Geometrically Closed Positive Varieties of Star-Free Languages*

Ondřej Klíma[1] and Peter Kostolányi[2][0000−0002−5474−8781]

[1] Department of Mathematics and Statistics, Masaryk University
Kotlářská 2, 611 37 Brno, Czech Republic
klima@math.muni.cz
[2] Department of Computer Science, Comenius University in Bratislava
Mlynská dolina, 842 48 Bratislava, Slovakia
kostolanyi@fmph.uniba.sk

Abstract. A recently introduced operation of geometrical closure on formal languages is investigated. It is proved that the geometrical closure of a language from the positive variety $\mathcal{V}_{3/2}$, the level 3/2 of the Straubing-Thérien hierarchy of star-free languages, always falls into the variety $\mathcal{R}_{LT}$, which is a new variety consisting of specific R-trivial languages. As a consequence, each class of regular languages lying between $\mathcal{R}_{LT}$ and $\mathcal{V}_{3/2}$ is geometrically closed.

Keywords: Language varieties · Geometrical closure · Straubing-Thérien hierarchy · R-trivial monoid

1 Introduction

A *geometrical closure* is an operation on formal languages introduced recently by Dubernard, Guaiana, and Mignot [8]. It is defined as follows: Take any language L over some k-letter alphabet and consider the set called the *figure* of L in [8], 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 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 from later paper [8], 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 a geometrical language is completely determined by its (geometrical) figure. In the particular case of binary alphabets, these (geometrical) figures were illustrated by plane diagrams in [8].

The class of all regular languages can be easily observed not to be geometrically closed – that is, one can find a regular language such that its geometrical closure is not regular [8] (see also the end of Section 2). One possible research

* The first author was supported by Grant 19-12790S of the Grant Agency of the Czech Republic. The second author was supported by the grant VEGA 2/0165/16.

aim could be to characterise regular languages L for which $\gamma(L)$ is regular, or to describe some robust classes of languages with this property. Another problem posed in [8] is to find some subclasses of regular languages that are geometrically closed. As we explain in Section 3, non-empty group languages have their geometrical closure equal to the universal language Σ^*. For this reason, it makes sense to look for more interesting geometrically closed subclasses among *star-free* languages, which are known to be "group-free". More precisely, a language L is *star-free* if and only if the syntactic monoid M_L of L is *aperiodic*, that is, if M_L does not contain non-trivial groups as subsemigroups.

It is well known that the star-free languages are 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) is formed by piecewise testable languages and the positive variety $\mathcal{V}_{3/2}$ is formed by polynomials built from languages of level 1. We refer to the survey paper by Pin [12] for an introduction to the Straubing-Thérien hierarchy of star-free languages and the algebraic theory of regular languages in general. This theory is based on Eilenberg correspondence between varieties of regular languages and pseudovarieties of finite monoids. Note that one well-known instance of Eilenberg correspondence, which plays an essential role in our contribution, is given by the pseudovariety of finite R-trivial monoids, for which the corresponding variety of languages is denoted by $\mathcal{R}$. Nevertheless, we emphasise that our contribution is rather elementary, and it does not use sophisticated tools developed in the algebraic theory of regular languages.

It was proved by Dubernard, Guaiana, and Mignot [8] 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 the above mentioned result in this contribution. Our approach is to concentrate on the form of languages that may arise as $\gamma(L)$ for L taken from $\mathcal{V}_{3/2}(\Sigma)$. In other words, we do not construct a concrete regular expression for $\gamma(L)$, but we determine what kind of expression exists for such a language. In particular, we introduce a new variety of languages $\mathcal{R}_{LT}$, which is a subvariety of the variety $\mathcal{R}$. 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 Section 4). 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.

We show that the geometrical closure of a language from the positive variety $\mathcal{V}_{3/2}$ always falls into the variety $\mathcal{R}_{LT}$. As a consequence, each class of regular languages lying between $\mathcal{R}_{LT}$ and $\mathcal{V}_{3/2}$ 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 $\mathcal{R}$.

2 Preliminaries

All automata considered in this paper 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 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 [12]. We recall this notion for a reader's convenience briefly. 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 $\mathcal{V}(\Sigma^*)$ is closed under quotients, finite unions and intersections, and the whole class is closed under preimages in homomorphisms. A positive variety $\mathcal{V}$ is a *variety* if each $\mathcal{V}(\Sigma^*)$ is closed under complementation. Note that an alphabet could be fixed in our contribution, so homomorphisms among different alphabets play no role, and we could consider lattices of languages [9] instead of varieties of languages. However, we prefer to stay in the frame of the theory of (positive) varieties of languages as a primary aim of this paper is to describe robust classes closed under geometrical closure.

