, as all (positive) varieties of commutative languages are geometrically closed by Lemma 5.22, and there is no smallest non-trivial (positive) variety of commutative languages.

7. Conclusions

We have characterised the recently introduced operation of geometrical closure on formal languages [11] in terms of three simple operations – the prefix closure, the commutative closure, and the prefix reduction – and used this characterisation to study the properties of the geometrical closure when applied to members of various classes of regular languages. We have fully characterised positive varieties all of whose languages have regular geometrical closures: a positive variety has this property if and only if it is contained in $\mathcal{W}$.

The main question that we have considered is that of identifying robust geometrically closed classes of regular languages. On the positive side, we have introduced a variety of languages $\mathcal{R}_{LT}$, corresponding to the pseudovariety of finite monoids **MK**, and we have proved that geometrical closures of languages from the intersection of $\mathcal{W}$ with the variety $\mathcal{SF}$ of all star-free languages always fall into $\mathcal{R}_{LT}$. As a consequence, we have seen that many natural classes of star-free languages are geometrically closed – examples of such classes include $\mathcal{V}_{3/2}$, $\mathcal{R}$, and the variety corresponding to **DA**. Moreover, we have seen that $\mathcal{W} \cap \mathcal{SF}$ is the largest geometrically closed positive variety of star-free languages and that **DA**-recognisable languages form the largest geometrically closed variety of star-free languages.

We have also seen that interesting geometrically closed classes are much rarer outside the universe of all star-free languages – this is demonstrated mainly by a result, according to which there is no geometrically closed positive variety of regular languages containing the variety corresponding to **DAb**. We have only slightly touched upon classes of languages not containing $\mathcal{R}_{LT}$ or the **DAb**-recognisable languages. We leave a more systematic study of geometrically closed classes of this kind open for future research.

Another open problem stems from the fact that the counterexamples used in our two main negative results – that is, Lemma 4.1 and Theorem 6.1 – make use of a three-letter alphabet. It would be interesting to know if a two-letter alphabet can be used instead, or if the landscape of binary languages is richer with geometrically closed classes.

Finally, one may ask how to effectively construct a regular expression (of the form characteristic for $\mathcal{R}_{LT}$ or $\mathcal{V}_{3/2}$) for the geometrical closure $\gamma(L)$ of a given language L from $\mathcal{W} \cap \mathcal{SF}$, or at least from $\mathcal{V}_{3/2}$, given a deterministic finite automaton $\mathcal{A}$ recognising L. Note that both the positive variety $\mathcal{W}$ [9] (and thus also its intersection with $\mathcal{SF}$) and the positive variety $\mathcal{V}_{3/2}$ [19, p. 725] are known to be decidable, but it is not even clear to us whether there is an efficient algorithm that, given a deterministic finite automaton recognising a language L from $\mathcal{V}_{3/2}$, computes a regular expression for this language L (of the form characteristic for the positive variety $\mathcal{V}_{3/2}$).

References

[1] J. Almeida. *Finite Semigroups and Universal Algebra*. World Scientific, 1994.

[2] M. Arfi. Opérations polynomiales et hiérarchies de concaténation. *Theoretical Computer Science*, 91(1):71–84, 1991.

[3] M.-P. Béal, J.-M. Champarnaud, J.-P. Dubernard, H. Jeanne, and S. Lombardy. Decidability of geometricity of regular languages. In *DLT 2012*, pages 62–72, 2012.

[4] B. Blanpain, J.-M. Champarnaud, and J.-P. Dubernard. Geometrical languages. In *LATA 2007*, pages 127–138, 2007.