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

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

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 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 alphabet Σ and each $L \in \mathcal{V}(\Sigma^*)$, the language

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

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

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, *i.e.* $[w] = \{u \in \Sigma^* \mid \Psi(u) = \Psi(w)\}$. Then we also write, for each language $L \subseteq \Sigma^*$,

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

and we call $[L]$ 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 alphabet Σ, the language $[L]$ belongs to $\mathcal{C}(\Sigma^*)$ 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 [8], 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 $\pi \subseteq F$.

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 $\|F\|$ the set $\{\|\pi\| \mid \pi \subseteq F\}$. Note that the language $\|F\|$ is prefix-closed.

Moreover, we put $\mathrm{fig}(L) = \Psi(\mathrm{pref}^{\uparrow}(L))$ for each $L \subseteq \Sigma^*$. The set $\mathrm{fig}(L) \subseteq \mathbb{N}^k$ is a *connex figure* in the sense of [8], *i.e.*, for each $\mathbf{x} \in \mathrm{fig}(L)$, there is a path π leading to $\mathbf{x}$ such that $\pi \subseteq \mathrm{fig}(L)$.

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

Note that the class of all regular languages is *not* geometrically closed, as observed in [8]. 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 via three simpler operations: the prefix closure, the commutative closure, and the prefix reduction. This characterisation is a key to our later considerations.

Proposition 2. *If L is a language over Σ, then $\gamma(L) = \mathrm{pref}^{\downarrow}\big([\mathrm{pref}^{\uparrow}(L)]\big)$.*

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 $\pi = [\mathbf{x}_0, \ldots, \mathbf{x}_n] \subseteq \Psi(\mathrm{pref}^{\uparrow}(L))$ 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 $\Psi(\mathrm{pref}^{\uparrow}(L))$, implying that w is in $\|\Psi(\mathrm{pref}^{\uparrow}(L))\| = \gamma(L)$. □

As a direct consequence of Proposition 1 and Proposition 2, we obtain the following sufficient condition, under which a positive variety of languages is geometrically closed.

Corollary 3. *Each positive variety of regular languages closed under prefix reduction and commutation is geometrically closed.*

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 $L \in \mathcal{V}(\Sigma^*)$. Let us observe that this is the case for L whenever $\mathrm{pref}^{\uparrow}(L) = \Sigma^*$. The proof of the following lemma is easy to see. We just note that by an *absorbing state* we mean a state p satisfying $p \cdot a = p$ for every $a \in \Sigma$.

Lemma 4. *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 4 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 geometrically closed as a consequence. 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 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 [12].)

In the rest of the paper, we move our attention to star-free languages.

4 Languages Recognised by LT-acyclic Automata

We now introduce the class of languages $\mathcal{R}_{LT}$, which plays a central role in our main result. For every alphabet Σ, the set $\mathcal{R}_{LT}(\Sigma^*)$ consists of languages which are finite unions of languages of the form

$$L = \Sigma_0^* a_1 \Sigma_1^* a_2 \dots a_n \Sigma_n^*, \tag{1}$$
$$\text{where } \Sigma_0 \subseteq \Sigma_1 \subseteq \dots \subseteq \Sigma_n \subseteq \Sigma \text{ and } a_i \in \Sigma \setminus \Sigma_{i-1} \text{ for } i = 1, \dots, n.$$

The previous definition is similar to definitions of other classes of languages that have already been studied in literature. First of all, if we omit the condition $\Sigma_0 \subseteq \Sigma_1 \subseteq \dots \subseteq \Sigma_n$, we get a definition of languages from the variety $\mathcal{R}$

corresponding to R-trivial monoids, which we recall in more detail later. Let us conclude here just that $\mathcal{R}_{LT} \subseteq \mathcal{R}$. Secondly, if we also require $a_i \in \Sigma_i$ in (1) for $i = 1, \ldots, n$, then we obtain a variety of languages considered by Pin, Straubing, and Thérien [13] and corresponding to a pseudovariety of finite monoids denoted $\mathbf{R}_1$. Finally, if we drop in (1) the condition $a_i \notin \Sigma_{i-1}$ and then we generate a variety, then we obtain the variety of languages corresponding to the pseudovariety **JMK** considered by Almeida [1, p. 236].

Since we want 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 loop. This means that 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. The defining condition means that one can number the states in Q as $1, \ldots, |Q|$ in such a way that the state $p \cdot a$, with $p \in Q$ and $a \in \Sigma$, is always greater than or equal to p. For this reason, these automata are called *extensive* in [11, p. 93]. It is known that they recognise precisely R-trivial languages [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 in a state p in an LT-acyclic automaton, then there is also an a-labelled loop in each state reachable from p. We may thus equivalently take $b \in \Sigma^*$ in the previous definition. The first aim of this section 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. *For a language L of the form (1), the automaton $\mathcal{D}_L$ is LT-acyclic.*

Proof. Let L be a language $L = \Sigma_0^* a_1 \Sigma_1^* a_2 \ldots a_n \Sigma_n^*$ of the form (1). For every $i = 1, \ldots, n$, we denote $\Gamma_{i-1} = \Sigma \setminus (\Sigma_{i-1} \cup \{a_i\})$ and we also put $\Gamma_n = \Sigma \setminus \Sigma_n$. Then it is an easy exercise to show that the automaton in Fig. 1 is the minimal automaton of L and that it is an LT-acyclic automaton. □

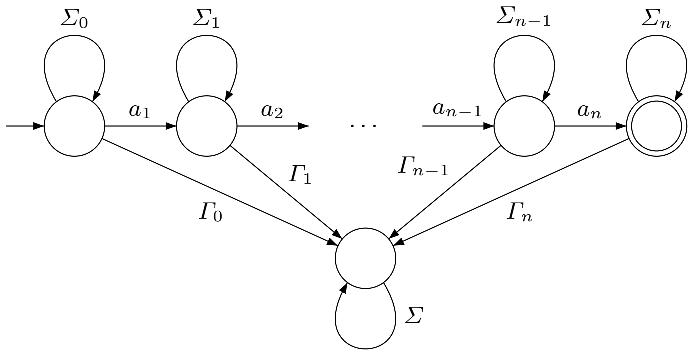

Fig. 1. An LT-acyclic automaton for the language of the form (1).

Lemma 6. *Let L, K be languages over an alphabet Σ recognised by LT-acyclic automata. Then $L \cup K$ is also recognised by an LT-acyclic automaton.*

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. It is a routine to check that a finite direct product of LT-acyclic automata is an LT-acyclic automaton. □

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

Lemma 7. *Let L be a language recognised by an LT-acyclic automaton. Then the minimal automaton of L is also LT-acyclic.*

Proof. Let $\mathcal{A} = (Q, \Sigma, \cdot, \iota, F)$ be an LT-acyclic automaton such that $\|\mathcal{A}\| = L$. The minimal automaton $\mathcal{D}_L$ is a homomorphic image of some subautomaton of $\mathcal{A}$ [14]. It is clear that a subautomaton of an LT-acyclic automaton is LT-acyclic. Thus we may assume that $\mathcal{A}$ has all states reachable from the initial state ι.

Let $\varphi\colon Q \to P$ be a surjective mapping, which is a homomorphism from the automaton $\mathcal{A}$ onto an automaton $\mathcal{B} = (P, \Sigma, \bullet, \varphi(\iota), \varphi(F))$. We claim that $\mathcal{B}$ is 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. Since the sequence $q', q' \cdot w, q' \cdot w^2, \ldots$ contains only finitely many states, there are natural numbers n and m such that $q' \cdot w^{n+m} = q' \cdot w^n = q$. Since $\mathcal{A}$ is acyclic, we have $q \cdot a = q$ for every letter a occurring in w. Consequently, $p \bullet a = \varphi(q) \bullet a = \varphi(q \cdot a) = \varphi(q) = p$. We showed that $\mathcal{B}$ is acyclic.