[5] A. Bouajjani, A. Muscholl, and T. Touili. Permutation rewriting and algorithmic verification. *Information and Computation*, 205(2):199–224, 2007.
[6] J. A. Brzozowski and F. E. Fich. Languages of $\mathscr{R}$-trivial monoids. *Journal of Computer and System Sciences*, 20(1):32–49, 1980.
[7] A. Cano, G. Guaiana, and J.-É. Pin. Regular languages and partial commutations. *Information and Computation*, 230:76–96, 2013.
[8] A. Cano Gómez and J.-É. Pin. Shuffle on positive varieties of languages. *Theoretical Computer Science*, 312:433–461, 2004.
[9] A. Cano Gómez and J.-É. Pin. A robust class of regular languages. In *MFCS 2008*, pages 36–51, 2008.
[10] L. Chaubard, J.-É. Pin, and H. Straubing. Actions, wreath products of C-varieties and concatenation product. *Theoretical Computer Science*, 356(1-2):73–89, 2006.
[11] J.-P. Dubernard, G. Guaiana, and L. Mignot. Geometrical closure of binary $V_{3/2}$-languages. In *LATA 2019*, pages 302–314, 2019.
[12] S. Eilenberg. *Automata, Languages, and Machines, Vol. B.* Academic Press, 1976.
[13] M. Gehrke, S. Grigorieff, and J.-É. Pin. Duality and equational theory of regular languages. In *ICALP 2008 (Track B)*, pages 246–257, 2008.
[14] G. Guaiana, A. Restivo, and S. Salemi. On the trace product and some families of languages closed under partial commutations. *Journal of Automata, Languages and Combinatorics*, 9(1):61–79, 2004.
[15] J. M. Howie. *Fundamentals of Semigroup Theory.* Clarendon Press, 1995.
[16] O. Klíma and P. Kostolányi. Geometrically closed positive varieties of star-free languages. In *LATA 2020*, pages 328–340, 2020.
[17] J.-É. Pin. Semigroupe des parties et relations de Green. *Canadian Journal of Mathematics*, 36(2):327–343, 1984.
[18] J.-É. Pin. *Varieties of Formal Languages.* North Oxford Academic Publishers, 1986.
[19] J.-É. Pin. Syntactic semigroups. In *Handbook of Formal Languages*, volume 1, chapter 10, pages 679–746. Springer, 1997.
[20] J.-É. Pin, H. Straubing, and D. Thérien. Small varieties of finite semigroups and extensions. *Journal of the Australian Mathematical Society*, 37(2):269–281, 1984.
[21] J. Sakarovitch. *Elements of Automata Theory.* Cambridge University Press, 2009.
[22] P. Tesson and D. Thérien. Diamonds are forever: the variety DA. In *Semigroups, algorithms, automata and languages*, pages 475–499. World Scientific, 2002.

Appendix A. Summary of Pseudovarieties Used

The table below briefly summarises, for convenience of the reader, the notation used for different pseudovarieties of finite monoids, as well as their characterisations in terms of pseudoidentities. The reader might consult [1] for a more comprehensive account.

Pseudovariety	Description	Pseudoidentities
A	Aperiodic finite monoids.	$x^{\omega+1} = x^\omega$
Com	Commutative finite monoids.	$xy = yx$
DA	Finite monoids such that all their regular D-classes are aperiodic semigroups.	$x^{\omega+1} = x^\omega$, $(xy)^\omega(yx)^\omega(xy)^\omega = (xy)^\omega$
DAb	Finite monoids such that all their regular D-classes are Abelian groups.	$(xy)^\omega = (yx)^\omega$, $(xy)^\omega xy = (xy)^\omega yx$
DG	Finite monoids such that all their regular D-classes are groups.	$(xy)^\omega = (yx)^\omega$
DS	Finite monoids such that all their regular D-classes are semigroups.	$((xy)^\omega(yx)^\omega(xy)^\omega)^\omega = (xy)^\omega$
J	Finite J-trivial monoids.	$x^{\omega+1} = x^\omega$, $(xy)^\omega = (yx)^\omega$
JMK	See [1, p. 236].	See [1, p. 236].
MK	A pseudovariety generated by monoids S^1 for finite semigroups S, in which all idempotents act as left zeros.	$x^\omega yx = x^\omega y$
R	Finite R-trivial monoids.	$(xy)^\omega x = (xy)^\omega$
$\mathbf{R}_1$	Finite idempotent R-trivial monoids [20].	$xyx = xy$
Sl	Finite semilattices.	$xy = yx$, $x^2 = x$

Table A.1: Pseudovarieties of finite monoids used in this article and their characterisations by pseudoidentities.

Several pseudovarieties listed in the table above are of the form **DV**, where **V** is some pseudovariety of semigroups. The pseudovariety **DV** consists of all finite monoids M such that all regular D-classes of M are subsemigroups of M that belong to **V**.