Now let $p \in P$ and $a \in \Sigma$ be such that $p \bullet a = p$. It follows from the previous paragraph that there is $q \in \varphi^{-1}(p)$ such that $q \cdot a = q$. Since $\mathcal{A}$ is LT-acyclic, we see that $(q \cdot b) \cdot a = q \cdot b$ for every $b \in \Sigma$. Thus $p \bullet ba = \varphi(q \cdot ba) = \varphi(q \cdot b) = p \bullet b$. We showed that $\mathcal{B}$ is an LT-acyclic automaton. In particular, it is true for $\mathcal{D}_L$. □

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

Lemma 8. *Let $\mathcal{A}$ be an LT-acyclic automaton over an alphabet Σ. Then $\|\mathcal{A}\|$ belongs to $\mathcal{R}_{LT}(\Sigma^*)$.*

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

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

We see that the set R is finite. Moreover, for each q in Q, let Σ_q denote the alphabet $\Sigma_q = \{c \in \Sigma \mid q \cdot c = q\}$. Then

$$L_w := \Sigma_{q_0}^* a_1 \Sigma_{q_1}^* a_2 \ldots a_n \Sigma_{q_n}^* \subseteq \|\mathcal{A}\|$$

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

$$\|\mathcal{A}\| = \bigcup_{w \in R} L_w.$$

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

The following theorem provides a summary of the previous lemmas.

Theorem 9. *For a language $L \subseteq \Sigma^*$, the following statements are equivalent:*
(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 and Lemma 6. The statement (ii) implies (iii) by Lemma 7. Finally, (iii) implies (i) by Lemma 8. □

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 [7]. Here we complete the previous characterisation by showing 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 a unique idempotent in the subsemigroup of M generated by s. It could be useful to know that, for a fixed finite monoid M, there is a natural number m such that $s^\omega = s^m$ for each $s \in M$.

Theorem 10. *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 pseudoidentities $(xy)^\omega x = (xy)^\omega$ and $x^\omega yx = x^\omega y$.
(iii) M_L satisfies the pseudoidentity $(xy)^\omega zx = (xy)^\omega z$.

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 [12, p. 692]). Elements of M_L are thus transitions of $\mathcal{D}_L$ determined by words from Σ^*. More formally, for $u \in \Sigma^*$, we denote by f_u the transition 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 9. In particular, the language L is R-trivial as we already mentioned. Hence, the monoid M_L is R-trivial, *i.e.*, M_L satisfies the pseudoidentity $(xy)^\omega x = (xy)^\omega$. Next, let x, y be mapped to elements in M_L which are given by words $v, w \in \Sigma^*$. We now need to check that $f_{v^m} f_w f_v = f_{v^m} f_w$. Since $\mathcal{D}_L$ 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. Since $\mathcal{D}_L$ is an LT-acyclic automaton, the loop labelled by a in 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 in the state $(p \cdot v^m) \cdot w$. The equality $f_{v^m} f_w f_v = f_{v^m} f_w$ follows.

Next, let us show that the pseudoidentity $(xy)^\omega zx = (xy)^\omega z$ is a consequence of pseudoidentities from item (ii). We may interpret x, y, z as arbitrary elements of any finite monoid M satisfying these pseudoidentities. Let m be such that

$s^\omega = s^m$ for each $s \in M$. Then we use the second pseudoidentity from (ii) repetitively, and we get

$$(xy)^\omega z = (xy)^\omega zxy = (xy)^\omega z(xy)^2 = \ldots = (xy)^\omega z(xy)^m = (xy)^\omega z(xy)^\omega . \quad (2)$$

By the first pseudoidentity from (ii), we get $(xy)^\omega z(xy)^\omega = (xy)^\omega z(xy)^\omega x$. Then we obtain $(xy)^\omega z(xy)^\omega x = (xy)^\omega zx$ using the equality (2). Thus we get $(xy)^\omega z = (xy)^\omega z(xy)^\omega = (xy)^\omega z(xy)^\omega x = (xy)^\omega zx$.

Finally, in order to prove that (iii) implies (i), suppose that M_L satisfies the pseudoidentity $(xy)^\omega zx = (xy)^\omega z$. Taking $z = 1$, it follows that M_L satisfies the pseudoidentity $(xy)^\omega x = (xy)^\omega$. Hence, L is R-trivial and $\mathcal{D}_L$ 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 $x \mapsto a$, $y \mapsto 1$, and $z \mapsto b$ in $(xy)^\omega zx = (xy)^\omega z$ gives us $f_{a^\omega} f_b f_a = f_{a^\omega} f_b$. Therefore, $p \cdot ba = p \cdot b$. So, we see that there is a loop labelled by a in the state $p \cdot b$. We proved that $\mathcal{D}_L$ is an LT-acyclic automaton and L belongs to $\mathcal{R}_{LT}(\Sigma^*)$ by Theorem 9. □

Corollary 11. *The class $\mathcal{R}_{LT}$ is a variety of languages corresponding to the pseudovariety of finite monoids $\mathbf{R}_{LT}$ given by*

$$\mathbf{R}_{LT} = [\![(xy)^\omega zx = (xy)^\omega z]\!] = [\![(xy)^\omega x = (xy)^\omega , x^\omega yx = x^\omega y]\!].$$

Let us also note that $[\![x^\omega yx = x^\omega y]\!]$ is known to describe the pseudovariety of finite monoids **MK**; cf. Almeida [1, p. 212], who attributes this result to Pin. Therefore, $\mathbf{R}_{LT} = \mathbf{R} \cap \mathbf{MK}$.

5 The Main Result

Let us now return to the geometrical closure and prove the main result of this paper: *each class of languages lying between the variety of languages $\mathcal{R}_{LT}$ and the positive variety $\mathcal{V}_{3/2}$ is geometrically closed*. This strengthens the result from [8] mentioned in the Introduction.

The route that we take to this result (Theorem 16) consists of three steps:

1. We recall that the class $\mathcal{V}_{3/2}$ is closed under commutation [10, 5]. Although it is not necessary to obtain our main result, we refine this observation by proving that a commutative closure of a $\mathcal{V}_{3/2}$-language is piecewise testable.
2. We prove that each commutative $\mathcal{V}_{3/2}$-language belongs to $\mathcal{R}_{LT}$.
3. We observe that the variety $\mathcal{R}_{LT}$ is closed under prefix reduction.

These three observations imply that the geometrical closure of a $\mathcal{V}_{3/2}$-language belongs to $\mathcal{R}_{LT}$, from which our main result follows easily.

Recall the result of Arfi [2], according to which a language belongs to $\mathcal{V}_{3/2}$ if and only if it is given by a finite union of languages $\Sigma_0^* a_1 \Sigma_1^* a_2 \ldots a_n \Sigma_n^*$, where $a_1, \ldots, a_n$ are letters from Σ and $\Sigma_0, \ldots, \Sigma_n$ are subalphabets of Σ. It follows by a more general result of Guaiana, Restivo, and Salemi [10], or of

Bouajjani, Muscholl, and Touili [5] that $\mathcal{V}_{3/2}$ is closed under commutation, and this observation is a first step to Theorem 16.

Let us show that a commutative closure of a $\mathcal{V}_{3/2}$-language is in fact piecewise testable.

Lemma 12. *A commutative closure of a $\mathcal{V}_{3/2}$-language is piecewise testable.*

Proof. Let an alphabet Σ be fixed. It is clear that if $L_1, \ldots, L_m \subseteq \Sigma^*$ are languages, then

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

As a result, it is enough to prove piecewise testability of $[L]$ for all languages $L = \Sigma_0^* a_1 \Sigma_1^* a_2 \ldots a_n \Sigma_n^*$, with $a_1, \ldots, a_n \in \Sigma$ and $\Sigma_0, \ldots, \Sigma_n \subseteq \Sigma$.

Let L be of this form. Denote $\Sigma' = \Sigma_0 \cup \ldots \cup \Sigma_n$, and $x = a_1 \ldots a_n$. We claim that

$$[L] = \{w \in \Sigma^* \mid \forall a \in \Sigma' : |w|_a \geq |x|_a;\ \forall b \in \Sigma \setminus \Sigma' : |w|_b = |x|_b\}. \qquad (3)$$

Indeed, if w is in $[L]$, then $\Psi(w) = \Psi(u)$ for some $u \in L$, while clearly $|u|_a \geq |x|_a$ for each a in Σ', and $|u|_b = |x|_b$ for each b in $\Sigma \setminus \Sigma'$. Conversely, let w in Σ^* be such that $|w|_a \geq |x|_a$ for each a in Σ', and $|u|_b = |x|_b$ for each b in $\Sigma \setminus \Sigma'$. Then $\Psi(w) = \Psi(v)$ for v in Σ^* given by $v = v_0 a_1 v_1 a_2 \ldots a_n v_n$, where v_i $(i = 0, \ldots, n)$ is given as follows: if $\Sigma_i \setminus (\Sigma_0 \cup \ldots \cup \Sigma_{i-1}) = \{b_1, \ldots, b_j\}$, then

$$v_i = b_1^{|w|_{b_1} - |x|_{b_1}} \ldots b_j^{|w|_{b_j} - |x|_{b_j}}.$$

The word v is in L by construction, hence w belongs to $[L]$.

It remains to observe that the language $[L]$ given by (3) is piecewise testable. However, this language is equal to

$$[L] = \bigcap_{a \in \Sigma'} (\Sigma^* a)^{|x|_a} \Sigma^* \cap \bigcap_{b \in \Sigma \setminus \Sigma'} \left((\Sigma^* b)^{|x|_b} \Sigma^* \cap \left((\Sigma^* b)^{|x|_b + 1} \Sigma^* \right)^C \right). \qquad (4)$$

The language on the right-hand side of (4) is piecewise testable. □

We now proceed to prove that the geometrical closure of each language from $\mathcal{V}_{3/2}$ belongs to $\mathcal{R}_{LT}$.

Lemma 13. *Every commutative language L from $\mathcal{V}_{3/2}$ belongs to $\mathcal{R}_{LT}$.*

Proof. If we take into account the proof of Lemma 12 and the fact that $\mathcal{R}_{LT}$ is closed under finite unions, it is enough to prove that every language of the form (3) belongs to $\mathcal{R}_{LT}$. We may also use the expression (4) for that language. For each letter $a \in \Sigma$ and a natural number m, we may write $(\Sigma^* a)^m \Sigma^* = ((\Sigma \setminus \{a\})^* a)^m \Sigma^*$. This shows that the language $(\Sigma^* a)^m \Sigma^*$ belongs to $\mathcal{R}_{LT}$. Since $\mathcal{R}_{LT}$ is a variety, we see that also the language $((\Sigma^* a)^m \Sigma^*)^C$ belongs to $\mathcal{R}_{LT}$. Altogether, the language (4) belongs to the variety $\mathcal{R}_{LT}$. □

Finally, let us observe that the variety $\mathcal{R}_{LT}$ is closed under prefix reduction.

Lemma 14. *Let L be a language from $\mathcal{R}_{LT}(\Sigma^*)$ for some alphabet Σ. Then* $\mathrm{pref}^{\downarrow}(L)$ *belongs to $\mathcal{R}_{LT}(\Sigma^*)$ as well.*

Proof. Let L be 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 $\mathcal{R}_{LT}(\Sigma^*)$. So we may assume that $\iota \in F$.

Now, simply saying, 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 put $p \bullet a = p \cdot a$ if $p \cdot a \in F$, and $p \bullet a = \tau$ otherwise. As $\mathcal{A}$ contains no cycle other than a loop, 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 equal to τ. As $\tau \bullet c = \tau$ for each c in Σ, this implies that $\mathcal{A}'$ is an LT-acyclic automaton and $\mathrm{pref}^{\downarrow}(L)$ belongs to $\mathcal{R}_{LT}(\Sigma^*)$ by Theorem 9. $\square$

Theorem 15. *Let Σ be an alphabet and $L \in \mathcal{V}_{3/2}(\Sigma^*)$. Then $\gamma(L) \in \mathcal{R}_{LT}(\Sigma^*)$.*

Proof. We have $\gamma(L) = \mathrm{pref}^{\downarrow}\left(\left[\mathrm{pref}^{\uparrow}(L)\right]\right)$ by Proposition 2. As $\mathcal{V}_{3/2}$ is a positive variety of languages, $\mathrm{pref}^{\uparrow}(L)$ belongs to $\mathcal{V}_{3/2}(\Sigma^*)$ whenever L belongs to this set by Proposition 1. The language $\left[\mathrm{pref}^{\uparrow}(L)\right]$ is thus a commutative $\mathcal{V}_{3/2}$-language by [10, 5]. (Note that the language $\left[\mathrm{pref}^{\uparrow}(L)\right]$ is actually commutative piecewise testable, by Lemma 12.) It follows by Lemma 13 that $\left[\mathrm{pref}^{\uparrow}(L)\right]$ belongs to $\mathcal{R}_{LT}(\Sigma^*)$, and by Lemma 14 that the language $\gamma(L) = \mathrm{pref}^{\downarrow}\left(\left[\mathrm{pref}^{\uparrow}(L)\right]\right)$ belongs to $\mathcal{R}_{LT}(\Sigma^*)$ as well. $\square$

We are now prepared to state the main result of this article merely as an alternative formulation of the theorem above.

Theorem 16. *Let $\mathcal{C}$ be a class of languages containing $\mathcal{R}_{LT}$, which is contained in $\mathcal{V}_{3/2}$. Then $\mathcal{C}$ is geometrically closed.*

There are many important (positive) varieties studied in the literature for which the main result can be applied.

Corollary 17. *The following classes are geometrically closed: the positive variety $\mathcal{V}_{3/2}$, the variety $\mathcal{R}$, the variety $\mathcal{R}_{LT}$, the variety of all* **JMK***-recognisable languages, the variety of all* **DA***-recognisable languages.*

The variety of all **DA**-recognisable languages 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 Straubing-Thérien hierarchy (see Section 5 in [15]).

6 Conclusions

We have introduced a new variety of languages $\mathcal{R}_{LT}$ and we have proved that geometrical closures of languages from $\mathcal{V}_{3/2}$ fall into $\mathcal{R}_{LT}$. As a consequence, we have seen that many natural classes of star-free languages are geometrically closed, namely those between the variety $\mathcal{R}_{LT}$ and the positive variety $\mathcal{V}_{3/2}$. On the contrary, the variety of all piecewise testable languages $\mathcal{V}_1$ is not geometrically closed. The example is not included in the paper due to space limitations.

There are some interesting questions in connection to the paper. First of all, one may ask how to effectively construct a regular expression for the geometrical closure $\gamma(L)$ for a given language L from $\mathcal{V}_{3/2}$. Note that it is effectively testable, for a given deterministic finite automaton $\mathcal{A}$, whether the language $\|\mathcal{A}\|$ belongs to $\mathcal{V}_{3/2}$ (see [12, p. 725]). It is not clear to us whether a regular expression for $\|\mathcal{A}\|$ can be effectively computed from $\mathcal{A}$.

Nevertheless, the main open question related to the topic is to clarify the behaviour of the geometrical closure outside the class $\mathcal{V}_{3/2}$.

References

1. Almeida, J.: Finite Semigroups and Universal Algebra. World Scientific (1994)
2. Arfi, M.: Opérations polynomiales et hiérarchies de concaténation. Theoretical Computer Science **91**(1), 71–84 (1991)
3. Béal, M.P., Champarnaud, J.M., Dubernard, J.P., Jeanne, H., Lombardy, S.: Decidability of geometricity of regular languages. In: DLT 2012. pp. 62–72 (2012)
4. Blanpain, B., Champarnaud, J.M., Dubernard, J.P.: Geometrical languages. In: LATA 2007. pp. 127–138 (2007)
5. Bouajjani, A., Muscholl, A., Touili, T.: Permutation rewriting and algorithmic verification. Information and Computation **205**(2), 199–224 (2007)
6. Brzozowski, J.A., Fich, F.E.: Languages of $\mathscr{R}$-trivial monoids. Journal of Computer and System Sciences **20**(1), 32–49 (1980)
7. Chaubard, L., Pin, J.É., Straubing, H.: Actions, wreath products of C-varieties and concatenation product. Theoretical Computer Science **356**(1-2), 73–89 (2006)
8. Dubernard, J.P., Guaiana, G., Mignot, L.: Geometrical closure of binary $V_{3/2}$-languages. In: LATA 2019. pp. 302–314 (2019)
9. Gehrke, M., Grigorieff, S., Pin, J.É.: Duality and equational theory of regular languages. In: ICALP 2008 (Track B). pp. 246–257 (2008)
10. Guaiana, G., Restivo, A., Salemi, S.: On the trace product and some families of languages closed under partial commutations. Journal of Automata, Languages and Combinatorics **9**(1), 61–79 (2004)
11. Pin, J.É.: Varieties of Formal Languages. North Oxford Academic Publishers (1986)
12. Pin, J.É.: Syntactic semigroups. In: Handbook of Formal Languages, vol. 1, chap. 10, pp. 679–746. Springer (1997)
13. Pin, J.É., Straubing, H., Thérien, D.: Small varieties of finite semigroups and extensions. Journal of the Australian Mathematical Society **37**(2), 269–281 (1984)
14. Sakarovitch, J.: Elements of Automata Theory. Cambridge University Press (2009)
15. Tesson, P., Thérien, D.: Diamonds are forever: the variety DA. In: Semigroups, algorithms, automata and languages, pp. 475–499. World Scientific (2002)